



HAL
open science

Le potentiel du web pour les enquêtes de mobilité urbaine

Caroline Bayart

► **To cite this version:**

Caroline Bayart. Le potentiel du web pour les enquêtes de mobilité urbaine. Economies et finances. Université Lumière - Lyon II, 2009. Français. NNT: . tel-00477745

HAL Id: tel-00477745

<https://theses.hal.science/tel-00477745v1>

Submitted on 30 Apr 2010

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

UNIVERSITÉ LUMIÈRE LYON II

FACULTÉ DE SCIENCES ECONOMIQUES ET DE GESTION

LE POTENTIEL DU WEB POUR LES ENQUÊTES
DE MOBILITÉ URBAINE

THESE POUR LE DOCTORAT DE SCIENCES ECONOMIQUES
MENTION ECONOMIE DES TRANSPORTS

Présentée et soutenue publiquement par

Caroline BAYART

Le 07 décembre 2009

MEMBRES DU JURY

M. Christophe BENAVENT, Professeur à l'Université Paris Ouest
M. Patrick BONNEL, Directeur de recherche à l'ENTPE, directeur de thèse
M. Jean-Louis CHANDON, Professeur à l'Université Aix-Marseille III, rapporteur
M. Yves CROISSANT, Professeur à l'Université de la Réunion
M. Yves CROZET, Professeur à l'Université Lyon II
M. François GARDES, Professeur à l'Université Paris I, rapporteur

Avant-propos

Avant de commencer cette thèse, je vais tenter de satisfaire au difficile et émouvant exercice d'écriture des remerciements.

La première personne que je tiens à remercier est Patrick Bonnel, qui a dirigé cette thèse. Il a su me guider et m'encourager tout au long de ces quatre années. Son regard critique, ses remarques rigoureuses et sa grande disponibilité ont largement contribué à la réalisation de ce travail.

Mes remerciements vont aussi à Yves Croissant, pour ses conseils et l'intérêt qu'il a porté à mes recherches. Je lui suis sincèrement reconnaissante d'avoir accepté de faire un long voyage pour participer au Jury.

Le Laboratoire d'Economie des Transports est un cadre privilégié pour effectuer une thèse. Je tiens à remercier chaleureusement l'ensemble de mes collègues pour les moments d'amitié partagés. Je remercie tout spécialement Aurélie, Louafi et Jorge, qui m'ont aidée par leurs compétences et leur gentillesse, Loulou, Pascal, Marie et Nathalie, pour leur sympathie et leur bonne humeur.

J'associe également à mes remerciements la Région Rhône-Alpes, pour son soutien financier, et l'IUT Lumière, qui m'a permis d'exercer les fonctions de moniteur puis d'ATER.

Enfin, cela va de soi, je remercie évidemment ma famille pour son irremplaçable et inconditionnel soutien. Laurent, qui a su supporter mon indisponibilité, Marie et Eva, toutes étonnées que leur maman ait enfin terminé 'son travail', cette thèse est aussi un peu la vôtre.

Sommaire

Avant-propos	ii
Introduction générale : un renouvellement des besoins sur les données de mobilité	1
I Analyse de la mobilité : enjeux et état des lieux	1
II De nouvelles exigences méthodologiques	3
III Des répondants de plus en plus réticents	4
IV Le potentiel des nouvelles technologies	9
V Le cadre de recherche	11
I La mise en ligne d'un questionnaire : fondements théoriques et empiriques	15
Introduction de la partie I	17
Chapitre 1 : Le potentiel du web pour les enquêtes ménages déplacements	18
I Apports et limites du web pour les enquêtes de mobilité	20
II Vers une méthodologie d'enquête mixte	36
III Les caractéristiques des internautes français	43
IV Conclusion	49
Chapitre 2 : La mise en place d'un module web dans l'enquête ménages déplacements de Lyon	51
I Considérations méthodologiques	52
II La rédaction des questions	66
III Administration du questionnaire en ligne	74
IV Conclusion	81
Chapitre 3 : Premiers résultats sur le comportement de réponse des internautes	83
I Un taux d'échec global à l'enquête en face-à-face non négligeable	84
II Une expérience encourageante	89
III Analyse de la non-réponse	93
IV Caractéristiques des connexions web	99
V Conclusion	103

Conclusion de la partie I	106
II Analyse comparative des échantillons web et face-à-face : quels enseignements ?	107
Introduction de la partie II	109
Chapitre 4 : Analyse comparative des répondants web et face-à-face	110
I Rappels théoriques sur les tests statistiques	112
II Qui sont les non-répondants à l'enquête en face-à-face ?	116
III Les ménages internautes : quelles spécificités ?	119
IV Les répondants web : quel profil ?	129
V Conclusion	134
Chapitre 5 : Analyse comparative de la mobilité	136
I Habitudes d'utilisation des modes de transport en semaine	138
II Caractéristiques des déplacements quotidiens	141
III Spécificités du web en termes de sorties quotidiennes	152
IV Conclusion	157
Chapitre 6 : Le redressement de l'échantillon en face-à-face : quels impacts sur la mobilité ?	159
I Les aspects théoriques du calage sur marges	160
II Application du calage sur marges à l'échantillon en face-à-face	162
III Des comportements de mobilité spécifiques	167
IV Conclusion	181
Conclusion de la partie II	185
III La comparabilité des données : enjeux et méthodes de correction	187
Introduction de la partie III	189
Chapitre 7 : Impact du mode d'enquête sur la mesure des comportements de mobilité	190
I La formalisation du problème	192
II Le biais de sélection de l'échantillon	194
III Modèle explicatif ne tenant pas compte du biais de sélection	206
IV Modèle explicatif incluant le biais de sélection	223
V Quantification de l'impact du mode d'enquête sur la mobilité	238
VI Conclusion	240

Chapitre 8 : Intérêt du modèle "Hurdle" pour la compréhension des comportements de mobilité	242
I Pour poser le problème	244
II Formalisation économétrique	246
III Application à l'équation de mobilité quotidienne	251
IV Conclusion	265
Conclusion de la partie III	270
Conclusion générale : le web, un média d'avenir pour les enquêtes ménages déplacements	271
I Portée et limites des choix méthodologiques	272
II Les principaux enseignements de l'approche empirique	275
III L'intérêt de l'approche économétrique	279
IV Les perspectives de recherche	282
Bibliographie	285
Annexes	309
I Evolution du nombre d'internautes en France	311
II Evolution des inégalités d'accès aux nouvelles technologies	312
III Guide d'entretien semi-directif	313
IV Verbatims	315
V Lettre avis envoyée aux non-répondants à l'enquête standard	318
VI Questionnaire web	319
VII Vagues de courriers web	333
VIII Répartition des répondants selon la réponse à la question : 'Souhaitez-vous communiquer votre adresse e-mail ?'	334
IX Répartition des personnes de 10 ans et plus par enquête et par caractéristique socio-économique	335
X Comparaison des marges dans l'échantillon et dans la population	336
XI Répartition des individus par enquête après redressement de l'échantillon face-à-face	337
XII Espérance d'une loi tronquée	338
XIII Statistiques descriptives - personnes mobiles de 18 ans et plus	339
Lexique	342
Table des figures	344
Liste des tableaux	349
Table des matières	353

Introduction générale : un renouvellement des besoins sur les données de mobilité

"All economically and technically advanced societies run on the same fuel. That fuel is information. The organizations and institutions within which we live and work have an unquenchable thirst for data [...] Survey research is one way and often the most effective and dependable way to obtain this information."
(Alreck et Settle, 1995).

Selon le Certu (1998), les enquêtes ménages déplacements sont indispensables pour connaître avec précision les comportements de mobilité de la population, ainsi que pour élaborer et évaluer les politiques de transport. Les modèles n'étant jamais meilleurs que les données à partir desquelles ils sont estimés (Stopher et Greaves, 2006), la qualité des données et des méthodes de collecte est très importante. Face aux changements organisationnels et culturels de la société, à l'exigence accrue des décideurs, au développement des nouvelles technologies, à la dérive des coûts des modes de recueil traditionnels et à la connaissance des facteurs qui viennent biaiser les résultats des études, les méthodes d'enquête évoluent.

I Analyse de la mobilité : enjeux et état des lieux

Les questions liées aux déplacements et aux modes de transport sont, dans le contexte actuel, cruciales, et Orfeuil *et al.* (2000) insiste sur le *"caractère essentiel du système de mobilité dans la définition des politiques de transport, à l'échelle des déplacements quotidiens"*. Nous allons d'abord définir le concept de mobilité (section I.1), avant de présenter la méthode traditionnellement utilisée, en France, pour l'estimer (section I.2).

I.1 Les composantes de la mobilité

Pendant de nombreuses années, on a caractérisé les comportements de déplacement des individus dans l'espace par un terme qui renvoie à une logique collective, celui de migration. Puis, l'individualisation croissante des pratiques a suscité l'utilisation d'un terme plus générique, emprunté aux sciences sociales, celui de mobilité (Massot et Orfeuil, 2008).

Les définitions de ce terme varient, selon les pays et les enquêtes. La mobilité urbaine concerne traditionnellement l'ensemble des déplacements dont l'origine et la destination appartiennent à une même agglomération. Elle est assez restrictive, car fondée sur les déplacements des seuls citadins dans leur ville (Orfeuil *et al.*, 2000), et n'est qu'une composante de la mobilité locale. La mobilité locale, concept défini dans l'enquête nationale transport, concerne l'ensemble des déplacements dont l'origine et la destination sont à moins de 100 kilomètres du domicile. Elle fait partie de la mobilité globale, celle-ci incluant également la mobilité de longue distance (voyages supérieurs à 100 km du lieu de résidence). Dans un récent ouvrage, Kaufmann (2008) propose une autre typologie de la mobilité. La mobilité résidentielle concerne les changements de domicile, selon un parcours qui se construit au fil du cycle de vie des ménages. Les changements d'établissement scolaire, de vie professionnelle, la cohabitation, la naissance des enfants ou le départ à la retraite sont des moments de transition dans le parcours de vie qui modifient les pratiques de mobilité. La mobilité quotidienne recense les déplacements réalisés à l'intérieur de l'aire urbaine, dans le cadre de la vie quotidienne. Enfin, la mobilité hybride regroupe les pendularités de longue distance (régions avec une organisation spatiale multi-polaire) et la bi-résidentialité (résidence dans deux logements distincts). Ces trois composantes de la mobilité sont liées, puisque le passage d'une phase à l'autre du cycle de vie familial modifie les programmes d'activité et les pratiques modales. Selon le même auteur, les activités de la vie quotidienne sont articulées autour de quatre grandes sphères, structurées par la rémunération et l'obligation. Durant la journée, l'individu passe d'une sphère à l'autre, et à chacune de ces sphères est associée une mobilité quotidienne (tableau 1).

	Activité rémunérée	Activité non rémunérée
Temps contraint	Sphère du travail	Sphère domestique
Temps non contraint	Sphère de l'engagement	Sphère du temps libre

TAB. 1 – Les quatre principales sphères d'activité
Source : Kaufmann (1999)

I.2 L'Enquête Ménages Déplacements : une méthodologie éprouvée

Les enquêtes ménages déplacements, réalisées par le Certu, visent à obtenir une image de la mobilité quotidienne et de son évolution. La méthodologie repose sur le principe suivant : recueillir les déplacements de l'ensemble des membres du ménage pendant un jour ouvrable. Ne considérant que les dépla-

cements internes à l'aire urbaine, elles s'intéressent donc à la mobilité quotidienne des habitants ¹. Ces enquêtes constituent un outil privilégié, largement utilisé par les collectivités et les autorités organisatrices des transports urbains en charge de la mobilité sur le territoire. La stabilité de la méthode depuis les années 70 autorise des comparaisons temporelles et spatiales.

Certaines critiques sont régulièrement formulées à leur rencontre. Ainsi, ces enquêtes restent conçues dans un but de comptabilisation des déplacements dans le temps et dans l'espace, et se prêtent difficilement à une analyse des activités. Vouloir identifier les modes complexes d'organisation quotidienne et les formes de mobilité qui en découlent obligerait à repenser les catégories traditionnellement utilisées pour le recueil des déplacements. Par ailleurs, le rapport à l'espace et au temps se transforme, et le vécu du temps de déplacement change. Il passe d'un temps de liaison à un temps social (travail dans le train, 'vivre ensemble' dans la voiture familiale pour les déplacements de loisirs...). Les interviews portent sur un jour unique de la semaine, pas nécessairement représentatif des emplois du temps individuels. Des actifs peuvent en effet être en congé ou en RTT le jour de référence, même s'il est légitime de penser que la part des personnes en congé le jour de l'enquête correspond à la part moyenne de personnes en congé un jour de semaine quelconque. La représentativité n'est donc établie qu'au niveau de l'échantillon (André-Poyaud *et al.*, 2008). Enfin, la notion de ménage est en constante évolution (familles recomposées, jeunes adultes vivant au domicile des parents en toute indépendance...). Les comportements sont donc parfois inadaptés au recueil des déplacements, tel que ce terme a été conçu il y a plus de trente ans, notamment en ce qui concerne la cohérence des déplacements entre les membres du ménage.

II De nouvelles exigences méthodologiques

Les enquêtes ménages déplacements sont une interface privilégiée d'échanges d'informations entre les usagers et les planificateurs des systèmes de transport ². Leur principal objectif est d'appréhender les comportements de mobilité d'une population donnée. Selon Stopher et Metcalf (1996) et Certu (1998), la collecte de données sur les comportements de déplacement permet aux décideurs de mieux comprendre l'utilisation des infrastructures de transport par les ménages et d'effectuer des prévisions sur leur fréquentation future. La finalité des enquêtes de mobilité a deux dimensions : l'alimentation de modèles ³ de prévision et la recherche de facteurs explicatifs des comportements. De nombreuses recherches mettent en évidence l'influence d'un certain nombre de facteurs sur

¹Dans ce cadre, les mobilités hybrides (selon la typologie de Kaufmann (2008) en augmentation ces dernières années sont éliminées, car considérées comme exceptionnelles et sortant du périmètre d'enquête.

²Se référer à Ressource System Group (2002) pour une présentation des différents types d'enquêtes transport : choix modal, études origine-destination des déplacements, études de transit ...

³Un modèle est une représentation simplifiée de la réalité, destiné à comprendre la formation de la demande et à prévoir son évolution (Guitton, 1964).

la mobilité, qu'il s'agisse de l'analyse des déterminants du choix modal ou de l'influence de la tarification (Bonnell et Le Nir, 1998). Par ailleurs, depuis les années 50, les données des enquêtes déplacements alimentent les modèles de planification stratégique et toute prise de décision dans le domaine sensible des transports (création d'infrastructure, programme de maintenance...) fait l'objet d'une évaluation a priori. La pertinence des politiques menées dépend pour une large part de la qualité des données utilisées pour évaluer les différents scénarios, et certains affirment qu'il est préférable de se passer de donner plutôt que d'utiliser des ressources peu fiables (Ampt, 1997). C'est pourquoi la collectivité doit compter sur des outils performants et des méthodes standardisées, qui garantissent la comparabilité des résultats (Chapleau *et al.*, 2001). Beaucoup de demandes nouvelles concernent les données d'enquêtes, en termes d'étendue, de qualité et de complétude (Stopher et Greaves, 2006). Les études deviennent plus complexes, avec le géocodage des adresses, des questionnaires basés sur les activités, des questions relatives au stationnement... Dans un contexte où les systèmes et technologies de l'information se déploient rapidement (puissance de calcul des ordinateurs, techniques pointues de modélisation...) et deviennent pratiquement des outils standard, des modèles de transport très exigeants sont développés. Ils requièrent des données précises sur les comportements de mobilité des ménages (comme le nombre moyen de déplacements quotidiens, le choix des modes de transport, les activités envisagées, la durée et la distance des déplacements et leur fréquence quotidienne), sur l'enchaînement des activités au cours de la journée et l'organisation de liens interpersonnels dans le ménage. Par ailleurs, les résultats des enquêtes ménages déplacements sont souvent utilisés pendant plusieurs années (10 ans pour Lyon) au cours desquelles la structure de l'agglomération évolue. Bien que les modèles de planification utilisent souvent les données géographiques à un niveau agrégé (zones fines), il est nécessaire que les répondants donnent les localisations exactes des lieux visités, pour permettre aux analystes de travailler sur des périmètres différents.

La communauté des transports a besoin d'une grande quantité de données "*on the quality and comprehensiveness of disaggregate travel data to support modern analytical procedures*" (Greaves, 2006). Or, les commanditaires d'études rencontrent de plus en plus de difficultés à recueillir rapidement des données de transport de bonne qualité. La quantité et la précision des données demandées impactent significativement la non-réponse partielle au questionnaire (Adler, 2003), ce qui rend difficile la constitution et la mise à jour de grandes bases de données représentatives sur les comportements de mobilité. Plus encore, les contraintes budgétaires limitent la constitution de larges échantillons pour les enquêtes en face-à-face, et compromettent l'analyse désagrégée des comportements de mobilité sur des zones géographiques restreintes.

III Des répondants de plus en plus réticents

Du côté des interviewés, les taux de réponse des enquêtes classiques tendent à décroître dans le temps (Atrostic et Burt, 1999; Keesling, 2002; Dillman,

2000). De nombreux facteurs, qui sont peu susceptibles d'évoluer dans le futur, expliquent le nombre croissant de non-réponses observé dans les enquêtes déplacements.

III.1 La difficulté de constituer une base de sondage

La résistance à la participation aux enquêtes est de plus en plus forte, et les premiers contacts sont très difficiles à obtenir. Depuis le développement du téléphone dans les années 80, ce premier contact ne se fait plus au domicile des interviewés. Ceci est encore plus vrai aujourd'hui, avec l'évolution des périmètres d'enquête et le développement des barrières physiques. Dans certaines zones géographiques, les ménages ouvrent leur porte sans trop de craintes. Dans d'autres, l'intervention humaine (concierge) ou technologique (digicodes, portail télécommandé, caméra devant les lotissements ...) permet de limiter l'intrusion d'étranger dans leur vie privée (Stopher et Greaves, 2006).

Le fort taux de pénétration du téléphone fixe a fait le succès des enquêtes téléphoniques. Mais la multiplication des enquêtes réalisées ces dernières années, notamment à but commercial, réduit leur niveau d'acceptation (Couper, 2000; Braunsberger *et al.*, 2007). Certains segments de répondants (les plus aisés) emploient des technologies nouvelles pour filtrer les appels (répondeurs, interphones, téléphones avec présentation du numéro entrant...) ⁴. Grâce au développement des nouvelles technologies, certains ménages ne possèdent plus de ligne fixe ⁵, mais uniquement des téléphones portables dont le numéro est parfois changeant, ou ont au contraire plusieurs lignes fixes à leur domicile, pour raccorder un fax ou un modem, ce qui rend infructueuses les tentatives de contact téléphonique (Christensen, 2004). La prise de contact est plus complexe et le coût de recrutement augmente le budget des études (Zmud, 2003; Cook *et al.*, 2000; Groves et Couper, 1998).

La diminution de la complétude des listes officielles d'habitants est problématique dans de nombreux pays. L'accroissement du nombre de possesseurs exclusifs de téléphones cellulaires, qui ne sont généralement pas inscrits dans les annuaires téléphoniques, et le nombre croissant de personnes qui refusent de voir leur numéro de téléphone publié dans ces mêmes registres compromettent la constitution d'échantillons représentatifs. Il est probable que la répartition géographique de ces individus ne soit pas uniforme sur le territoire. Couper (2002) met en évidence le caractère paradoxal du développement des nouvelles technologies de l'information et de la communication. La diffusion des téléphones mobiles rend leurs utilisateurs théoriquement joignables à tout moment. Cependant, ces derniers acceptent rarement d'être dérangés pour convenir d'un rendez-vous avec un enquêteur. Les personnes très occupées sont par définition difficiles à interroger, ce qui pose un gros problème de représentativité puisque le taux de réponse peut être lié au comportement de mobilité (Ettema *et al.*,

⁴Selon Zmud (2003), une plus forte proportion de répondeurs et de dispositifs permettent de filtrer les appels chez les individus jeunes, diplômés et possédant d'importantes ressources.

⁵Selon De La Godelinai et Mercier (1992), les ménages qui ne disposent pas d'un abonnement téléphonique ont un profil particulier.

1996).

Le problème de la constitution d'une base de sondage représentative reste valable pour les enquêtes en face-à-face, puisqu'il n'est pas possible d'utiliser les bases du recensement pour sélectionner des ménages.

III.2 Une certaine lassitude des enquêtes

Une certaine lassitude des enquêtes, associée à l'appréhension de révéler des informations personnelles, tend à accroître le taux de refus. Les individus semblent en effet moins disponibles et ont peu envie de participer aux enquêtes. Une première explication peut être l'abondance des démarches de sondage et d'enquêtes commerciales. L'autre est liée au niveau perçu de difficulté, que les individus associent à une enquête à laquelle on leur demande de répondre⁶, ce que Ampt (2003) qualifie de 'Respondent burden'. En jouant sur les éléments qui affectent cette perception de difficulté, variable selon le moment de la sollicitation et les individus concernés, il est possible d'augmenter le taux et la qualité des réponses. La première source de difficultés est physique et culturelle (illétrisme, handicap...) et échappe souvent aux concepteurs de l'enquête. D'autres sont plus faciles à anticiper et à éviter, comme la difficulté de compréhension des questions, d'accès et de manipulation de l'outil de recueil. Les difficultés émotionnelles sont en revanche peu présentes dans les enquêtes de mobilité, par définition factuelles.

La pénibilité d'une enquête peut aussi être mesurée par le temps et les coûts auxquels doit faire face l'individu pour y répondre. Sont inclus les contacts préalables à l'enquête, le temps passé à prendre connaissance du questionnaire et de ses instructions, à rechercher l'information demandée et à la transmettre (saisie des réponses sur le web, mise sous pli pour une enquête postale...). Les questionnaires se complexifient, pour recueillir davantage de données, et la pénibilité augmente (longueur, effort de mémoire...) (Bradley, 2004). Ampt (2000) montre que cette pénibilité est relative et varie selon les individus. Elle peut être minimisée si le thème de l'étude est important, si l'enquêté peut choisir le moment opportun pour répondre à l'enquête, si le questionnaire proposé est simple et ergonomique et si on fait appel à son altruisme. Un autre moyen de réduire la lassitude des enquêtés est d'offrir un mode de réponse alternatif, au risque de se heurter au problème de la comparabilité des données.

III.3 Des répondants au profil particulier

La complexité du recueil de données de mobilité et le coût associé à ces études ne permet pas de recenser les déplacements de l'ensemble des habitants d'une agglomération. Un sondage est réalisé, auprès d'un échantillon jugé représentatif de la population, dont le taux de réponse et le profil des répondants

⁶Les questions que se posent les personnes interrogées sont notamment : 'est-ce le bon moment pour répondre?', 'cette étude est-elle importante pour moi?', 'existe-t-il des pressions extérieures qui m'incitent à répondre au questionnaire?', 'ai-je des difficultés physiques, intellectuelles ou émotionnelles à le faire?', 'me sentirai-je mieux après avoir répondu?'

vont conditionner la qualité (Bonnell et Le Nir, 1998). Les non-réponses représentent une large part de l'échantillon de l'enquête ménages déplacements de Lyon (47% en 2006 contre 35% en 1994-1995, taux qui était déjà en croissance par rapport à l'enquête précédente (1985-1986)). La non-réponse est un concept complexe, qui a deux sources principales : la difficulté à contacter les répondants et la faible volonté des individus à répondre au questionnaire. Les facteurs explicatifs de la non-réponse généralement mis en avant sont le sujet de l'étude, son contexte, le design et l'administration de l'enquête, les caractéristiques des répondants... La difficulté vient de l'interaction de ces divers éléments (Atrostic et Burt, 1999).

De nombreuses techniques existent pour tenter de limiter la non-réponse. L'information préalable et les relances auprès des enquêtés sont couramment utilisées, de même que la formation et l'encadrement des enquêteurs. Une autre voie concerne la réduction de la pénibilité de l'étude, en diminuant sa durée et en améliorant l'ergonomie du questionnaire (Ampt, 2000). Malgré l'intérêt indéniable de ces efforts, les participants aux enquêtes se font rares et cette propension à la non-réponse tend à diminuer la confiance que l'on peut accorder aux résultats des enquêtes en termes de représentativité de la population étudiée (Cobanoglu *et al.*, 2001). Selon Dillman et Christian (2005), les non-répondants sont moins éduqués, plus masculins et plus jeunes que les répondants, et aucun mode ne semble performant pour obtenir la participation des personnes peu éduquées. De Heer et Moritz (2000) et Richardson (2003) ont montré que les non-répondants sont davantage des ménages disposant de revenus très élevés ou faibles, des individus hyper mobiles ou au contraire peu mobiles, des jeunes célibataires ou des personnes très âgées, des non-utilisateurs de la voiture particulière, des citadins et des ménages avec enfants. Kalfs et Van Evert (2003) rajoutent que les utilisateurs de transport en commun, et les personnes ayant des schémas d'activité simples ou au contraire relativement complexes ont tendance à moins répondre aux enquêtes déplacements. Zmud (2003) montre que les caractéristiques des déplacements en termes de modes et de motifs varient entre les répondants et les non-répondants. Par ailleurs, l'auteur affirme qu'il existe des différences sur les données socio-économiques entre les répondants et les non-répondants, qui impactent directement les comportements de mobilité ⁷.

Si la non-réponse n'est pas aléatoire, alors certains profils de répondants vont être sur ou sous-représentés dans l'enquête, et générer un biais important. C'est notamment pour tenter de réduire ces biais que les méthodes de redressement se sont développées ⁸. Elles conduisent toujours à postuler que les non-répondants ayant certaines caractéristiques socio-économiques se comportent

⁷Par exemple, les jeunes (moins de 25 ans) sont sur-représentés parmi les répondants, contrairement aux 25-44 ans, dont l'activité professionnelle ne laisse que peu de temps pour répondre aux enquêtes.

⁸La méthode de redressement communément utilisée est la postratification, qui ajuste les poids des répondants pour les rendre égaux à ceux de la population. Les variables socio-démographiques supposées fortement corrélées à la variable d'intérêt sont souvent utilisées pour définir les poids de la population, de façon à augmenter la représentativité des groupes qui ont moins de chance de répondre à l'enquête (Baines *et al.*, 2007).

comme les répondants ayant les mêmes caractéristiques. De nombreux travaux permettent toutefois de douter de la validité de cette hypothèse (Richardson et Ampt (1994); Richardson (2003); Murakami (2004)) et il est légitime de penser que la non-réponse pèse sur la qualité des enquêtes déplacements. Au final, certains segments de voyageurs manquent dans les résultats, comme ceux au style de vie 'pressé', qui n'ont pas de temps pour répondre. Beaucoup de ménages qui ne répondent pas voyagent plus que la moyenne, ou sont de plus grand ménages. Ils sont difficiles à joindre et perçoivent les enquêtes comme pénibles, puisque le nombre de déplacements à déclarer est élevé (NCHRP, 2006). Une autre solution explorée pour tenter de réduire la non-réponse⁹ consiste à laisser le choix aux répondants de remplir le questionnaire à un moment et par un média qui leur conviennent (Stopher *et al.*, 2004b). Les protocoles d'enquêtes mixtes se développent depuis quelques années, et recourent à présent aux nouvelles technologies telles que le web ou le GPS.

III.4 Une précision des données remise en cause

La capacité des individus à rapporter avec exactitude des informations précises sur les déplacements effectués durant un laps de temps relativement long (généralement une journée entière dans les enquêtes de mobilité) est remise en cause dans de nombreux travaux (Stopher *et al.*, 2007). Il semble que le nombre de déplacements soit souvent sous-estimé par les répondants. Wolf *et al.* (2003) expliquent ce constat par la durée de l'étude, le manque de mémoire des individus et la sélectivité de leur déclaration, les petits déplacements à pied ou à vélo étant davantage omis. Ces informations manquantes ou erronées peuvent réduire l'utilité d'une enquête, puisque les déplacements omis, volontairement ou pas, impactent négativement le niveau de mobilité et peuvent affecter la distribution géographique des lieux visités. L'analyste ne sait pas a priori si les données manquantes ont une distribution similaire ou différente de celle des données renseignées (Adler, 2003).

Dans une enquête déplacements menée au Kansas en 2004, on a pu comparer les déplacements réellement effectués en voiture par les enquêtés, à l'aide d'un GPS, aux déplacements déclarés. Globalement, on remarque une sous-estimation du nombre de déplacements réalisés, essentiellement des petits arrêts au cours de déplacements plus longs. Interrogés au sujet de ces écarts, les répondants ont dit ne pas avoir considéré ces déplacements comme importants, ou tout simplement les avoir oubliés. Il s'agissait essentiellement de la dépose/récupération de passagers, d'arrêts pour prendre du carburant ou déposer une lettre à la poste... d'une durée moyenne de 5,7 minutes. Cette

⁹Nous excluons ici l'incitation monétaire, fréquemment utilisée dans les pays anglo-saxons, à l'inverse de ce qui se pratique en France. L'étude de la littérature suggère pourtant que la promesse d'un cadeau ou d'une rémunération est le moyen le plus efficace pour encourager la coopération des individus sélectionnés pour répondre à une enquête (Tooley (1996) et Ressource System Group (2002)). Mais aucun résultat ne confirme cet effet pour les enquêtes web (Alsnih, 2004; O'Neil et Penrod, 2001). Kalfs et Van Evert (2003) et Andrews *et al.* (2003) précisent que si les incitations monétaires peuvent convaincre des groupes traditionnellement réticents, l'impact sur la qualité des données n'est pas encore connu.

sous-déclaration variait selon les caractéristiques socio-économiques des ménages, les couples bi-actifs et bi-motorisés aux revenus moyens sous-estimant le plus leurs déplacements (Murakami, 2004). Par ailleurs, si les distances sont généralement bien retranscrites, les répondants ont tendance à arrondir les heures de début et de fin des déplacements, ce qui change leur durée (Rietveld, 2002).

Les enquêtes de mobilité requièrent des informations détaillées. Certains répondants ne sont pas capables de décrire les lieux fréquentés à l'aide du questionnaire, ou refusent de livrer l'information, jugée trop personnelle ou demandant un effort de mémoire trop important. Certaines fois, il n'est pas possible d'identifier précisément les origines et destinations des déplacements, en fonction des informations fournies. Il peut donc exister un écart significatif entre les réponses données par les enquêtés et leurs comportements de mobilité réels (Ressource System Group, 2002). Le problème est plus aigu dans les enquêtes auto-administrées, puisqu'aucun interviewer n'est présent pour relancer le répondant et l'inciter à faire un effort de mémoire (Stopher *et al.*, 2007). Les informations collectées ne sont donc pas toujours de très bonne qualité ¹⁰.

IV Le potentiel des nouvelles technologies

L'augmentation de la non-réponse et la complexification des enquêtes sont peu compatibles avec la qualité des données souhaitées. Aujourd'hui, de nombreux professionnels rencontrent des difficultés pour mener des enquêtes de mobilité représentatives des populations étudiées. La pénibilité du questionnaire, associée aux préoccupations croissantes des individus sur le respect de leur vie privée rend les modes d'enquête traditionnels (téléphonique et postal) moins efficaces. Par ailleurs, les enquêtes en face-à-face, bien que souvent de bonne qualité, ont un coût très élevé que les commanditaires d'études ne sont pas toujours prêts à payer.

IV.1 Une évolution des modes de recueil de données

Les interviews en face-à-face étaient très populaires dans les années 50 et 60. Les enquêtes téléphoniques les ont peu à peu supplantées dès les années 70, atteignant leur apogée dans les années 90. Les combinaisons du téléphone et face-à-face se sont alors développées (enquête réalisée en face-à-face et suivi effectué par téléphone). Les outils disponibles pour recueillir des données d'enquête sont en perpétuelle évolution, suivant celle des technologies de l'information (Couper, 2002). Les progrès de l'informatique des 30 dernières années ont permis les enquêtes assistées par ordinateur et plus récemment les enquêtes web (De Leeuw, 2005). Les méthodes assistées par ordinateur complètent depuis plusieurs années les modes de recueil traditionnels. Il ne s'agit pas de

¹⁰Dans les enquêtes déplacements réalisées en face-à-face en Australie, Stopher et Stecher (2005) note une sous-estimation des déplacements de 7 à 12% par rapport au GPS. Les répondants ont également des difficultés à donner des réponses précises sur les lieux et les temps de parcours.

nouveaux médias de recueil, mais d'une aide apportée par l'ordinateur pour améliorer l'efficacité des enquêtes et la qualité des données (Wermuth *et al.*, 2001; Couper, 2002). L'enquêteur lit les questions au répondant, dans l'ordre où elles apparaissent à l'écran, et saisit les réponses en temps réel. Le Computer Assisted Telephone Interviewing (CATI), apparu dans les années 70, est le plus ancien. Il a contribué au succès des enquêtes téléphoniques, en gérant notamment la sélection des individus par une numérotation aléatoire des deux derniers chiffres des numéros de téléphones. Le Computer Assisted Personal Interviewing (CAPI) facilite depuis la fin des années 80 le travail des enquêteurs à domicile, qui peuvent montrer des éléments multimédias aux répondants, et celui des gestionnaires d'enquête, puisque le processus de saisie et de codage des réponses est automatisé. Depuis une dizaine d'années, le développement des nouvelles technologies permet de développer rapidement de nouveaux modes de recueil de données (Global Positioning System, téléphone mobile, internet ...) ¹¹, ou offrent des opportunités intéressantes pour améliorer la précision des réponses (géocodage). La technologie a eu un impact positif sur les enquêtes de mobilité, en permettant d'obtenir à budget constant des données de meilleure qualité (Chapleau *et al.*, 2001). Par ailleurs, des études ont montré que les répondants apprécient les enquêtes assistées par ordinateur (Couper et Rowe, 1996; Beckenbach, 1995).

IV.2 L'essor des enquêtes web

Le développement du web et des navigateurs génère la standardisation des interfaces graphiques, ce qui rend possible la réalisation d'études assistées par ordinateur : CAWI ou Computer Assisted Web Interview (ce qui semble être une tautologie (De Leeuw, 2005)). Les interfaces utilisateurs sont riches et dynamiques. Les traitements lourds peuvent être dirigés vers de puissants groupes de serveurs. Ils supportent le traitement de questionnaires complexes, l'accès et la mise à jour de bases de données conséquentes, des analyses statistiques et la création d'images graphiques en temps réel et d'une manière transparente pour le répondant (Stanton, 1998). Les enquêtes web se sont développées à l'origine pour palier à certaines limites des enquêtes téléphoniques et postales, arrivées à leur phase de maturité, et permettre d'accéder à un grand nombre d'individus (Cobanoglu *et al.*, 2001). Elles ont gagné en popularité au début des années 90, et semblent avoir depuis un réel potentiel comme nouveau moyen de collecte de données. Concernant le taux de réponse, certains auteurs avancent que la non-réponse augmente lorsqu'un mode est automatisé (Atrostic et Burt, 1999). Il est cependant envisageable qu'un questionnaire web bien construit facilite le travail de l'enquêteur et génère un taux de retour équivalent à celui des méthodes plus traditionnelles. Si l'accès aux nouvelles technologies augmente rapidement, certains groupes ne sont pas représentés (bas revenus, faible éducation...), et les réponses aux questionnaires web ne sont donc pas

¹¹A l'origine, le GPS a été utilisé dans les enquêtes de mobilité pour vérifier la précision des informations sur les déplacements enregistrées par des modes de recueil traditionnels (Wolf *et al.*, 2003).

toujours représentatives.

Même si dans le domaine des transports les enquêtes assistées par ordinateurs font encore figure d'exception, il est légitime de se demander jusqu'à quel point les nouvelles technologies peuvent remplacer ou suppléer les modes d'enquête traditionnels¹². Selon Huang (2006), il est nécessaire de réfléchir sur le rôle de l'outil. Les enquêtes web ne sont pas un simple avancement technologique par rapport aux enquêtes papier, mais un format totalement nouveau. La possibilité d'utiliser le multimédia donne des opportunités au concepteur d'enquêtes, notamment pour le design du questionnaire (Llieva et Healey, 2002).

V Le cadre de recherche

Face au déclin des taux de réponse des enquêtes ménages déplacements et à l'évolution des nouvelles technologies, il est nécessaire que les recherches sur les méthodes d'enquêtes explorent des alternatives aux modes de recueil traditionnels et travaillent sur la mise en place de protocoles d'enquête mixtes (Wilson, 2004). La combinaison du web et du face-à-face semble prometteuse, car elle permet de toucher suffisamment de répondants sans renoncer à la qualité des réponses, mais se heurte à certains problèmes, tels que l'inégalité d'accès à internet et la comparabilité des données. Bien que l'utilisation pragmatique des données d'enquête transport rende peu nombreuses les opportunités de financement de tels projets (Zmud, 2003), nous avons pu tester un protocole d'enquête mixte lors de l'Enquête Ménages Déplacements menée sur l'agglomération lyonnaise en 2005-2006 par le SYTRAL.

Dans le cadre d'un contrat de recherche co-financé par le PREDIT, le SYTRAL et la région Rhône-Alpes, le Laboratoire d'Economie des Transports a proposé de mener une enquête par le web auprès des non-répondants au questionnaire traditionnel, administré en face-à-face. L'objectif était d'accroître la précision des données, en limitant le nombre de non-réponses et en analysant des comportements jusqu'ici inconnus. Ceci représente un enjeu considérable pour les études de modélisation et de prospective menées dans les grandes agglomérations françaises, puisque les ménages qui refusent de répondre en face à face, ou qui restent injoignables, représentent une large part de l'échantillon (47% à Lyon en 2005-2006), en croissance par rapport à l'enquête précédente (33% 1985-1986). En revanche, l'inclusion de l'enquête en ligne dans le contrat du SYTRAL ne nous a pas permis de construire l'expérimentation en toute autonomie.

Une partie de cette recherche consiste à réfléchir sur la possible intégration du web dans les enquêtes de mobilité, à explorer plus largement la combinaison de plusieurs modes de collecte de données (face à face et web), à tester puis à proposer une méthodologie adaptée au recueil de données sur les déplacements quotidiens par internet. Un autre objectif est de décrire la population ayant répondu à l'enquête web, ainsi que caractériser ses comportements de

¹²Selon Stopher *et al.* (2004a), grouper Internet et le GPS est une piste intéressante pour les enquêtes ménages déplacements.

déplacements. Une attention particulière est portée à l'identification et à la caractérisation des principaux biais souvent rencontrés avec ce nouveau média (population plus jeune, plus masculine, au niveau d'éducation et aux revenus plus élevés qu'en moyenne...). Enfin, un lourd travail sur la comparabilité des données est nécessaire, aucune expérience sur la comparabilité des données obtenues par le web avec celles collectées en face à face n'ayant été relatée dans la littérature à ce jour, les seules expériences connues concernant la comparaison du web avec les enquêtes postales. Un des enjeux de cette thèse sera donc de mener une analyse comparative des comportements de mobilité des individus qui ont répondu à l'enquête par l'intermédiaire d'Internet, avec ceux soumis au questionnaire standard. Pour aller plus loin dans l'analyse, nous développons des techniques économétriques permettant, d'une part, de tenir compte des différences socio-économiques subsistant au sein des échantillons web et face-à-face et, d'autre part, d'identifier un effet média.

-

Afin de répondre à nos objectifs, nous avons organisé notre travail en trois parties, chacune incluant des éléments théoriques et des résultats empiriques, issus de l'enquête ménages déplacements menée à Lyon en 2006.

Dans la première partie, nous nous interrogeons sur le potentiel des enquêtes web dans le secteur des transports. Les atouts et les limites de ce nouveau média sont analysés, ainsi que sa possible intégration dans des protocoles d'enquêtes mixtes (**chapitre 1**). La considération de nouveaux modes d'enquête exige de mener une réflexion approfondie sur les objectifs visés, la population de référence et la rédaction du questionnaire. Internet est bien plus qu'une évolution technique permettant d'améliorer la version papier des enquêtes. Il s'agit d'un outil innovant qui présente de riches opportunités pour les concepteurs de questionnaires (Couper *et al.*, 2001). Nous présentons une revue de la littérature, concernant la réalisation de questionnaires sur internet, ainsi que la démarche suivie pour la création du questionnaire web de l'enquête ménages déplacements de Lyon (**chapitre 2**). Si le développement du web promet un renouveau des méthodes d'enquête, il reste à valider l'hypothèse que ce nouveau média soit adapté aux études de déplacements et permette de recueillir des données de mobilité détaillées. Le **chapitre 3** présente les premiers résultats de l'enquête web, en termes de taux et de comportements de réponse. Il est alors possible de faire une première évaluation de la qualité des données saisies en ligne.

Utiliser plusieurs modes de recueil n'est pas sans risque, mais il existe peu d'expériences, avec une méthodologie contrôlée, qui compare différents modes de recueil de données de déplacements. Avec la baisse du taux de réponse aux modes traditionnels et le développement des nouvelles technologies, la comparabilité devient une préoccupation croissante des enquêtes de mobilité (Bonnel, 2003). Dans la deuxième partie, nous adoptons une approche descriptive et analytique. Le principal objectif est de confronter, à un niveau agrégé, les données issues des échantillons web et face-à-face. Nous étudions les différences observées en termes de profil de répondants (**chapitre 4**) et de comportement

de mobilité (**chapitre 5**). Les principaux indicateurs utilisés pour le suivi de la mobilité sont : l'accès aux modes de transport (possession du permis, nombre de véhicules à libre disposition ...), l'intensité de la mobilité (nombre de déplacements par jour), l'attractivité des modes de transport (parts modales), la taille du territoire pratiqué (distance parcourue) et le budget temps consacré aux déplacements (Orfeuil *et al.*, 2000). Les résultats montrent qu'il existe des différences socio-économiques entre les répondants web et face-à-face, qui peuvent impacter leur niveau de mobilité quotidienne. A l'aide d'outils statistiques, nous redressons l'échantillon face-à-face, afin de réduire ces différences et d'obtenir des échantillons comparables (**chapitre 6**).

Dans la troisième partie, nous cherchons à modéliser les données, pour mettre en évidence de manière plus formalisée les facteurs explicatifs des différences constatées entre les deux échantillons. L'approche est économétrique et s'effectue à un niveau désagrégée. Dans un premier temps, nous établissons, par enquête, quels sont les facteurs explicatifs les plus significatifs pour rendre compte du niveau de mobilité des répondants. Une fois sélectionnés ces principaux facteurs, nous spécifions un modèle en deux étapes (Heckman), mieux adapté pour mettre en évidence l'existence d'un biais de sélection des individus, les amenant à refuser l'enquête en face-à-face et à répondre sur le web, et impactant le niveau de mobilité (**chapitre 7**). Les techniques économétriques standard permettent d'identifier, à partir d'un ensemble de variables explicatives, lesquelles ont une influence significative sur la mobilité. Grâce à des modèles plus spécifiques (Poisson Hurdle), nous pouvons distinguer, par enquête, les facteurs explicatifs de la décision de mobilité de ceux qui influencent le niveau de mobilité (**chapitre 8**). Les caractéristiques socio-économiques tels que le niveau d'éducation, le lieu de résidence, l'exercice d'une profession et la position dans le cycle de vie, entre autres critères, sont utilisés pour expliquer les comportements d'immobilité et le niveau de mobilité des individus ayant effectué au moins un déplacement durant la période de référence. Nous nous intéressons au final sur la pertinence du rapprochement des deux bases de données, web et face-à-face.

**

Première partie

La mise en ligne d'un questionnaire : fondements théoriques et empiriques

Introduction de la partie I

Les taux de participation aux enquêtes sont en nette diminution, et les enquêtes ménages déplacements suivent la même tendance. Même si certaines procédures permettent de réduire le taux de non-réponse et que les méthodes de redressement permettent de limiter le biais introduit par la non-réponse totale, il reste nécessaire de postuler que les comportements de mobilité des ménages qui refusent de participer à l'enquête sont identiques à ceux des ménages ayant des caractéristiques socio-économiques similaires. Or, de nombreux travaux (Bonnel, 2003; Ampt, 1997; Sammer, 1997) ont montré que les comportements étaient généralement corrélés au fait de ne pas répondre, ou au nombre de tentatives nécessaire pour atteindre un répondant. C'est pour essayer de limiter ce biais de non-réponse, et ainsi la représentativité statistique des résultats obtenus, que le Laboratoire d'Economie des Transports a proposé de réaliser une enquête par le web auprès des non-répondants à l'enquête ménages déplacements réalisée à Lyon en face-à-face en 2006.

Pour réaliser ce test, nous avons adapté le questionnaire papier, traditionnellement utilisé dans les enquêtes ménages en face-à-face, en tenant compte de plusieurs contraintes. D'une part, des contraintes méthodologiques, imposées par le Certu (1998) pour toute réalisation d'enquêtes ménages déplacements (thèmes, vocabulaire, ordre et formulation des questions...). D'autre part, des contraintes techniques, imposées par le cahier des charges de l'enquête de Lyon. Le web étant un module additionnel à l'enquête en face-à-face, nous n'avons pu choisir ni le prestataire en charge de l'administration de l'étude en ligne, ni le logiciel utilisé pour le développement du questionnaire.

L'objectif de cette partie est d'évaluer le potentiel du web pour les enquêtes ménages déplacements, principalement dans le cadre d'une méthodologie d'enquête mixte (Chapitre 1). Suivent des réflexions sur la construction du questionnaire web, les choix opérés par rapport à sa version papier et l'administration de l'enquête en ligne (Chapitre 2). Nous présenterons finalement les premiers résultats de l'option web de l'enquête ménages déplacements de Lyon, en termes de taux et de comportement de réponse (chapitre 3). Grâce à la méthodologie employée, il est en effet possible d'étudier la progression des internautes dans le questionnaire, et comprendre pourquoi certaines personnes ont préféré répondre en ligne plutôt que de participer à l'enquête face-à-face.

Chapitre 1 : Le potentiel du web pour les enquêtes ménages déplacements

"Technology by itself is not inherently good or bad. It is how technology is harnessed in the service of human endeavor that determines its effect."
(Couper, 2002).

Un des principaux enjeux de l'économie des transports est de comprendre comment les comportements de mobilité varient selon les caractéristiques socio-économiques des individus. Des études récurrentes, comme les enquêtes ménages déplacements, viennent alimenter de larges bases de données qui permettent aux analystes de rédiger des recommandations pertinentes. Avec la démocratisation de l'informatique et de l'accès à Internet, les enquêtes web sont promises à un rapide développement. Couramment utilisées dans certains domaines comme le marketing (historiquement, les premières enquêtes en ligne abouties sont apparues aux USA en 1997), il semble intéressant de s'interroger à présent sur la pertinence de ce nouveau média pour les enquêtes de mobilité, et de déterminer sous quelles conditions l'utilisation du web peut augmenter la précision des données produites, en tenant compte des non-répondants aux enquêtes traditionnelles. Si les enquêtes web présentent de nombreux atouts (faible coût, rapidité d'administration, personnalisation du questionnaire...), leurs principales limites résident essentiellement dans sa diffusion et son niveau d'appropriation. Le taux de pénétration d'internet dans la population française progresse régulièrement, mais ce média n'est pas encore accessible à tous et les internautes ont des caractéristiques sociodémographiques particulières, susceptibles d'impacter leurs comportements de mobilité. L'utilisation exclusive du web pour mener des enquêtes auprès de l'ensemble de la population ne permet donc pas une généralisation des résultats par les techniques d'inférence statistique.

La qualité des décisions en politique des transports dépend fortement de la qualité des informations disponibles. La difficulté croissante pour obtenir des données d'enquêtes représentatives de la population visée et la complexité croissante des données nécessaires à l'alimentation de modèles de plus en plus sophistiqués ne permettent généralement plus de recueillir toutes les données

selon une méthodologie unique. Une des solutions consiste aujourd'hui à intégrer le web dans un protocole d'enquête plus large. L'idée est de réussir à capter de nouveaux individus à l'aide d'un média attractif, tout en conservant la méthode de recueil de données traditionnellement utilisée dans les enquêtes ménages déplacements.

Nous revenons d'abord sur les apports du web comme méthode d'enquête, avant de nous intéresser aux limites de ce nouveau média (section I). Nous étudions plus spécifiquement le potentiel du web dans les protocoles d'enquêtes mixtes, ainsi que l'impact sur la qualité des données recueillies (section II). Enfin, l'analyse des caractéristiques sociodémographiques de la population d'internautes et de l'évolution du taux de pénétration d'internet dans les ménages français clôturent ce chapitre (section III).

I Apports et limites du web pour les enquêtes de mobilité

I.1 Une grande diversité d'enquêtes

Les enquêtes électroniques ont pris des formes très variées, suivant les développements des technologies de la communication¹³. Il est utile de distinguer les enquêtes conduites sur Internet, de celles diffusées par mail (message incluant des questions, ou e-mail avec lettre avis et questionnaire en pièces jointes). Les enquêtes diffusées par e-mail sont peu coûteuses, simples, rapides, personnalisées, et encouragent la réponse (Tse, 1998). Mais les messages peuvent être rapidement effacés, ne sont pas matérialisés ni anonymes et retiennent rarement l'attention des destinataires (Kittleson, 1995). Une enquête web est un questionnaire interactif parfois très complexe, auquel les individus peuvent répondre en se connectant sur internet (Stanton, 1998). Les personnes qui disposent d'un ordinateur, équipé d'un navigateur et d'une connexion internet, doivent se rendre sur un site web, compléter le questionnaire et renvoyer leurs réponses. La base de données est implantée sur un serveur et les données sont transmises électroniquement à l'analyste (Alvarez et Van Beselaere, 2003).

Les enquêtes web peuvent être très diverses, suivant le mode d'échantillonnage, l'accessibilité du questionnaire, sa présentation et les modes de recueil alternatifs éventuellement proposés. Leur qualité est par conséquent très inégale. Des classifications ont été données dans la littérature. Couper (2000), Bradley (1999) et Hojman *et al.* (2004) discriminent les enquêtes web selon le mode de recrutement des répondants (échantillon probabiliste ou non probabiliste), qui n'offre pas les mêmes possibilités de contrôle des données. Les méthodes probabilistes reposent sur deux principes : la nécessité de disposer

¹³Les premières enquêtes électroniques consistaient en une disquette envoyée aux répondants. Ces derniers introduisaient la disquette dans l'ordinateur pour lire le fichier et répondre au questionnaire, puis enregistraient leurs réponses et renvoyaient le support aux administrateurs de l'enquête (Couper et Nichols, 1998).

d'une base de sondage, c'est-à-dire d'une liste complète et actualisée des individus qui appartiennent à l'univers de l'étude, et une probabilité connue de désignation de ces unités de sondage. Ces méthodes ne garantissent pas la représentativité de l'échantillon, puisque les non-réponses peuvent menacer la valeur inférentielle de l'étude, mais il est possible d'évaluer la non-réponse et d'ajuster les données. Il y a deux moyens de mener des enquêtes probabilistes en ligne, en restreignant l'échantillon à ceux qui possèdent un accès internet ou en utilisant plusieurs modes de recueil de données, afin de toucher l'ensemble de la population (tableau 2). Dans la plupart des cas, les commanditaires d'études ne disposent pas de base de sondage et utilisent des méthodes empiriques, qui nécessitent certains arbitrages pour constituer l'échantillon. L'inférence statistique est impossible, bien que la représentativité de l'échantillon puisse être améliorée en s'efforçant de reproduire au mieux les conditions du hasard pour la sélection des répondants. Schillewaert *et al.* (1998) et Schaefer et Dillman (1998) passent en revue les différents moyens disponibles pour collecter des données sur le web et proposent quelques pistes d'amélioration des procédures de recrutement.

Alsnih (2005) considère l'accessibilité au questionnaire, qui peut être ouvert à tous, sécurisé par un code d'accès personnel ou caché sur un site ('pop-up'). Dillman et Bowker (2001) insistent sur la présentation, et distinguent les questionnaires interactifs (contenant une question par page) des questionnaires passifs (permettant de visualiser l'ensemble des questions sur une seule page, grâce à l'ascenseur)¹⁴. Dans les questionnaires interactifs, le répondant doit valider chaque réponse en cliquant sur un bouton, ce qui permet de récupérer les réponses sur le serveur au fur et à mesure qu'elles sont saisies, mais rend difficile le retour en arrière et quasi impossible la correction des réponses déjà validées (Alvarez et Van Beselaere, 2003). Si proposer plusieurs questions sur une même page est plus motivant pour l'enquêté, le temps de chargement des pages web est élevé et les biais de mesure plus importants, les répondants ayant tendance à cocher systématiquement la même modalité (Couper *et al.*, 2001). Dans un questionnaire passif, les réponses sont transmises à la fin de la saisie. Il existe donc un risque de perte totale des données, en cas de rupture de la connexion internet, mais les réponses sont facilement modifiables par l'internaute et les difficultés techniques moins prononcées.

Malhotra (1999) oppose la constitution d'un échantillon 'interne', où les répondants sont sélectionnés sur internet, à celle d'un échantillon 'externe', où les répondants ne sont pas recrutés sur le web (par exemple dans l'annuaire), mais sont invités à remplir un questionnaire en ligne. Enfin, des enquêtes sont exclusivement réalisées grâce à Internet, notamment lorsqu'il s'agit d'appréhender le comportement ou l'opinion d'une population d'internautes, alors que d'autres combinent plusieurs modes de recueil de données (Bonnell, 2004).

Il existe une multitude d'enquêtes web, qui se différencient par leur niveau de qualité (Couper, 2001). Le choix du type d'enquête et son implication sur la méthode d'échantillonnage dépend au final de l'objectif de l'étude. Dans

¹⁴Dans ce cas, le temps de réponse est plus court et les non-réponses partielles moins nombreuses (Gunn, 2002).

ce rapport, nous considérerons uniquement les enquêtes menées sur Internet, conformément à la méthodologie utilisée pour l'enquête ménages déplacements de Lyon. Les répondants web sont des ménages initialement recrutés de façon aléatoire, pour participer à l'enquête-en-face à face. Cependant, le questionnaire en ligne n'est proposé qu'à ceux qui refusent de recevoir un enquêteur à domicile ou qui restent injoignables. L'enquête web n'est donc pas représentative de la population.

Méthode	Type d'étude	Mode de recrutement
Méthodes non probabilistes	Sondage 'divertissement'	Recueil d'opinion, non représentatif
	Enquêtes ouvertes et non restreintes	Participation à une étude offerte sur plusieurs sites
	Panels de volontaires	Individus recrutés pour répondre à des études ciblées
Méthodes probabilistes	Enquête systématique	Invitation de tous les visiteurs d'un site web
	Enquête par sondage basé sur des listes	Individus recrutés à partir d'une liste d'e-mail
	Enquêtes mixtes	Possibilité de répondre en ligne ou par d'autres modes
	Panels d'internautes prérecrutés	Recrutement probabiliste et proposition d'étude en ligne
	Echantillon aléatoire, accès web fourni	Recrutement probabiliste et étude en ligne (équipement fourni)

TAB. 2 – Les différentes méthodes de sélection de l'échantillon pour réaliser une enquête web

Source : D'après Couper (2000)

I.2 Pourquoi un tel engouement pour les enquêtes web ?

Les enquêtes déplacements sont généralement longues et coûteuses à mettre en oeuvre. Il s'agit le plus souvent d'interroger un large échantillon de ménages, ou d'individus, au sujet de leurs déplacements quotidiens, à l'aide d'un questionnaire relativement lourd et administré en France en face-à-face, à tous les membres du foyer. Les implications en termes de disponibilité des répondants et du nombre d'enquêteurs sont importantes. L'augmentation du taux de non-réponse a apporté de nombreux questionnements sur cette méthodologie, les comportements de mobilité des non-répondants ne pouvant être considérés comme identiques à ceux des individus interrogés. Dans ce contexte, l'utilisation d'internet est apparue comme une voie intéressante. L'absence de registre officiel d'adresses e-mail et la difficulté d'administration des enquêtes par e-mail (problèmes de téléchargement, d'affichage ...) ont laissé la place aux enquêtes web (Dillman *et al.*, 1998a).

I.2.1 Un coût marginal quasiment nul

Le coût des études est directement lié à la main d'oeuvre employée, c'est-à-dire au nombre d'enquêteurs (Bonnell, 2003). Les coûts générés par une enquête web ne sont donc pas très élevés par rapport à une enquête traditionnelle en face-à-face (Cleland, 1996). Il n'est pas nécessaire de recruter des enquêteurs, ni de dupliquer les questionnaires sur un support papier. Les coûts engendrés par l'affranchissement des lettres, les communications téléphoniques ou les déplacements des enquêteurs au domicile des interviewés sont inexistantes. Il est également possible de modifier aléatoirement l'ordre des questions, sans supporter les coûts relatifs à la création de documents papier. De plus, le codage et la saisie des réponses se font par l'enquêté, directement sur l'ordinateur, sans recours à des opérateurs de saisie. Il en est de même pour le tri, la classification et la vérification des questionnaires. Le stockage physique des questionnaires n'est plus nécessaire, ce qui représente encore des économies en termes d'espace et de manutention (Keesling, 2002). Enfin, les relances, qui sont un moyen efficace de réduire la non-réponse, peuvent être automatisées sans générer de dépenses importantes. Il faut relativiser les avantages financiers des enquêtes web, et garder en mémoire que leur coût varie selon le nombre de relances effectuées auprès des non-répondants et le mode de recrutement mis en place. S'il existe une base de sondage disponible (liste d'adresses e-mail), alors le coût de mise en oeuvre de l'enquête web est marginal. En revanche, si la base d'échantillonnage doit être construite, le coût de réalisation de l'enquête web est quasi-identique à celui des autres modes.

L'utilisation de ce nouveau média permet ainsi d'augmenter le nombre de contacts à moindre frais, et de diminuer le coût moyen par contact, le coût des enquêtes web augmentant de façon non linéaire avec la taille de l'échantillon (Crawford *et al.*, 2002). Les coûts variables sont faibles et correspondent essentiellement au support technique mis à la disposition des répondants. Les coûts fixes, non négligeables, varient selon la longueur et la complexité de la programmation du questionnaire. La trame peut toutefois être stockée, puis réutilisée ultérieurement. Le web est actuellement le média le moins coûteux¹⁵, en particulier pour de gros échantillons (Schonlau *et al.* (2001), Couper (2001) et Anderson *et al.* (2006)), et le développement de logiciels spécifiques ces dernières années devrait contribuer à réduire ces coûts fixes (Bonnell, 2004).

I.2.2 L'interactivité du questionnaire

L'interactivité du questionnaire, traditionnellement utilisée dans les enquêtes administrées par ordinateur au téléphone ou en face-à-face (CATI et CAPI), est également possible dans des enquêtes auto-administrées sur le web.

¹⁵En Australie, environ 350\$ sont dépensés par ménage répondant en face-à-face à une enquête de mobilité (Stopher et Greaves, 2006). Selon Crawford *et al.* (2002), le coût pour une étude auprès de 3500 répondants est 40% plus cher par la poste que par le web. Une expérience menée par Braunsberger *et al.* (2007) montre que le coût d'une enquête web est environ égal à 30% de celui d'une enquête téléphonique. Enfin, Cobanoglu *et al.* (2001) mettent en évidence que le web est moins cher que le fax et la poste quand le nombre de personnes dépasse les 200, d'autant plus si on ajoute le coût de codage des réponses.

L'assistance fournie par l'ordinateur décharge l'enquêté du travail de sélection des questions et de décryptage de la structure du questionnaire, généralement nécessaire dans le cas d'enquêtes auto-administrées. Il s'agit d'un mode dynamique, qui sollicite le répondant en cas d'incohérence des réponses, renseigne sur le temps restant pour terminer le questionnaire et fournit des informations précises avant chaque question. Le web autorise une grande interactivité au niveau de la personnalisation des questions et des réponses. Grâce aux nombreux filtres, les individus ne se voient pas poser des questions qui ne les concernent pas (Pitkow et Recker, 1995); la clarté du questionnaire venant réduire leur niveau de frustration (Gunn, 2002; Alvarez et Van Beselaere, 2003). Il est ainsi possible de s'adapter à la situation de chaque répondant et de cibler le questionnaire sur les questions pertinentes au regard de cette situation. La personnalisation du contact est également un atout intéressant des enquêtes web. En ne posant que quelques questions jugées 'qualifiantes' aux répondants, avant de soumettre l'intégralité du questionnaire aux individus sélectionnés, le gestionnaire d'étude peut segmenter l'échantillon de façon très précise. Par ailleurs, le web offre la possibilité de rendre aléatoire l'ordre des questions, sans complexifier l'administration du questionnaire. Enfin, diverses avancées technologiques voient le jour, comme l'intégration d'animations sonores ou visuelles (photos, vidéos, couleurs) et l'ajustement d'échantillon en cours d'étude. En ce qui concerne plus précisément les enquêtes de mobilité, des programmes permettent de géocoder instantanément les adresses saisies en liant le questionnaire web à des bases de données externes (Kreitz, 2001).

I.2.3 Des données de qualité

Selon Zmud (2003), augmenter la qualité d'une enquête consiste à réduire la non-réponse totale et partielle, ainsi que l'erreur de mesure. Dans les enquêtes web, les risques d'erreur sont plus faibles, qu'il s'agisse des erreurs de saisie ou du biais que peuvent générer les tierces personnes sur les réponses de l'enquêté (Krosnick et Chang, 2001). Des contrôles automatiques et dynamiques tout au long du questionnaire garantissent la validité des réponses, en limitant les erreurs et les incohérences dans la saisie (Adler *et al.*, 2002). L'enquêté est relancé, afin de corriger les réponses détectées comme suspectes, avant de pouvoir progresser dans le questionnaire (Sarvis, 1991). Les questionnaires obtenus sont ainsi apurés et aucun traitement ad-hoc n'est nécessaire ¹⁶.

Le répondant, libéré de sa relation avec l'enquêteur et les autres membres du ménage, peut plus facilement se livrer. Les données saisies sont transcrites dans un fichier sans intervention humaine et ne souffrent pas du biais de l'interviewer. Les normes de comportement, les préoccupations des jugements et la désirabilité sociale sont donc réduites en ligne (Mc Daniel et Gates, 2004) ¹⁷.

¹⁶Certains programmes vont plus loin dans la personnalisation du contact et proposent des corrections au répondant (Franceschini, 2000).

¹⁷Le biais de désirabilité sociale se définit par la tendance pour une personne à donner des réponses selon les espérances culturelles perçues des autres. Ceci se produit surtout dans les interviews en face-à-face, puisque l'enquêteur est présent (Robinson *et al.*, 1991; De Leeuw et Van Der Zowen, 1992).

Des études montrent que les réponses fournies par internet sur des sujets sensibles ou des comportements peu socialement désirables sont plus franches que lorsque l'étude est menée par internet (Joinson, 1998; Tourangeau et Smith, 1996). Selon Cobanoglu *et al.* (2001), les enquêtes web donnent des réponses de meilleure qualité que les enquêtes menées par fax ou par voie postale. Les réponses fournies aux questions ouvertes sont plus complètes et plus pertinentes sur le web (Yun et Trumbo, 2000). Lozar Manfreda et Vehovar (2002a) montrent que la fréquence d'utilisation du web impacte positivement la qualité des réponses, en particulier pour les questions difficiles. Enfin, Stanton (1998) utilise la variance des items comme indicateur de la qualité des données d'enquête. Ses recherches mettent en évidence que le schéma de covariance interne des données web est similaire à celui des données collectées par des modes plus traditionnels.

I.2.4 Des données collectées rapidement

Les études en ligne sont simples à mener, qu'il s'agisse de la diffusion du questionnaire, des relances ou de la phase de suivi, car le mode de communication est quasiment instantané (Schmidt, 1997; Weible et Wallace, 1998). Le répondant entre directement ses réponses, le support informatique supprimant tout délai dû renvoi du questionnaire, par rapport à une enquête postale (Alsnih, 2004). La centralisation de l'enquête sur un serveur rend possible un suivi régulier des réponses et la mise en place d'ajustements en temps réel, le gestionnaire de l'étude pouvant accéder à tout moment et de façon sécurisée aux données collectées. De plus, les données saisies sont déjà sous format électronique, et sont accessibles 'immédiatement' par les analystes. Certains logiciels d'enquêtes en ligne permettent même de générer automatiquement des analyses statistiques simples pour visualiser rapidement les premiers résultats d'une étude. Le caractère dynamique du web permet de générer des statistiques intermédiaires et d'envoyer un retour aux répondants (Schmidt, 1997). La diffusion de l'enquête, la collecte des données et le traitement de l'information sont donc plus rapides que dans les modes traditionnels.

Les protocoles mixtes montrent que les enquêtes web génèrent des réponses plus rapides que les autres médias. Dans la plupart des cas, la différence de temps de réponse entre une enquête web (envoyée par mail) et une enquête postale est de 6 à 10 jours, expliquée pour beaucoup par l'efficacité du service postal (Cobanoglu *et al.*, 2001; Kwak et Radler, 2000; Schaefer et Dillman, 1998; Bates, 2001). Franceschini (2000) rapporte que plus de deux tiers des réponses web parviennent à l'analyste avant la réception d'un questionnaire par voie postale. Potoglou et Karanoglou (2008) précisent que l'avantage du web en termes de vitesse de réponse dépend également beaucoup du mode de recrutement des individus. Si ces derniers sont contactés par un média technologique (internet, e-mail . . .), alors la réponse est très rapide, mais si le recrutement a été fait par un mode non technologique (poste, téléphone . . .), alors la réponse est plus longue et le coût plus élevé.

I.2.5 Un mode moins contraignant

Les enquêtes administrées, particulièrement dans le cas du face-à-face, sont souvent perçues par les répondants comme une intrusion dans la sphère privée, ce qui pèse sur le taux de réponse. Le caractère peu intrusif et non contraignant du web en terme de disponibilité temporelle permet de toucher davantage d'individus (Alsnih (2005) et Ampt (2003)). La connexion en ligne est souvent personnelle, et habituelle, aussi bien dans le foyer qu'au travail et la technologie utilisée permet au répondant de remplir le questionnaire par partie, les données recueillies étant stockées dans une base et rappelables ultérieurement par l'intermédiaire d'un mot de passe. L'enquêté est donc libre de répondre à l'enquête à son domicile ou au bureau, à un horaire qui lui convient.

A la suppression de la barrière temporelle s'ajoute une flexibilité spatiale. Les enquêtes web autorisent le recueil de données dans des zones géographiques relativement étendues, et permettent d'enquêter des personnes ne pouvant pas être jointes facilement. Qu'il s'agisse d'individus peu libres en journée, ne souhaitant pas recevoir un enquêteur, ou de personnes fortement contraintes comme les professions libérales, le web semble mieux adapté que les modes traditionnels. Les jeunes sont également attirés par ce nouveau média, qui leur est familier (Stopher *et al.*, 2004b), et l'utilisation du web est un moyen d'augmenter le taux de réponse auprès de ces classes d'âge. Enfin, mener une enquête en face-à-face ou par téléphone génère des coûts salariaux importants, qu'il est nécessaire de prévoir en amont et de borner en limitant la durée du recueil de données. Il est en revanche possible de proposer une enquête en ligne de façon récurrente ou quasi-permanente ¹⁸ sans alourdir le budget de l'étude, la base de données s'alimentant automatiquement des nouveaux questionnaires remplis par les internautes. Le couplage du web avec les nouvelles technologies telles que le GPS facilite également le recueil de données de mobilité sur une longue période.

I.2.6 Le suivi du comportement de réponse

Les enquêtés diffèrent non seulement par leurs réponses aux variables d'intérêt, mais aussi par leur attitude de complétion du questionnaire. Grâce aux enquêtes web, il est à présent possible d'analyser le comportement de réponse des enquêtés (Bosnjak et Tuten, 2001). Il s'agit par exemple du temps passé sur chaque question, des moments d'interruption partielle ou totale du questionnaire, ainsi que de la navigation à travers les pages ¹⁹.

Certains auteurs, intéressés par le comportement des internautes lorsqu'ils remplissent un questionnaire en ligne, distinguent les 'optimizers', qui portent une grande attention aux questions et fournissent des réponses très précises, des

¹⁸Par exemple, l'observatoire Sierra (programme de recherche initié par les responsables des transports en France) a mis au point une enquête automatisée sur Internet, afin de suivre les comportements des français en termes de mobilité quotidienne.

¹⁹Des conditions techniques doivent être respectées, comme la présentation d'une question par page, le téléchargement des pages sur le serveur et le calcul du temps passé sur les différentes parties du questionnaire

'satisficers', qui cherchent à répondre brièvement, en limitant leur effort (Krosnick et Alwin, 1991).

I.3 Des limites importantes

L'utilisation du web dans les protocoles d'enquête semble intéressante. Les expériences sont positivement évaluées selon leur coût, le temps de réponse, le design du questionnaire et la qualité des informations recueillies. Toutefois, comme le mode postal et téléphonique au temps de leur reconnaissance comme médias d'enquête, le web présente quelques limites méthodologiques, qui modèrent son intérêt (Sills et Song, 2002). Dans la littérature, les auteurs distinguent quatre types de biais potentiels pour les enquêtes (Dillman et Bowker, 2001; Alsnih, 2005) : le biais de couverture, le biais d'échantillonnage, le biais de non-réponse et le biais de mesure. Certains auteurs (Couper, 2001) sont allés plus loin, prenant également en considération les biais d'imputation des valeurs manquantes, technique largement utilisée dans les enquêtes auto-administrées, afin de palier à la non-réponse partielle, ou les biais de pondération des individus, puisque les enquêtes web ne sont généralement pas représentatives de l'ensemble de la population (figure 1).

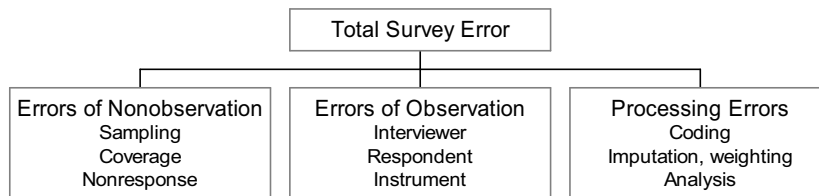


FIG. 1 – Les différentes sources de biais dans les enquêtes web
Source : M.P. Couper (2001)

I.3.1 Le biais de couverture

Afin d'effectuer une enquête de qualité, dont les résultats seront généralisables à l'ensemble de la population, il est nécessaire de sélectionner l'échantillon en effectuant un tirage aléatoire dans la population de référence, chaque individu devant avoir exactement la même chance que les autres de participer²⁰. Cette condition n'est pas toujours respectée dans les enquêtes web, puisque répondre à un questionnaire en ligne présuppose de disposer d'un ordinateur et d'une connexion Internet. Bien que l'équipement informatique des ménages français et le taux de pénétration d'Internet sur le territoire soient en forte progression²¹, tous les ménages de l'agglomération lyonnaise n'ont pas accès à internet (tableau 3), et il est impossible de leur attribuer, a priori, une probabilité connue et non nulle de participer à l'enquête.

²⁰Bien qu'il soit possible de constituer un échantillon stratifié

²¹Il est très difficile d'établir la pénétration réelle d'internet, car les données varient selon que l'on considère les individus ou les ménages, la connexion au domicile ou sur le lieu de travail, ou même la capacité individuelle à naviguer sur le web.

I Apports et limites du web pour les enquêtes de mobilité

Zones géographiques	Population (est. 2009)	Accès Internet (27 sept 2009)	Croissance (2000-2009)	Pénétration (% pop.)
France	62 150 775	42 050 465	394,70%	67,70%
Union Européenne	803 850 858	402 380 474	282,90%	50,10%

TAB. 3 – Accès à internet en France et au sein de l’Union européenne
Source : <http://www.internetworldstats.com/stats.htm>
(site visité le 27/09/09)

Au vu des chiffres, la proportion de ménages équipés d’une connexion web reste encore faible pour permettre le lancement d’études de mobilité à l’aide de ce seul média, si la représentativité statistique de l’ensemble de la population est visée²². L’erreur de couverture, biais le plus important rencontré lors d’une enquête web à grande échelle, est fonction du taux de pénétration d’internet dans la population et des différences entre les internautes et les individus n’ayant pas accès au web sur les variables d’intérêt (Groves, 1989). Ce biais est présent, même pour une enquête générale auprès des internautes, puisqu’il n’existe pas de liste officielle et exhaustive des adresses e-mail des individus. Couper (2001) propose une écriture formalisée de l’erreur de couverture. Avec Y_t , Y_c et Y_{nc} les valeurs de la variable d’intérêt sur l’ensemble de la population, sur la population couverte par internet et sur la population n’ayant pas accès au web, N_t et N_{nc} le nombre d’individus de la population totale et le nombre d’individus ayant accès au web, on obtient :

$$Y_c = Y_t + \frac{N_{nc}}{N_t}(Y_c - Y_{nc}) \quad (1)$$

Plus encore, il est possible de différencier les internautes selon leur équipement multimédia (navigateur, type de connexion ...) et leur maîtrise de l’usage d’Internet. Bradley (1999) distingue ainsi 13 classes d’internautes selon l’équipement informatique et l’usage d’internet (fréquence de lecture des e-mails, nombre de messages en attente dans la boîte aux lettres, fréquence de visites de sites web, connexion internet à domicile ou sur le lieu de travail, nombre d’adresses e-mail utilisées...). Un rapport du Ressource System Group (2002) définit la capacité de navigation des internautes par le nombre de pages web visualisables sur une période de temps. Cet indicateur dépend notamment du matériel disponible (ordinateur, navigateur ...) et du serveur sur lequel les requêtes sont lancées. Même si les ménages auxquels nous nous intéressons possèdent une connexion Internet, il n’est pas établi que l’ensemble des individus qui composent ce ménage soient à l’aise avec l’utilisation du web (Gunn, 2002; Lazar et Preece, 1999). Les enquêtes web ne sont accessibles qu’aux individus familiers d’internet (Abdel-Aty, 2003). Pour les internautes peu expérimentés, la capacité de répondre aux questions à l’écran avec un clavier et une souris est limitée. Selon le degré de complexité du questionnaire, certains ne pourront pas télécharger les pages, d’autres auront des difficultés pour répondre

²²Certaines catégories de la population sont difficiles à joindre par le web (personnes âgées, peu qualifiées ...) (Roos et Marchand, 2006).

aux questions (utilisation de menu déroulant, validation des réponses, saisie des données dans des champs spécifiques, accès aux pages suivantes ...). Il est donc possible que non seulement les caractéristiques sociodémographiques des internautes diffèrent de celles de l'ensemble de la population, mais qu'elles varient également selon le type d'équipement possédé et le degré de familiarité avec le web (Alsnih, 2005).

I.3.2 Le biais d'échantillonnage

L'objectif d'une enquête est d'estimer avec précision la distribution d'une variable dans la population de référence, à partir de données obtenues sur une petite proportion d'individus. Afin de permettre l'inférence statistique des résultats, l'échantillon utilisé pour l'enquête doit être représentatif de la population étudiée. Le biais d'échantillonnage vient des différences, en termes de caractéristiques socio-économiques mais aussi sur les variables d'intérêt, entre les personnes qui constituent l'échantillon et celles de la population de référence (Couper, 2000).

Certaines méthodes de recrutement, telle que celle utilisée pour les enquêtes ménages déplacements en France, s'appuient sur une procédure d'échantillonnage de la population très pointue, alors que d'autres méthodes ne contrôlent pas l'échantillon au-delà du respect de certains quotas. Lors de la réalisation d'une enquête web à grande échelle, il est quasiment impossible d'utiliser les méthodes d'échantillonnage probabilistes classiques, puisqu'à la différence d'autres médias²³, aucune base de sondage n'est disponible (Alvarez et Van Belsaere, 2003; Andrews *et al.*, 2003). Un certain pourcentage d'internautes changent régulièrement de fournisseur d'accès et donc d'adresse mail (Lang, 2002), d'autres conservent leur ancienne adresse, mais ne l'utilisent plus et ne vérifient pas régulièrement leurs messages (Bradley, 1999) et d'autres possèdent de multiples adresses, professionnelles et personnelles. Au final, un document répertoriant l'ensemble des adresses mail serait de faible recours pour mener des enquêtes, puisque difficilement actualisable et que les e-mail considérés comme intrusifs sont directement effacés par les internautes (Cho et LaRose, 1999). De plus, les internautes ont un profil socio-économique particulier (Solomon, 2001; Romano et Himmelmann, 2002). Ils sont plus jeunes, actifs, davantage masculins et avec un haut niveau de diplôme et de revenus (Myles et Tibert, 1998; Braunsberger *et al.*, 2007; Kwak et Radler, 2000; Bates, 2001; Deutskens *et al.*, 2006; Haraldsen *et al.*, 2002). Vehovar et Lozar Manfreda (2002) et Badagliacco (1990) précisent que la fréquence d'utilisation de l'ordinateur et d'internet, corrélée au sexe et au niveau d'éducation sont les plus importants facteurs qui influencent la participation aux enquêtes web. Ces dernières sont plus facilement acceptées chez ceux qui sont familiers avec l'ordinateur (à l'école, au travail ou à la maison), et parmi les répondants qui savent l'utiliser 'correctement' (Wright *et al.*, 1998)²⁴. Certains types de ménages, les plus modestes ou ceux qui ne disposent pas d'un accès internet au domicile ou sur

²³Fichier des abonnés à France Telecom pour les enquêtes téléphoniques ou fichier des adresses du recensement de l'INSEE pour les enquêtes postales.

²⁴La non familiarité avec l'outil génère un effet 'big brother' amenant plus de refus (Lo-

leur lieu de travail, se trouvent mécaniquement exclus d'une enquête web. Ils représentent pourtant une part non négligeable de la population et ont parfois des comportements spécifiques qu'il est délicat d'ignorer ²⁵. Myles et Tibert (1998) précisent qu'une enquête web est une enquête auprès d'individus qui s'auto-sélectionnent et que toute représentativité de la population de référence est accidentelle et non le reflet d'une méthode d'échantillonnage ²⁶.

Si la sélection des répondants est aléatoire, la précision des résultats est directement liée à la taille de l'échantillon ²⁷. Les conducteurs d'études en ligne ont été tentés de compenser le biais de couverture et d'échantillonnage en interrogeant de larges échantillons ²⁸. L'idée est qu'un large échantillon diminue les chances d'exclure systématiquement certains segments de la population. Mais un taux de réponse important, obtenu sur un échantillon non probabiliste, ne garantit pas l'obtention de données de qualité (Dillman et Bowker, 2001). Selon Fowler (1993) la pertinence des données dépend de la façon dont l'échantillon est construit, et non de sa taille. Ainsi, il ne sert à rien d'augmenter le nombre de répondants afin d'augmenter le degré de précision des résultats, si le mode de sélection des individus est biaisé. Dans certains cas, un plus petit échantillon bien sélectionné conduit à des résultats de meilleure qualité (Couper, 2001).

Lors d'une enquête web, les répondants ne sont donc généralement pas représentatifs de la population étudiée, ce qui interdit toute généralisation des résultats à l'ensemble de la population par inférence statistique (Resource System Group, 2002). La construction d'un échantillon probabiliste nécessite alors le recours à d'autres médias pour sa construction, par exemple un recrutement par téléphone où l'enquêteur encourage le répondant à remplir le questionnaire en ligne (Alsnih, 2005).

I.3.3 Le biais de non-réponse

Même si l'échantillon est totalement aléatoire, certains individus ne veulent ou ne peuvent répondre à l'enquête, d'autres ne terminent pas le questionnaire. Le biais de non-réponse est fonction du taux de non-réponse à l'enquête et des différences entre les répondants et les non-répondants sur la variable d'intérêt (Keeter *et al.*, 2000; Groves et Couper, 1998) ²⁹. Dans une enquête

zar Manfreda et Vehovar, 2002b).

²⁵Par exemple en termes d'utilisation de la voiture particulière pour les enquêtes de mobilité.

²⁶Au mieux, les chargés d'études vont comparer les données socio-économiques agrégées déclarées par les répondants web avec celle de la population de référence, pour vérifier que l'échantillon web peut être considéré comme le reflet possible de la population (Lang, 2002).

²⁷L'erreur d'échantillonnage diminue de moitié, lorsque la taille de l'échantillon est quadruplée.

²⁸Cependant, les déterminants de la taille de l'échantillon sont plus souvent d'ordre budgétaire (consommation d'argent, de temps, d'espace et d'énergie) que statistique (Roscoe, 1975; Alreck et Settle, 1995).

²⁹Il n'est pas démontré que le taux de réponse et le biais de non-réponse soient corrélés, des taux de réponse hauts et bas pouvant amener à un biais de non-réponse (Baines *et al.*,

web, certains groupes sociodémographiques sont sur ou sous-représentés dans l'échantillon (Witte *et al.*, 2000). Par exemple, les hommes jeunes refusent fréquemment de répondre aux enquêtes traditionnelles, alors que les enquêtes web touchent majoritairement des individus jeunes, masculins, citadins, diplômés et ayant des revenus élevés (Alsnih, 2004; Conn, 2005; Murakami et Morris, 2003). Alsnih (2005) précise que les familles nombreuses et les personnes au statut socio-économique élevé qui échappent aux modes traditionnels répondent plus volontiers sur le web. L'analyse de la littérature reste assez évasive sur les facteurs pouvant diminuer le taux de réponse des enquêtes web. Il est certain que le sujet de l'enquête et les caractéristiques sociodémographiques des individus sélectionnés pour y répondre vont largement impacter le taux de réponse (Sills et Song, 2002). Des problèmes techniques, tels que la basse vitesse du modem, la lenteur du navigateur, les interruptions survenant en cours de remplissage du questionnaire ou des temps de chargement parfois longs amènent le répondant à abandonner la saisie du questionnaire, sans que les raisons précises de ces comportements soient connues de l'administrateur (Dillman et Bowker, 2001). Il faut également prendre en compte des facteurs plus personnels, comme la familiarité des individus avec le média internet et leur attitude face à la sécurité et à la confidentialité des données délivrées (Couper, 2000).

Solomon (2001) suggère deux moments cruciaux, où les internautes risquent d'abandonner la saisie des réponses : à l'apparition d'une grille complexe de questions - réponses, ou lorsqu'on leur demande de saisir leur adresse e-mail ou des données trop personnelles. Dans le cas spécifique des enquêtes de mobilité, les abandons risquent de survenir lors du recueil des déplacements, fastidieux pour les hyper-mobiles, et lors du recueil du revenu, question toujours sensible dans les enquêtes françaises. Il n'est pas évident d'établir que ces comportements de non-réponse, totale ou partielle, ne soient pas liés aux pratiques de mobilité (Bonnell, 2004) et tout non-répondant à une enquête est en effet une source possible de biais qui doit être étudiée (Couper *et al.*, 1999). Ce biais reste toutefois difficile à estimer, dans la mesure où on ne peut identifier la population totale d'internautes, ni connaître les caractéristiques ou le comportement de mobilité de ceux qui restent 'hermétiques' à l'enquête³⁰. Il a par ailleurs peu de sens si la méthode d'échantillonnage retenue est non probabiliste.

I.3.4 Le biais de mesure

L'erreur de mesure correspond à une déviation des réponses par rapport à leur vraie grandeur (Couper, 2000). En cas de faible motivation à répondre au questionnaire ou de mauvaise compréhension des questions, les interviewés peuvent aisément omettre certaines réponses ou 'cliquer' sans réellement réfléchir à la signification d'une question un peu vague. Des répondants 'professionnels' peuvent également biaiser les réponses. Il s'agit d'individus qui répondent systématiquement aux enquêtes, de façon très rapide, sans réellement lire les

2007).

³⁰(Kalfs et Van Evert, 2003) conclut que les enquêtes sur les non-répondants sont utiles pour la recherche, mais trop coûteuses pour être mises en oeuvre à grande échelle.

questions ni penser à leurs réponses (Nielsen, 2005).

La deuxième source du biais de mesure, commune aux enquêtes auto-administrées, vient de l'absence de la relation enquêteur-enquêté, qui ne permet pas au répondant d'obtenir des compléments d'information sur le sens des questions ou les termes employés. Elle empêche également toute analyse du comportement non-verbal du répondant, souvent riche d'enseignements pour l'enquêteur et l'analyste (soupirs en cas de lassitude, expressions du visage marquant l'incompréhension...). Les conséquences sur la pertinence des données recueillies ne sont pas neutres.

Enfin, de par sa construction, le questionnaire peut conduire à un biais de mesure, les stimuli utilisés étant susceptibles de modifier le comportement des répondants. Le design du questionnaire (placement des questions, typographie, couleur, échelles de réponse...) peut motiver l'enquêté à saisir ses réponses, ou au contraire le décourager dans sa démarche (Couper *et al.*, 1999; Dillman *et al.*, 1998b). L'apparence du questionnaire en ligne varie parfois fortement selon le matériel informatique et certains dysfonctionnements techniques, non anticipés, peuvent également causer des erreurs dans la collecte des données par internet (temps de chargement des pages, non disponibilité du serveur, problème d'affichage...) (Dillman et Bowker, 2001; Alsnih, 2004).

I.3.5 Les diverses sources d'incertitudes

Trois autres contraintes doivent être considérées par l'administrateur d'une enquête web : le coût, les problèmes techniques et le contrôle du répondant. Le temps nécessaire pour télécharger les pages web a en effet un coût pour l'enquêté, qu'il s'agisse du temps passé devant l'écran ou du montant de la communication, pour ceux qui ne bénéficient pas de forfaits illimités³¹. En règle générale, si les capacités des individus face à l'ordinateur diffèrent substantiellement, il existe une durée limite du questionnaire à ne pas dépasser pour éviter les abandons. En ce qui concerne le commanditaire, la mise en place d'une enquête web génère des coûts fixes parfois importants. Le développement du questionnaire nécessite du temps et des compétences spécifiques, sa gestion en ligne occasionne des coûts de maintenance du réseau et d'assistance des enquêtés (Schonlau *et al.*, 2001; Watt, 1999).

Les problèmes techniques constituent une autre limite des enquêtes web (Truell, 2003). Les coupures de réseau peuvent interrompre la saisie des réponses sans que le gestionnaire de l'enquête n'en soit informé, et conduire à la perte définitive des répondants. La présentation du questionnaire peut varier selon les paramètres de l'ordinateur, et apparaître fort désagréable à certains internautes, qui vont quitter le questionnaire avant d'avoir terminé la saisie des réponses.

Enfin, dans une enquête web, il est difficile de contrôler 'celui' qui se cache derrière l'ordinateur. Cette limite est propre au caractère auto-administré du questionnaire. Aucun enquêteur ne valide l'identité de la personne interro-

³¹Les enquêtes web longues à télécharger augmentent le coût pour l'enquêté (Swoboda *et al.*, 1997).

gée, qui peut facilement mentir sur ses caractéristiques sociodémographiques principales, telles que le sexe et l'âge (Nielsen, 2005) ou reporter des données relatives à un autre membre du ménage (Lang, 2002). Le répondant contrôle le temps et la façon de remplir le questionnaire. Il peut être endormi, en colère, intoxiqué, contrarié ou dans un autre état d'esprit l'empêchant de déclarer des réponses justes et honnêtes. Stanton (1998) montre que cette variation de motivation génère des réponses manquantes et biaise les résultats de l'enquête. Les internautes peuvent également 'zapper' entre différents programmes et avoir plusieurs pages web ouvertes, ce qui génère des processus cognitifs plus superficiels (Alvey, 2004). Le contrôle des réponses et du profil du répondant est donc incertain et complexifie l'analyse des données. Si la connexion n'est pas sécurisée par un mot de passe, l'administrateur s'expose par ailleurs à une duplication possible des questionnaires, si plusieurs personnes répondent successivement.

I.3.6 La sécurité des données

Les études sérieuses garantissent la confidentialité des réponses (seuls les analystes ont accès aux données), mais l'anonymat n'est pas toujours respecté (les analystes peuvent identifier les répondants). En France, la CNIL impose des règles très strictes sur la collecte des données et leur utilisation. Des techniques sont mises au point, afin de conserver l'anonymat des répondants. Par exemple, les adresses exactes ne figurent généralement pas dans les enquêtes de mobilité, et la codification géographique se fait à un niveau suffisamment élevé pour que les répondants ne soient pas identifiables. Mais les modèles sophistiqués actuellement utilisés pour prévoir la demande de transport nécessitent des données désagrégées, et les enquêtés déclarent des informations de plus en plus personnelles, comme l'adresse de leur lieu de travail (Murakami et Morris, 2003). Bien que des auteurs aient montré que certains internautes valorisaient la personnalisation du contact, l'intrusion de l'informatique dans la vie privée n'est pas toujours rassurante (Yun et Trumbo, 2000; Conn, 2005). Chaque ordinateur laisse des traces sur les sites visités, puisqu'il est possible d'enregistrer l'adresse IP de la machine utilisée par le répondant à l'aide d'un programme informatique adapté (Berry, 2004). Les données sont également susceptibles d'être interceptées pendant leur transfert de l'ordinateur du répondant vers le serveur dédié à l'enquête et Cho et LaRose (1999) ont montré que les chercheurs commettent davantage de violations de la vie privée des répondants sur le web qu'avec les modes de recueil plus traditionnel. Il semble donc préférable de garantir la confidentialité des données en recourant à des serveurs sécurisés³². Les réponses anonymes semblent préférables pour obtenir la confiance des répondants, mais leurs croyances concernant l'anonymat sur internet, avec ou sans accès contrôlé, ne sont pas connues (Stanton, 1998).

³²Les sites web sécurisés ont une adresse qui commence par 'https' (et non par 'http', comme la plupart des sites). Si une lettre est utilisée pour avertir les répondants que le questionnaire peut être rempli en ligne, alors ces derniers sont susceptibles de ne pas saisir correctement l'adresse du site sécurisé. Lors d'une expérience menée en laboratoire, Berry (2004) a montré que ce type d'erreur concerne 13% des individus.

L'utilisation de codes d'accès et de mots de passe permet de restreindre l'accès aux données, mais peut augmenter la suspicion des répondants. Enfin, il peut être utile de mettre en place des pare-feux pour garantir la sécurité des fichiers. Certains internautes refusent en effet de se connecter sur des sites inconnus ou de répondre à un questionnaire en ligne,

I.4 Un avenir prometteur

Les modes d'enquêtes traditionnels tels que le téléphone, le postal ou le face-à-face sont plus populaires que le web (Huang et Liaw, 2005). Si les études en ligne sont encore modestes, puisqu'elles ne représentent que 10% du chiffre d'affaires total du secteur des études, la diffusion des technologies de l'information et de la communication dans les entreprises, le monde universitaire et les particuliers, font du web un outil prometteur pour les commanditaires d'études. Les expériences montrent que le web peut remplacer avantageusement un certain nombre de modes traditionnels, en arrivant à des résultats identiques (Roos et Marchand, 2006; Pitkow et Recker, 1995; Stanton, 1998). En 2005, AC Nielsen a mené outre-Atlantique une enquête d'Usages et Attitudes sur deux échantillons, web et téléphone, qui a abouti à des résultats comparables³³. Abdelwahab et Abdel-Aty (2003) ont comparé la qualité et la précision des données provenant d'une étude postale, avec celles d'une étude web, et a conclu à la supériorité de ces dernières³⁴. Selon Dillman et Bowker (2001), il existe peu de littérature sur la variation de la qualité des réponses d'un individu, selon qu'il réponde sur le web ou par un mode plus traditionnel (postal, face-à-face). Plusieurs auteurs (Benfield et Szlemko, 2006; Schillewaert et Meulemeester, 2005) ont montré que le mode d'enquête, qu'il repose sur une version papier ou électronique, ne modifie pas en théorie la qualité des données fournies par le répondant. Les compétences informatiques des individus, associées à la connaissance du sujet d'étude et à leur aisance relative au remplissage d'un questionnaire peuvent cependant altérer leur capacité à répondre en ligne et donc impacter la qualité des données recueillies.

Le web est un média restreint par les capacités informatiques des répondants (Conrad *et al.*, 2003). Huang et Liaw (2005) ont mené une expérience auprès d'étudiants, pour analyser les perceptions des répondants web. Ils postulent que les déterminants de l'intention de répondre à une enquête en ligne sont : la capacité des individus à saisir leurs réponses, leur appréhension concernant l'utilisation future des données, leur motivation intrinsèque pour répondre en ligne et la pertinence du web pour l'étude. Les résultats montrent qu'une fois que les enquêtés estiment qu'ils sont en mesure de remplir le questionnaire web, le nombre de questions, la vitesse de la connexion et le design de l'interface sont les principaux facteurs qui déterminent la saisie des réponses (figure 2). L'anxiété face à l'utilisation de l'ordinateur impacte en revanche négativement la propension à répondre en ligne et la performance individuelle dans la saisie

³³Le web a généré des réponses de qualité, même en ce qui concerne les questions ouvertes.

³⁴Pourtant, de nombreuses études rapportent un taux de réponse par voie postale supérieur au taux de réponse en ligne (Schaefer et Dillman, 1998; Morris et Adler, 2003).

des réponses. Il est difficile aux commanditaires d'études de contrôler certains déterminants tels que les compétences des répondants, la vitesse de connexion des internautes ou l'accessibilité des ménages à une connexion internet. En revanche, une réflexion sur la prise de contact et la construction du questionnaire peut influencer positivement l'intention des enquêtés de répondre en ligne.

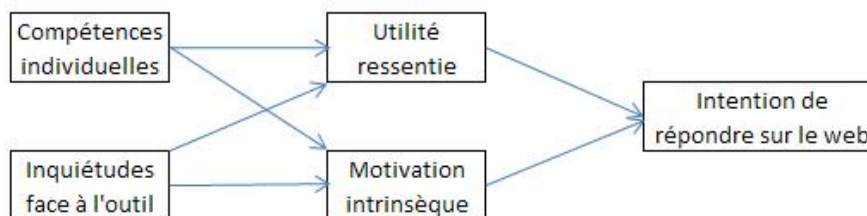


FIG. 2 – Modèle conceptuel des déterminants de l'intention de répondre en ligne

Source : Huang and Liaw (2005)

Si le coût et la rapidité sont les deux avantages les plus souvent cités par les partisans des enquêtes web (Witt, 1998), ce nouveau mode de collecte présente d'autres atouts intéressants, comme la souplesse du questionnaire et une réactivité inégalée des gestionnaires d'études³⁵. Les problèmes de l'utilisation du web sont de trois sortes : échantillonnage, consistance des réponses et la motivation des participants (tableau 4). Afin de réduire ces biais, Une attention particulière doit être portée à la conception des questionnaires web, les développeurs étant davantage des techniciens du web que des professionnels des méthodes d'enquête. L'enquête idéale, support au développement d'un modèle de prévision, permettrait sans doute de couvrir l'ensemble de la population, pour un rapport coût-qualité intéressant, et prévoirait des stratégies pour limiter les non-réponses et assurer la cohérence des données recueillies (Morris et Adler, 2003). Mais il est difficile de réunir l'ensemble de ces conditions à l'aide d'un seul mode d'enquête.

Avantages (+)	Limites (-)
Fort potentiel à faible coût	Visualisation différente selon l'équipement
Réponse plus rapide	Sensible au niveau d'expertise informatique
Envoi de rappels	Sécurité des données sur le serveur
Vérification dynamique des incohérences	Confidentialité des réponses
Questions aléatoires	Comportement des Internaute
Personnalisation	Méthodes de sélection de l'échantillon
Dispersion géographique	Pas de contrôle de l'identité du répondant
Affichage progressif, interactivité	Biais de compréhension des questions

TAB. 4 – Avantages et limites du média web pour la réalisation d'enquêtes

Source : Gunn (2002)

³⁵En comparaison avec les enquêtes auto-administrées sur papier, les questionnaires web sont plus conviviaux, mieux structurés, plus personnalisés, moins chers, et permettent de limiter les non-réponses partielles et les erreurs (Nowack, 1997; Weible et Wallace, 1998).

II Vers une méthodologie d'enquête mixte

Face à la résistance de certains segments de la population à participer aux enquêtes, aux réponses incomplètes ou peu fiables et aux difficultés croissantes pour établir le premier contact avec un enquêteur, la combinaison de méthodologie différente au cours d'une même enquête semble être une solution intéressante. Il s'agit de combiner le meilleur des modes, l'augmentation du taux de couverture permettant d'améliorer la qualité des données, sans grever le coût de l'étude (Dillman et Tarnai, 1988; Murakami et Morris, 2003; Wermuth *et al.*, 2001). La question de la mixité des modes d'enquête n'est pas récente ³⁶, mais elle mérite d'être reconsidérée au regard des récents développements technologiques qui impactent le domaine des études (Stopher et Jones, 2003). Par exemple, la facilité avec laquelle le script d'un questionnaire peut être transféré d'un média à l'autre encourage l'utilisation de différents modes au cours d'une même enquête (Dillman *et al.*, 2001). Si les technologies de l'information et de la communication actuellement disponibles (internet, téléphones mobiles, GPS, PDA ...) offrent la possibilité d'augmenter la qualité des données d'enquête et de réduire la lassitude des répondants, elles ne peuvent cependant pas se substituer totalement aux modes plus traditionnels, tels que le face-à-face, le postal ou le téléphone (Murakami et Morris, 2003).

II.1 Une diversité de protocoles d'enquête

Au cours d'une étude, il y a trois phases de contact possibles avec les répondants : la phase de recrutement, la phase d'interview et la phase de suivi. Dans les enquêtes transport, il arrive fréquemment que des médias différents soient employés pour chacune de ces phases, par exemple le postal pour le recrutement et l'information des répondants et le web pour l'enquête. L'utilisation de plusieurs modes pour le recueil de données suscite en revanche plus d'interrogations. Cette mixité peut prendre différentes formes, suivant l'objectif poursuivi : réduire la lassitude des enquêtés, en leur laissant le choix du mode de réponse, ou faire face à une diminution du taux de réponse, en proposant un mode alternatif aux non-répondants. Ces deux protocoles sont schématisés dans la figure 3.

Chaque mode peut être appliqué à un segment particulier de la population cible, un mode pouvant être optimal pour un segment mais pas pour un autre. Par exemple, dans l'enquête nationale transport allemande, l'enquête est conduite par CATI, pour la partie de l'échantillon pour laquelle il est possible d'obtenir un numéro de téléphone, et par voie postale pour le reste (Bonnell et Armoogum, 2005). Cette pratique n'est toutefois pas sans risque sur la qualité des données, puisque les effets de modes peuvent créer des différences dans les

³⁶Les enquêtes postales et face-à-face sont les plus anciennes, et leur combinaison constitue les premières formes d'enquêtes mixtes (De Leeuw, 2005). Pour réduire le biais de couverture au démarrage du téléphone, une enquête mixte permettrait de bénéficier des coûts attractifs des enquêtes par téléphone, tout en maintenant une bonne couverture de la population étudiée (Groves et Lepkowski, 1985). En recherche marketing, les hybrides du téléphone et du web sont devenus très populaires (Oosterveld et Willems, 2003) et (De Leeuw, 2005).

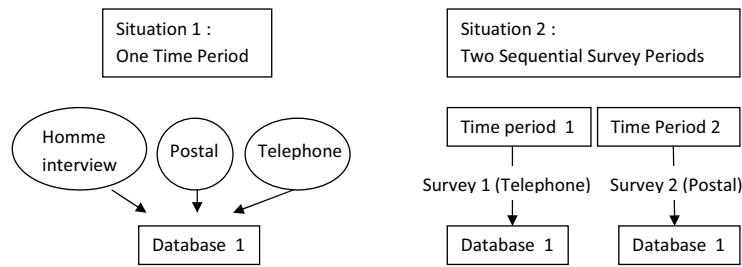


FIG. 3 – Deux situations de mixité du recueil de données
Source : J. Morris and T. Adler (2003)

réponses des groupes. Il peut également y avoir des vagues séquentielles d'enquêtes sur la même population, à l'aide de modes différents (Morris et Adler, 2003; Dillman et Christian, 2005). Dans l'enquête de mobilité Belge MOBEL, les questionnaires sont envoyés par courrier et des contacts téléphoniques sont réalisés auprès des ménages répertoriés dans les annuaires pour les motiver à participer (Hubert et Toint, 2003). L'intérêt du protocole d'enquête séquentiel mixte est que ceux qui répondent au second mode sont différents de ceux qui répondent au premier. L'espoir est que cette seconde population rassemble les non-répondants restants (Baines *et al.*, 2007). Le choix peut également être laissé à chaque enquêté du média utilisé pour répondre à l'enquête. C'est le cas dans le protocole mis en place par Keesling (2002), où une mention accompagnait le questionnaire envoyé par voie postale, précisant qu'il était possible de remplir le même questionnaire en se connectant sur le site web dédié au projet. Cette méthode est souvent utilisée pour réduire l'erreur de couverture, mais génère d'autres problèmes comme l'auto-sélection des individus et l'erreur de mesure due aux modes utilisés (Lozar Manfreda *et al.*, 2001). Enfin, lors d'enquêtes longitudinales, des répondants sont interrogés à plusieurs reprises à l'aide de différents médias.

Dans notre étude, le processus est séquentiel, puisque l'enquête web n'est proposée qu'aux individus qu'il n'aura pas été possible de joindre durant la première vague d'interviews en face-à-face.

II.2 Pour diminuer le taux de non-réponse

Face à la réticence de plus en plus vive de la population à répondre aux enquêteurs, il devient difficile de réaliser une enquête de mobilité en face-à-face de bonne qualité, les non-répondants ayant probablement des comportements de déplacements différents des ménages qui acceptent de se soumettre à l'interview. De plus en plus d'enquêtes reposent sur des protocoles complexes, associant plusieurs modes ou méthodologies, pour augmenter le taux de réponse global, accroître le taux de couverture de la population cible ou améliorer la qualité des données produites (Couper, 2000; Gunn, 2002; Dillman *et al.*, 2001; Alsnih, 2004; Groves et Kahn, 1979). Qu'il s'agisse de la phase de recrutement ou d'interview, multiplier les modes de contact est un moyen sûr de toucher davantage d'unités statistiques (Goyder, 1987). Non seulement les individus

apprécient de disposer de plusieurs moyens pour communiquer sur leurs habitudes ³⁷, mais tous les modes ne permettent pas de toucher les mêmes segments de la population (les 'hypers mobiles' par exemple sont rarement présents à leur domicile aux créneaux horaires pendant lesquels les enquêteurs prennent les rendez-vous et viennent administrer le questionnaire. Les répondants à un média ne sont donc pas forcément les mêmes que les répondants à un autre média (Bonnel, 2003; Bayart *et al.*, 2009; Shettle et Mooney, 1999).

L'utilisation d'internet dans un processus séquentiel apparaît comme un moyen efficace de capter des répondants qui échappent aux modes traditionnels, et de réduire le biais de non-réponse si les répondants web sont représentatifs de l'échantillon restant qui n'a pas répondu au premier mode (Baines *et al.*, 2007).

II.2.1 En limitant le niveau d'effort requis

Avant de participer à une enquête, l'individu va évaluer le niveau d'effort requis. Il va notamment s'interroger sur le temps pris par l'interview, l'importance du sujet, les difficultés physiques, intellectuelles ou émotionnelles ressenties pour répondre au questionnaire, le bien être procuré par ses réponses, sa volonté de livrer des informations personnelles et les éventuelles pressions externes qu'il peut subir (Ampt, 1997). A cela peut s'ajouter la nature de l'étude, les renseignements factuels (données socio-économiques) étant plus facilement fournis que les réponses de nature comportementale ou attitudinale, qui réfèrent davantage à l'émotionnel (Richardson *et al.*, 1995). L'évaluation de ces différents paramètres va influencer les comportements individuels de réponse / non-réponse. Mais l'enquêté va également prendre en compte ses expériences passées, allant des sollicitations aux participations effectives. Pour améliorer le taux de réponse, il est donc nécessaire de considérer l'ensemble des éléments qui affectent la perception de difficulté de l'enquête par le répondant, et donc son processus de choix d'y participer ou non. Cette étape est d'autant plus délicate que le niveau de difficulté ressenti dépend des individus et du moment du contact.

La mise en place d'une enquête sur Internet va jouer essentiellement sur les dimensions temporelles, physiques, intellectuelles et interpersonnelles de l'étude. Le temps dont disposent les individus pour répondre au questionnaire est en effet limité et fortement contraint par le type d'activités dans lesquelles ils sont impliqués (travail, école, loisirs, occupations diverses au domicile ...). De manière générale, les questionnaires auto-administrés sont plus flexibles, puisque l'individu choisit lui-même le moment le plus approprié pour y répondre. Les entretiens en face-à-face essaient de lever cette contrainte temporelle, en laissant choisir à l'interrogé le créneau horaire du rendez-vous, mais les possibilités sont limitées, et dégager une plage d'une heure trente dans un emploi du temps surchargé n'est pas toujours une tâche facile.

³⁷Une étude sur les non-répondants menée en Australie par l'Institute of Transport Studies montre que les personnes préfèrent avoir le choix du mode de réponse et que cette option impacte positivement le taux de réponse et la qualité des données.

II.2.2 En conservant la pertinence des réponses obtenues

Si l'interactivité du web permet de relancer le répondant en cas de non-réponse à une ou plusieurs question, des effets en termes de comportements de réponse ont également été remarqués. Nous considérerons plus particulièrement dans cette section les effets de modes qui pèsent sur la qualité des données (Gunn, 2002). Dans le cas de questionnaires administrés, la présence de l'enquêteur génère un biais, puisque le répondant risque de positionner son discours en fonction de l'impact positif qu'il espère produire sur son interlocuteur ('social compliance') (De Leeuw, 1992; Schuman et Presser, 1981). Bien que ce phénomène s'observe surtout lors de question d'opinion ou d'attitude, il est possible de le remarquer également dans les enquêtes de déplacements, le domaine des transports étant par nature très sensible³⁸. Ce biais est fortement réduit dans le cas d'une enquête web, puisque l'interaction entre l'enquêteur et l'enquêté est virtuelle.

Par ailleurs, lorsqu'une liste de modalités est lue par l'interviewer, le répondant a généralement tendance à sélectionner le dernier choix ('recency effect'), car il n'a pas le temps de fixer chaque choix dans une mémoire de long terme avant que le prochain ne soit donné. Dans les modes auto-administrés, le processus cognitif est plus profond, et les premiers items sont davantage choisis (Krosnick et Alwin, 1987; Schuman et Presser, 1981). Cet effet survient surtout dans les modes oraux tels que le téléphone ou le face-à-face, mais est peu présent dans les enquêtes web. Enfin, ce mode de recueil assisté par ordinateur permet une vérification instantanée des réponses fournies (Schonlau *et al.*, 2001).

Si l'utilisation exclusive du web dans les enquêtes déplacements semble prématurée, il est en revanche possible de combiner efficacement ce média avec un mode de recueil plus traditionnel, au risque de modifier les mécanismes cognitifs sollicités chez les répondants.

II.3 Le web : un mode complémentaire du face-à-face ?

Quatre grandes générations de modes d'enquête se sont succédées, s'adaptant au fil des années aux évolutions technologiques : le face-à-face, le postal, le téléphone et l'internet. Chacun offre des avantages incontestables, mais également des limites non négligeables, dont le commanditaire doit tenir compte avant de faire son choix. De nombreux pays tels que le Canada, la Suisse et les USA utilisent depuis plusieurs années le téléphone³⁹, moins coûteux, pour réaliser leurs enquêtes de mobilité (Richardson *et al.*, 1995). En France, le mode dominant reste le face-à-face. Il semble opportun de s'interroger à présent sur le potentiel du web, comme complément aux enquêtes traditionnelles.

³⁸Par exemple, l'enquête étant financée par le Sytral, les individus peuvent avoir tendance à sur valoriser les transports en commun.

³⁹Aux USA, le mode postal est également utilisé.

II.3.1 Deux modes aux atouts différents

Le face-à-face est un mode d'enquête relativement flexible, puisque les enquêteurs fixent à l'avance des rendez-vous avec les répondants, en fonction de leurs disponibilités. Il s'agit d'un mode adapté aux enquêtes lourdes et complexes. L'enquêteur facilite la tâche du répondant, montre certains éléments (utilisation possible de visuels et de sons dans les protocoles d'enquête), et relance éventuellement l'entretien. Il peut s'assurer que les répondants comprennent les questions et que toutes sont identiquement interprétées, (Conrad et Schober, 2000; Dillman et Christian, 2005), sa voix et ses gestes agissant comme un paralangage qui donne des informations au répondant. La couverture de l'échantillon peut être totale, mais le coût horaire des enquêteurs augmente le budget de l'étude et limite souvent son étendue géographique. D'un autre côté, l'influence de l'enquêteur, propre aux enquêtes administrées, peut venir biaiser les réponses. Le face-à-face est également qualifié d'intrusif, puisque l'enquêteur vient au domicile des personnes interrogées. Ces dernières ne sont pas toujours disponibles aux créneaux proposés, selon leur activité professionnelle, ni intéressées par le sujet de l'étude, et réunir l'ensemble des personnes du ménage pour recevoir un enquêteur est souvent un exercice difficile. Enfin, des problèmes de sécurité limitent la réalisation d'enquêtes à domicile dans certaines zones urbaines ⁴⁰.

Les avantages et les limites du web comme média d'enquête ont été détaillés dans les sections précédentes. Concernant le mode d'administration du questionnaire, les enquêtes web sont à mi-chemin entre le mode administré et le mode auto administré. L'environnement s'apparente aux enquêtes postales et le questionnement met en scène un enquêteur virtuel (Haraldsen *et al.*, 2002). Le web et le face-à-face sont donc des modes d'enquête aux caractéristiques très diverses (média utilisé, forme et longueur du questionnaire ...), qui impactent directement leurs performances (tableau 5).

Critères	Mode d'enquête	
	face-à-face	web
Couverture	++++	+
Taux de réponse	++++	+++
Qualité des données	++++	++
Langage	++++	++
Complexité des questions	+++	+++
Coûts	+	++++
Qualité du contrôle	++	++++

TAB. 5 – Comparaison des enquêtes en face à face et par web
 Source : A partir de Morris et Adler (2003), adapté de Ettema *et al.* (1996)

⁴⁰Les enquêtes ménages déplacements en face-à-face ont été remplacées par des enquêtes téléphone aux USA, car trop chères et trop dangereuses à mener dans certaines aires urbaines (Stopher et Metcalf, 1996).

II.3.2 Intérêt de la mixité du web et du face-à-face

Nous avons par exemple montré dans la section I qu'il n'est pas recommandé de réaliser une enquête web exclusive auprès de l'ensemble de la population, dans le but de recueillir des données représentatives (Bonnel, 2004). En effet, les nouvelles technologies ne sont pas à la disposition de l'ensemble des habitants, et employer seulement le web élimine la possibilité pour certaines personnes d'être sélectionnées (Dillman *et al.*, 1999). Par ailleurs, le problème des spams et les mesures préventives utilisées pour y mettre fin rendent difficile le contact d'un répondant sur le web (Rose *et al.*, 2004).

Dillman (2000) et Dillman *et al.* (1995) montrent en revanche que l'utilisation du web dans un protocole d'enquête est un moyen efficace d'augmenter le taux de réponse global, et concluent que les enquêtes web peuvent être utilisées de façon complémentaire aux méthodes traditionnelles. Il est donc intéressant de réfléchir sur la possible intégration du web dans les enquêtes de mobilité et d'explorer plus largement la combinaison de plusieurs modes de collecte de données, telles que le web et le face à face (Couper, 2000; Gunn, 2002; Comley, 1996). Introduire le web dans un protocole d'enquête contribue à la diffusion de ce média, et à la conversion des ménages autrefois fidèles aux modes traditionnels tels que le face-à-face (Cobanoglu *et al.*, 2001). En proposant le web en complément du face-à-face, la population générale est mieux cernée, puisque certains individus se reconnaissent davantage dans l'utilisation du web et sont par conséquent plus disposés à livrer des informations ⁴¹.

Dans les enquêtes ménages déplacements, il est demandé aux répondants de se souvenir des déplacements effectués les jours précédents la passation du questionnaire. Cet important effort de mémoire ne permet pas toujours de collecter de manière exhaustive les petits déplacements ou de restituer les chaînes de déplacements. Si des interviews en face-à-face permettent de recueillir des informations très précises grâce aux relances et à la vigilance de l'enquêteur, leur coût ne rend pas possible un fréquent renouvellement de l'enquête. A contrario, les enquêtes web souffrent des biais propres aux enquêtes administrées, une défaillance de la mémoire du répondant pouvant fausser les réponses, mais sont peu coûteuses. Une méthodologie faisant appel au web et du face-à-face permet d'exploiter les atouts des deux modes. Il est donc envisageable de continuer à mener des enquêtes ménages déplacements en face-à-face pour recueillir des informations précises sur l'évolution des comportements de mobilité des ménages, et de compléter ces données avec une enquête web afin d'améliorer le taux de réponse et de toucher des individus qui ne répondent pas en face-à-face.

⁴¹Les technologies de l'information et de la communication se diffusent progressivement dans la population. Les 'adopteurs précoces' ont un profil particulier, souvent sous-représenté dans les enquêtes traditionnelles. L'utilisation du web est donc un moyen de réduire certains biais d'échantillonnage (Cf. section I.3.3).

II.3.3 Des problèmes de comparabilité

Proposer plusieurs modes de recueil de données n'est pas sans risque sur la validité de l'enquête, des variations dans les questionnaires utilisés pouvant influencer les réponses (Christensen, 2004). Il est donc préférable de construire un questionnaire simple, qui sera adaptable facilement sur le web sans nécessiter de changements majeurs, et induira peu de différence sur les réponses (Cobanoglu *et al.*, 2001). A l'inverse, un questionnaire web contenant de nombreux filtres risque d'aboutir à une version papier très complexe et difficilement administrable (Alsnih, 2004). Il existe des différences culturelles sous-jacentes au choix du mode d'enquête (Morris et Adler, 2003; Jeavons, 1999), et les ménages choisissant de répondre par Internet peuvent avoir des comportements de mobilité différents de ceux qui répondent en face-à-face. Selon une étude de Ressource System Group (2002), les répondants web sont plutôt jeunes, perçoivent des revenus élevés et ont une mobilité supérieure aux répondants par voie postale ou par téléphone (entre 7.4 et 10.1 déplacements quotidiens par ménage, ce qui représente une mobilité plus élevée de 18% à 45% par rapport aux autres modes.). Les ménages qui répondent sur internet sont de plus grande taille et ont un statut socio-économique plus élevé que les répondants par voie postale ou par téléphone (Arentze *et al.*, 2004), ce qui impacte positivement leur niveau de mobilité.

Dans le cas d'une enquête en face-à-face, la communication est essentiellement auditive, alors que l'utilisation du web favorise le langage visuel, sous contrôle du répondant (Dillman, 2000) (textes, graphiques et informations numériques (Couper *et al.*, 2001). La taille et le type des lettres, mais aussi leur couleur, les animations et les symboles sont autant de détails de présentation qui ajoutent de la compréhension au texte, mais qui peuvent éloigner du sens des questions. Ces différents signaux, qui envoient des messages conflictuels aux répondants, affectent le taux de réponse et la qualité des informations saisies (Dillman *et al.*, 1999). L'effet est particulièrement important dans les questionnaires web, où le potentiel du multimédia autorise la combinaison de plusieurs langages. Ainsi, la combinaison d'une enquête web et d'une enquête en face-à-face pose des problèmes de comparabilité, le support de communication entre l'enquêteur et l'enquêté étant très différent (Dillman *et al.*, 2001). Les effets de modes varient selon le type de questions (faibles pour les questions socio-démographiques et élevés pour les questions d'attitude et d'opinion), puisque certaines appellent des réponses préformatées, alors que d'autres amènent le répondant à écouter la question, l'échelle de réponse et à réfléchir (Dillman et Christian, 2005).

Dans les protocoles d'enquête mixte, la question de l'intégration des données se pose : les données collectées avec des modes différents peuvent-elles être combinées et comparées ? La structure des questions change quelquefois, et le temps des enquêtes n'est pas fixe. Après une sérieuse réflexion, le chercheur doit décider si les effets de modes prévus sont assez sérieux pour ne pas recourir à la mixité des modes d'enquête ou si les avantages d'une telle démarche nécessitent d'en prendre le risque. Le but est d'optimiser la procédure de recueil de données et de réduire les biais pour un temps et un budget donné. La

mixité des modes donne une opportunité de compenser la faiblesse de chaque mode à un coût acceptable. Il y a donc moins de coût et d'erreur qu'avec un seul mode. Le trade-off se fait entre les coûts et les erreurs, particulièrement l'erreur d'échantillonnage (De Leeuw, 2005), afin de sélectionner le meilleur protocole d'enquête.

III Les caractéristiques des internautes français

Internet est l'adoption technologique la plus rapide de l'histoire, et l'utilisation de ce nouveau média s'inscrit pleinement dans la logique de l'évolution des modes de communication des individus. Internet est en effet devenu progressivement un outil d'expression, de consommation et de divertissement (commerce en ligne, 'chat' ...). En 2006, environ six adultes sur six possédaient un ordinateur à domicile (57%), et 43% étaient équipés d'une connexion internet chez eux (Bigot et Croutte, 2007)⁴². Ces taux sont passés respectivement à 64% et 53% en 2007, les 10 points de gagnés en taux de pénétration d'internet représentant plus de 4 millions de personnes équipées (annexe I). Si les internautes ont un profil de plus en plus varié, il est encore légitime de supposer que certaines catégories de la population, davantage concernées que d'autres par la 'fracture numérique', n'ont pas un accès direct à ce nouveau moyen de communication⁴³. Afin de mieux comprendre quels segments de la population sont potentiellement les plus concernés par les enquêtes web, il est nécessaire de considérer trois composantes : la possession d'un ordinateur, la capacité à se connecter à internet et le niveau de familiarité des individus avec le web.

III.1 L'ordinateur, un bien de consommation largement diffusé

Dans cette section, les données concernent l'année de réalisation de l'enquête ménages déplacements sur le web, c'est-à-dire 2006. La progression des indicateurs observée depuis confortent nos hypothèses sur la légitimité des enquêtes web.

Près des deux tiers de la population était équipée d'un ordinateur à domicile (60%) en 2006, et une personne interrogée sur 10 déclarait en posséder plusieurs (Bigot et Croutte, 2007). Le rythme de progression, faible durant des années, s'est nettement accéléré depuis 1997, grâce notamment à l'attrait d'internet (figure 4)⁴⁴.

⁴²Enquête réalisée en juin 2007, auprès de 2015 personnes de 18 ans et plus et 215 individus de 12 à 17 ans, en face-à-face, au domicile. Les échantillons sont représentatifs et sélectionnés selon la méthode des quotas.

⁴³Dans une étude comparative menée sur le web et par téléphone, Flemming et Sonner (1999) ont montré que les personnes qui choisissaient de répondre en ligne étaient plutôt des hommes, jeunes et diplômés, dont les réponses sur des questions d'opinion publique variaient significativement par rapport aux répondants par téléphone.

⁴⁴Il y a donc un effet d'entraînement très net, puisque dès qu'une personne possède un équipement technologique (de type téléphone mobile, GPS, ordinateur, internet...), sa probabilité d'en acquérir d'autres s'accroît. Dans ce contexte, internet apparaît comme le prin-

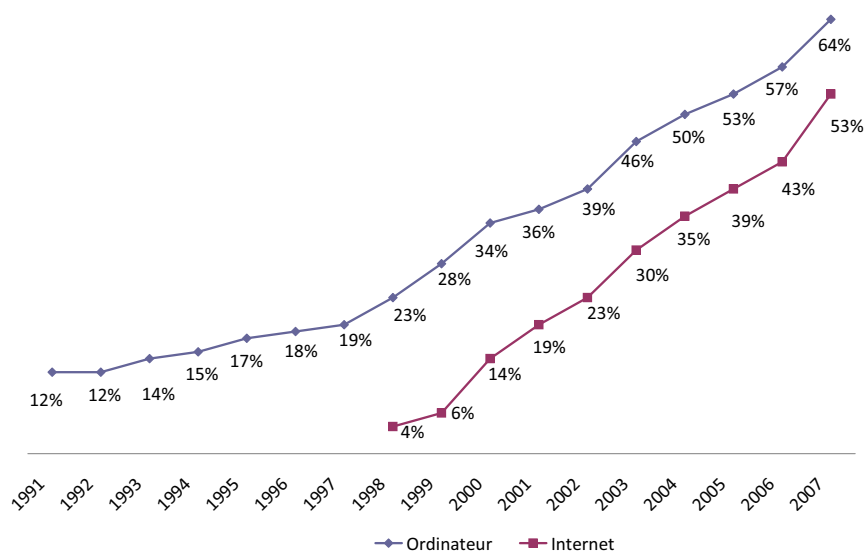


FIG. 4 – Evolution du taux de pénétration de l'ordinateur et d'internet dans les ménages français

Source : CREDOC, enquête "Conditions de vie et Aspirations des Français" (2007)

Des chercheurs ont montré la relation entre les variables sociodémographiques et la peur de l'ordinateur (Igarria et Parasurman, 1989). Globalement, le niveau de diplôme est le critère le plus déterminant pour expliquer la possession d'un ordinateur (86% des diplômés du supérieur), suivi du nombre de personnes dans le foyer (80% des ménages de 4 personnes et plus) et du revenu⁴⁵. L'accès à l'ordinateur est quasiment généralisé chez les cadres supérieurs (89%) et les professions intermédiaires (86%). Les personnes de plus bas statut social tendent à avoir moins d'expérience et d'attitude positive avec l'ordinateur (Shashaani, 1994). Vient ensuite l'âge, puisque 83% des jeunes de 12-17 ans disposaient d'un ordinateur, contre seulement 17% des plus de 70 ans (annexe II). Si les répondants âgés sont moins confortables dans l'utilisation des applications informatiques, les jeunes ont une attitude positive et une certaine familiarité avec l'ordinateur. Ils ont davantage confiance dans la technologie (Wright *et al.*, 1998). L'effet de l'âge va cependant s'estomper, avec les cohortes de personnes familières avec l'informatique qui progressent (Couper et Rowe, 1996).

Depuis 2006, il est intéressant de noter que les progressions les plus importantes du taux de pénétration de l'ordinateur se font au sein des catégories les plus démunies, et que les inégalités d'équipement se réduisent (annexe II).

cipal levier dans l'équipement informatique des ménages.

⁴⁵C'est dans les foyers les plus favorisés que le taux de multi équipement est le plus élevé, puisque 41% des cadres sup possèdent plus d'un ordinateur.

III.2 Une progression rapide des connexions internet dans les ménages français

Lors de la réalisation de l'enquête web de 2006, 43% des personnes de 18 ans et plus étaient connectées à internet à leur domicile, soit environ 21 millions d'adultes sur l'ensemble du territoire, et ces chiffres n'ont cessé de progresser depuis ⁴⁶. De façon de plus en plus systématique, la présence au domicile d'un micro ordinateur se traduit par une connexion à internet (75% des personnes qui disposent d'un ordinateur ont également une connexion internet) et on observe les mêmes disparités en termes de catégories socodémographiques (figure 5).

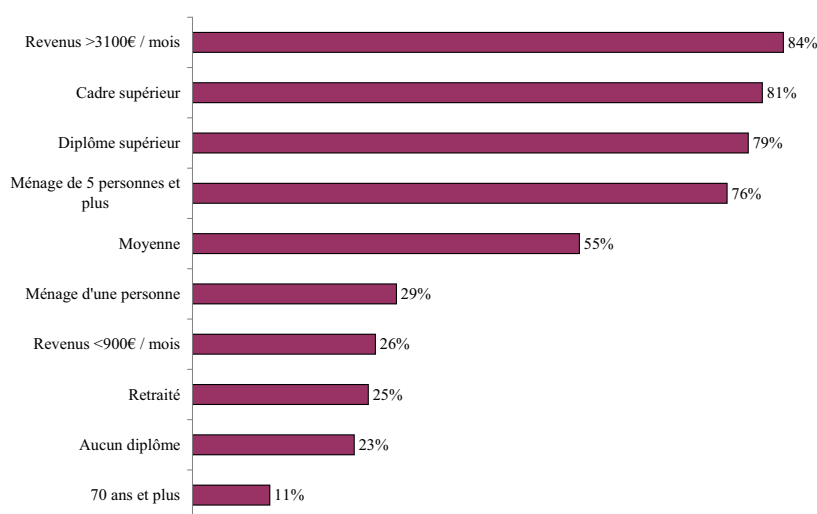


FIG. 5 – Taux d'équipement à internet au domicile
Source : CREDOC, enquête "Conditions de vie et Aspirations des Français" (2007)

Le web reste peu accessible aux personnes âgées (6% des plus de 70 ans et 14% des retraités seulement sont connectés chez eux), aux individus peu qualifiés (un non diplômé sur 5 seulement dispose d'une connexion internet à domicile, contre près des trois quarts des diplômés du supérieur), et aux personnes disposant d'un faible niveau de revenu (22% sont connectées) (annexe II).

Cependant, un effet de rattrapage s'opère, car les progressions les plus importantes du taux de pénétration d'internet sont à mettre au profit de catégories plutôt sous-équipées, comme les 60 ans et plus, les personnes au foyer et les revenus moyens. Nous assistons à une normalisation du sexe et de l'âge, et les personnes âgées sont de plus en plus sensibilisées. Cette évolution rapide des internautes génère une meilleure représentativité des enquêtes web (Mediametrix, 2000).

⁴⁶En juillet 2008, la proportion de français de plus de 11 ans qui se sont connectés au moins une fois à internet a passé la barre des 60%, soit 31,77 millions de personnes (Médiamétrie).

III.2.1 Des français au bord de la route du haut débit

En 2006, 88% des connexions à internet des particuliers étaient à haut débit, alors qu'en 2004 cela ne concernait qu'une connexion sur deux (55%). Ces connexions se font à 95% grâce à l'ADSL. Le taux d'accès au haut débit le plus bas est enregistré chez les retraités, alors que les travailleurs indépendants et les habitants des communes de plus de 100 000 habitants ont un accès quasiment total au haut débit (95%). Cette différence d'accès résulte de l'inégalité d'équipement en ordinateur, mais également de l'architecture du réseau téléphonique français, puisque 2% des lignes téléphoniques sont inéligibles à l'ADSL, 8% ne permettent pas un accès à plus de 512Kbits/seconde et 27% ne permettent pas un accès à plus de 2Mbits/seconde⁴⁷. Des travaux sont en cours, et c'est selon le critère géographique que le rattrapage est le plus spectaculaire : les ruraux souffraient en 2005 d'un retard d'accès au haut débit de 27 points par rapports à l'agglomération parisienne. Il était de 18 points en 2006 et est seulement de 8 points aujourd'hui.

III.2.2 Quels accès hors du domicile ?

Internet est un outil de communication largement utilisé par les entreprises. L'accès personnel au web sur le lieu de travail n'est pas linéaire. Son évolution a connu une pause en 2005 et 2006 (38% des actifs), puis est repartie en 2007 (44% des actifs). La situation reste fortement contrastée entre les groupes sociaux, puisque plus des trois quarts des cadres supérieurs ont accès à internet sur leur lieu de travail en 2006 (74%), contre seulement 17% des ouvriers. Seule une minorité d'actifs utilisent internet à des fins personnelles sur le lieu de travail, mais les individus concernés sont de plus nombreux et assidus (connexions quotidiennes). Plus de la moitié des étudiants utilisent internet sur leur lieu d'études, mais ces connexions sont ponctuelles.

D'autres solutions existent pour naviguer sur internet, notamment dans des lieux publics (cybercafé, wi-fi, liaison avec le téléphone mobile ...). Il serait néanmoins illusoire de penser que les individus qui n'ont pas accès à internet à leur domicile ou sur leur lieu de travail/études profitent de ces évolutions technologiques pour naviguer régulièrement sur internet. En 2006, les connexions dites 'nomades' sur le web ne concernait que 17% de la population (principalement des étudiants de moins de 25 ans, des personnes diplômées du supérieur et ayant des forts revenus), et ne progressent pas. Le domicile reste donc le lieu de connexion privilégié à internet, suivi du lieu de travail (tableau 6).

III.3 Des usages très variés, selon les caractéristiques des internautes

Sur l'ensemble du territoire, 41% des individus utilisent quotidiennement un ordinateur chez eux, 41% n'ont pas d'ordinateur ou ne l'utilisent jamais

⁴⁷Ces foyers, très nombreux en zones rurales, mais l'arrivée de la fibre optique et du très haut débit, telle qu'annoncée par les trois plus gros opérateurs de télécommunication français, va réduire cette fracture.

	2005	2006	2007
Connexion au domicile	37%	42%	51%
Connexion au travail ou études	25%	27%	31%
Dans cyber-café, bibliothèques	11%	13%	12%
Dans lieu public (wi-fi)	2%	5%	4%
Sur le téléphone mobile	7%	6%	6%
Total	527%	55%	62%

TAB. 6 – Lieux de connexion des internautes français
Source : CREDOC (population de 12 ans et plus)

(personnes âgées, non diplômées, retraités et possédant de faibles revenus) et 19% en ont un mais l'utilisent plus rarement. Si le sexe n'est pas prédictif, l'âge, le niveau de diplôme (80% des diplômés du supérieurs l'utilisent tous les jours, vs. 13% des non diplômés) et le revenu semblent déterminants. Le lieu de résidence révèle quelques écarts, puisqu'à Paris et dans les grandes agglomérations il ya deux fois plus d'utilisateurs quotidiens que de personnes qui ne l'utilisent jamais ⁴⁸. Chez les actifs, la polarisation est très grande : 71% des cadres utilisent l'ordinateur tous les jours au travail, alors que 79% des ouvriers ne l'utilisent jamais. Les étudiants se trouvent dans une situation intermédiaire, puisque 45% d'entre-eux utilisent un ordinateur sur leur lieu d'études en 2006.

Parmi les personnes qui disposent d'un ordinateur et d'une connexion à internet au domicile, le taux d'utilisation dépasse 90% depuis 2005. La part de connexion quotidienne est très importante (34% des français de 12 ans en 2006) et, depuis 2007, le nombre d'internautes qui se connectent chaque jour dépasse celui des non internautes. La fréquence de connexion à internet varie selon les caractéristiques sociales, l'âge et le diplôme étant les facteurs les plus discriminants : plus on est jeune diplômé, plus on est assidu. Il est également intéressant de noter une corrélation positive entre les différents types de connexion, puisque 76% des individus qui se connectent dans la sphère professionnelle se connectent également chez eux.

La compétence web des internautes dépend largement des services utilisés (navigateur, e-mail, chat, création de blog ...) (Haraldsen *et al.*, 2002). La diversité des accès aux services et activités proposés par internet dépend souvent de l'âge. Ainsi, les adolescents (12-17 ans) créent beaucoup de sites personnels ou blogs (47%), jouent en réseau (43%), et écoutent la radio (47%) ou regardent la télévision sur le net (23%), alors qu'une faible proportion des adultes sont familiers avec ces activités. En revanche, les achats sur le web progressent beaucoup et touchent tous les internautes (47% ont fait au moins un achat en ligne en 2006). Mais ce sont les démarches administratives ou fiscales par internet qui rencontrent le plus de succès (49% des internautes).

La crainte concernant la protection des données individuelles limite cependant considérablement la navigation sur le web (24% des internautes en 2006,

⁴⁸Cela s'explique en partie par la corrélation du lieu de résidence avec d'autres variables sociodémographiques précédemment citées.

et 29% en 2008 sont sensibles à ce problème). Trois internautes sur quatre souhaiteraient pouvoir s'opposer à la collecte des traces de leur passage. Cette crainte concerne surtout les cadres supérieurs, les 40-59 ans et les diplômés du supérieur, les jeunes y étant moins sensibles. Viennent ensuite la complexité de l'outil et son manque d'utilité, pour les plus de 70 ans, les retraités et les non diplômés. Enfin, malgré les offres commerciales des opérateurs, toujours plus nombreuses, le coût reste un frein pour les personnes à faibles revenus et les jeunes (12-17 ans) qui ne disposent pas d'une connexion illimitée à internet.

III.4 Vers une réduction des inégalités

Les différents taux d'équipement en ordinateur et en internet connaissent un rythme de progression soutenu. Mais des inégalités perdurent, et l'ensemble des groupes n'ont pas identiquement accès à ces technologies. Selon la figure 6, les inégalités sont non négligeables pour l'équipement en ordinateur et internet à domicile.⁴⁹ Les indicateurs d'inégalité évoluent toutefois à la baisse, et en ce qui concerne le taux de pénétration d'internet, les inégalités sont deux fois moins importantes aujourd'hui qu'en 2000. Si le 'fossé numérique' s'est réduit, l'accès à un ordinateur et à internet ne garantit pas une utilisation fréquente de ce nouveau média.

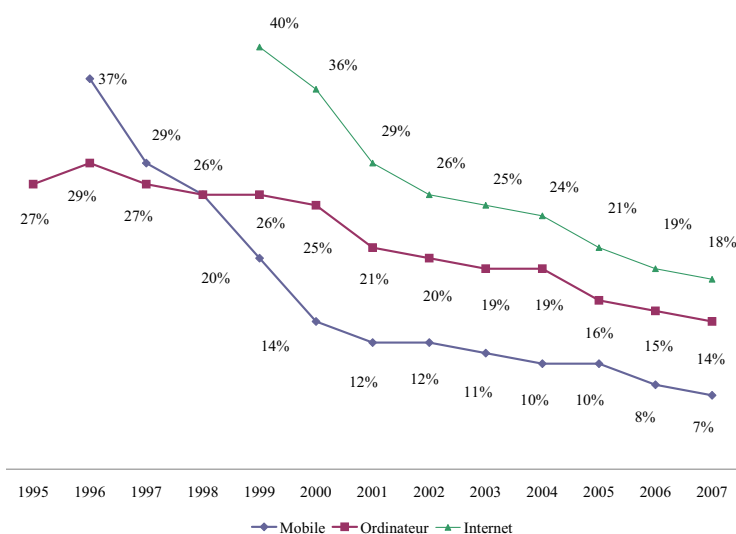


FIG. 6 – Evolution des inégalités d'équipement des ménages français - Indice de Gini

Source : CREDOC, enquête "Conditions de vie et Aspirations des Français" (2007)

Il faut cependant relativiser ces évolutions. Aux USA par exemple, le taux de pénétration d'internet est très élevé (74%). Ceci n'empêche pas qu'un biais

⁴⁹L'indice de Gini est un indicateur synthétique d'inégalités qui varie entre 0 et 100%. Un coefficient de Gini égal à zéro signifie qu'il n'y a pas d'inégalité. Un coefficient égal à 100% signifie qu'un seul groupe détient tout l'équipement et que les autres en sont privés. Entre 0 et 100%, l'inégalité est d'autant plus forte que l'indice de Gini est élevé.

de couverture subsiste, des personnes (bas revenus, faible niveau d'études...) n'ayant toujours pas accès à internet. Si les enquêtes web permettent de toucher des individus qui ne répondent pas aux modes traditionnels, elles excluent toutefois certains groupes sociodémographiques.

IV Conclusion

Les méthodes d'enquête doivent tenir compte de l'évolution des modes de communication entre les individus, puisque la réalisation d'une étude implique l'existence de communications interpersonnelles (Murakami et Morris, 2003). Internet est un espace naturel de communication pour les jeunes générations. C'est un média hautement interactif, dont les récentes applications ont remis l'expression écrite au goût du jour. Ainsi, le développement des blogs, de la messagerie électronique et des forums font que les internautes sont habitués à s'exprimer par l'intermédiaire de leur ordinateur. Les enquêtes web apparaissent comme un moyen de palier à la crise de légitimité que semblent subir le secteur des études, en offrant aux répondants un mode de collecte de données plus proche de leur mode de vie et de communication et en autorisant une grande interactivité, malgré l'absence d'enquêteur pour introduire les questions.

L'adoption rapide d'internet par les foyers français en fait outil intéressant pour les gestionnaires d'enquêtes, qui peuvent interroger rapidement et à moindre coût un grand nombre d'individus. Ils sont capables de concevoir des questionnaires dont le design des questions et des zones de réponses s'adaptent parfaitement aux besoins de l'étude, de lancer quasi-instantanément l'enquête à une large échelle et d'automatiser une grande partie des traitements statistiques (Rose *et al.*, 2004). Le développement du haut-débit en France laisse présager que des enquêtes pourront être menées à grande échelle ces prochaines années, même si internet ne sera peut-être jamais totalement accessible (contraintes de coût, incapacité à naviguer...). En effet, au fil des mois, davantage de foyers sont équipés d'une connexion internet, et les internautes deviennent de plus en plus familiers avec les transactions sécurisées sur le web (qu'il s'agisse de données financières ou personnelles). Avec l'augmentation du nombre des enquêtes en ligne, la population cible s'élargit (Cho et LaRose, 1999) et des études complexes, utilisant des interfaces toujours plus conviviales, vont pouvoir se développer. Par ailleurs, le faible coût des études en ligne permet d'en augmenter la fréquence et de limiter l'impact d'événements spéciaux comme les 'violences urbaines', les grèves ou les attentats sur les résultats (Sharp *et al.*, 2004). Un accent particulier devra cependant être mis sur la confidentialité des données, avec des procédures de contrôle sur le point clé de l'identité du répondant (Michalowska, 2000).

Mais des doutes sur la représentativité des échantillons ont alimenté la polémique autour des enquêtes en ligne, et les interrogations des chercheurs et des professionnels des études rappellent celles posées en leur temps par le développement du recueil de données par téléphone ou par minitel. Les inéga-

lités en termes d'équipement informatique et d'accès au haut-débit (bien que ce dernier soit de plus en plus choisi par les internautes, au détriment du bas débit) compromettent l'utilisation exclusive du web comme mode de recueil de données sur l'ensemble de la population, la représentativité de l'échantillon n'étant plus assurée. Les expériences d'enquêtes web menées jusqu'à aujourd'hui reposent essentiellement sur des méthodes non probabilistes, ce qui limite la qualité des données.

L'objectif des enquêtes ménages déplacements est de fournir régulièrement des données récentes et détaillées pour alimenter les modèles de plus en plus sophistiqués des planificateurs, et permettre aux chercheurs de réaliser des analyses comportementales. Ces enquêtes ne sont pas seulement restreintes à une population d'internautes et nécessitent la construction d'un échantillon qui permette d'appliquer les techniques d'inférence statistique. Il semble que la solution optimale consiste aujourd'hui à intégrer le web dans un protocole d'enquête plus large. Il peut par exemple être utilisé en complément, pour améliorer la performance des modes traditionnels (téléphone, voie postale ou face-à-face). L'utilisation des nouvelles technologies peut en effet potentiellement diminuer la lassitude des répondants et augmenter le taux de réponse global de l'étude, sans grever le coût de l'enquête ⁵⁰. Le couplage du web avec un mode de recueil plus traditionnel, tel que le face-à-face, est une solution intéressante pour améliorer la représentativité et la qualité des enquêtes ménages déplacements, en produisant une image plus fidèle du comportement de mobilité de l'ensemble de la population. L'idée est de réussir à capter de nouveaux individus, tout en conservant la méthode traditionnelle de recueil des données, qui permet de toucher une large partie de la population de référence, et d'étudier en profondeur les comportements de déplacements des ménages.

Dans le chapitre suivant, nous présentons la méthodologie retenue pour tester la réalisation d'une enquête web en complément de l'enquête ménages déplacements, effectuée en face-à-face.

⁵⁰Sauf si l'enquête web nécessite la construction d'un échantillon qui autorise la généralisation des résultats à l'ensemble de la population

Chapitre 2 : La mise en place d'un module web dans l'enquête ménages déplacements de Lyon

"Ne jamais oublier que l'enquête, à un moment donné, échappe à ses concepteurs. Elle n'est dès lors que le résultat de la relation établie entre enquêteur et enquêté et la qualité de cette relation est grandement facilitée par la mise à disposition d'un bon outil de recueil [le questionnaire]."
(Certu, 1998), à propos des enquêtes ménages déplacements..

Notre recherche consiste à tester puis à proposer une méthodologie d'enquête web adaptée au recueil de données sur les déplacements quotidiens. L'enjeu est conséquent, puisqu'il n'existe pas d'expérience similaire en France en ce qui concerne les enquêtes de mobilité, et que la comparabilité des données avec l'enquête standard réalisée en face-à-face n'a pas été éprouvée à ce jour. La conception du questionnaire web est délicate, puisqu'il s'agit non seulement de respecter les contraintes de la version standard des enquêtes ménages déplacements élaborée par le Certu (1998), mais également de tenir compte des spécificités du web en tant que nouveau média d'enquête. En effet, si le web permet d'améliorer la qualité de l'enquête, il est en revanche important de ne pas accroître la charge pour l'enquêté (Bonnell, 2004). Le design affecte non seulement les taux de réponse et d'abandon, mais aussi la qualité des réponses obtenues (Couper *et al.*, 2001). Un arbitrage doit donc être fait entre convivialité et difficulté d'administration du questionnaire.

La première partie de notre recherche a consisté à modifier et clarifier le modèle de questionnaire utilisé pour les enquêtes ménages déplacements standard, afin de s'assurer que chaque question génère une réponse valide et appropriée au format de l'enquête web. Plusieurs sociétés ont été impliquées dans ce projet (Sofreco, en charge de l'enquête traditionnelle en face-à-face et responsable de l'enquête web, Affimétrie, pour la gestion, la mise en place et le suivi de l'enquête en ligne, et Conversoft, développeur du questionnaire web, à partir du logiciel Converso) et ont imposé certaines contraintes techniques et méthodologiques.

Nous analysons d'abord les difficultés relatives à l'élaboration d'un questionnaire web, et les principales recommandations suggérées par la littérature

(section I). Puis, nous présentons en détail la méthode suivie pour la rédaction des questions (section II), avant d'expliquer d'administration de l'étude en ligne (section III).

I Considérations méthodologiques

I.1 Quelques recommandations

La plupart des recommandations concernant la structure du questionnaire web sont adaptées de celles qui s'appliquent aux autres modes d'enquête, telles que la teneur de l'introduction, la présence de sections et de filtres, l'usage limité des questions ouvertes, la mise en place de pré-tests et le rôle des incitations matérielles (Gaddis, 1998). L'utilisation d'Internet génère toutefois des préoccupations nouvelles. Dillman *et al.* (1998a) recensent par exemple pas moins de onze principes à respecter pour concevoir des questionnaires web de qualité. Andrews *et al.* (0003) présentent cinq éléments relatifs à la présentation du questionnaire et l'administration de l'enquête, essentiels pour obtenir des résultats de bonne qualité : la capacité d'accès avec différents équipements, l'accès sécurisé afin d'empêcher les soumissions multiples, la personnalisation des questions, les sauvegardes partielles et les remerciements en fin de questionnaire. Nous proposons dans cette section une revue de la littérature, suivie de l'application concrète dans l'enquête ménages déplacements web réalisée à Lyon.

I.1.1 Un questionnaire accessible au plus grand nombre

Pour limiter les problèmes techniques et augmenter le taux de réponse, il est nécessaire de faciliter l'accès au questionnaire (Morrel-Samuels, 2003). Du point de vue technique, il est évident que le questionnaire doit être conçu de façon à rester identique quel que soit le type de matériel utilisé par l'internaute, et à pouvoir être facilement adaptable sur d'autres supports. Les questions doivent être simples, ne pas nécessiter de compétences particulières pour y répondre, et doivent être conçues de façon à être téléchargeables par différents systèmes d'exploitation (Mac, Windows) et avec différentes versions de navigateurs (Netscape, Internet Explorer...), des problèmes techniques pouvant impacter négativement le taux de réponse (Bishop, 1997). Les internautes les moins expérimentés ont des navigateurs plus anciens. Les plus expérimentés ont des explorateurs 'libres', dernière génération, ou plus complexes et quelquefois non compatibles avec la version optimisée du questionnaire (Haraldsen *et al.*, 2002). Le concepteur doit également tenir compte des limitations liées à la taille de l'écran, qui varie selon l'équipement des internautes et donne des présentations différentes du questionnaire (Couper *et al.*, 2001; Dillman et Bowker, 2001). Des recherches montrent que les questionnaires simples et rapides à télécharger génèrent de meilleurs taux de réponse (Solomon, 2001).

Lors de l'administration du questionnaire, il est probable que des erreurs de saisie de l'identifiant ou du mot de passe, ou plus en amont une confusion

dans le nom du site hébergeant le questionnaire, contraignent le répondant à l'abandon. Il est recommandé d'utiliser des identifiants et mots de passe courts, afin de limiter le risque d'erreur dans la saisie. Selon Schleyer et Forrest (2000) et Couper *et al.* (2001), les répondants dont les identifiants ou mots de passe contiennent des caractères ambigus ont significativement moins de chance de réussir à accéder au questionnaire web que les autres (la lettre 'O' peut être confondue avec le chiffre '0', la lettre 'I' avec le chiffre '1'). Suivant la même logique, l'adresse du site qui héberge le questionnaire doit être simple et ne pas contenir de caractères spéciaux.

Enfin, beaucoup d'enquêtes internet requièrent Java pour fonctionner. Une application est chargée et lancée sur l'ordinateur du répondant, puis le serveur reçoit les réponses du navigateur du répondant et lui envoie les données appropriées au format HTML. Mais les temps de chargement sont relativement longs, ce qui génère des frustrations auprès des répondants. Le désavantage d'utiliser un programme complexe pour une enquête web est que les répondants ne possèdent pas forcément la bonne version de l'application, et que d'autres peuvent vite se retrouver face à un bug (Stopher *et al.*, 2004a). Ce risque est fort sur un environnement internet, où ces facteurs sont difficiles à contrôler (pare-feu).

Pour l'enquête web de Lyon, nous n'avons eu aucun contrôle sur l'attribution des codes personnels (identifiants et mots de passe) générés par la société Sofreco. En revanche, nous avons choisi un nom de site simple et non ambigu (www.enquete.let.fr). Nous n'avons pas opté pour le langage Java, compte-tenu des nombreux inconvénients développés ci-dessus, et nous nous sommes assurés par l'intermédiaire de tests techniques de l'accessibilité et de la présentation du questionnaire pour différents types d'équipement informatique.

I.1.2 Une introduction de qualité

Lorsqu'il accède au questionnaire, le répondant doit trouver une introduction accueillante, motivante et rassurante, en ce qui concerne la confidentialité des données et la sécurité du site (Dillman *et al.*, 1998b; Lazar et Preece, 1999; Zmud, 2003). Dès les premières pages, des informations sur les manipulations requises pour naviguer dans le questionnaire sont fournies (par exemple le moyen d'accéder à la page suivante), et restent accessibles par la suite. L'objectif est de motiver le répondant, non d'accroître sa lassitude en dressant une liste importante d'instructions. Les instructions concernant des questions précises sont divulguées au fur et à mesure de l'avancement dans le questionnaire, le moment le plus approprié étant celui où l'internaute va en avoir besoin (Dillman *et al.*, 1998a) ⁵¹.

Le temps nécessaire pour répondre à l'étude est également indiqué, pour permettre aux internautes pressés de reporter leur participation à l'étude à un moment plus opportun, ce qui réduit leur frustration (Alsnih, 2005). Une barre de progression indique le temps restant avant la fin du questionnaire, bien

⁵¹Le rapport du Ressource System Group (2002) précise que les instructions doivent figurer sur le même écran que la question concernée, afin de simplifier la tâche des répondants.

que son bénéfice soit largement discuté dans la littérature. Certaines études n'y sont pas favorables et Couper *et al.* (2001) met en garde les concepteurs d'enquête. Si cet outil informe le répondant sur sa progression dans le questionnaire et entretient la motivation, l'utilisation d'éléments graphiques trop sophistiqués peut élever inutilement le temps de chargement des pages web et générer au final des abandons. Enfin, il est important de préciser à quelle date les résultats seront disponibles et qui aura accès aux données (Conn, 2005), pour lever les éventuels doutes sur la confidentialité des données.

Dans l'enquête web de Lyon, les deux premières pages donnent des indications concernant la navigation dans le questionnaire ('*Pour sélectionner vos réponses, il suffit de cliquer sur la puce.*') et rassurent le répondant sur la confidentialité et la sécurité des données ('*Vos réponses resteront strictement confidentielles, et leur exploitation sera réalisée de façon anonyme.*'). Le temps nécessaire pour répondre au questionnaire est également précisé ('*Cette enquête prend moins de 20 minutes.*') et une barre de progression permet à l'enquêté de suivre l'avancement du questionnaire (Cf. VI).

I.1.3 Une présentation adaptée au média web

Lorsque le contact avec le répondant a été établi, le questionnaire joue un important rôle de motivation et son design est un élément clé du succès de l'enquête (Andrews *et al.*, 0003). Le questionnaire doit être accueillant et ergonomique, pour donner envie aux internautes d'y répondre (Hojman *et al.*, 2004). Les internautes sont familiers de ce nouveau média et ont des attentes précises en termes de design des pages web. Si l'étude ne fait pas rapidement une bonne impression aux répondants, ils peuvent quitter le questionnaire en fermant la fenêtre web par un simple 'clic'. La mise en page, les questions et les modalités de réponse vont affecter le taux de réponse, le taux d'abandon et les réponses elles-mêmes (Dillman *et al.*, 1998a; Zmud, 2003). Selon Alvarez et Van Beselaere (2003), la validité et la pertinence des résultats d'une enquête web dépendent de la manière dont l'enquête est conçue et administrée. Les éléments du questionnaires qui sont difficiles à comprendre, qui nécessitent un temps de téléchargement important ou qui semblent inintéressants pour les enquêtés diminuent les chances de réponse. Ainsi, trop de questions ouvertes, une présentation complexe ou au contraire trop simpliste découragent les répondants. Suivant le type d'étude et le public visé, certains designs sont plus appropriés (Couper, 2000).

Des auteurs suggèrent que le questionnaire web propose les questions d'une manière différente de la version papier, puisque les individus appréhendent différemment les questionnaires lorsqu'ils sont face à un ordinateur (Dillman et Bowker, 2001). Les concepteurs d'enquêtes web tirent partie des logiciels performants et de la capacité de traitement des ordinateurs pour créer des questionnaires très sophistiqués. Par ailleurs, les capacités graphiques des explorateurs de type 'Internet Explorer' ou 'Netscape' permettent d'agrémenter

les enquêtes web de documents multimédia ⁵². Les conséquences sont toutefois assez paradoxales. D'un côté, les internautes sont séduits par l'attractivité des questionnaires, de l'autre la complexité de l'outil rend son accès et son utilisation difficile pour certains répondants (Dillman *et al.*, 1998b; Couper *et al.*, 2001; Dillman *et al.*, 1998a). La plupart des possibilités techniques requièrent en effet des ordinateurs performants, des logiciels récents et un temps de téléchargement conséquent pour pouvoir fonctionner. Avec le développement des capacités multimédia, l'écart entre ce que les concepteurs d'enquête peuvent mettre en ligne et ce à quoi les internautes peuvent accéder et répondre se creuse ⁵³. Les possibilités multimédia du web ont donc parfois un effet paradoxal, puisqu'elles rendent les questionnaires en ligne plus difficiles à administrer.

Selon Smith (1997), l'aspect visuel du questionnaire est très important, puisque ce sont les aspects non verbaux qui influencent les réponses aux questionnaires web. Or, ce que voit le répondant sur son ordinateur ne correspond pas forcément à ce que voit le concepteur de l'enquête. Andrews *et al.* (2003) montrent que les phrases courtes sont plus facilement lisibles à l'écran, les internautes ne lisant pas tout le texte, mais balayant du regard l'écran, à la recherche de mots clé. Il est également important que les concepteurs d'enquête ne se focalisent pas seulement sur la manière dont opèrent les ordinateurs, mais également sur la façon dont les répondants espèrent que l'enquête va se dérouler. Dillman *et al.* (1998a) précisent que lorsqu'un répondant est sollicité, il va se remémorer ses expériences passées. Les questionnaires en rupture totale avec les méthodes traditionnelles risquent de dérouter les internautes et de conduire à un échec ⁵⁴. Enfin, la standardisation de la présentation donne un rythme et assure une certaine continuité du questionnaire. A contrario, de fréquents changements de formats rendent la lecture et la compréhension du questionnaire plus difficiles. Il est donc préférable de faire varier la police de caractères avec parcimonie, pour créer un contraste sans dérouter les répondants. La forme des réponses peut également impacter les résultats. Couper *et al.* (2001) précise que sélectionner une valeur à l'aide d'un 'bouton radio' requiert moins d'effort que saisir une valeur dans une case vide. Les non-réponses sont moins importantes, puisque le répondant est obligé de choisir une valeur, et le traitement des données est simplifié, puisque les réponses sont prédéfinies par le concepteur de l'étude.

Les contraintes techniques que nous détaillons en section I.2.5 ne nous ont pas permis d'optimiser la présentation du questionnaire web lors de l'enquête ménages déplacements de Lyon. Ce dernier étant développé à partir d'un logi-

⁵²Selon (Bachmann *et al.*, 2000), des couleurs, des images de bonne qualité et des graphiques complexes peuvent être utilisés sur le web, ainsi que des commentaires audio et vidéo.

⁵³Le risque est de décourager les personnes les moins familières avec l'informatique, de produire un biais de non-réponse important et de recueillir des données de faible qualité.

⁵⁴Il est recommandé que la forme des questions, tout en respectant les contraintes propres à ce média, se rapproche du format conventionnel que les internautes sont habitués à trouver dans les enquêtes papier (Dillman *et al.*, 1998a).

ciel CATI, il n'a pas été possible de bénéficier de tous les avantages des logiciels CAWI (ergonomie, interface graphique...).

I.1.4 Une progressivité des questions

Une enquête web est par définition auto-administrée. Le rôle essentiel du créateur de l'enquête est d'anticiper au mieux la façon dont les individus vont appréhender le questionnaire, afin de les encourager à saisir l'ensemble des réponses souhaitées (Dillman *et al.*, 1998a). Selon Mertler (2002), les répondants doivent être en mesure de remplir le questionnaire web avec la même aisance que la version papier. La proximité des éléments est importante, et il faut veiller à grouper les questions et les items qui sont proches, pour faciliter le travail de l'interviewé (Conn, 2005). Le questionnaire web doit être concis, structuré par parties (Gunn, 2002; Alsnih, 2005), reliées par quelques phrases de transition. Des termes tels que 'finalement' peuvent être employés en introduction de la dernière partie, pour indiquer au répondant qu'il est proche de la fin du questionnaire, et éviter ainsi les abandons tardifs, grande source d'insatisfaction pour l'enquêté et l'administrateur de l'enquête (Gunn, 2002; Alsnih, 2004).

Concernant le contenu des questions, une progressivité doit être respectée (Frery, 1996). Il est judicieux de commencer par des questions simples et en relation avec le sujet de l'étude, afin d'encourager le répondant à rentrer dans le questionnaire (Ampt, 2003)⁵⁵. Puis, apparaissent les questions plus complexes, et en dernier lieu les questions personnelles. La connaissance des caractéristiques individuelles est nécessaire pour permettre le traitement statistique des réponses, mais ce type de question peut inquiéter l'individu, surtout dans le cas d'un questionnaire auto-administré, et conduire à l'abandon ou à des non-réponses partielles. Si nombre d'auteurs recommandent de les placer en dernier, Andrews *et al.* (2003) mettent en garde l'effet de surprise que peut provoquer l'apparition de questions personnelles à la fin du questionnaire, ce sentiment de 'trahison' pouvant conduire à un abandon.

Le questionnaire web de Lyon est structuré en quatre parties (Cf. section II.1). Afin de ne pas inquiéter le répondant au point que ce dernier abandonne l'étude, nous avons choisi de commencer le questionnaire par des questions simples sur la motorisation, le lieu d'habitation, et la composition du ménage. Suivent des questions concernant les principales caractéristiques socio-économiques des répondants. Ceci permet d'établir plus rapidement un lien avec les pratiques de déplacement et de renvoyer les questions relatives à l'équipement et aux ressources du ménage dans la dernière partie du questionnaire.

⁵⁵Ainsi, la première question, qui concerne directement le sujet de l'étude, doit être facilement compréhensible et s'adresser à l'ensemble de l'échantillon (Gaddis, 1998; Dillman *et al.*, 1998a).

I.1.5 Limiter les non-réponses

Les internautes ouvrent généralement plusieurs pages web et exercent différentes tâches simultanément (courrier électronique, chat, messengers ...). Leur niveau de concentration et d'attention n'est pas optimal, la durée de réponse au questionnaire plus longue et la proportion de non-réponses partielles parfois élevée (Lozar Manfreda et Vehovar, 2002b). Le problème de la non-réponse partielle est plus important lorsque l'enquête est réalisée sur le web qu'en face-à-face, car aucun enquêteur ne vient relancer le répondant⁵⁶. S'il est possible, en cas de non-réponse, d'avertir le répondant par un message d'alerte ou de forcer la réponse en empêchant le passage à la question suivante, ces choix méthodologiques ne sont pas neutres. D'un côté, nous risquons d'obtenir des réponses fausses ou incohérentes, voire des abandons, puisque le questionnaire est auto-administré et que l'enquêteur n'est pas présent pour motiver l'interviewé (Dillman *et al.*, 1998b; Dillman, 2000; Abraham *et al.*, 1998). Les répondants font davantage d'erreurs quand la pénibilité augmente, et les messages d'erreur renforcent cette pénibilité. De l'autre, nous pouvons penser que les individus qui font la démarche de répondre comprennent plus facilement la rigidité de la structure du questionnaire. Au final, il est utile de décider d'un niveau de tolérance, afin d'éviter de demander des confirmations trop fréquentes ou des corrections qui sèment le doute et augmentent la pénibilité et le caractère inquisiteur des questions (Bonnell, 2003)⁵⁷.

La solution intermédiaire consiste à laisser un échappatoire par la présence d'une modalité 'Ne Sait Pas' ou 'Autre, à préciser' aussi souvent que possible. Les recherches sur le rôle de ces échappatoires dans les enquêtes sont nombreuses (Krosnick et Fabrigar, 1997; Schuman et Presser, 1981; Schuman et Scott, 1989). Elles montrent que proposer une modalité 'ne sait pas' augmente la part des répondants qui la choisissent et nuit à la qualité des données puisque l'analyste est privé d'information. Par ailleurs, la réponse 'Ne Sait Pas' est peu pertinente, si les modalités sont exhaustives. Elle reste en revanche utile pour réduire la frustration d'individus qui ne souhaitent pas s'exprimer sur une partie du questionnaire, quelle qu'en soit la raison. Dans les modes administrés, l'enquêteur peut avoir comme consigne de relancer l'individu qui ne veut pas répondre, et de ne proposer la modalité 'ne sait pas' qu'en dernier ressort. Cette démarche est moins évidente lors des enquêtes administrées, bien que le web autorise une utilisation modérée de ces 'non-réponses'. Il est en effet possible de relancer les répondants à l'aide d'une alerte, avant de proposer explicitement un échappatoire ou de proposer la modalité 'ne sait pas' (éventuellement en surbrillance, pour diminuer son impact visuel) (De Rouvray et Couper, 2002). L'impact de ces différentes formulations sur le comportement de réponse des internautes n'est pas encore clairement démontré. De Rouvray et Couper (2002) rapportent que le taux de données manquantes est plus faible

⁵⁶Haraldsen *et al.* (2002) montrent, dans une expérience où les réponses ne sont pas forcées, 98.6% des questionnaires contiennent au moins une erreur et 82.6% au moins une non-réponse à une question principale.

⁵⁷Des relances simples sont plus efficaces. Elles incitent le répondant à fournir des réponses complètes, sans accroître la pénibilité du questionnaire (Zuckerberg *et al.*, 1999).

lorsqu'une relance est prévue, combinée avec une modalité 'ne sait pas' et il semblerait que la proportion de réponse 'ne sait pas' reste inchangée, selon que les répondants visualisent ou non la relance. On peut donc supposer que cette technique permette aux individus qui auraient évité la question de fournir des réponses pertinentes.

En encourageant les enquêtés à donner une réponse à chaque question, améliorons-nous la qualité des données ou augmentons-nous le 'bruit' dans les résultats ? Converse (1974) rappelle que *'instead of browbeating our respondents into giving opinions they did not feel they possessed . . . we explicitly invited respondents who had non opinion to report this fact directly'*.

Dans l'enquête web de Lyon, nous avons choisi de proposer des modalités 'Ne sait pas' ou 'Non-réponse' aux questions complémentaires concernant l'équipement en moyens de communication (téléphone fixe et portable, connexion internet à domicile et type de connexion internet à domicile) et le revenu des ménages. Ces questions sont situées à la fin du questionnaire, et risquent d'apparaître trop curieuses auprès des répondants qui viennent de renseigner le cycle de déplacements, sujet central de l'étude.

I.1.6 Les sauvegardes partielles des réponses

Nous savons que les problèmes techniques pénalisent souvent les enquêtes web, et que des perturbations de la connexion peuvent se produire aux dépens du gestionnaire d'enquête. La navigation dans le questionnaire s'interrompt et la saisie des réponses n'est plus possible (Alsnih, 2005). Les individus sont également susceptibles d'être dérangés momentanément au domicile ou au travail, et contraints de suspendre la saisie des réponses. Il est donc indispensable que les répondants puissent reprendre le questionnaire ultérieurement, sans que l'ensemble des données initialement saisies soient perdues. En structurant le questionnaire, il est possible d'enregistrer les réponses au fur et à mesure⁵⁸. Ces sauvegardes partielles demandent peu de temps de chargement et, en cas d'interruption de la saisie, l'internaute est en mesure de rentrer à nouveau dans le questionnaire à l'endroit où il l'avait quitté.

Une des solutions consiste à affecter un mot de passe personnel à chaque individu, qui reste valide durant toute la durée de l'enquête. Pour accéder au questionnaire en ligne, le répondant doit saisir à chaque connexion son identifiant et son mot de passe. Cette identification permet par ailleurs de contrôler l'échantillon, puisqu'un même individu ne peut remplir plusieurs questionnaires (Conn, 2005; Couper *et al.*, 2001). Le risque est que les interviewés ne comprennent pas le caractère aléatoire de l'affectation des mots de passe et perçoivent cette mesure comme une atteinte à la confidentialité de l'étude, ce qui peut modifier leurs réponses (Stanton et Rogelberg, 2001).

⁵⁸Certains questionnaires proposent même un bouton 'click-save-exit' sur chaque page, permettant au répondant de quitter le questionnaire à tout moment sans perdre les informations déjà saisies (Gaddis, 1998; Morrel-Samuels, 2003).

Dans l'enquête web de Lyon, un identifiant et un mot de passe figurent sur la lettre avis envoyée aux ménages sélectionnés. Ces informations sont nécessaires pour accéder au questionnaire et permettent d'enregistrer les réponses saisies. Ces sauvegardes interviennent à plusieurs endroits du questionnaire (Cf. section III.3).

I.1.7 Les filtres et tests de cohérence : personnalisation des questions et qualité des données recueillies

De nombreux filtres sont présents dans les questionnaires web. Ils permettent d'éviter la redondance des questions ou les questions non pertinentes, qui allongeraient inutilement la durée du questionnaire et limiteraient sa crédibilité. Leur rôle est encore plus fort que dans le cas d'un questionnaire administré en face-à-face ou par téléphone, puisque l'interviewé ne connaît pas a priori le questionnaire et ne sait pas comment naviguer à l'intérieur de celui-ci. Par ailleurs, il est fréquent que des tests de cohérence opèrent au fur et à mesure de la présentation des questions. L'objectif est de vérifier la structure du fichier et de détecter des contradictions éventuelles dans les réponses saisies. Soit l'internaute valide les informations saisies et est dirigé vers la suite de l'enquête, soit il modifie sa réponse. La forme de la relance, privilégie la souplesse et l'ergonomie pour l'enquêté.

Dans l'enquête web de Lyon, nous avons choisi de placer des filtres aussi souvent que possible, afin de personnaliser le questionnaire. Des tests de cohérence sont également utilisés, surtout dans la partie relative aux déplacements, pour s'assurer de la pertinence des réponses. Par exemple, si le répondant déclare ne pas être sorti de chez lui durant la période d'enquête ou si l'heure de fin d'un déplacement est postérieure à celle de fin de l'enquête, une alerte est envoyée pour demander des corrections. Il est également important de s'assurer que les personnes qui ne déclarent pas de déplacements soient des vrais immobiles. Les utilisateurs de données assurent en effet qu'une large proportion de non mobiles témoigne de la faible qualité d'une enquête ⁵⁹. Pour certaines questions, l'utilisation de tests de cohérence est plus délicate. Ainsi, lorsqu'un répondant ne déclare pas son domicile comme lieu de destination du dernier déplacement, il est probable qu'il n'ait pas correctement assimilé la définition du terme 'déplacement'. Mais il se peut également qu'il n'ait pas passé la nuit chez lui. Dans ce cas, nous n'avons pas placé de relance. L'objectif est d'obtenir des données de qualité, pas de ralentir inutilement la durée de l'enquête.

I.1.8 La mise en oeuvre du pré-test

Avant de lancer une enquête, il est fortement conseillé de vérifier sur un petit échantillon de la population si les principaux objectifs de l'étude sont atteints et si les résultats sont bien ceux espérés (Zmud, 2003). Selon Oppenheim

⁵⁹Au moment du codage, il peut être intéressant de différencier les personnes qui ne se sont pas déplacées de celles qui refusent de dévoiler des informations sur leurs comportements de mobilité.

(1992), '*Survey piloting is the process of conceptualizing and re-conceptualizing the key aims of the study, and making preparations for the fieldwork and analysis so that not too much will go wrong and nothing will have been left out*'. Il s'agit de tester la bonne compréhension des questions, la faisabilité technique de l'étude par rapport aux différents équipements existants sur le marché, d'estimer de façon réaliste le temps de réponse et le bon positionnement des filtres (Alsnih, 2004). Les fréquentes erreurs rattrapées grâce au pré-test sont : le biais dans la formulation des questions ⁶⁰, les erreurs de rédaction, les questions inappropriées, les instructions manquantes, le vocabulaire technique non défini, le manque de place pour répondre aux questions ouvertes, et le manque de techniques pour maintenir la motivation du répondant (Andrews *et al.*, 2003). Les tests techniques aident au contrôle des procédures de stockage de l'information, à la gestion et à la récupération des données collectées et à la compatibilité des navigateurs (Benfield et Szlemko, 2006). Nous avons vu dans la section I.1.1 qu'il est recommandé de vérifier la présentation du questionnaire pour des versions anciennes de navigateurs (Dillman et Bowker, 2001). Il est également souhaitable de tester la capacité du serveur sur lequel va être hébergé le questionnaire, en envoyant simultanément plusieurs requêtes sur le site web dédié à l'enquête.

Les pré-tests peuvent dans un premier temps concerner des personnes familières avec l'étude, pour déceler les principaux problèmes, puis être menés sur une plus grande échelle, auprès d'individus de la population étudiée. Selon Morris et Adler (2003), au moins trente questionnaires doivent être remplis pour pouvoir tirer des conclusions du pré-test, les individus concernés étant ensuite exclus de la base de sondage ⁶¹. Afin de tester diverses versions du questionnaire, il arrive que plusieurs pré-tests soient menés successivement.

Pour l'enquête web de Lyon, nous avons réalisé deux séries de pré-tests, tant pour évaluer la faisabilité technique du questionnaire que tester sa bonne compréhension par les internautes. L'objectif était de pouvoir améliorer sensiblement la qualité du questionnaire, avant de lancer l'étude auprès des non-répondants à l'enquête standard. Nous avons d'abord réalisé un test de validation en interne auprès de dix personnes familières des enquêtes ménages déplacements, afin d'évaluer la pertinence de la méthodologie retenue pour le recueil des déplacements individuels. Dans un deuxième temps, la version finale du questionnaire a été diffusée auprès de personnes non familières des enquêtes de déplacements, aux caractéristiques socio-économiques diverses. Des entretiens semi-directifs (téléphoniques ou en face-à-face) ont été menés auprès de ces répondants, suivant des thèmes déterminés à l'avance (annexe III) :

- *Evaluation de la durée du questionnaire* : minuter le questionnaire, pour avoir une estimation fiable du temps de réponse, évaluer l'impact de la barre de progression, s'interroger sur la 'lourdeur' du questionnaire et vérifier si certaines questions doivent être supprimées.

⁶⁰Par exemple si on obtient une grande proportion de réponse 'Autre' (Pratt, 2003).

⁶¹Un pré-test diffère d'une enquête pilote. Cette dernière est menée sur un petit échantillon de la population qui sera enquêtée, afin de tester chaque élément de l'enquête.

- *Evaluation du contenu du questionnaire* : tester la compréhension des questions et la pertinence des réponses, vérifier l'utilisation du bon vocabulaire, vérifier l'impact des instructions sur le bon déroulement du questionnaire, identifier les causes possibles d'abandon.
- *Evaluation de la structure du questionnaire* : vérifier que les textes d'accroche initiale, de conclusion et de liaison sont compréhensibles et suffisants, contrôler la cohérence du cheminement du questionnaire et valider la présentation du questionnaire.

D'une façon générale, aucun problème de fonctionnalité du questionnaire web n'a été formulé dans les pré-tests (Les verbatims recueillis sont présentés en annexe IV). Ils ont permis de valider la durée moyenne du questionnaire (15 minutes en moyenne), temps acceptable par les répondants (*'Un peu plus de 10 minutes pour remplir le questionnaire. Honnêtement, cela ne semble ni long, ni difficile à remplir'*), malgré une certaine lourdeur des questions (*'Le questionnaire n'est pas long, à condition bien sûr de ne pas avoir une vie trépidante avec plein de déplacements en journée'*). L'ajout de la barre de progression n'est visiblement pas un atout du questionnaire (*'La présence de la barre est attrayante au début, puis nous ne l'utilisons plus'*). Les sauvegardes partielles des réponses semblent en revanche appréciées, bien qu'un répondant précise qu'*'il faudrait renseigner l'internaute sur les enregistrements intermédiaires des données, afin de ne pas le décourager en cas d'interruption de la saisie'*. Concernant la compréhension des questions, on constate une *'bonne compréhension du questionnaire'*. Si la formulation semble claire, elles *'demandent des réponses super détaillées des fois, ce qui est un peu fastidieux'*. La saisie des adresses est globalement lourde et certains précisent qu'*'il est difficile de se souvenir des adresses exactes des lieux, et nous avons tendance à ne pas renseigner tous les champs'*. La pénibilité de la partie sur les déplacements (*'La partie où il faut décrire l'ensemble des déplacements effectués est assez rébarbative et il est certainement difficile de se souvenir avec exactitude des horaires exacts...'*) peut générer des abandons ou l'omission de déplacements (*'Lorsque nous connaissons le temps nécessaire pour saisir un déplacement en totalité, nous avons tendance à omettre certains déplacements'*). Les textes d'accroche, de liaison et de conclusion ne posent pas de problèmes. A contrario, certaines questions apparaissent comme trop personnelles (*'Que viennent faire les questions sur les téléphones et l'Internet ? L'enquête porte-t-elle bien sur les transports ou recherche-t-on à nous vendre quelque chose derrière ?'*).

L'évaluation du design du questionnaire montre que *'les aides visuelles et les instructions sont claires dans l'ensemble'*, mais que le nombre de validations et les relances sont trop nombreuses. Dans l'ensemble, *'c'est fonctionnel et ça se remplit facilement au moins'*. Ces remarques ont conduit à certaines simplifications du questionnaire, mais ont surtout mis en évidence un manque d'ergonomie (*'Pourquoi l'exemple avec les symboles (maison, travail ...) ne sont ils pas repris ensuite, ce serait plus attractif ?'*, *'Lorsque nous saisissons un déplacement, il faudrait proposer la saisie du retour de façon automatique. Cela éviterait de tout ressaisir'*) et des problèmes de mise en page du question-

naire (*'Mauvaise adaptation des pages du questionnaire à la taille de l'écran, et nécessité d'utiliser les ascenseurs horizontaux et verticaux'*). L'utilisation d'un logiciel adapté aux enquêtes web semble incontournable pour que l'internaute ait l'impression de se trouver face à n'importe quel site web lui demandant des informations pour traiter sa demande (comme les sites d'information/réservation de compagnies ferroviaires, aériennes ou d'agences de voyages, ou des sites de vente en ligne...). Au final, les pré-tests ont permis de mettre en ligne un questionnaire efficient et attractif, et ont révélé quelques problèmes techniques qui auraient pu entraver la réalisation de l'étude (la lenteur du serveur internet ou le temps d'attente relativement long pour le téléchargement des pages du questionnaire, à cause du grand nombre de répondants essayant d'accéder à l'enquête en même temps).

I.2 Les principales difficultés rencontrées

Les internautes ont des attentes concernant le graphisme ou l'ergonomie des questionnaires web auxquels ils acceptent de se soumettre, et il est fort probable qu'un design 'au rabais' ait un impact très négatif sur le taux de réponse final (Morrel-Samuels, 2003). Une fois prise en compte les recommandations détaillées ci-dessous, la mise en ligne de l'enquête ménages déplacements réalisée sur le web pose quatre problèmes majeurs : la longueur et la sophistication du questionnaire, le nombre de personnes sollicitées et de fortes contraintes légales.

I.2.1 Des interviews en face-à-face traditionnellement longues

Les études de mobilité réalisées traditionnellement en face-à-face sont généralement longues, car les réponses doivent être très précises et concerner l'ensemble du ménage. Les interviews durent une heure et trente minutes en moyenne, chaque personne du ménage de 5 ans et plus étant enquêtée personnellement. L'utilisation du web nécessite d'abord de considérer certaines caractéristiques propres à ce média, comme le temps de téléchargement des pages, proportionnel à la durée du questionnaire, ou la durée de concentration du répondant. Par ailleurs, les personnes concernées par les enquêtes en ligne sont souvent pressées et ne veulent pas passer trop de temps à télécharger le questionnaire ni à saisir leurs réponses ⁶². Il a été démontré que de longs questionnaires auto-administrés produisaient des taux de réponse plus faibles (Dillman *et al.*, 1993), même si certains auteurs doutent de l'existence d'une relation significative entre la longueur du questionnaire et le taux de réponse (Bogen, 1996). D'autres (Michalowska, 1999) ont insisté sur la nécessité d'utiliser le média web en tant que tel et de ne pas transposer directement en ligne les versions de questionnaires papier ou téléphone.

Nous avons allégé le questionnaire standard, afin d'adopter une longueur

⁶²Les internautes ont une capacité variable à utiliser leur ordinateur, mais il existe un temps 'limite' (entre 15 et 20 mn) à ne pas dépasser pour une enquête web (Gunn, 2002).

raisonnable, en supprimant les questions d'opinion, qui ne paraissaient pas indispensables à notre recherche ⁶³. En revanche, nous avons conservé les questions d'ordre sociodémographiques, nécessaires pour identifier le profil des répondants, et évité les simplifications abusives, sources de biais dans les réponses. Les intitulés de certaines questions ont également été modifiés, pour les rendre plus compréhensibles. En effet, si dans un questionnaire administré en face-à-face ou par téléphone l'enquêteur peut expliciter certains termes non compris par les répondants, dans le cas d'une enquête web, l'internaute se retrouve seul face à son ordinateur et doit être en mesure de comprendre rapidement la question posée. Pour ce travail, nous nous sommes inspirés du questionnaire CERTU Enquête Déplacements Villes Moyennes, réalisé par téléphone. Ce questionnaire est relativement court, puisqu'il est administré en vingt minutes en moyenne, et présente l'avantage d'être facilement transposable sur le web, puisque les interviews téléphoniques sont assistées par ordinateur (CATI).

I.2.2 Un questionnaire standard relativement complexe

Pour encourager le maximum de réponses, les atouts du web (interactivité, possibilité graphiques et sonores ...) ne doivent pas être utilisés de façon abusive, des expériences comparatives montrant un taux de réponse plus élevé pour un questionnaire simple que pour un questionnaire plus sophistiqué (Dillman *et al.*, 1998b). Un questionnaire complexe peut dérouter les internautes, notamment ceux qui ne maîtrisent pas encore suffisamment l'utilisation d'Internet. Par ailleurs, en présence de plusieurs modes d'enquêtes, la similarité des questionnaires est importante. Une construction trop complexe, qui peut apparaître comme conviviale et personnalisée en ligne est difficile à retranscrire sur papier, que ce soit pour le répondant dans le cas d'une enquête postale, ou pour l'enquêteur dans le cas d'une enquête en face-à-face non assistée par ordinateur. Les questionnaires trop disparates rendent très difficile la comparabilité des enquêtes et il est préférable de conserver une version électronique proche de la version papier.

Concernant la présentation du questionnaire, proposer l'ensemble du questionnaire sur une seule page, l'ascenseur permettant aisément le déplacement du curseur sur l'écran, permet d'obtenir un temps de réponse plus court et moins de non-réponses partielles, mais peut affecter la qualité des données, en produisant davantage de réponses 'similaires' (Gunn, 2002). D'un autre côté, faire figurer une question par page permet d'étudier a posteriori les comportements de réponse des individus et surtout de personnaliser le questionnaire en fonction des réponses déjà fournies. Cela permet d'écourter le questionnaire, car il est possible d'utiliser des filtres pour ne retenir que les questions pertinentes pour l'individu interrogé ⁶⁴. L'utilisation d'une seule question par

⁶³Le questionnaire web a ensuite été validé par l'ensemble des acteurs impliqués, le CERTU, le SYTRAL et le CETE fin septembre 2005.

⁶⁴Si toutes les questions sont visibles, le répondant qui ne maîtrise pas l'usage des filtres risque de surestimer le temps de réponse et de ne pas comprendre pourquoi certaines questions ne lui sont pas posées.

page nécessite toutefois de donner un peu de flexibilité dans la navigation du questionnaire, à l'aide de boutons 'précédent' et 'suivant', et de permettre une correction/validation rapide des réponses saisies à l'aide de boutons 'Effacer' et 'Envoyer' (Smith, 1997). La mise en page du questionnaire, la rédaction des questions, le vocabulaire utilisé et la qualité du graphisme sont autant de détails qui permettent de conserver l'attention des internautes jusqu'à la fin du questionnaire (Michalowska, 1999).

Le questionnaire de l'enquête ménage est relativement complexe, des données très précises sur les déplacements et leur enchaînement étant demandées aux répondants. Notre objectif a été de développer un questionnaire convivial, pour attirer de nouveaux répondants, tout en restant simple, pour ne pas alourdir le travail de réponse et décourager les internautes inexpérimentés. Les contraintes techniques n'ont pas permis de choisir la présentation des questions. Une seule question figure sur chaque écran, l'internaute devant choisir une modalité et valider sa réponse pour accéder à l'écran suivant.

I.2.3 Un seul répondant sélectionné dans le ménage

Pour analyser la cohérence des déplacements, qui s'inscrivent dans une logique au sein du ménage, il faut interroger tous les membres et déterminer un jour de référence pour le recueil des données (Certu, 1998). Les grands ménages répondent moins facilement à ce type d'enquête, ce qui génère un biais puisque leur mobilité et leurs caractéristiques sociodémographiques diffèrent de ceux des autres répondants (De Heer et Moritz, 2000). Pour certains modes de recueil, notamment auto-administrés, il est par définition difficile d'imposer le jour de référence des déplacements et d'obtenir les réponses pour l'ensemble des membres du ménage (Bonnell, 2003). Par ailleurs, les recherches montrent que le nombre de déplacements est sous-estimé lorsqu'il est rapporté par une tierce personne (Sharp *et al.*, 2004; Richardson *et al.*, 1995) et que ces pratiques ne permettent pas d'obtenir les réponses détaillées (Stopher et Jones, 2003).

Techniquement, il aurait été possible de préciser ces contraintes aux individus dans la lettre de présentation de l'étude, puis de leur laisser le choix de suivre ou non les directives. Concrètement il semble difficile de demander à l'ensemble des personnes du ménage de se connecter le même jour pour remplir tour à tour le questionnaire web. Cette restriction peut avoir un effet repoussoir sur les répondants potentiels (Bonnell et Le Nir, 1998), et générer des erreurs de mesure (réponses rapportées, ne concernant pas le même jour de référence...). Enfin, il ne faut pas oublier que l'enquête web est destinée aux personnes non joignables ou refusant de recevoir un enquêteur à domicile, et qu'un des principaux attraits du média web est sa souplesse. Dans notre démarche exploratoire, nous avons opté pour le remplissage d'un seul questionnaire par foyer (par toute personne de 11 ans et plus), quelle que soit la composition de celui-ci. Cette souplesse contraste fortement avec l'enquête en face-à-face et ne permet pas de connaître avec certitude l'identité du répondant.

I.2.4 Un cadre législatif précis

Avant le lancement d'une enquête auprès du grand public, il est nécessaire d'effectuer des déclarations à la C.N.I.L., qui s'intéresse à la constitution de l'échantillon et à l'exploitation des données.

Dans la mesure où le fichier relatif à l'échantillon utilisé pour l'enquête web existe déjà (interviews prévues en face-à-face, mais non réalisées), une nouvelle déclaration C.N.I.L. de constitution du fichier des enquêtés n'a pas été nécessaire. En revanche, le questionnaire web contient des questions spécifiques et a fait l'objet d'une déclaration C.N.I.L. de traitement des données. De plus, lors des enquêtes ménages déplacements précédentes, il était impossible de connaître les adresses précises des lieux visités par les enquêtés, puisque l'agglomération Lyonnaise était découpée en zones. Or, dans le cadre de notre recherche sur la faisabilité d'un questionnaire web, nous souhaitions obtenir l'intégralité des informations sur les origines et les destinations des déplacements. Légalement, il est possible de disposer des coordonnées précises d'un individu, si ce dernier n'est pas identifié, c'est-à-dire de gérer deux bases de données, une avec les adresses et une avec les noms et numéros de téléphone. Il est également intéressant de demander aux individus qui le souhaitent d'indiquer leur adresse mail à la fin du questionnaire, pour pouvoir les recontacter en cas de besoin (phase de suivi). Les questions posées n'étant pas relatives à l'état de santé du répondant, la C.N.I.L. ne s'est pas opposée à cette demande. Enfin, la loi informatique et liberté est mentionnée dans la lettre avis envoyée aux enquêtés, et rappelée en bas de la première page web du questionnaire, où un lien permet d'accéder directement au site de la C.N.I.L.

I.2.5 Contraintes de réalisation

Tous les logiciels d'enquête proposent désormais un module web, dont la qualité est très inégale selon les prestataires. Les interfaces CATI réécrites pour les enquêtes CAWI sont souvent peu performantes et les meilleurs produits sont ceux conçus spécifiquement pour administrer des enquêtes en ligne. Ils permettent notamment de visualiser les réponses saisies en temps réel, de traduire le script du questionnaire en plusieurs langages et d'exercer un réel contrôle sur le design des pages du questionnaire et sur la façon dont elles apparaissent aux enquêtés (Burke et James, 2006). Par ailleurs, il est important de définir dès la conception de l'étude le format sous lequel les données saisies seront enregistrées et vers quels logiciels elles pourront ensuite être exportées, en vue d'analyses statistiques et économétriques.

Dans l'enquête ménage déplacements de Lyon, le web constitue une annexe au cahier des charges. Ainsi, nous n'avons pas pu choisir le prestataire chargé du développement du questionnaire en ligne. SOFRECO, qui gère la réalisation de l'enquête face-à-face, a développé le questionnaire web en utilisant l'interface CATI du logiciel 'Converso', adapté aux enquêtes téléphoniques. Cette contrainte n'a pas permis d'apporter au questionnaire web toute la souplesse

que ce média peut offrir aux enquêtes de mobilité. Les données validées par les enquêtés alimentent une base de données au format 'Converso', peu répandu, et une importante phase de mise en forme du fichier fut nécessaire avant de l'exporter vers des logiciels de gestion de bases de données. Les arbitrages technologiques sont un des facteurs clés de succès d'une enquête web. Faire l'impasse d'un logiciel approprié pour l'administration du questionnaire et ne pas assez anticiper les besoins en termes de traitement des données nuisent au bon déroulement de l'étude.

II La rédaction des questions

Face à un ordinateur, les individus n'appréhendent pas une enquête de la même manière que si elle leur est soumise en face-à-face (Dillman et Bowker, 2001). Ce résultat est particulièrement important dans les enquêtes de mobilité urbaine, où le recueil des données de déplacements reste difficile à mettre en oeuvre. La perspective de comparabilité des données ne doit pas conduire à figer le questionnaire, mais au contraire à l'adapter aux comportements des enquêtés face à chaque type de média et à la présence ou non d'un enquêteur. Sur le web, il est indispensable que le questionnement apparaisse comme le plus 'naturel' possible pour l'enquêté. Afin de faciliter la tâche des enquêtés, et de pouvoir comparer les résultats d'enquêtes menées à des horizons temporels différents ou par plusieurs modes de recueil, il est préférable de conserver le mode de rédaction des questions et des modalités de réponses couramment utilisées (Stopher *et al.*, 2004b). Enfin, même si l'enquête web est simplifiée par rapport à la version face-à-face, il faut maintenir un noyau de questions permettant de recueillir des données de mobilité précises et de les croiser avec des caractéristiques sociodémographiques pertinentes (Pratt, 2003).

II.1 Présentation du questionnaire web

Notre questionnaire s'intitule 'Enquête web sur les déplacements dans l'agglomération lyonnaise'. L'introduction reprend des éléments de la lettre avis envoyée aux enquêtés, en insistant sur l'importance des informations recueillies et sur la confidentialité des données. La charte graphique du questionnaire web est basée sur celle que le Sytral a utilisé dans sa campagne de communication, pour conserver une certaine homogénéité⁶⁵. Le principal changement concerne la structure du questionnaire. En effet, le questionnaire standard CERTU est organisé en quatre parties, et recueille successivement :

- les caractéristiques du ménage et du logement ;
- les caractéristiques socio-économiques de chaque personne composant le ménage ;
- les déplacements de chaque personne du ménage âgée de 5 ans et plus, effectués la veille du jour de l'enquête ;

⁶⁵Toutefois, le logo du Sytral n'apparaît pas à l'écran, pour ne pas faire croire aux répondants que l'enquête s'intéresse uniquement aux déplacements effectués en transports en commun.

- des opinions auprès d'une personne du ménage âgée de 16 ans et plus, après tirage au sort.

Le questionnaire web est structuré en quatre parties, ou 'blocs', conformément au questionnaire C.A.T.I., utilisé pour les enquêtes téléphoniques. Nous distinguons d'abord les questions relatives au ménage, celles concernant la personne interrogée, puis celles concernant les données de déplacements, avant de terminer par des questions complémentaires sur l'équipement multimédia du ménage et le niveau de revenu (annexe VI).

II.2 La collecte d'informations sur les caractéristiques du ménage et du répondant

II.2.1 Les caractéristiques générales du ménage

Par définition, l'unité statistique des enquêtes ménages déplacements est le ménage. Dans les interviews en face-à-face ou par téléphone, plusieurs personnes répondent au questionnaire, et fournissent des renseignements sociodémographiques. Ce n'est plus le cas pour l'enquête web, où un seul individu remplit le questionnaire. Il est important que l'analyste connaisse les caractéristiques des différentes personnes du ménage et leur relation avec le répondant, pour appréhender les comportements de mobilité (Pratt, 2003). La première question permet donc de définir l'âge, le sexe et l'occupation des personnes qui composent le ménage, en se basant sur les renseignements fournis par le répondant. Les prénoms sont également demandés, bien qu'ils n'aient pas d'utilité directe, afin de personnaliser le questionnaire et de faire le lien avec les questions individuelles. L'adresse du domicile est recueillie avec précision, car c'est un élément explicatif fort des comportements de mobilité, mais n'aura pas à être ressaisie par le répondant dans la partie relative à ses déplacements. Concrètement, un cadre détaillé est proposé à l'enquêté (n°, rue, lieu dit, code postal, commune et département) et une relance apparaît lorsqu'il manque des informations essentielles.

Certains renseignements concernant le ménage sont reportés à la fin de l'interview. Il s'agit de questions sur l'équipement téléphonique et les revenus annuels nets du ménage. Les premières ne sont pas directement liées à la mobilité et peuvent être ressenties comme une intrusion dans la sphère privée. La seconde est traditionnellement très sensible, et risque d'effrayer les répondants. Connaître l'équipement des ménages en moyens de communication est particulièrement important dans cette enquête, qui s'adresse à des internautes. Nous avons donc rajouté une question par rapport au questionnaire standard, afin de connaître le type de connexion Internet possédée au domicile. La distinction se fait entre haut et bas débit. La notion de forfait limité est également intéressante, puisque les internautes qui en bénéficient sont moins contraints, mais reste peu discriminante, puisque les offres haut débit proposées sur le marché permettent majoritairement un accès illimité au web.

Par analogie avec l'enquête standard, nous avons placé la question concernant les revenus annuels du ménage à la fin de l'étude. Cette variable, souvent

utilisée dans les enquêtes ménages déplacements, est plus utile pour appréhender la mobilité des membres du ménage que le niveau de ressources personnelles du répondant (Pratt, 2003). Les classes de revenus sont similaires à celles utilisées dans le questionnaire en face-à-face, afin de faciliter les comparaisons⁶⁶.

II.2.2 Les caractéristiques du répondant

Avant de passer aux renseignements individuels, il est utile de vérifier quel individu 'se cache' réellement derrière l'ordinateur. Ayant demandé à l'internaute de saisir les prénoms des membres du ménage, en commençant par lui-même, il est peu délicat et inconsistant de lui redemander son identité. L'ordinateur génère donc automatiquement le premier prénom saisi, et demande simplement une confirmation.

Une fois pris en compte le nombre de voitures possédées par le ménage, le renseignement individuel le plus pertinent dans le cadre d'une enquête de mobilité est la possession du permis de conduire. Puis, viennent des questions concernant le niveau d'étude, le type d'emploi occupé, et l'adresse exacte du lieu de travail ou d'étude. Cette adresse est enregistrée dans le système et est utilisable de nouveau dans le recueil des déplacements pour les trajets impliquant le lieu de travail. Contrairement au lieu de résidence, il est possible que certaines personnes ne connaissent pas les coordonnées du lieu de travail avec exactitude. Une zone de saisie 'en clair' permet aux individus d'indiquer des renseignements divers (nom d'établissement, lieu public le plus proche, croisement de rues . . .), qui seront utilisés pour une codification a posteriori. Enfin, nous recueillons la profession de l'individu enquêté. Dans un souci d'adaptation du contenu des questions au média utilisé, nous proposons un menu simplifié pour le choix de la profession principale, complété si besoin par une liste détaillée.

II.2.3 La fréquence d'usage des modes de transport

Diverses questions concernant la fréquence d'utilisation des modes de transport dans les déplacements quotidiens complètent le questionnaire individuel. Comme dans l'enquête face-à-face, nous distinguons la voiture particulière, conducteur ou passager, les transports collectifs et les deux roues. La distinction de puissance entre les deux roues à moteur est maintenue (bien que le nombre de réponses pour chacune de ces modalités soit trop faible pour être exploité statistiquement), puisque les usages de ces deux types de véhicules ainsi que leurs impacts environnementaux sont différents.

Des questions concernent l'utilisation de la voiture pour se rendre sur le lieu de travail ou sur le lieu d'étude, ainsi que les éventuelles difficultés de stationnement rencontrées par le conducteur. En effet, les déplacements quotidiens pour motif travail ou étude sont nombreux, et la possibilité de pouvoir

⁶⁶Dans notre étude, la modalité 'Ne Sait Pas' est largement choisie, quel que soit le mode d'enquête, de nombreux individus refusant de dévoiler des informations liées à leur pouvoir d'achat (Cf. Chapitre 3 section III.3).

garer ou pas un véhicule à proximité de son lieu d'activité principale impacte fortement l'utilisation de la voiture particulière.

Au final, les questions du bloc 'Ménage' et 'Individu' restent relativement proches de celles utilisées dans l'enquête ménages déplacements réalisée par téléphone dans les villes moyennes et en face-à-face (thèmes, ordre d'apparition, formulation . . .). La différence principale réside dans le tableau de composition du ménage, que les internautes doivent remplir au début du questionnaire.

II.3 La collecte d'informations sur les déplacements individuels

Les enquêtes ménages déplacements requièrent des données précises concernant les lieux fréquentés par le répondant durant la période de référence de l'étude. Dans le cas de Lyon, cette période ne dure que 24h, mais la mémoire des individus est fortement sollicitée. La partie du questionnaire dédiée aux déplacements doit donc être conviviale et motivante, et permettre le recueil de données précises.

II.3.1 Une approche mixte déplacements-activités

Deux approches sont possibles pour le recueil des données de mobilité : une basée sur les déplacements, l'autre sur les activités. Dans l'approche basée sur les déplacements, les répondants doivent renseigner tous les champs relatifs à un déplacement, avant de passer au suivant. Dans l'approche par activité, l'individu liste les activités effectuées dans la journée, ainsi que leur horaire de départ, puis renseigne pour chaque activité les données de localisation et les caractéristiques des déplacements entre ces activités, lorsque la localisation change (le mode de transport, la durée, et éventuellement le nombre de personnes participant au déplacement). Selon Ressource System Group (2002), les enquêtés préfèrent les questionnaires qui proposent une approche par activité. Proposer un cheminement des activités effectuées la journée de référence permet en effet de recueillir les déplacements d'une manière naturelle pour l'interviewé et l'effort de mémoire est moins important que dans le cas d'un recueil chronologique des déplacements. Des travaux anciens (Jones *et al.*, 1980) ont montré que l'approche par les activités génère davantage de déplacements que celle par les déplacements. D'autres, plus récents, confirment cette hypothèse ⁶⁷. Stopher et Jones (2003), précisent que l'approche par activité fournit des informations plus justes et doit être mise en oeuvre même si l'objectif de l'enquête est le recueil de données concernant les déplacements. Nous avons élaboré le questionnaire web de l'enquête ménages déplacements de Lyon en nous basant notamment sur des expériences australienne ⁶⁸ et autrichienne ⁶⁹

⁶⁷Sous-estimation des déplacements (environ 20 à 30%), en particulier les petits déplacements, lors d'un recueil à l'aide d'un journal de déplacements (Wolf *et al.*, 2003; Pendyala, 2003).

⁶⁸Postal Survey, Two-Day Travel Diary, Department 'Infrastructure, Planning and Natural Resources', Institute of Transport Studies, University of Sydney, NSW 2006.

⁶⁹Postal Traffic survey, Household and individual questionnaires, Austria, 2002.

d'enquêtes auto-administrées et sur l'acquis des travaux conduits sur les schémas d'activités depuis plus de trente ans (Jones *et al.*, 1980; Bonnel, 1985, 1995; Stopher, 1998). Le cas de l'enquête postale réalisée par l'université de Sydney illustre cette approche. Un cahier a été envoyé aux répondants, pour qu'ils notent durant deux jours l'ensemble de leurs déplacements, en raisonnant en termes d'activités. L'idée sous-jacente est que les déplacements ne sont pas une fin en soi, mais un moyen de réaliser une activité (loisir, travail, accompagnement...). Toutefois, cette approche est en partie en rupture avec celle qui est utilisée en France dans les enquêtes ménages déplacements en face à face, où le recueil porte sur l'enchaînement des déplacements. Afin de pouvoir comparer les résultats des deux enquêtes, nous avons retenu une solution intermédiaire, avec une entrée activité et une saisie de l'enchaînement des déplacements. Pour simplifier le travail du répondant, nous lui proposons un cheminement dans ses déplacements quotidiens, à l'aide d'une série de questions (horaires, lieux, activités, moyens de transport...). A posteriori, l'analyse des résultats nous conduit à remettre en cause ce choix. L'oubli de certains déplacements, notamment le retour au domicile en fin de journée, par quelques internautes aurait probablement pu être évité par une approche centrée sur les pratiques d'activités, les déplacements découlant des changements de localisation.

II.3.2 Le déplacement, une notion difficile à appréhender

Au début de la partie sur les déplacements, nous indiquons sous forme de liste les consignes les plus importantes pour la saisie des données. L'objectif n'est pas d'être exhaustif, mais de valoriser les réponses de l'internaute, d'insister sur la richesse et la précision des données fournies, et d'expliquer le moyen d'obtenir rapidement de l'aide. Puis un exemple de cycle de déplacements est donné, à titre purement illustratif. Malgré le risque de suggérer des réponses, sa présence rend compréhensible le questionnaire, prend peu de place (1 page) et exploite les possibilités multimédia du web. Nous avons développé un exemple, en prenant garde à varier les modes et les motifs de déplacement, afin de ne pas influencer l'enquêté (figure 7).

Concrètement, le répondant doit renseigner précisément l'endroit où il se trouve au début de la période de référence, ainsi que l'activité effectuée. Ensuite, les questions suivent le cheminement suivant, jusqu'à ce que l'heure de départ du dernier déplacement soit postérieure à l'heure de fin de la période d'enquête (figure 8) ⁷⁰.

D'un côté, Internet est un média qui nécessite des questionnaires concis. De l'autre la forme auto-administrée requiert des instructions suffisamment détaillées, aucun enquêteur ne venant apporter de précision au répondant ou organiser la saisie des réponses. La notion de 'déplacement', au coeur de l'étude n'est pas facilement comprise et assimilée par les répondants, et les résultats des enquêtes de mobilité ne peuvent pas être plus précis que la mémoire des

⁷⁰S'agissant de recueillir des déplacements, les heures de début et les heures de fin doivent être renseignées. Si nous avons opté pour un recueil basé sur les activités, seule l'heure de début aurait été demandée, l'heure de fin d'une activité étant par définition égale à l'heure de début de l'activité suivante.

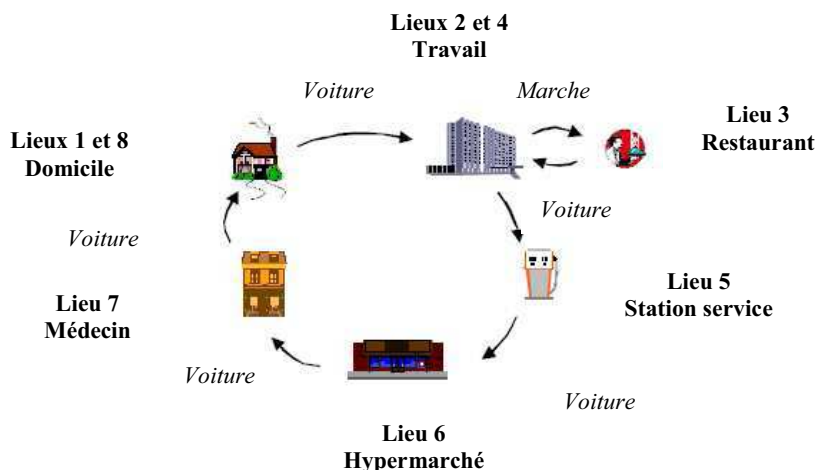


FIG. 7 – Exemple d'un cycle de déplacements quotidiens
Source : Questionnaire web (2005)

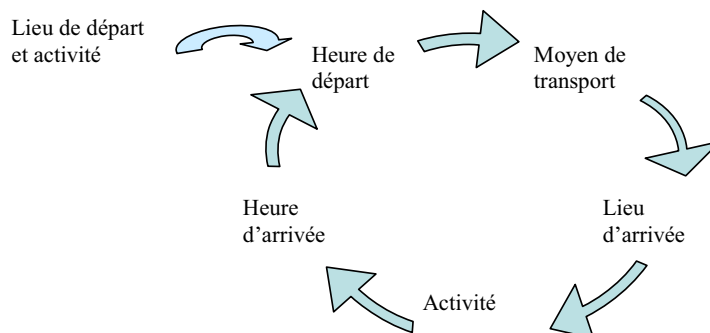


FIG. 8 – Cycle des questions relatives à un déplacement
Source : Questionnaire web (2005)

répondants. Plus les enquêtés vont pouvoir être 'aidés' dans leur effort de mémoire, plus juste vont être les réponses (Zmud, 2003). Pour permettre le recueil de données de qualité et éviter les biais de mesure dus à des interprétations différentes de ce concept, les enquêteurs renseignent les individus à l'aide de leur manuel d'instruction lors des enquêtes en face à face. Or, le caractère auto-administré d'une enquête web ne permet pas cet échange. Il est toujours possible de fournir la définition du déplacement et de l'illustrer à travers quelques exemples. Certaines enquêtes soulignent les termes pouvant prêter à confusion dans le libellé des questions et y associent une définition à l'aide d'un lien hypertexte. L'internaute n'a alors qu'à cliquer pour voir s'afficher des précisions à l'écran ⁷¹. Mais il n'est pas possible de fournir une aide en ligne regroupant tous les cas de figure prévus dans le manuel de l'enquêteur,

⁷¹Conrad *et al.* (2003) montrent que les répondants en ligne tirent relativement peu profit de cette possibilité, surtout si la démarche nécessite plus d'un 'clic' ou si les précisions concernent des termes courants (exemple : ménage, résidence...). L'option est davantage utilisée lorsque les informations semblent utiles aux répondants, et les données recueillies sont de meilleure qualité.

au risque de complexifier le questionnaire, ce qui générerait l'effet inverse de celui souhaité.

La définition du déplacement est loin d'être évidente "*Il n'est donc pas si simple de donner de cette unité de mesure de la mobilité qu'est le déplacement une définition opératoire. C'est en réalité à travers les instructions qui sont données aux enquêteurs que se précise dans ses détails cette définition* (Bonnafous et Puel, 1983)". Dans les enquêtes en face à face, les enquêteurs disposent de la définition suivante : 'Le déplacement est le mouvement d'une personne, effectué pour un certain motif, sur une voie publique, entre une origine et une destination, selon une heure de départ et une heure d'arrivée, à l'aide d'un ou plusieurs moyens de transport' (Cete, 2005) ⁷². Dans l'enquête web, l'assimilation de la notion de 'déplacement' par l'internaute est difficile à contrôler et il n'est pas certain que l'aide en ligne ⁷³ soit comprise et prise en compte.

II.3.3 Le choix du mode et de l'activité

Nous avons jugé préférable de simplifier la liste des activités proposées dans le questionnaire en face à face, en effectuant d'abord des regroupements par thèmes, puis en demandant à l'enquêté des précisions dans des sous-menus pour les modalités études, achats et loisirs (figure 9). En ce qui concerne les tournées professionnelles, une note précise à l'internaute les informations à fournir.

Quelle était votre activité principale à cet endroit ? ▼		
Travail		
Etudes		
Achats		
Démarches personnelles (administratif, santé, recherche d'emploi...)		
Loisirs ou visites		
Restauration hors du domicile		
Aller chercher / déposer quelqu'un		
Tournée professionnelle		
Autre Précisez		

Si études ▼	Si achats ▼	Si loisirs ou visites ▼
Collège	Grand magasins et galeries marchandes	Activités sportives, culturelles ou associatives
Lycée	Hypermarchés et supermarchés	Promenade, « lèche-vitrines »
Universités et grandes écoles	Petits et moyens commerces, marchés	Visite à des parents ou à des amis

FIG. 9 – Choix de l'activité exercée
Source : Questionnaire web (2005)

⁷²En France, l'organisme central, le CERTU, assure l'homogénéité technique des enquêtes ménages.

⁷³Illustration de la définition donnée par le CERTU dans le guide méthodologique (Certu, 1998) à l'aide d'exemple.

La liste des modes de transport a été simplifiée et limitée aux modes les plus fréquemment utilisés pour les déplacements urbains ⁷⁴. Une modalité autre est également nécessaire, afin de recueillir les modes plus rarement utilisés. Par ailleurs, dans la version face à face, on distingue les transports collectifs selon l'autorité organisatrice. Il n'est pas raisonnable de conserver cette typologie dans l'enquête web, puisque les internautes ne disposent pas toujours de l'information, qui peut par ailleurs être reconstruite a posteriori à partir de la localisation de l'origine et de la destination du déplacement effectué. Au final, nous avons retenu seulement 13 modes de transport, principalement en effectuant des regroupements ⁷⁵ ou des suppressions ⁷⁶ dans la liste des modalités proposées en face-à-face. Nous avons conservé la question relative au paiement du stationnement pour les individus déclarant un déplacement en voiture particulière, cette contrainte ayant généralement une influence sur le choix du mode de transport.

II.3.4 La chronologie des déplacements

Les enquêtes ménages déplacements commencent généralement entre 3h et 4h du matin, et se terminent à la même heure un ou plusieurs jours après (Stopher *et al.*, 2004b). A cette heure, le niveau de trafic est minimal sur les principaux axes routiers ⁷⁷. Peu de déplacements seront interrompus si on choisit ce créneau pour borner la période de référence et les individus ne se retrouveront pas dans la situation délicate d'essayer de saisir un déplacement ayant commencé ou pris fin en dehors de cette tranche horaire.

L'usage est de recueillir les heures de départ et d'arrivée des déplacements sous la forme 0h/24h, afin de permettre un calcul rapide de la durée des activités (Pratt, 2003). Cette codification pose cependant des problèmes pour les déplacements dont l'aller se fait avant minuit et le retour après minuit, puisque la durée d'activité calculée est négative ⁷⁸. Dans l'enquête web, nous avons demandé aux répondants de remplir deux champs, un pour le jour et un pour l'heure du déplacement (plus un champ pour les minutes). Ainsi, le début et la fin des déplacements se situent soit le jour 1 (jour de référence de l'étude) entre 04h et 24h00, soit le jour 2 (jour de fin de l'étude) entre 0h00 et 3h59. Quels que soient les horaires de début et de fin des déplacements se situant dans la période de référence, la durée calculée reste donc positive.

⁷⁴Dans la version face à face, l'avion fait partie de la liste des modalités, mais la codification est effectuée par l'enquêteur. En revanche dans une version auto-administrée, proposer explicitement l'avion comme mode de transport pour des déplacements urbains peut sembler fantaisiste (même si la modalité est tout à fait possible, puisque les déplacements ayant leur origine ou leur destination dans le périmètre d'enquête sont supposés être tous saisis).

⁷⁵Car interurbain, transport scolaire ou transport employeur.

⁷⁶Les rollers, skates et trottinettes, qui ne peuvent être assimilés aux bicyclettes, mais aussi le transport fluvial ou maritime, aérien.

⁷⁷Ce constat se vérifie en France, aux Etats-Unis et en Australie, mais peut varier entre minuit et 05h00 selon les pays.

⁷⁸Par exemple, un déplacement de 22h00 à 02h00 durerait -20 heures.

III Administration du questionnaire en ligne

III.1 L'importance de la prise de contact

Le premier contact entre l'étude et le répondant est très important, puisqu'il va déterminer en grande partie son choix de se soumettre ou non au questionnaire (Ampt, 2003). Généralement, une lettre avis est envoyée aux ménages de l'échantillon, et les recherches ont mis en évidence une augmentation du taux de réponse aux enquêtes web, grâce à ce type de pré-notification (Alsnih, 2004; Mehta et Sivadas, 1995). Cette lettre permet d'informer les individus de l'enquête à venir, de la légitimer en insistant sur l'objectif poursuivi et les instances impliquées ⁷⁹, de rassurer sur le caractère confidentiel du traitement des données, de fournir une estimation du temps nécessaire pour remplir le questionnaire (20 minutes maximum), ainsi que les coordonnées d'un contact (une adresse e-mail ou un numéro de téléphone gratuit, si possible) pour les éventuelles questions ⁸⁰. Dans les enquêtes web, cette lettre avis est fortement recommandée (Sills et Song, 2002), bien qu'elle soit souvent envoyée par mail lorsque l'adresse électronique des individus est connue. Contacter les individus à de multiples reprises par mail est ni coûteux ni chronophage, mais lorsque l'adresse électronique est inconnue, les méthodes classiques de prise de contact par téléphone ou voie postale restent la seule solution. La lettre mentionne précisément l'adresse du site où est hébergé le questionnaire, l'identifiant et le mot de passe attribué, et insiste sur le caractère simple et rapide de l'enquête (Annexe V). L'option web peut être mise en avant, même s'il est préférable de réduire l'anticipation de la difficulté de remplissage du questionnaire, plutôt que de développer les atouts du web (Haraldsen *et al.*, 2002).

Les individus supportent des coûts pour répondre aux enquêtes (temps, concentration, frais de connexion ...). La théorie de 'l'échange', avancée par Sudman et Bradburn (1996) montre que ces derniers sont d'autant plus motivés à se soumettre au questionnaire que le retour espéré est supérieur aux efforts engagés. (Dillman, 1978) précise que ce retour peut être de nature intangible, comme la satisfaction de faire partie d'un groupe sélectionné ou le sentiment que les réponses fournies contribuent largement au succès de l'étude. Un moyen d'augmenter le seuil de tolérance aux enquêtes est donc d'insister sur l'importance de l'étude et sur la nécessaire participation de tous pour garantir la qualité des résultats. Les techniques de personnalisation, qui donnent de l'importance au répondant vont dans ce sens. Il est également conseillé de communiquer sur le caractère ponctuel de l'étude et de fixer une date limite de présence du questionnaire en ligne ⁸¹. Enfin, Alsnih (2004) recommande de

⁷⁹Les personnes sollicitées répondent plus facilement lorsque la demande émane d'une autorité publique (Zmud, 2003).

⁸⁰Une étude menée par le bureau du Census des Etats-Unis montre que la moitié des répondants ont contacté l'assistance pour obtenir de l'aide sur l'utilisation des identifiants et mots de passe ou pour résoudre des problèmes de configuration de leur navigateur (Nichols et Sedivi, 1998).

⁸¹Stopher *et al.* (2004b) rajoute que la publicité autour de l'étude (télévision, radio, presse...) motive la coopération des individus, notamment pour les enquêtes en face-à-face.

préciser qu'aucune contribution monétaire ne sera demandée, pour distinguer l'étude des démarches de télémarketing.

Dans certains protocoles d'enquête, le choix du mode de collecte est donné aux individus. Cette option est valorisante, ce qui augmente le taux de réponse et génère des questionnaires de meilleure qualité (Alsnih, 2004). En effet, les individus se sentent plus ou moins proches d'un mode particulier, et ne peuvent parfois pas être touchés par certains médias. Pour l'Enquête Ménages de Lyon, ce choix n'a malheureusement pas été possible. Il s'agit d'une enquête séquentielle, où les non-répondants au premier mode doivent être informés du nouveau média disponible par une nouvelle lettre avis (Kalfs et Van Evert, 2003). Les enquêtes transport comprennent généralement une phase de recrutement et une phase d'interview et des protocoles sophistiqués combinent des contacts off-ligne et en ligne. Au moment du repérage, les enquêteurs essaient d'obtenir un numéro de téléphone, utilisable pour les relances après l'envoi du premier courrier (si aucun numéro n'est disponible, les relances sont postales). Les individus refusant de recevoir un enquêteur à domicile se voient proposer oralement, puis par écrit, de répondre au questionnaire web (annexe V). Une lettre est également envoyée aux ménages qu'il est impossible de joindre par téléphone, après 8 tentatives à des jours et horaires différents (ménages déclarés non-répondants). Ces lettres reprennent l'argumentaire développé dans la première lettre avis (présentation de l'étude, choix aléatoire des ménages, importance des réponses, confidentialité des données ...) et insistent sur le web, nouveau média disponible qui satisfait davantage aux contraintes temporelles des ménages. Pour encourager les individus, la durée restreinte de l'enquête web (15 mn) et un numéro de téléphone permettant d'obtenir des informations complémentaires sont indiqués dans les courriers. La sélection du répondant sur le web parmi les membres du ménage s'opère grâce à l'âge (plus de 11 ans).

III.2 La codification des adresses

Les enquêtes ménages déplacements recueillent des données sur la localisation des origines et des destinations des déplacements. De façon générale, les informations géographiques obtenues sont cruciales, car la modélisation de la demande de transport évolue vers des approches complexes, telles que la microsimulation. Par ailleurs, le zonage peut changer entre deux enquêtes ménages déplacements, et l'identification précise des adresses permet de réaliser des analyses sur de nouveaux périmètres. Avant de lancer une étude, il est donc nécessaire de s'interroger en amont sur la précision des données que l'on souhaite obtenir.

III.2.1 Le recueil de données précises

Le géocodage est une étape clé du traitement des données de mobilité. Cette méthode consiste à traduire les informations saisies en données directement utilisables par les modèles de planification stratégique. Il s'agit d'affecter une zone géographique à chaque localisation fournie par le répondant, selon

une cartographie très précise (Cambridge Systematics, 1996). Dans la plupart des enquêtes, le géocodage est effectué manuellement, ce qui implique un lourd travail, une bonne connaissance du périmètre d'enquête de la part des chargés d'études et conduit inévitablement à des erreurs (Greaves, 1998, 2003). Pour simplifier la saisie des adresses et le traitement de l'information lors de la réalisation d'une enquête web, il est possible de s'appuyer sur les systèmes d'information géographique et les nouvelles technologies de l'information et de la communication. Des fonctions déployables aisément à l'aide de menus déroulants (nom de rue, de ville . . .) ou de cartes peuvent être envisagées. Des bases de données interactives permettent de codifier en temps réel les lieux d'origine et de destination des déplacements, en déchargeant une large part du travail sur l'enquêté. Selon l'expérience de Adler *et al.* (2002), proposer un géocodage en direct sur le web permet d'obtenir un taux de géocodage de 100% (au final, l'individu clique sur une carte, mais renseigne la localisation). Ceci permet aux répondants peu familiers des ordinateurs de renseigner facilement les adresses des lieux visités ⁸², sans recourir à l'assistance technique (Adler *et al.*, 2001). Dans une enquête déplacements menée au Michigan en 2001, les distances entre deux adresses données par les répondants étaient systématiquement calculées. Lorsque la durée du déplacement saisie ne correspondait pas à une fourchette de valeur acceptable au regard de la distance théoriquement parcourue, une alerte permettait au répondant de corriger sa réponse (Murakami, 2004). D'autres études font appel à des logiciels qui interagissent directement avec les répondants, suggérant des alternatives lorsque l'adresse saisie n'est pas reconnue (Ressource System Group, 2002). Ces outils, bien que complexes, facilitent l'effort de mémoire et de saisie des enquêtés, et permet de réduire la non-réponse partielle et les erreurs de géocodage. Mais l'usage de grosses bases de données sur le web est à éviter, à cause du risque de blocage du serveur et du temps de téléchargement. De plus, si l'utilisation d'une cartographie est relativement lourde au niveau de l'affichage des pages web, cette méthode suppose également une bonne capacité des internautes à se situer sur une carte (Alsnih, 2005; Wermuth *et al.*, 2001).

Pour l'enquête Ménages Déplacements de 2006, le périmètre du SCOT de Lyon a été divisé en 856 zones géographiques appelées zones fines. Lors de l'interview en face-à-face, les localisations des activités sont directement codées par l'enquêteur. Il est cependant difficile de mettre en place ce zonage sur le web, et de demander à l'internaute de choisir la zone correspondant à l'origine et à la destination de ses déplacements. L'abandon du zonage amène au recueil de données réelles, plus précises, le répondant saisissant l'adresse exacte des lieux visités. Ces renseignements sont ensuite transformés en coordonnées longitudinales et latitudinales, conformément à l'adresse générée par le logiciel de géobornage, qui recherche l'adresse officielle se rapprochant le plus de celle donnée par l'enquêté ⁸³. Cette méthode autorise ensuite toute modification du

⁸²Se référer à Ressource System Group (2002); Adler *et al.* (2001) pour les résultats des méthodes utilisées par les répondants pour saisir les adresses des lieux fréquentés.

⁸³Le référentiel utilisé est le Lambert II étendu.

zonage du territoire (Adler *et al.*, 2001).

Dans certains cas, les internautes sont dans l'incapacité de saisir certaines adresses, soit parce que leur mémoire leur fait défaut, soit parce qu'il ne veulent pas révéler l'information (par exemple, une personne n'a pas renseigné l'adresse du domicile, élément par définition parfaitement connu). De plus, si le lieu d'origine ou de destination d'une activité se situe en dehors du périmètre de référence, l'adresse saisie est souvent très peu précise, car il s'agit d'un lieu inhabituel pour le répondant. Nous avons choisi de demander aux individus de saisir les adresses en clair, puis de les recoder a posteriori (géo-bornage) à l'aide de dictionnaires d'adresses. Depuis une quinzaine d'années, il est d'usage de raisonner en localisation fonctionnelle. Les lieux n'ont donc plus seulement une adresse, mais un nom, une fonction (enseigne, nom d'établissement ...). Pour permettre aux répondants web de situer le mieux possible la destination de leurs déplacements, une zone de saisie libre est proposée. Ils peuvent y indiquer notamment des noms de points d'intérêt (les gares, les universités, les écoles, les centres hospitaliers, les mairies ...) ou des intersections de rues. Une base concernant les points d'intérêt sera utilisée en deuxième recours, si le premier codage ne permet pas de définir des coordonnées précises.

III.2.2 Le géobornage : des résultats mitigés

Suite à l'administration d'une enquête à 600 individus, Adler *et al.* (2001) ont trouvé que près de 80% des adresses du domicile étaient parfaitement renseignées, les autres 20% préférant situer le lieu sur une carte afin que leur adresse précise ne figure pas dans la base de données. Environ la moitié des lieux de travail étaient identifiés à l'aide d'un nom de rue et la majorité des autres lieux (activités hors du domicile) à l'aide du nom d'établissement ou de points d'intérêts. Stopher *et al.* (2004b) recommandent que dans une enquête de qualité, 100% des adresses du domicile des répondants, 95% des adresses de leur lieu de travail habituel ou d'études et 90% des autres localisations soient correctement géocodées. Il existe plusieurs niveaux de géocodage, et le SCOT de Lyon a été divisé en 859 zones fines, qu'il est possible de regrouper en 158 zones agrégées. Ainsi, lorsque les adresses saisies ne sont pas assez précises pour qu'une zone fine soit affectée (exemple : seule la ville est renseignée), alors on fait correspondre une zone agrégée. En face-à-face, l'enquêteur peut montrer des cartes aux enquêtés, afin de les aider à situer les lieux fréquentés. Mais bien que les technologies disponibles autorisent la visualisation de cartes sur internet, les contraintes de l'expérimentation et les outils mis à disposition par Conversoft ne l'ont pas permis.

Les résultats du géobornage des adresses sont présentés dans le tableau 7. Il en ressort que l'adresse du domicile est correctement restituée. 64% des répondants indiquent les coordonnées précises de leur lieu de résidence, et 29% se contentent de livrer le nom de la rue. En ce qui concerne le lieu de travail ou d'études habituel, le niveau de précision est un peu plus faible (44% d'adresses exactes et 50% de coordonnées relatives au nom de rue), ces données étant probablement jugées plus confidentielles par les répondants. Les adresses des autres lieux fréquentés (achats, loisirs, accompagnement ...) sont moins biens

III Administration du questionnaire en ligne

retranscrites, bien que 38% soient exactes et 43% identifiés grâce au nom de la rue. Le géobornage des adresses est donc possible a posteriori, et conduit à des données de bonne qualité. Le temps de traitement n'est pas négligeable, et il est nécessaire de desserrer progressivement les contraintes du logiciel de géobornage pour arriver à identifier une adresse précise en fonction de la qualité des renseignements saisis.

Géobornage	Lieu du domicile	Lieu de travail	Autres lieux
Numéro exact	58%	39%	38%
Numéro le plus proche	6%	5%	4%
Centroïde rue	29%	51%	43%
Centroïde arrondissement	0%	1%	1%
Ville	7%	2%	13%
Non renseigné	0%	2%	1%
Total	369	282	1075

TAB. 7 – Qualité de géobornage des adresses saisies par les répondants
Source : EMD web - Lyon 2006

III.2.3 Un calcul des durées et distances parcourues proche de la réalité

Le géobornage permet d'obtenir une adresse exacte pour chaque lieu fréquenté et donc des coordonnées longitudinales et latitudinales. La distance de chaque déplacement est ensuite estimée à vol d'oiseau, à partir des coordonnées de l'origine $(x_{\text{départ}}, y_{\text{départ}})$ et de la destination $(x_{\text{arrivée}}, y_{\text{arrivée}})$ du déplacement, par la formule suivante :

$$Distance = \sqrt{[(y_{\text{arrivée}} - y_{\text{départ}})^2 + (x_{\text{arrivée}} - x_{\text{départ}})^2]} \quad (2)$$

La distance est fournie en mètre. Si l'adresse exacte des lieux visités n'est pas renseignée, alors la distance à vol d'oiseau en mètres entre les centroïdes des zones d'origine et de destination des déplacements est calculée. Au final, les distances calculées des déplacements recueillis dans le cadre de l'enquête web sont de bonne qualité, comme le montre le tableau 8.

Calcul de distance	Effectif	Pourcentage
Vol d'oiseau	1047	95%
Vol oiseau centroïde de zones	59	5%
Total	1108	100%

TAB. 8 – Précision du calcul de distance
Source : EMD web - Lyon 2006

Dans 95% des cas, la distance vol d'oiseau calculée des déplacements internes au périmètre d'enquête est précise, grâce au géobornage.

III.3 La validation du questionnaire

Nous avons montré dans la section I.1.6, qu'il était préférable de prévoir, des validations intermédiaires du questionnaire, afin de permettre un stockage des réponses. Dans l'enquête web, nous sollicitons quatre types de validation : une après chaque bloc ménage et individu, une après chaque déplacement, puis une validation finale. Les questions relatives au ménage et à l'individu sont peu nombreuses et rapides. En revanche, la saisie d'un déplacement est relativement longue, surtout lorsqu'il faut renseigner des adresses différentes de celle du lieu de domicile ou du lieu de travail, et demande un effort de mémoire plus important. Afin d'éviter un lourd travail de ressaisie en cas d'interruption de la connexion, une validation est requise à la fin de chaque déplacement, et les données sont sauvegardées dans la base de l'administrateur. Ces validations intermédiaires réduisent également le niveau de fatigue de l'enquêté, qui en cas d'erreur ou d'oubli peut revenir au déplacement concerné pour le corriger (Ressource System Group, 2002).

Il est impossible d'anticiper le moment de connexion des répondants, et donc le jour de référence utilisé pour la saisie des déplacements. A l'écran, le jour de référence apparaît automatiquement et correspond au dernier jour de la semaine avant le jour de la première connexion du répondant. Cette méthode pose deux problèmes. Si beaucoup de personnes se connectent durant le week-end, les déplacements du vendredi seront sur représentés et l'effet mémoire sera plus important pour les internautes qui répondent le dimanche ou le lundi. Par ailleurs, il n'est pas évident de se rappeler tous les déplacements effectués la veille ou l'avant veille, selon leur ordre chronologique, et il est fort probable que des répondants souhaitent modifier, ajouter ou supprimer des déplacements. Proposer un résumé des informations saisies sous forme graphique ou textuelle a l'effet d'une récompense pour certains répondants (Adler, 2003). Un schéma dynamique qui s'agrémente au fur et à mesure des réponses saisies permet de visualiser le cheminement des déplacements, mais pose des contraintes de faisabilité liées aux étapes de validation des données et au langage retenu par l'administrateur. Ce dernier a recommandé la production d'un tableau synthétique des réponses saisies à la fin du bloc 'déplacements', pour garantir une bonne présentation des données, quel que soit le type de matériel utilisé et le nombre de déplacements saisis. Il suffit alors à l'internaute de vérifier ces informations, puis de modifier les données ou d'insérer un nouveau déplacement. La suppression étant irréversible, une confirmation est demandée à l'enquêté lors de l'activation de cette commande, pour éviter des erreurs de manipulation pouvant conduire au découragement et à l'abandon du questionnaire. Une validation générale est également prévue à la fin du bloc 'déplacements', après que l'internaute ait vérifié l'ensemble de ses réponses.

III.4 L'imputation des données

III.4.1 Les valeurs aberrantes

Avant le traitement des données, il est nécessaire d'effectuer quelques tests pour vérifier la cohérence des réponses. Certains tests sont réalisés automatiquement, mais d'autres n'ont pas été prévus dans le script du questionnaire web. Par exemple, l'heure de début d'un déplacement doit être supérieure à l'heure de fin du déplacement précédent, mais les répondants ont la possibilité de rajouter deux déplacements 'oubliés' à la fin du questionnaire, sans que leur cohérence ne soit vérifiée (le logiciel CATI ne permet pas d'effectuer ce test).

La durée des déplacements est également un indice de la qualité des données. Une mauvaise compréhension de la définition d'un déplacement peut conduire à inclure la durée de l'activité dans celle du déplacement, et donc générer des durées aberrantes. Enfin, un nombre non négligeable d'individus ne retournent pas à leur domicile à la fin de la période d'enquête. Il semble que le dernier retour au domicile soit un déplacement fréquemment omis, car considéré comme peu important, ou par lassitude du répondant.

III.4.2 Les valeurs manquantes

La non-réponse partielle se définit comme l'impossibilité d'obtenir des données réelles et complètes de l'ensemble des répondants (Zmud et Arce, 2000). Cette non-réponse vient souvent grever la qualité de l'étude, et peut être minimisée en optimisant la construction et l'administration du questionnaire. Sur le web, il est fréquent que des alertes encouragent les répondants à répondre et dans notre étude, les réponses sont forcées. En effet, les répondants doivent impérativement sélectionner une modalité de réponse à chaque question, pour pouvoir accéder à la question suivante et donc progresser dans le questionnaire. Les répondants ont également la possibilité de saisir leur adresse e-mail à la fin du questionnaire, mais pour des raisons de délais (les résultats de l'enquête web ont été disponibles fin juillet, soit 3 mois après son lancement), nous n'avons pas utilisé cette information pour les relancer. Malgré les nombreuses alertes mises en place tout au long du questionnaire, les premiers tris effectués sur la base de données laissent apparaître un certain nombre de valeurs manquantes, qui ne permettent pas une exploitation satisfaisante des données. Ces valeurs manquantes s'expliquent par un refus du répondant de communiquer l'information (l'adresse de son domicile), un manque de connaissance ne lui permettant pas de le faire (l'adresse précise d'un magasin) ou une question filtrée en amont du questionnaire (la possession du permis de conduire pour les moins de 16 ans).

III.4.3 Les techniques d'imputation

Si les données manquantes peuvent être imputées suite à des analyses précises, il est toujours possible d'introduire des biais. En effet, le chargé d'étude peut être tenté d'effectuer des imputations 'logiques', alors que le comportement du répondant est en réalité inconsistant avec son profil ou ses déclarations

antérieures (Adler, 2003). Les procédures d'imputation permettent cependant de construire une information probable à partir d'une information incomplète, plutôt que de la rejeter en totalité (Bonnell et Le Nir, 1998). Nous avons imputé certains résultats de l'enquête web, bien que les données individuelles soient par définition difficiles à prédire par la plupart des méthodes d'imputation courantes (Stopher et Jones, 2003).⁸⁴.

Dans la plupart des cas, la durée des déplacements peut-être calculée à l'aide des heures et des minutes de début et de fin. Pour les déplacements internes au périmètre d'enquête pour lesquels la durée est inconnue ou aberrante, nous imputons la valeur à partir de la distance du déplacement⁸⁵. Pour cela, nous retenons la vitesse moyenne des déplacements réalisés avec le même mode de transport, en considérant seulement les déplacements dont la durée et la distance sont inférieures ou égales à la valeur du 90^e centile de durée et de distance, afin de tenir compte de certaines valeurs probablement aberrantes. Au final, il y a 63 déplacements enquêtés pour lesquels la durée a du être imputée. Les vitesses moyennes de référence par mode figurent dans le tableau 9.

Mode	Minimum	Maximum	Moyenne
Deux-roues	12,16	33,31	17,73
Marche à pied	0,2	9,4	2,90
Transport en commun	0,24	20,94	6,03
Vélo	3,03	16,55	8,08
Voiture conducteur	1,12	66,18	14,46
Voiture conducteur + TC	4,34	22,64	12,10
Voiture passager	3,32	55,17	14,95
Voiture passager + TC	4,54	4,54	4,54

TAB. 9 – Vitesses moyennes de référence par mode en km
Source : EMD web - Lyon 2006

IV Conclusion

Les enquêtes en ligne sont parfois risquées, mais une méthodologie adaptée aux objectifs et à la population étudiée peut générer une étude moins coûteuse, plus rapide et de bonne qualité. A l'issue de cette expérience, nous pouvons tirer certains enseignements relatifs à la conduite d'une enquête ménages déplacements sur le web.

La construction du questionnaire est un facteur clé de succès d'une enquête web. Selon Dillman *et al.* (1998b), il faut considérer les limites technologiques

⁸⁴Les méthodes utilisant les régressions multivariées et la maximisation de la vraisemblance sont statistiquement performantes et permettent d'utiliser les données d'enquêtes à des fins de modélisation (Acock, 1997).

⁸⁵L'imputation n'a pas été faite pour les déplacements ayant au moins une extrémité hors du périmètre d'enquête.

du web, ainsi que celles des utilisateurs, et comprendre comment les répondants s'attendent à ce que l'enquête se déroule. Dans les protocoles d'enquête mixtes, il est illusoire de chercher une standardisation quel que soit le mode utilisé, pour garantir la comparabilité des données, et préférable d'harmoniser la perception et la compréhension du questionnaire, plutôt que de le reproduire à l'identique. Nous avons suivi cette recommandation dans la conception du questionnaire web de Lyon, mais nous n'avons pas été assez loin en ce qui concerne le recueil des déplacements. L'analyse des résultats montre qu'il faut abandonner la logique déplacements et privilégier une logique activité pour le recueil des données de mobilité, les déplacements apparaissant alors 'naturellement' à chaque changement de localisation.

Par ailleurs, le recueil des localisations est un élément clé des enquêtes déplacements. Il est important de pouvoir localiser précisément les déplacements et de produire des données comparables à celles de l'enquête en face à face et compatibles avec le zonage du territoire. Si la mise en place d'un géocodage en ligne n'est pas envisageable (coût des bases de données, contraintes techniques...), il est indispensable de trouver d'autres moyens de recueillir les données. L'expérience de l'enquête web montre qu'il est possible de recueillir des adresses et de les géoborner a posteriori. Le traitement de l'information a toutefois nécessité un lourd travail, les adresses saisies étant parfois trop imprécises pour autoriser un traitement automatique. Il est donc souhaitable de tester une validation en temps réel de l'information.

Enfin, une enquête web est différente des autres enquêtes assistées par ordinateur. Il n'est donc pas judicieux d'utiliser des logiciels conçus pour des enquêtes CATI ou CAPI (logiciel d'aide à la réalisation d'enquêtes téléphoniques ou en face à face) lors de la réalisation d'une enquête web, et préférable de recourir à un logiciel CAWI (logiciel d'aide à la réalisation d'enquêtes web). Ces logiciels proposent une ergonomie du questionnaire aussi proche que possible des environnements que les internautes ont l'habitude de trouver sur la toile et permettent d'enregistrer les réponses dans des bases de données fonctionnelles et performantes. La tâche est d'autant plus délicate que les chercheurs en méthodes d'enquête ne sont pas familiers des enquêtes de déplacements, les planificateurs de transport ne connaissent pas les problèmes relatifs aux enquêtes et les ingénieurs et concepteurs d'enquêtes ne connaissent ni les problèmes de la recherche ni ceux de la politique des transports (Adler, 2003).

Les résultats de notre étude permettent de formuler des recommandations précises concernant la conduite d'enquête de mobilité sur le web. Le caractère exploratoire de cette recherche et les délais n'ont cependant pas permis de définir un cahier des charges suffisamment précis, à l'image de ce qui est disponible pour la réalisation des enquêtes en face à face.

Chapitre 3 : Premiers résultats sur le comportement de réponse des internautes

"Les outils nouveaux (le téléphone portable, les prothèses qu'on a créées, l'ordinateur, l'Internet, la mémoire de Google) nous changent, on est déjà des mutants."

Joël de Rosnay (2007).

Lors d'une enquête auto-administrée, l'enquêté se retrouve seul face au questionnaire. Il est donc impossible d'obtenir des informations précises sur son comportement de réponse. La situation est différente dans le cas d'une enquête assistée par ordinateur, comme les enquêtes CAPI ou CAWI (web). L'interface informatique utilisée pour la saisie des réponses autorise en effet le recueil de données comportementales. Nous avons saisi cette opportunité, et disposons d'un certain nombre de résultats, qui permettent de comprendre la manière dont les répondants ont appréhendé le questionnaire et de tirer des enseignements pour la mise en ligne de prochaines enquêtes.

L'enquête web concerne les individus ayant refusé de répondre en face-à-face ou restant injoignables. Nous présentons les résultats globaux de l'enquête face-à-face, en termes de taux de réponse, afin de définir la cible des répondants potentiels sur le web (section I). Puis, nous analysons en détail le taux de réponse à l'enquête web, ainsi que les raisons qui ont poussé les individus refusant de répondre en face-à-face à remplir le questionnaire en ligne (section II).

Deux relances étaient prévues, pour motiver les individus à répondre en ligne. Il est intéressant d'analyser l'impact de ces relances sur le taux de réponse et la qualité des réponses obtenues, au regard d'expériences similaires détaillées dans la littérature (section III). Un des principaux objectifs de notre recherche consiste à diminuer le biais de non-réponse constaté dans les enquêtes auprès du grand public. Avec le web, il est possible d'analyser le profil des individus selon le degré de remplissage du questionnaire et d'identifier les questions qui génèrent de nombreux abandons. Par ailleurs, si les réponses sont forcées, certaines questions offrent une modalité 'Ne sait pas' ou 'Non-réponse', utile pour ceux qui ne veulent pas se prononcer. Ces modalités sont des échap-

patoires, et nous étudions dans quelle mesure elles pèsent sur la non-réponse partielle.

Enfin, le web est un média peu contraignant, puisque les individus ont la possibilité de se connecter plusieurs fois pour remplir le questionnaire, à tout moment de la journée. La connexion sécurisée, à l'aide d'identifiants et de mots de passe permet d'analyser l'heure, la durée et le nombre de connexions des internautes (section IV).

I Un taux d'échec global à l'enquête en face-à-face non négligeable

I.1 Le calcul du taux de réponse

Selon (Ettema *et al.*, 1996), le taux de réponse du questionnaire, ainsi que la distribution des données est un indicateur de la qualité d'une enquête. (Stopher *et al.*, 2004b) précisent qu'un taux de réponse élevé permet de réduire a priori l'incidence du biais de non-réponse. Le taux de réponse à une enquête peut être formulé comme suit (Groves, 1989) :

$$\text{Taux de réponse global} = \frac{I}{I + P + NC + R + NI} \quad (3)$$

Avec :

- I : le nombre d'interviews complètes ;
- P : le nombre d'interviews partielles ;
- NC : le nombre de ménages éligibles à l'étude, mais non joignables ⁸⁶ ;
- R : le nombre de ménages éligibles à l'étude, mais refusant de répondre ;
- NI : le nombre d'autres ménages éligibles mais non-répondants (problèmes techniques, maladie ...)

Selon Ampt (1997), il est utile de distinguer l'échantillon brut, constitué de l'ensemble des ménages sélectionnés pour répondre à l'enquête, de l'échantillon net, obtenu en retirant les unités statistiques non pertinentes, sans affecter la qualité de l'enquête au regard de ses principaux objectifs. A contrario, les ménages qui ne peuvent être joints après plusieurs visites ou appels, ou dont le seul contact reste un répondeur viennent peser sur le taux de réponse, car leur absence de l'échantillon nuit à la qualité des données.

Dans l'enquête ménages déplacements réalisée en face-à-face, nous estimons un taux de réponse après contact :

$$\text{Taux de réponse} = \frac{\text{Nombre de ménages répondant à l'enquête}}{\text{Nombre de ménages contactés}} \quad (4)$$

⁸⁶Cette définition pose problème, car il est parfois difficile de faire la différence entre des ménages non éligibles et des ménages éligibles mais non joignables.

I.2 Estimation de la cible théorique de l'enquête web

Le taux d'échec global (1 - taux de réponse global) est calculé en additionnant le taux de refus et le taux de rebuts. Les refus concernent les ménages qui ont refusé de répondre à l'enquête en face-à-face. Les rebuts sont de plusieurs sortes :

- les ménages impossibles à joindre (jamais présents à leur domicile, ne veulent pas répondre au téléphone...);
- les ménages absents de longue durée (ne trouvent pas de moment pour répondre en famille pendant la durée de réalisation de l'enquête)⁸⁷;
- les ménages inéligibles (résidences secondaires, les logements vacants et les logements qui n'ont plus l'usage d'habitation);
- les autres cas de rebuts.

A la fin de la période d'enquête, 6 375 interviews ont été réalisées en face-à-face sur le SCOT de Lyon, pour un total de 11 951 ménages contactés, soit un taux de réponse global de 53%. Les résultats sont détaillés dans le tableau 10.

Statut	Effectifs	%
Refus	3 716	
	Taux de refus	31 %
Impossibles à joindre	1 531	
Absents de longue durée	44	
Autres cas de rebuts	301	
	Taux de rebuts	16 %
	Taux d'échec	47 %
Total sans autres rebuts	5 291	
Total avec autres rebuts	5 592	

TAB. 10 – Taux de réponse global enquête face-à-face sur le périmètre du SCOT de Lyon

Source : SOFRECO (suivi semaine 19 - 2006)

Seuls les ménages refusant de répondre à l'enquête standard, impossibles à joindre ou absents de longue durée ont été contactés par courrier pour répondre à l'enquête web. La cible théorique de l'enquête web est donc de 5 291 ménages. En réalité, seuls 4 335 ménages ont reçu un courrier les invitant à répondre sur le web. Les envois ont débuté avant la fin de l'enquête en face-à-face, et certains repérages complémentaires ont été réalisés sans qu'un double de la fiche adresse ne soit conservé par la société en charge de l'étude. On estime donc que 956 ménages non-répondants ou non joignables lors de l'enquête en face-à-face et pouvant potentiellement répondre sur le web n'ont pas été sollicités (18% de la cible théorique).

⁸⁷Les ménages sont considérés comme absents de longue durée, si l'enquêteur dispose d'informations explicites.

I.3 Des disparités spatiales

Le périmètre d'enquête (S.C.O.T. de Lyon) a été segmenté en 9 zones, selon la logique de couronnes et la segmentation Est-Ouest correspondant à la coupure du Rhône et à des différences sociologiques assez marquées. Nous observons que le taux d'échec global de l'enquête en face-à-face est inégalement réparti entre ces 9 zones (tableau 11 et figure 10), avec un pourcentage particulièrement élevé pour les zones 1 (Hypercentre), 3 (1ère couronne est), 4 (1ère couronne ouest) et 6 (2ème couronne ouest).

Pour les zones 1 (hypercentre), 4 (1ère couronne ouest) et 6 (2ème couronne ouest), c'est un fort taux de rebus (21%, vs. un taux moyen de 16% sur l'ensemble du périmètre) qui fait chuter le taux de réponse à l'enquête en face-à-face. Pour la zone 3 (1ère couronne est), en revanche, c'est le nombre important de refus qui pèse sur le niveau de qualité de l'enquête (42% des personnes contactées refusent d'être interviewées). Deux communes enregistrent des résultats particulièrement mauvais : Bron et Vaulx-en-Velin (45%). Les interviews en face-à-face ont débuté à la fin de l'année 2005, lorsqu'est survenue la crise des banlieues, et les communes de Bron et Vaulx-en-Velin regroupent un certain nombre de quartiers dits 'sensibles'.

Zones	Total	Taux de refus	Taux de rebus	Taux d'échec global
Hypercentre	1 690	30 %	21 %	51 %
Lyon-Villeurbanne	3 249	29 %	16 %	45 %
1re couronne Est	1 535	42 %	10 %	53 %
1re couronne Ouest	1 851	30 %	21 %	51 %
2è couronne Est	1 148	33 %	8 %	41 %
2è couronne Ouest	890	28 %	21 %	49 %
3è couronne Est	1 091	29 %	10 %	38 %
Externe	497	23 %	16 %	39 %
Total	11 951	31 %	16 %	47%

TAB. 11 – Taux de réponse enquête face-à-face par zone sur le périmètre du SCOT de Lyon

Source : SOFRECO (suivi semaine 19 - 2006)

I.4 Les raisons de la non participation à l'enquête en face-à-face

Il est intéressant de connaître les raisons pour lesquelles des individus échappent à l'enquête standard, mais sont motivés pour participer à l'étude en ligne. A la fin du questionnaire web, les répondants avaient la possibilité de s'exprimer sur ce point. Six modalités sont proposées : l'attrait du web comme média de réponse, le refus des interrogés de recevoir un enquêteur à leur domicile, le manque de disponibilité des ménages sur les créneaux proposés, l'absence de temps pour répondre à l'étude, l'absence de contact avec un enquêteur ou d'éventuelles autres raisons. Les réponses pouvaient être multiples,

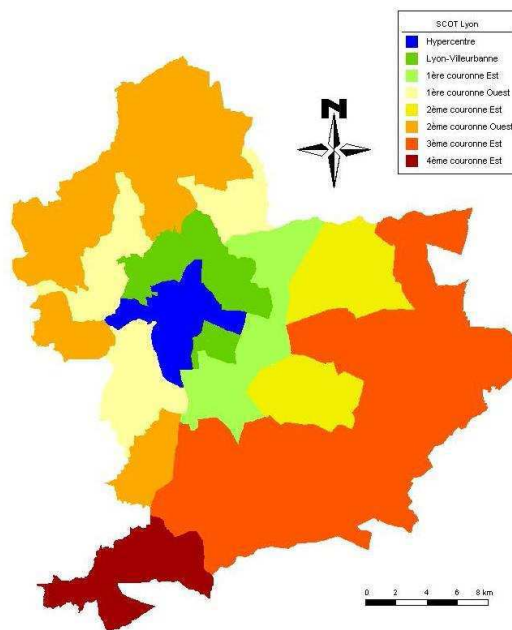


FIG. 10 – Les 9 zones du périmètre du SCOT de Lyon
Source : EMD Lyon 2006

et un champ textuel permettait aux répondants de saisir librement un commentaire. Au final, 365 personnes ont renseigné cette question et 33 ont ajouté un commentaire. Les résultats figurent dans le tableau 12.

L'engouement pour le web semble évident (*'c'est plus facile sur le web'*), même si certains enquêtés prétendent seulement avoir *'reçu un courrier me demandant une réponse par Internet'*. La mixité des modes d'enquêtes permet donc de toucher un plus grand nombre de personnes (*'vous ne m'avez pas laissé le choix. Je ne peux vous répondre que par le net'*). Les enquêtes à domicile sont en effet difficiles à réaliser (*'je n'ai pas envie d'être dérangé chez moi'*), et certaines personnes mettent en avant la *'surabondance de démarcheurs par téléphone'*. Le manque de disponibilité des répondants impacte fortement la non-réponse à l'enquête standard, qu'il s'agisse de contraintes professionnelles (*'les enquêteurs n'ont jamais réussi à me contacter compte tenu de mes horaires de travail et ont fini par m'adresser un courrier', 'quelqu'un est venu et je n'étais pas là. Etudiante, pas facile pour les horaires', 'probablement l'enquêteur n'a pas réussi à me joindre pour fixer RDV car souvent absente ou rentre tard le soir', 'entreprise ouverte du lundi au samedi inclus'*), de congés (*'en vacances lorsque l'enquêteur est passé'*) ou des exigences méthodologiques de l'étude (*'l'enquêteur proposait que les 4 personnes soient présentes !! mission impossible', '...et en plus vous exigez que toute la famille soit présente au moment de l'enquête!!!', 'j'ai reçu un enquêteur à mon domicile mais j'étais seule et l'enquête n'a pas été retenue'*). Et même lorsque les personnes sont disponibles, elles sont parfois dans l'incapacité de répondre aux enquêtes (*'Maman (je réponds pour elle) est sous tutelle et ne répond pas aux enquêteurs', 'c'est ma mamie qui m'a demandé de le faire pour elle', 'ma femme rentrait de cure*

I Un taux d'échec global à l'enquête en face-à-face non négligeable

lorsque l'enquêtrice est passée; elle n'était pas au courant de sa venue et j'étais absent à ce moment; si j'avais été là, je l'aurais reçue et aurais répondu au questionnaire', 'Je suis en appartement avec mon frère! Peut-on considérer cela comme un ménage?')

Raisons invoquées	Occurrences	Ménages concernés
Préfère répondre en ligne	145	40%
Ne souhaite pas recevoir un enquêteur	124	34%
Peu disponible aux créneaux proposés	107	29%
Aucun contact avec un enquêteur	101	28%
Pas le temps	75	21%
Autre	26	7%
Total	578	

TAB. 12 – Les différentes raisons du choix du web comme média de réponse
Source : EMD web - Lyon 2006

Par ailleurs, selon les déclarations recueillies, quelques enquêteurs manquent de professionnalisme. Certains ne soignent pas assez la phase de recrutement (*'étudiante nous a laissé un message sur 1/2 page écolier déchirée, numéro personnel, ...', 'J'ai eu un seul message d'un enquêteur, qui me demandait de le rappeler. C'est très bête, mais je n'avais pas l'intention de payer le téléphone pour répondre à votre enquête', 'L'enquêteur s'est présenté à mon domicile sans avoir pris de rendez-vous préalable'*). D'autres négligent les rendez-vous avec les ménages (*'J'avais convenu d'un rendez-vous avec un enquêteur, mais il ne s'est pas présenté, ni téléphoné pour annuler le rendez-vous', 'Deux rendez-vous pris avec vos enquêteurs ont été annulés par eux. J'avais rappelé pour refixer un rendez vous et vous n'avez jamais rappelé. C'est plus pratique de répondre sur internet', 'rendez-vous manqué avec votre représentant'*). Les sollicitations incessantes jouent négativement sur le niveau d'acceptation des études, ce qui se traduit également par des commentaires virulents (*'Gaspillage d'argent, nous avons reçu 4 courriers', 'je vois pas trop à quoi sert cette enquête ... on es en vacances, donc pas beaucoup de déplacements, donc vous ne comprenez pas nos galères d'attendre longtemps un bus alors qu'il ne passe pas pour X raisons!!!', 'les enquêtes de ... comme celle la me font ..., je perd mon temps!'*). Enfin, certains répondants ont compris tardivement l'importance de l'étude (*'Après plusieurs courriers j ai compris que cette enquête était importante'*, ce qui légitimise l'utilisation des relances. D'autres ont été interpellés par le sujet de l'étude (*'je suis intéressée par la protection de l'environnement donc les déplacements urbains et je suis en train d'adhérer au PDE de mon entreprise'*) ou espèrent peser sur l'amélioration de leurs conditions de transport quotidiennes (*'volonté que Charbonnières centre soit mieux desservie dans le futur'*).

II Une expérience encourageante

II.1 Les taux de réponse constatés dans les protocoles d'enquêtes mixtes

Nous avons vu dans la section I, que le taux de réponse à une enquête se définit comme le ratio du nombre d'interviews complètes, par le nombre d'unités éligibles à l'enquête. Dans le cas de l'enquête web, il est difficile de connaître précisément le nombre de ménages éligibles, c'est-à-dire en mesure de se connecter à internet.

Les expériences d'enquêtes mixtes incluant le web étant restreintes dans le domaine des transports, nous faisons référence à des études réalisées dans d'autres domaines, tels que la santé, l'éducation ou la connaissance. Les recherches de Stanton (1998) montrent que les enquêtes web génèrent des taux de réponses relativement instables. Il peut être faible, de l'ordre de 8% (Smith, 1997) ou très élevé, atteignant les 90% (Zhang, 1999). Cependant, peu d'enquêtes web atteignent un taux de réponse de plus de 50%, et selon Falconer et Hodgett (1999) et Burke et James (2006) il semble que ce taux se situe davantage entre 10 et 35%. Schuldt et Totten (1994) et Kittleson (1995) précisent que lorsque le web est utilisé comme mode de recueil principal, le taux de réponse peut même se situer en dessous de celui d'une enquête postale. Schaefer et Dillman (1998) montrent que dans un protocole multimode, le taux de réponse à l'enquête web augmente. Cobanoglu *et al.* (2001) concluent à l'existence de grandes variations en termes de taux et de vitesse de réponse pour les enquêtes web. Comparer les taux de réponse de différents modes est complexe, puisqu'ils dépendent de la procédure et de la zone d'enquête. Dans un protocole d'enquête mixte, il n'est pas toujours possible de maintenir une procédure strictement identique entre deux enquêtes, puisque chaque mode a des caractéristiques particulières (Bonnell, 2003).

Le tableau 13 donne quelques résultats concernant le taux de retour de différents médias utilisés dans des enquêtes mixtes, et il semble que le taux de réponse soit plus faible lorsque le questionnaire est rempli sur le web.

Enquête	Description	Taux de réponse
Vehovar et al. 2000	Enquête sur l'usage du e-commerce en Slove- nie. Répondants qui ont accès à internet.	Téléphone 52%
		Postal 39%
		Fax 32%
		Web 26%
Cobanoglu et al. 2001	Enquête auprès de 300 professeurs hospitaliers (USA)	Postal 26.39%
		Fax 17%
		E-mail/Web 44.2%
Crawford et al. 2002	Enquête auprès d'étudiants sur l'usage des stupéfiants	Postal 40%
		Web 63%
Huang 2006	Enquête auprès de collégiens	Papier 85%
		Web 63%

TAB. 13 – Etude comparée des taux de réponse dans des enquêtes mixtes
Source : Adapté de M. Lang, 2002

Les comparaisons sont cependant risquées, car les variables et les disciplines diffèrent. Dans certaines enquêtes le choix du média est laissé au répondant alors que dans d'autres ce choix est imposé. Le processus d'enquête varie parfois selon les modes et les taux de couverture des médias utilisés sont très différents. Il ne fait cependant aucun doute que les recherches futures permettront de rendre le taux de réponse aux enquêtes web comparable à celui des autres modes traditionnels, bien que ces procédures augmenteront substantiellement le coût de l'étude (Dillman et Bowker, 2001).

II.2 Une réactivité importante

L'enquête web s'adressant aux non-répondants à l'enquête traditionnelle, la diffusion du questionnaire en ligne était fortement liée au calendrier des interviews en face-à-face. Le lancement du terrain en face-à-face a eu lieu début novembre 2005, avant d'être suspendu pendant deux semaines, suite aux événements dans les banlieues françaises et à une perturbation du réseau de transport en commun. 4 335 courriers ont été envoyés aux répondants potentiels sur le web par les quatre bureaux d'enquêtes principaux de l'agglomération, en deux vagues, de manière à limiter le délai entre le refus et la relance. La première a débuté en avril 2006 (1 838 ménages contactés) et la seconde s'est déroulée à la mi-mai (2 497 ménages contactés). Pour chaque vague, deux relances étaient prévues auprès des ménages qui ne s'étaient pas connectés sur le site pour répondre à l'enquête en ligne⁸⁸. Comme la figure 11 le montre, la réactivité des internautes est très forte, et l'envoi des lettres avis ou des relances génère une augmentation du nombre de connexions les jours suivants.

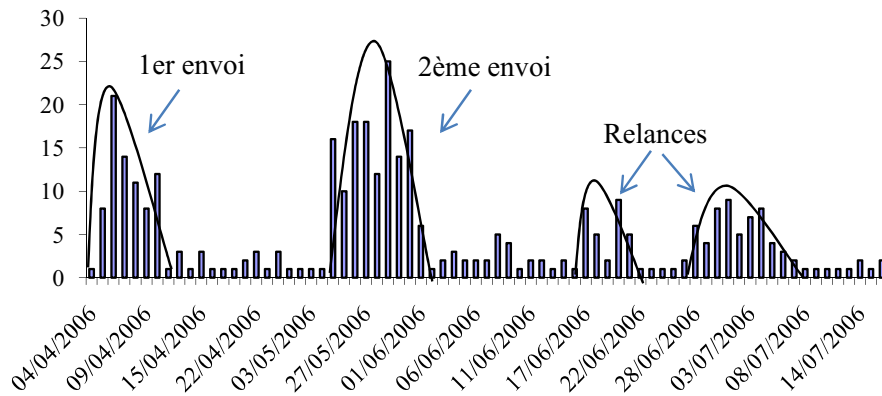


FIG. 11 – Dates de connexion des internautes
Source : Questionnaire web (2005)

II.3 Un taux de réponse satisfaisant

Au final, 536 individus ont accédé au site pour tenter de répondre à l'enquête (678 connexions enregistrées), ce qui représente un taux de connexion de

⁸⁸Les relances ont eu lieu les semaines du 22 mai et du 26 juin 2006 pour la première vague, les semaines du 12 et du 26 juin 2006 pour la deuxième vague.

Premiers résultats sur le comportement de réponse des internautes

12,4%. Ce chiffre est satisfaisant, puisque les ménages concernés par l'enquête web sont ceux qu'il n'avait pas été possible d'atteindre durant la première phase de l'enquête (face-à-face). De plus, le mode de recrutement n'est pas optimisé, puisqu'il est plus encourageant d'avoir à cliquer sur un lien pour répondre à une étude, que de suivre les instructions qui figurent sur une lettre pour pouvoir accéder au questionnaire en ligne (Conn, 2005). Parmi ces 536 contacts, 11 ont effectivement répondu au questionnaire, mais les données saisies ont été perdues suite à un problème informatique au démarrage de l'enquête. Ces 11 ménages de la vague 1 ont été relancés, puis classés en 'refus'. Cependant, tous les ménages qui se sont connectés n'ont pas terminé le questionnaire. Dans l'estimation du taux de réponse, nous ne retenons que les interviews 'exploitables', c'est-à-dire contenant au moins le cycle de déplacements de la personne interrogée. Après apurement, nous disposons de 364 interviews complètes (la personne interrogée est allée jusqu'à l'écran de validation finale du questionnaire) et de 369 interviews 'exploitables'. La population que nous allons étudier se limite alors à 369 ménages (ou individus), enregistrant un total de 1 108 déplacements, soit un taux de réponse de 8,5% (tableau 14). Ces chiffres sont cohérents avec les résultats de Lozar Manfreda et Vehovar (2002b), qui obtiennent 70% de questionnaires complets, parmi ceux ayant commencé à le remplir.

Nombre de connexions	Nombre de ménages	Bloc ménage	Bloc personne	Bloc déplacements	Validation finale
1	434	354	333	290	288
2	73	64	64	58	55
3	22	21	21	19	19
4	4	4	3	1	1
5	3	1	1	1	1
Total	536	444	422	369	364
% abandons	NA	17%	21%	31%	32%

TAB. 14 – Taux d'abandon cumulé du questionnaire web
Source : EMD web - Lyon 2006

On peut se demander si la taille de l'échantillon web est suffisante pour mener une analyse de la mobilité individuelle, c'est-à-dire si les indicateurs statistiques retenus se rapprochent de ceux réellement observés dans la population étudiée (Hill, 1998). Selon Roscoe (1975), il est important de respecter quelques règles avant de se lancer dans des analyses statistiques. Le nombre acceptable de répondants permettant de produire des analyses de qualité dépend du type d'étude⁸⁹. Dans les recherches expérimentales, où des variables explicatives sont manipulées dans le but d'évaluer la variation d'une variable

⁸⁹Selon Gay et Diehl (1992), de larges échantillons augmentent la probabilité d'obtenir des résultats significatifs aux analyses statistiques. En effet, pour de grands échantillons, une très petite différence observée entre deux moyennes peut apparaître significative, et être difficilement exploitable. Les recherches effectuées ne sont alors pas pertinentes. D'un autre côté, une différence significative obtenue entre deux petits échantillons est souvent complexe à expliquer, mais plus utile pour la recherche effectuée.

d'intérêt, les échantillons doivent contenir au moins 30 observations pour bénéficier des avantages du théorème central limite. Si des analyses sont menées sur des sous-échantillons, alors cette règle s'applique aux sous-échantillons considérés. Enfin, dans des analyses multivariées, la taille de l'échantillon doit être au moins dix fois supérieure au nombre de variables concernées. L'enquête ménages déplacements de Lyon est le fruit d'une recherche expérimentale et une expérience pilote. Au regard des règles citées ci-dessus, les 369 questionnaires recueillis en ligne représentent un effectif suffisant pour mener une analyse comparative de la mobilité.

II.4 La conversion des non-répondants

Le profil des répondants aux enquêtes web diffère de celui des non-répondants, même parmi ceux qui ont un accès individuel à internet. Les répondants, aussi nombreux soient-ils, peuvent constituer un échantillon non représentatif de la population (Stanton, 1998). Selon Crawford *et al.* (2002), différentes raisons expliquent le fort taux de non-réponse observé sur le web. Des individus ne sont pas dans la cible, n'ont pas consulté leur messagerie, ont rencontré des problèmes techniques, ont perdu la lettre avis, ont oublié de répondre, n'ont pas pris le temps de le faire, n'ont pas intérêt pour l'étude, craignent pour la confidentialité des données, ont une adresse e-mail erronée... Dans la littérature, on regroupe classiquement les refus en deux catégories (Sudman et Bradburn, 1996). Les 'hard refusals' concernent des individus réfractaires aux enquêtes, qu'il va être très difficile de faire changer d'avis. Les 'soft refusals' peuvent accepter de participer à l'étude après une ou plusieurs relances, en fonction de leurs convictions.

Couper et Rowe (1996) recensent les principales causes de refus invoquées par les répondants lorsqu'ils sont sollicités pour participer à une enquête. Le premier facteur est temporel ('je suis trop occupé', 'je n'ai pas le temps'), le second vient du manque d'intérêt des individus pour le sujet étudié ('je ne suis pas intéressé par...', 'je ne sais rien sur...'). L'auteur va plus loin et remarque des différences sociodémographiques entre les répondants, selon le motif de non participation. Ceux qui se disent trop occupés sont davantage diplômés, possèdent des hauts revenus, habitent dans de grandes agglomérations et subissent des contraintes temporelles importantes (actifs). Les individus plus vieux et moins éduqués refusent de répondre en invoquant plus souvent le manque d'intérêt pour le sujet. Si ces motifs peuvent être considérés comme une manière polie de décliner l'invitation de l'enquêteur, le fait de citer le manque de temps peut avoir un impact sur les enquêtes de mobilité, le temps étant une variable explicative du nombre de déplacements effectués.

Andrews *et al.* (2003) énumèrent quelques caractéristiques techniques qui impactent le taux de réponse aux enquêtes web : les éventuels problèmes techniques, la présence de caractères ambigus dans le mot de passe, les problèmes de confidentialité des données, les sauvegardes partielles des réponses, le placement des questions personnelles dans le questionnaire, le mode de contact (protocole en plusieurs étapes, mixité des modes, personnalisation de la lettre-

avis) les relances. . . Groves *et al.* (1992) complètent la liste avec les facteurs socio-économiques, les répondants ayant un niveau d'éducation et de vie plus élevés que les autres. Enfin, certains individus ont une propension à répondre aux enquêtes, alors que d'autres ne l'ont pas ou bien sont lassés d'être sollicités (Bickart et Schmittlein, 1999). Trois domaines de recherches sont particulièrement florissants dans la littérature : les leviers du taux de réponse (Yu et Cooper, 1983; Dillman, 2000), l'estimation et la correction du biais de non-réponse et la différence de comportement entre les répondants et non-répondants (Couper et Rowe, 1996).

Lors de l'enquête web de Lyon, un certain nombre de personnes ont accepté de livrer des informations personnelles. En fin de questionnaire, il est demandé aux répondants de bien vouloir saisir leur adresse e-mail. Cette question n'est pas obligatoire, mais 40% répondent positivement et livrent une adresse exploitable, s'exposant ainsi au risque d'être recontactés par le gestionnaire de l'enquête. Parmi eux, 15% communiquent leur adresse professionnelle, même s'ils déclarent posséder une connexion internet à domicile. Une investigation sur les caractéristiques socio-économiques de ces répondants montre que ce sont davantage les hommes actifs, diplômés du supérieur et occupant un emploi de cadre ou une profession intellectuelle supérieure qui saisissent volontairement leur adresse e-mail (annexe VIII). Ces personnes ne sont pas réticentes à livrer des informations personnelles, puisque 90% déclarent leur niveau de revenus (vs. 76% parmi ceux ne saisissant pas leur adresse e-mail). Les ménages concernés semblent avoir des ressources plus importantes, même si les faibles effectifs considérés ne nous permettent pas de conclure de manière significative. En revanche, il n'y a pas de différence selon l'âge.

III Analyse de la non-réponse

III.1 Non-réponse totale et non-réponse partielle

Groves et Couper (1998) distinguent les non-réponses totales ('unit non response'), lorsqu'un individu ne participe pas à l'étude, des non-réponses partielles ('item non response'), si l'individu participe à l'étude mais laisse quelques questions sans réponse. La non-réponse totale est la conséquence d'une absence de motivation du sujet, qui ne veut pas ou ne peut pas répondre à l'enquête (questionnaire non reçu, pas de temps pour répondre, incapacité à trouver les informations demandées, problèmes techniques . . .). La non-réponse partielle survient lorsque les individus ne sont pas confortables avec certaines questions, mais sont motivés et en mesure de répondre au reste de l'étude. Il est parfois difficile de respecter cette césure. Des individus qui accèdent au questionnaire et prennent connaissance des questions sans y répondre sont généralement classés comme non-répondants et ceux qui abandonnent la saisie du questionnaire viennent alimenter la part des non-réponses partielles. Mais il est difficile de distinguer un abandon en cours de remplissage de la volonté d'éviter certaines questions.

Dans les enquêtes web, contrairement aux autres modes, le processus de ré-

ponse des individus est tracé automatiquement. Il est donc possible de recueillir des données qui permettent d'expliquer en profondeur les variations potentielles dans la participation aux enquêtes. Ces données, complémentaires aux questions de l'étude et récoltées passivement, sont qualifiées de méta-données (Bosnjak et Tuten, 2001). Pour recueillir des informations pertinentes, il est nécessaire que chaque question apparaisse séparément sur un écran, que l'internaute ne soit pas contraint de saisir une réponse pour accéder à l'écran suivant et que chaque page soit téléchargée indépendamment sur le serveur. Ces conditions sont difficiles à remplir et l'enquête web menée à Lyon ne les respecte pas toutes. Les méta-données collectées permettent cependant d'établir un premier diagnostic des abandons et de la non-réponse partielle.

III.2 Les causes d'abandon

Il est nécessaire de comprendre quelles sont les principales variables amenant des individus motivés pour participer à l'étude à interrompre le remplissage du questionnaire. Le taux d'attrition par section du questionnaire, ou le nombre de personnes qui commencent une enquête mais ne la terminent pas, peut révéler des jugements systématiques par certains groupes de la population et pèse sur la qualité de l'étude (Andrews *et al.*, 0003). Jeavons (1999) a mis en évidence trois principaux moments où les individus risquent d'abandonner la saisie des réponses : la première question, l'apparition d'une grille complexe de questions et la saisie de leur adresse e-mail. Certains autres écueils sont à éviter, comme les échelles de réponses trop complexes, qui demandent un grand effort de compréhension et qui aboutissent souvent à des abandons ou à des erreurs de mesure (Dillman, 1998). Il en est de même des questions ouvertes, qui alourdissent la tâche du répondant et sont difficilement analysables (Frary, 1996), ou des questions trop personnelles placées au début du questionnaire, qui effrayent les répondants. Les abandons volontaires et ceux relatifs à des problèmes techniques sont souvent difficiles à distinguer.

Etant donné la longueur du questionnaire utilisé pour l'enquête ménages déplacements de Lyon, et son caractère auto-administré, certains individus n'ont pas pu (modem bas-débit, navigateur inadéquat et autres contraintes techniques...) ou pas voulu (questions parfois très personnelles, qui demandent un important effort de mémoire...) répondre entièrement à l'enquête et ont interrompu prématurément la connexion. Le taux d'abandon progresse par paliers (tableau 14). Deux phases semblent importantes : la validation du bloc 'Ménage' et celle du bloc 'Déplacements' (respectivement 17% et 10% d'abandons). Le graphique 12 montre l'évolution du nombre d'abandons en fonction de la progression dans le questionnaire. Un répondant sur cinq interrompt la connexion après avoir saisi son identifiant et son mot de passe. Le reste des abandons est assez progressif, bien que deux moments soient critiques : la question sur le nombre de personnes composant le ménage (14%) et l'introduction de la partie relative à la saisie des déplacements (16%). Il est difficile de trouver des facteurs socio-économiques explicatifs de ces comportements, au regard des données recueillies.

Premiers résultats sur le comportement de réponse des internautes

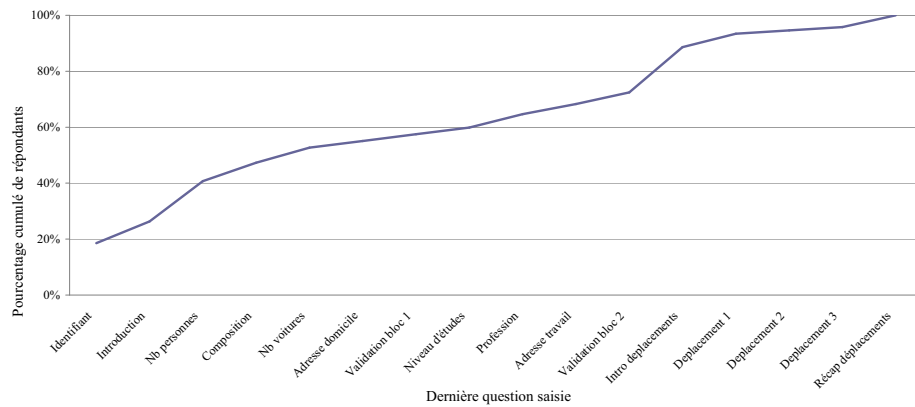


FIG. 12 – Evolution du nombre d'abandons selon la progression dans le questionnaire

Source : EMD web - Lyon 2006

Par ailleurs, le logiciel CATI utilisé pour l'enquête web impose au répondant de sélectionner une modalité à chaque question, afin de pouvoir accéder à l'écran suivant. Cette contrainte technique consiste à forcer les réponses, puisque les individus qui ne répondent pas aux questions sont relancés. Ces relances permettent de diminuer la non-réponse partielle, mais peuvent jouer sur les abandons.

Nombre de connexions	Nombre de ménages	Pourcentage de ménages	Durée moyenne par connexion (mn)
1	143	86%	12
2	16	10%	13
3	3	2%	11
4	3	2%	8
5	2	1%	10
Total	167	100%	12

TAB. 15 – Comportement de réponse des répondants partiels

Source : EMD web - Lyon 2006

Grâce au web, il est possible de recueillir des informations sur les répondants partiels au questionnaire, c'est-à-dire ceux qui n'ont pas terminé la saisie des réponses (tableau 15). La majorité d'entre-eux n'ont effectué qu'une connexion, mais 24 personnes (soit 15% des répondants partiels) ont accédé plusieurs fois au questionnaire web. Parmi les personnes qui se sont connectées plusieurs fois, seulement 42% l'ont fait le même jour et 50% dans un délai de 48h. Une hypothèse est le manque de motivation des répondants, qui n'ont pas mesuré l'importance de l'étude. Il se peut également que suite à des problèmes rencontrés pour saisir leurs réponses, ils aient tenté de joindre l'assistance technique à leur disposition. Les données disponibles ne nous permettent toutefois pas de conclure. Le temps de connexion est stable et relativement faible, les répondants partiels passant en moyenne 12 minutes devant leur ordinateur à

chaque connexion. Ce temps est moins élevé que celui observé chez les personnes ayant terminé le questionnaire (18 minutes). En revanche, aucune différence sociodémographique est notée, et l'âge moyen des répondants partiels ayant complété le tableau de composition du ménage est identique à celui des répondants (44 ans).

III.3 Un taux de non-réponse partielle modéré

Certains auteurs utilisent la non-réponse partielle comme critère d'évaluation de la qualité d'une étude (Atrostic et Burt, 1999). Le nombre de valeurs manquantes ou erronées varient selon le type de variables. Il est intéressant d'estimer la qualité d'une enquête non pas en s'intéressant au nombre de valeurs manquantes totales, mais en ne retenant que celles liées aux questions stratégiques pour une enquête déplacements⁹⁰. Cette non-réponse peut s'avérer problématique dans des enquêtes de mobilité, lorsque les informations saisies ne permettent pas au chargé d'études de localiser précisément les lieux visités. Le répondant peut ne pas être capable de se remémorer une adresse précise, surtout si ce lieu n'est pas celui de son domicile ou de son travail, et la possibilité offerte par le questionnaire, de renseigner l'intersection de rues ou le point d'intérêt le plus proche (mairie, école...) ne permet pas toujours un géobornage précis a posteriori. Il se peut également que les adresses saisies, bien que précises, ne soient pas identifiées par le logiciel de géobornage. Cela concerne notamment les adresses d'entreprises dans des zones industrielles, celles de centres commerciaux ou d'habitations dans des lieux-dits. Enfin, le répondant peut juger que les informations relatives aux lieux visités sont trop personnelles et préférer ne pas remplir certains champs du questionnaire. Les recommandations que nous avons formulées ci-dessus sur le design du questionnaire adapté, la formulation des questions, l'utilité du prétest et la prise en compte des thèmes sensibles tels que le revenu sont un moyen efficace de limiter les non-réponses partielles dans les enquêtes.

Face à des questionnaires relativement longs et à des personnes souvent pressées, certaines questions restent sans réponse. Ce phénomène est particulièrement vrai, lorsque le protocole d'enquête ne prévoit pas l'obligation de répondre aux questions ou lorsqu'un échappatoire (sous la forme 'NSP') est présent. Les questionnaires incomplets ou les réponses peu fiables se multiplient, nuisant à la qualité des données et à la pertinence des résultats obtenus (Morris et Adler, 2003). La structure du questionnaire web utilisé à Lyon impose le choix d'une réponse pour passer à la page suivante. L'analyse de la non-réponse partielle est donc complexe, bien que l'étude des réponses de type 'Autre', 'Ne sais pas' ou 'Non-réponse', assimilées à des non-réponses partielles, donne des pistes intéressantes. On peut se demander si ces modalités pèsent sur le taux de non-réponse partielle. De Rouvray et Couper (2002) postulent que proposer explicitement un choix de ne pas répondre favorise la non-réponse. Les détails des non-réponses de l'enquête web sont donnés dans le tableau 16.

⁹⁰Stopher *et al.* (2004b) définissent un indicateur statistique des valeurs imputées.

Premiers résultats sur le comportement de réponse des internautes

Nombre de non-réponses	Nombre de personnes	Pourcentage des non-réponses	Pourcentage des répondants (N=369)
0	259	-	70%
1	71	65%	19%
2	16	15%	4%
3	8	7%	2%
4	15	14%	4%
Total	110	100%	100%

TAB. 16 – Répartition des répondants selon le nombre de non-réponses partielles au questionnaire

Source : EMD web - Lyon 2006

Le tableau 17 présente le nombre de non-réponses aux questions prévoyant un échappatoire.

Intitulé de la question	Nombre de non-réponses
Possession du permis	1
Téléphone fixe	28
Téléphone portable	29
Connexion internet	21
Connexion haut débit	6
Revenus	90

TAB. 17 – Occurrence des non-réponses aux questions proposant la modalité de réponse 'Ne sait pas' ou 'Non-réponse'

Source : EMD web - Lyon 2006

Les modalités 'ne sait pas' et 'non-réponse' ont été utilisées de manière modérée par les répondants. 25% des personnes ne se sont pas prononcées sur le niveau de revenu du ménage, résultat conforme à nos attentes et qui encourage la présence d'un échappatoire dans les questions relatives aux gains financiers des individus. Les autres questions concernées par la non-réponse sont relatives à l'équipement du ménage en moyens de communication. Qu'il s'agisse de la possession d'un téléphone fixe, d'un téléphone portable, ou d'une connexion internet au domicile, le nombre de non-réponse est similaire. En revanche, les internautes connaissent le type de connexion dont ils disposent, et sont favorables à délivrer l'information.

Bien que la réponse aux questions soit forcée, il semble que quelques individus aient réussi à progresser dans le questionnaire en omettant certaines réponses. Il s'agit notamment des questions relatives aux caractéristiques du répondant (sexe, âge et occupation des personnes), aux adresses du domicile et du lieu de travail ou d'études et de celles concernant l'habitude d'utilisation des différents modes de transport. Ces oublis sont marginaux et ne concernent qu'une ou deux personnes selon les questions.

III.4 Des relances efficaces

Prévoir des relances est un bon moyen de réduire le biais de non-réponse. Keeter *et al.* (2000), Kalfs et Van Evert (2003) et Goyder (1985) montrent qu'il est possible d'augmenter le taux de réponse en déployant des efforts pour contacter les personnes difficilement joignables. Pour être efficaces, ces relances doivent être programmées dès la conception de l'étude, et effectuées à différents moments de la journée. En effet, plus les individus sont sollicités, plus ils ont des chances d'accepter le questionnaire, bien qu'il existe un seuil de 6 tentatives au-delà duquel l'impact sur la non-réponse est marginal (Harpuder et Stec, 1999) ⁹¹.

Les répondants web sont des non-répondants à l'enquête standard. Ils ont donc été sollicités plusieurs fois (8 au maximum) pour fixer un rendez-vous avec un enquêteur à domicile. Après réception du courrier les informant de la mise à disposition du questionnaire en ligne, deux lettres de relance étaient prévues, à un mois d'intervalle. Elles ont permis de récupérer certains individus qui n'avaient pas répondu au questionnaire web, suite à l'envoi du premier courrier (annexe VII).

Si les relances semblent efficaces pour augmenter le taux de réponse à l'enquête, qu'en est-il de l'impact sur la qualité des données ? Les travaux de Brög et Meyburg (1980, 1982) montrent suite à l'étude des résultats d'une enquête de mobilité réalisée par voie postale en Allemagne, que les répondants aux relances déclarent moins de déplacements que les autres et que les immobiles sont sur-représentés. L'hypothèse avancée est que les hyper mobiles répondent rapidement à une étude dont le sujet les intéresse (les déplacements), alors que les individus à faible mobilité ne se sentent pas concernés par le sujet et attendent plusieurs relances pour répondre au questionnaire. Les résultats de l'enquête web de Lyon vont dans ce sens, comme le montre le tableau 18.

		Nb de personnes	Mini	Maxi	Moyenne	Ecart-type	% mobiles
Total répondants	Total	369	0	15	2,97	2,33	81%
	Mobiles	299	1	15	3,71	2,02	100%
Répondants immédiats	Total	235	0	15	3,02	2,33	83%
	Mobiles	196	1	15	3,62	2,08	100%
Répondants relance 1	Total	60	0	12	3	2,57	75%
	Mobiles	45	1	12	4	2,18	100%
Répondants relance 2	Total	74	0	8	2,78	2,08	78%
	Mobiles	58	1	8	3,55	1,67	100%

TAB. 18 – Evolution du nombre de déplacements selon le nombre de relances effectué

Source : EMD web - Lyon 2006

⁹¹Selon Adler (2003), on observe un plus fort taux de non-réponse partielle chez les individus qui répondent au questionnaire après plusieurs relances.

IV Caractéristiques des connexions web

Au-delà du taux de réponse global, d'autres facteurs peuvent être mesurés pour évaluer la qualité de l'enquête : la dernière page complétée, le nombre de pages vues, le nombre de cases cochées, le nombre et la durée des connexions (Dillman *et al.*, 1998b).

IV.1 Des moments de connexions privilégiés

Un des avantages du web est la liberté de réponse laissée à l'internaute, qui peut choisir le lieu et le moment le plus opportun pour répondre. L'expérience menée par Hojman *et al.* (2004) montre que la répartition des connexions n'est pas uniforme sur la journée et met en évidence deux pics de connexion : un le matin entre 9h00 et 10h00 et un le soir entre 19h00 et 20h00.

En ce qui concerne l'enquête ménages déplacements de Lyon, nous remarquons une grande disparité entre les jours de semaine et le week-end ⁹². La semaine, les internautes se connectent de 9h00 à 23h00, avec trois pics qui se distinguent dans la journée : le midi entre 11h00 et 13h00, l'après-midi entre 15h00 et 17h00 et le soir entre 19h00 et 22h00 (figure13). Il est probable qu'un nombre important de connexions soit réalisé sur le lieu de travail, d'autant que 20% des répondants déclarent ne pas posséder de connexion internet à leur domicile.

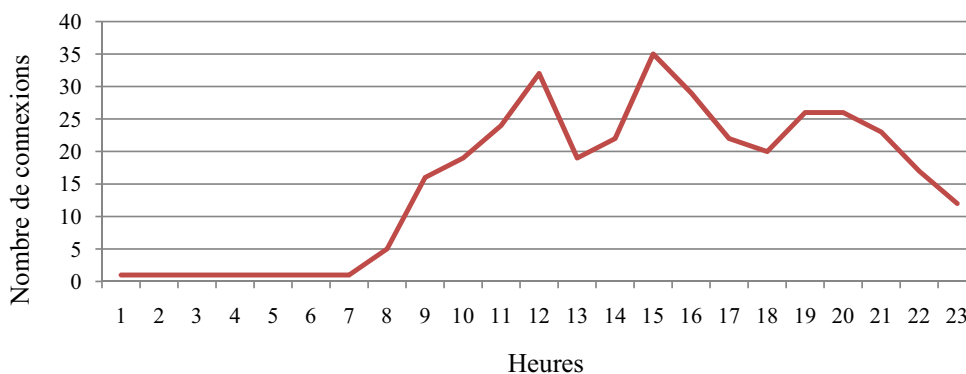


FIG. 13 – Heures de connexion des internautes la semaine
Source : EMD web - Lyon 2006

Le week-end, la répartition horaire des connexions est totalement différente (figure 14). Un seul gros pic entre 18h00 et 00h00 diffère avec le reste de la journée, où l'on constate un nombre de connexions faible mais constant (environ 5 à 10 par heure).

⁹²Bradley (1999) postule que le profil des enquêtés varie selon l'heure de remplissage du questionnaire, mais des analyses sur les enquêtes transports devraient être menées pour valider ou infirmer ce résultat.

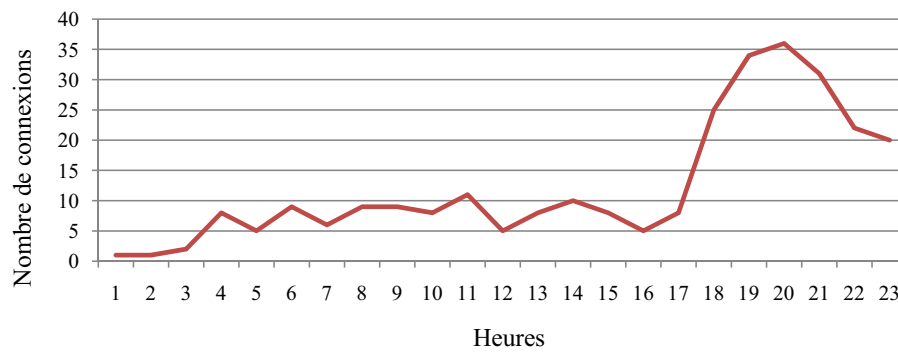


FIG. 14 – Heures de connexion des internautes le week-end
Source : EMD web - Lyon 2006

IV.2 Une durée de connexion étroitement liée au nombre de déplacements saisis

Le questionnaire web a été simplifié par rapport au questionnaire de l'enquête face-à-face, et plus ou moins basé sur celui des enquêtes C.A.T.I., réalisées par téléphone dans les villes moyennes. Cependant, le caractère peu ergonomique du site, la complexité des questions et la non familiarité supposée des individus avec le vocabulaire utilisé dans les enquêtes de mobilité laissaient penser que le temps de réponse moyen serait plus élevé sur le web. Le tableau 19 montre que le temps moyen de réponse au questionnaire web est proche de celui constaté lors des enquêtes par téléphone (23 mn, vs. 20 mn). Les disparités sont toutefois importantes (entre 5 mn et plus d'une heure et demi).

Statistique	Durée du contact (mn)
Minimum	4,9
Maximum	96,1
1er Quartile	14,1
Médiane	18,9
3ème Quartile	28,3
Moyenne	22,6
Ecart-type (n)	12,6
Effectif total	369

TAB. 19 – Durée de connexion globale du questionnaire web
Source : EMD web - Lyon 2006

Une analyse plus précise du comportement des internautes en termes de remplissage du questionnaire est possible, puisque des chronos insérés dans le questionnaire permettent de connaître le temps effectif passé sur les blocs de questions 'Ménages', 'Personnes' et 'Déplacements'. Les données du tableau 20 concernent les individus ayant terminé la saisie du questionnaire ou validé le bloc 'Déplacements'. Nous remarquons que la partie la plus consommatrice

Premiers résultats sur le comportement de réponse des internautes

de temps est celle qui concerne les déplacements, puisqu'en moyenne plus de 9 mn y sont consacrées. Ceci s'explique par le nombre et la précision des questions, qui demandent un effort de mémoire important (adresse des lieux visités, horaires de départ et d'arrivée ...). Ici encore, des disparités sont observées, puisque 25% des répondants prennent moins de 5 mn pour renseigner leurs déplacements, alors que 25% y consacrent environ 12 mn. Pour les autres parties du questionnaire, les temps de réponse sont plus courts (entre 2,5 et 4 mn en moyenne) et les différences moins marquées (écart-type d'environ 2 mn). Les blocs ménage et personne regroupent des questions plus 'habituelles' pour les interviewés, et proposent souvent des listes à cocher pour les réponses, ce qui simplifie considérablement le travail de l'internaute et minimise son temps de connexion.

Statistique	Bloc		Bloc	
	Ménage 1	Personne	Déplacements	Ménage 2
Minimum	0,2	0,2	0,6	0,8
Maximum	18,0	20,9	53,9	16,6
1er Quartile	2,4	1,9	4,4	1,5
Médiane	3,2	2,7	7,6	1,9
3ème Quartile	4,4	3,7	11,8	2,7
Moyenne	3,8	3,2	9,2	2,4
Ecart-type (n)	2,6	2,3	7,4	1,8
Effectif total	369	369	369	369

TAB. 20 – Durée de connexion du questionnaire web par bloc de réponses (en mn)

Source : EMD web - Lyon 2006

Le temps de saisie des réponses au questionnaire web est au final lié au nombre de déplacements déclarés, comme le montre la figure 15.

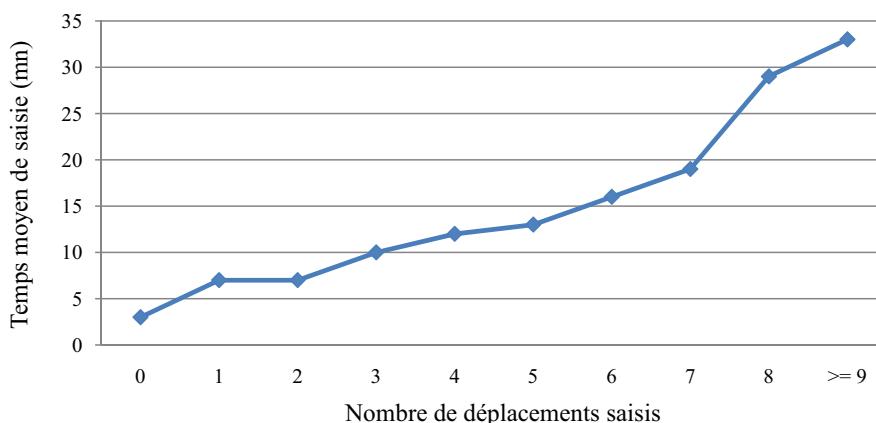


FIG. 15 – Temps moyen de saisie des réponses en ligne (mn)

Source : EMD web - Lyon 2006

IV.3 Peu de connexions multiples

Il est également intéressant de considérer l'impact du nombre de connexions sur la qualité des données saisies (tableau 21). Malgré la possibilité offerte aux ménages de remplir le questionnaire en plusieurs parties, on observe une connexion unique chez 80% des répondants. 21% des enquêtés (78 personnes) se sont connectés à plusieurs reprises avant de terminer le questionnaire, interrompus par des problèmes techniques ou un manque de disponibilité, dont près des trois quarts (73%) ne se sont connectés que deux fois seulement. Le nombre de répondants ayant effectué trois connexions ou plus est donc marginal. La sauvegarde intermédiaire des résultats, rendue possible grâce à une connexion sécurisée par un identifiant et un mot de passe rassure les répondants, mais est peu utilisée. Le temps moyen de connexion est stable, quel que soit le nombre de tentatives nécessaire pour arriver à la fin du questionnaire, et est en moyenne égal à 18 minutes.

Nombre de connexions	Nombre de ménages	Pourcentage des ménages	Durée moyenne par connexion (mn)
1	291	79%	19
2	57	15%	15
3	19	5%	15
4	1	0%	24
5	1	0%	15
Total	369	100%	18

TAB. 21 – Comportement de réponse selon le nombre de connexions
Source : EMD web - Lyon 2006

Il semble donc y avoir une durée maximum de connexion acceptable par les répondants, qui choisissent de se reconnecter à un autre moment si le questionnaire n'est pas terminé durant ce laps de temps. L'analyse des dates de connexion est également riche d'enseignement : 60% des connexions multiples ont lieu le même jour, 20% se font dans les deux jours qui suivent la connexion initiale et environ huit personnes sur 10 terminent de remplir le questionnaire en trois jours, quel que soit le nombre de connexions effectué. Ce constat rassure sur la qualité des données recueillies, puisque la mémoire individuelle concernant les déplacements est plus aléatoire au fur et à mesure que le temps passe. Il est en revanche délicat de cerner les raisons de ces connexions multiples, bien que les répondants qui terminent le questionnaire lors d'une connexion unique soient légèrement plus jeunes que les autres (42 ans, vs. 45 ans), ce qui peut traduire une certaine aisance avec l'outil internet. La littérature met en garde les chercheurs sur les différences d'équipement informatique des ménages, mais le tableau 22 montre que la proportion de ménages équipés d'une connexion internet à domicile, et parmi eux la part des connexions haut débit, est stable suivant que l'individu se soit connecté une seule ou plusieurs fois pour remplir le questionnaire.

Type de connexion	Unique	Multiple	Total
Connexion internet au domicile	82%	77%	80%
Dont connexion haut débit	93%	90%	92%

TAB. 22 – Equipement internet selon le nombre de connexions
Source : EMD web - Lyon 2006

V Conclusion

Les résultats de l'enquête web, bien que modestes, restent encourageants. Mais au-delà du taux de réponse global, il est nécessaire de s'interroger sur la composition du groupe d'internautes qui n'avaient pas accepté de répondre à la première enquête en face-à-face, ainsi qu'à la qualité des données recueillies. Certaines erreurs de la part des enquêteurs semblent avoir pesé sur le taux de non-réponse à l'enquête en face-à-face. Nombre d'individus préfèrent cependant répondre aux enquêtes en ligne. Il s'agit de personnes jeunes, ayant un haut niveau d'éducation et de revenus. Tous les internautes qui se sont connectés sur le site n'ont pas achevé la saisie des réponses. Des problèmes techniques peuvent être à l'origine de quelques abandons, mais nous avons identifié des questions clés, à partir desquelles des répondants quittent le questionnaire. Il s'agit essentiellement de questions personnelles, comme la saisie de l'adresse du domicile, ou de l'enregistrement des déplacements effectués durant la période de référence, étape fastidieuse.

Le web est un média très réactif. Les relances semblent efficaces et le temps entre la sollicitation du répondant et sa première connexion sur le site est relativement court. Bien que certains remplissent le questionnaire par partie, le temps moyen acceptable par connexion est 20 minutes. Par ailleurs, la présence d'échappatoires dans le questionnaire ne pèse pas trop sur la non-réponse partielle, et permet de ne pas frustrer les individus ne souhaitant pas se prononcer sur certaines questions. Enfin, un climat de confiance s'installe avec certains répondants, qui acceptent de livrer leur adresse e-mail personnelle à la fin du questionnaire.

Si le bilan de la conduite de l'enquête apparaît positif, il n'en reste pas moins que l'analyse des résultats doit faire face à quelques difficultés. D'une part, il convient d'analyser les caractéristiques de la population des internautes qui risquent de se différencier de l'ensemble de la population, compte tenu du taux de pénétration et d'aisance d'internet dans la population française. D'autre part, il faut s'assurer de la comparabilité des comportements de déplacements saisis quel que soit le média. Ces analyses sont conduites dans la partie suivante.

Conclusion de la partie I

Dans cette partie, nous avons mis en évidence les nombreux atouts du web en tant que mode de recueil de données (coût modéré, interactivité du questionnaire, administration rapide de l'enquête...). Cependant, si ce média est attractif et moins contraignant en termes de disponibilité temporelle, tous les individus n'ont pas accès à Internet et les compétences des internautes sont inégales. Actuellement, il est impossible de réaliser une enquête en ligne auprès d'un échantillon représentatif de l'ensemble de la population. La généralisation des réponses fournies par les méthodes d'inférence statistique reste donc problématique, même si dans le cas de l'enquête ménages déplacements de Lyon le biais d'échantillonnage est réduit, puisque les personnes qui se voient proposer un questionnaire web ont été auparavant recrutées par une méthode probabiliste.

Le questionnaire traditionnellement utilisé pour les interviews en face-à-face ne peut être diffusé en l'état sur internet, puisque la perception des questions et l'organisation du questionnaire dépendent du potentiel du média utilisé. Le questionnaire papier de l'enquête en face-à-face a donc été simplifié, afin de l'adapter à l'enquête web, de nature auto-administrée. Il était également essentiel de le rendre interactif et conforme aux attentes des internautes, habitués à participer à des enquêtes en ligne depuis plusieurs années.

Le taux de réponse global de l'enquête web est de 8.5%. Ce chiffre est faible, en comparaison du taux de réponse de l'enquête ménages standard menée en face-à-face (53%, et 50% sur le territoire du SCOT de Lyon, concerné par l'enquête web), mais reste encourageant, étant donné qu'une partie seulement des ménages ciblés peuvent se connecter à Internet, au domicile ou sur leur lieu de travail (50% des ménages français disposent d'une connexion Internet) et que l'option web a seulement été proposée aux ménages qui refusaient de répondre à l'enquête traditionnelle ou qui restaient injoignables. Par ailleurs, étant donné le caractère exploratoire de cette étude, nous nous sommes heurtés à quelques difficultés, notamment l'utilisation d'un logiciel peu approprié aux enquêtes de mobilité. Des améliorations dans le design et l'ergonomie du questionnaire nous laisseraient ainsi espérer un taux de réponse plus avantageux.

Si le web peut augmenter la qualité des enquêtes, il ne peut remplacer totalement les autres modes. La recherche doit montrer quels groupes sont plus à même de fournir des réponses de bonne qualité et par quel média (Kalfs et Van Evert, 2003). La collecte de données par différents médias est une solution intéressante, pour augmenter le taux de réponse, mais nécessite une

réflexion sur l'intégration des données, dans le but d'améliorer la représentativité de l'enquête. En effet, si la mise en place d'une enquête web en parallèle de l'enquête ménages déplacements permet d'appréhender des comportements de mobilité peu représentés jusqu'ici (individus actifs, aux horaires décalés...), la question de la comparabilité des données reste entière. Le choix du média, du mode d'administration du questionnaire et les caractéristiques des répondants ont probablement une influence non négligeable sur les comportements de mobilité. Il est donc nécessaire de comparer les caractéristiques et les comportements de mobilité des individus qui répondent en ligne avec ceux qui préfèrent recevoir un enquêteur à domicile.

Deuxième partie

Analyse comparative des échantillons web et face-à-face : quels enseignements ?

Introduction de la partie II

Face à la difficulté de recueillir des données de mobilité précises et représentatives de la population à un coût acceptable, les commanditaires d'étude se tournent vers des protocoles d'enquête plus complexes, associant plusieurs modes. La combinaison de différents médias apparaît comme un moyen d'améliorer la qualité des données produites à moindre coût, en permettant une augmentation du taux de réponse global. En effet, certains individus se reconnaissent davantage dans l'utilisation du web et sont par conséquent plus disposés à livrer des informations en ligne. Cependant, proposer plusieurs modes de recueil de données n'est pas sans risque sur la validité de l'enquête, des variations dans les questionnaires utilisés pouvant influencer les réponses (Cobanoglu *et al.*, 2001). Dans les protocoles d'enquête mixtes incluant le web, il est par exemple fréquent que les internautes diffèrent dans leurs réponses des autres, car les possibilités de communication et le contexte d'usage d'internet sont particuliers (Lozar Manfreda et Vehovar, 2002b). Enfin, les internautes ont probablement des caractéristiques particulières (position dans le cycle de vie, niveau d'éducation et de revenus, motorisation...) qui peuvent avoir une influence sur leur niveau de mobilité. La question de la comparabilité des données dans le temps et entre les différents modes reste entière.

L'objectif de cette partie est de mener une analyse comparative des comportements de mobilité des individus qui ont répondu à une enquête web, avec ceux soumis à un questionnaire en face-à-face. Nous cherchons d'abord à savoir si les internautes et les ménages auxquels ils appartiennent ont un profil spécifique en termes de caractéristiques socio-démographiques (chapitre 4). Puis, nous étudions la différence de mobilité déclarée des répondants web et face-à-face, selon des indicateurs tels que le nombre, le mode et le motif des déplacements, les budgets temps et distance quotidiens des répondants... (chapitre 5). Enfin, nous redressons l'échantillon face-à-face, de façon à le rendre comparable à l'échantillon web, du point de vue des principales caractéristiques socio-économiques des individus. Une fois réduites les différences entre les répondants, nous cherchons à savoir s'il existe encore des écarts sur les comportements de mobilité déclarés (chapitre 6). Ces résultats nous permettent de conclure sur la comparabilité des données des deux échantillons web et face-à-face.

Chapitre 4 : Analyse comparative des répondants web et face-à-face

"La volonté aboutit à un ajournement, l'utopie ; la science aboutit à un doute, l'hypothèse."
(Hugo, 1901).

Un de nos objectifs consiste à évaluer le potentiel du web pour les enquêtes de mobilité. Il est donc intéressant de voir si certains biais de représentativité de l'enquête en face-à-face peuvent être 'corrigés' à l'aide d'une enquête web. Nous ne connaissons pas a priori la cible des répondants à l'enquête web, puisqu'il s'agit de ménages qu'il n'a pas été possible d'enquêter en face-à-face. Après quelques rappels théoriques sur les tests statistiques (section I), nous allons comparer les données socio-économiques des répondants à l'enquête standard à celles de la population de référence⁹³, afin d'évaluer la représentativité de l'enquête face-à-face (section II).

Puis, nous allons comparer les résultats obtenus sur le web avec ceux donnés par l'enquête en face-à-face. Il s'agit plus précisément de caractériser le profil des deux échantillons, face-à-face et web, à l'aide des données socio-économiques, avant de pouvoir comparer leurs comportements de déplacement. D'un modèle marqué par le partage des biens au niveau du groupe familial, nous passons progressivement à un modèle marqué par la possession individuelle (voiture, téléphone portable...). Pour rendre la présentation plus lisible, nous raisonnons d'abord sur les données du ménage (section III), puis sur les données concernant les personnes (section IV).

L'enquête web a été menée auprès des non-répondants à l'enquête face-à-face résidant sur le périmètre du SCOT de Lyon. Nous restreindrons donc nos analyses à ce territoire, qui accueille 64% de la population de l'aire urbaine de Lyon⁹⁴, mais ne représente que 14% de sa superficie⁹⁵. Par ailleurs, si

⁹³Les données de la population de référence sont fournies par l'INSEE (Recensement Général de la Population de 1999.).

⁹⁴Aire urbaine au sens de l'INSEE : 40% des actifs au moins vont travailler dans un pôle urbain, lui-même défini par son volume minimum d'emplois.

⁹⁵Ce périmètre, situé au coeur de l'aire métropolitaine Lyonnaise est particulièrement intéressant pour les enquêtes de mobilité, puisqu'une partie correspond au périmètre des transports urbains du Sytral, desservi par le réseau des Transports en Commun Lyonnais (TCL).

l'INSEE recense les caractéristiques sociodémographiques de tous les résidents, l'enquête ménages standard ne s'intéresse qu'aux individus de 5 ans et plus et l'enquête web ne contient que des répondants âgés d'au moins 10 ans (suivant la consigne figurant au début du questionnaire). Dans ce chapitre, nous avons restreint la comparaison des données sociodémographiques à tous les individus de 10 ans et plus, afin de tenir compte des différences entre les trois bases de données.

I Rappels théoriques sur les tests statistiques

Les tests statistiques donnent une règle permettant de décider si l'on peut rejeter une hypothèse, en fonction d'observations relevées sur des échantillons. Ils répondent tous à une procédure constante, composée de deux éléments : un tableau de prise de décision, qui inclut divers risques d'erreur, et une séquence d'étapes dont les statistiques utilisées varient selon la situation. La démarche consiste à poser une hypothèse, puis à conduire une expérience et enfin à analyser la compatibilité de cette hypothèse avec les observations issues de l'expérience. Deux hypothèses sont généralement formulées :

- L'hypothèse nulle : hypothèse dont on cherche à savoir si elle peut être rejetée, souvent définie comme une absence de différence entre les deux populations et notée H_0 ;
- L'hypothèse alternative : hypothèse concurrente, notée H_1 .

Nous pouvons décider que H_0 est soit acceptée soit rejetée, et en réalité (mais nous ne le savons pas) elle est vraie ou fausse. Quatre situations sont donc possibles (tableau 23). Pour prendre une décision, on définit une zone de rejet de l'hypothèse nulle, construite sur une base de vraisemblance en probabilité. Lorsque le résultat du test appartient à la région de rejet, on rejette H_0 et on conclut que le test est significatif au risque α . Si le résultat du test n'appartient pas à la région de rejet, on ne peut pas rejeter H_0 et on conclut que le test est non significatif.

Etat du monde	Décision	
	H_0 acceptable	H_0 rejetée
Si H_0 est vraie	Décision correcte Seuil de confiance = $1 - \alpha$	Erreur de première espèce Seuil du test = α
Si H_0 est fausse	Erreur de deuxième espèce Probabilité = β	Décision correcte Puissance du test = $1 - \beta$

TAB. 23 – Résultats possibles d'un test d'hypothèses
Source : Wonnacott et Wonnacott (2000)

Lorsque le test est significatif, il est d'usage de quantifier le degré de signification du test, ou valeur du risque de 1ère espèce. Il s'agit de la plus petite taille du test qui aurait permis, avec ces données, de rejeter le test. Le risque de première espèce α est la probabilité a posteriori que l'on a de rejeter l'hypothèse nulle quand elle est vraie. Il est généralement fixé à 5%, ce qui signifie

que l'on accepte de faire une erreur dans un cas sur vingt. Lorsque le test n'est pas significatif, on se retient d'affirmer qu'il n'existe pas de différence entre les deux populations et on conclut que les différences observées, s'il y en a, ne sont pas statistiquement significatives. Il faut tenir compte du risque de deuxième espèce β (souvent inconnu), qui représente la probabilité de ne pas rejeter l'hypothèse nulle alors qu'elle est fautive. La notion de puissance du test est définie à partir du risque de seconde espèce $(1-\beta)$. Elle indique la probabilité d'admettre à raison l'existence de la relation.

En fonction du type de variables concernées, plusieurs tests statistiques sont disponibles. On présentera successivement le fonctionnement du test du Khi-deux (I.1), du test de comparaison de deux proportions (I.2) et du test de comparaison de deux moyennes (I.3), qui sont utilisés dans les chapitres suivants.

I.1 Test du Khi-deux

Ce test permet de comparer une distribution observée à une distribution théorique ⁹⁶. Grâce à la statistique du χ^2 de Pearson, il est possible de tester l'indépendance entre les lignes et les colonnes du tableau croisé des distributions observées et théoriques, en mesurant à quel point le tableau est éloigné (au sens du χ^2) de ce que l'on pourrait obtenir en moyenne, en conservant les mêmes sommes marginales. Soit un échantillon de taille n , dont les valeurs sont réparties en k classes distinctes. La statistique est donnée par :

$$\chi^2 = \sum_{i=1}^k \frac{(O_i - E_i)^2}{E_i} \quad (5)$$

avec $O_i (i = 1, \dots, k)$ les effectifs observés et $E_i (i = 1, \dots, k)$ les effectifs théoriques ou espérés.

On montre que cette statistique suit une loi du χ^2 à $(r-1)(c-1)$ degrés de liberté, avec r le nombre de lignes et c le nombre de colonnes du tableau de contingence. Ce résultat étant asymptotique, il faut vérifier avant d'utiliser ce test que :

- la taille de l'échantillon n est supérieure ou égale à 20 ;
- aucune somme marginale n'est inférieure à 5 ;
- au moins 80% des effectifs calculés sont supérieurs à 5.

L'hypothèse à tester est :

H_0 : Les lignes et les colonnes du tableau sont indépendantes.

L'hypothèse alternative est :

H_1 : Il existe un lien entre les lignes et les colonnes du tableau.

Si le résultat du test appartient à la région de rejet, on rejette l'hypothèse nulle d'indépendance (généralement au risque $\alpha = 5\%$). On conclut à une

⁹⁶Les variables considérées doivent prendre deux modalités ou plus, et être catégorielles.

liaison entre les lignes et les colonnes du tableau, c'est-à-dire à l'existence d'une différence statistiquement significative entre la distribution théorique et la distribution observée, ou plus généralement entre les deux distributions. Dans le cas contraire, on affirme que le test n'est pas significatif.

I.2 Test de comparaison de deux proportions

Ce test permet de comparer deux proportions issues de deux échantillons différents. Soit n_1 le nombre d'observations vérifiant une certaine propriété pour un échantillon E_1 de taille N_1 , et n_2 le nombre d'observations vérifiant la même propriété pour un échantillon E_2 de taille N_2 . Nous avons :

- $p_1 = \frac{n_1}{N_1}$, la proportion de l'échantillon E_1 vérifiant la propriété ;
- $p_2 = \frac{n_2}{N_2}$ la proportion de l'échantillon E_2 vérifiant la propriété.

Soit D la différence (exacte, minimale ou maximale) supposée entre les deux proportions. Classiquement, D est fixée à 0. Le test bilatéral correspond au test de la différence entre $(p_1 - p_2)$ et D .

L'hypothèse à tester est :

$$H_0 : p_1 - p_2 = D$$

L'hypothèse alternative est :

$$H_1 : p_1 - p_2 \neq D$$

Si les deux échantillons sont indépendants, la différence $p_1 - p_2$ suit une loi normale dont la moyenne est :

$$\mu = (p_1 - p_2) \tag{6}$$

et la variance :

$$s^2 = \frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2} \tag{7}$$

Le paramètre :

$$Z = \frac{p_1 - p_2}{\sqrt{\left(\frac{p_1(1-p_1)}{n_1}\right) + \left(\frac{p_2(1-p_2)}{n_2}\right)}} \tag{8}$$

suit approximativement une loi normale centrée-réduite, lorsque les conditions suivantes sont vérifiées :

- n_1 et n_2 sont grands (supérieurs à 30) ;
- $\min(n_1 ; n_2) * p > 5$ et $\min(n_1 ; n_2) * (1 - p) > 5$, avec p la proportion commune aux deux populations sous H_0 , si elle est connue ;
- $\min(n_1 ; n_2) * p' > 20$ et $\min(n_1 ; n_2) * (1 - p') > 20$, si p n'est pas connue.

p' est un estimateur de la proportion commune. Il se calcule à l'aide de la

formule :

$$p' = \frac{n_1 * p_1 + n_2 * p_2}{n_1 + n_2} \quad (9)$$

Si le résultat du test appartient à la région de rejet, on rejette l'hypothèse nulle d'absence de différence entre les deux proportions. Dans le cas contraire, on affirme que le test n'est pas significatif. La zone de rejet est $|z| > 1,96$, pour un risque $\alpha = 5\%$.

I.3 Test de comparaison de deux moyennes

Ce test permet de comparer deux moyennes issues de deux échantillons différents. Soit :

- un échantillon E_1 , comprenant n_1 observations, de moyenne \bar{x}_1 et de variance s_1^2 ;
- un échantillon E_2 indépendant de E_1 , comprenant n_2 observations, de moyenne \bar{x}_2 et de variance s_2^2 .

D est la différence supposée entre les moyennes (D vaut 0 lorsque l'on suppose l'égalité). Le test bilatéral correspond au test de la différence entre $(\bar{x}_1 - \bar{x}_2)$ et D.

L'hypothèse à tester est :

$$H_0 : \bar{x}_1 - \bar{x}_2 = D$$

L'hypothèse alternative est :

$$H_1 : \bar{x}_1 - \bar{x}_2 \neq D$$

Si les deux échantillons sont indépendants, la différence $\bar{x}_1 - \bar{x}_2$ suit une loi normale dont la moyenne est :

$$\mu = (\bar{x}_1 - \bar{x}_2) \quad (10)$$

et la variance :

$$s^2 = \sigma'^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right) \quad (11)$$

Avec :

$$\sigma'^2 = \frac{n_1 * s_1^2 + n_2 * s_2^2}{n_1 - 1 + n_2 - 1} \quad (12)$$

Le paramètre :

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sigma' \sqrt{\left(\frac{1}{n_1}\right) + \left(\frac{1}{n_2}\right)}} \quad (13)$$

suit approximativement une loi de Student à $n_1 + n_2 - 2$ ddl, lorsque les conditions suivantes sont vérifiées :

II Qui sont les non-répondants à l'enquête en face-à-face ?

- normalité : dans les deux populations la distribution du caractère x est normale. Ceci se vérifie si les tailles des échantillons sont assez grandes (30 observations) ;
- homoscélasticité : x_1 et x_2 ont la même variance, autrement dit $s_1^2 = s_2^2$

Si le résultat du test appartient à la région de rejet, on rejette l'hypothèse nulle d'absence de différence entre les deux moyennes. Dans le cas contraire, on affirme que le test n'est pas significatif. La zone de rejet est $|z| > 1,96$, pour un risque $\alpha = 5\%$.

II Qui sont les non-répondants à l'enquête en face-à-face ?

En comparant les caractéristiques socio-économiques des répondants avec ceux de la population de référence, il est possible de déduire les caractéristiques des non-répondants (Armoogum et Madre, 1997). L'objectif de cette partie est d'obtenir des informations sur les non-répondants à l'enquête standard, en raisonnant par différence avec les données de l'INSEE. Il est nécessaire de préciser que cette comparaison ne tient pas compte d'un décalage temporel, puisque les données utilisées par le CETE de Lyon pour la stratification géographique de l'échantillon face-à-face de l'enquête ménages déplacements proviennent du recensement général de la population de 1999, et ont été redressées selon l'évolution constatée entre les fichiers FILOCOM ⁹⁷ de 1999 et de 2003.

Pour l'analyse des caractères sociodémographiques, nous ne tenons compte que des données INSEE de 1999, puisque nous n'avons pu acquérir les données du fichier FILOCOM 2003 à ce jour. Dans le recensement de 1999, le SCOT de Lyon regroupe 524 032 ménages et 1 249 216 personnes. L'échantillon de répondants à l'enquête en face-à-face restreint au SCOT de Lyon compte 6 428 ménages et 15 251 personnes. Après redressement, nous comptabilisons 546 764 ménages et 1 243 026 personnes.

II.1 Des ménages de taille modeste, mais davantage motorisés

La comparaison des données de l'enquête ménages déplacements standard avec celles de l'INSEE permet de conclure que la répartition du lieu de résidence des ménages sur le territoire du SCOT de Lyon est inchangée. Ce résultat est conforme à nos attentes, puisque les données de l'enquête ménages déplacements sont redressées selon une stratification géographique. En re-

⁹⁷Le fichier FILOCOM (Fichier des LOGements par COMmunes) est un fichier construit par la Direction Générale des Impôts, pour les besoins du Ministère de l'Ecologie, de l'Energie, du Développement Durable et de l'Aménagement du Territoire. Il est constitué par le rapprochement du fichier de la taxe d'habitation, du fichier foncier, du fichier des propriétaires et du fichier de l'impôt sur les revenus des personnes physiques, et ne concerne que la France métropolitaine (à l'exclusion des DOM).

vanche, la comparaison du nombre moyen de personnes dans le ménage laisse apparaître des différences, la taille moyenne des ménages de l'enquête en face-à-face étant légèrement inférieure à celle constatée par l'INSEE (2,27 vs. 2,38 personnes), comme le montre le tableau 24. Il est en revanche impossible de tester la fiabilité statistique de cette observation, puisque les écarts-types des distributions de la taille moyenne des ménages ne sont pas disponibles pour les données de l'INSEE.

	INSEE	Enquête standard
Nombre de ménages	524 032	546 763
Nombre de personnes	1 249 216	1 243 026
Taille du ménage	2,38	2,27

TAB. 24 – Taille moyenne des ménages
Source : EMD Lyon 2006 et INSEE - RGP 1999

Les deux distributions (EMD06⁹⁸ et INSEE) des ménages selon le nombre de voitures font également apparaître des différences statistiquement significatives selon le test du Khi-deux (p -value < 0,01%). Les ménages de l'enquête ménages déplacements interrogés en face-à-face semblent davantage motorisés, puisque 30,8% possèdent 2 voitures ou plus, contre seulement 25,1% des ménages de l'INSEE (p -value < 0,01%)(figure 16). Inversement, 25% des ménages de l'INSEE sont dépourvus de voiture particulière, vs. 22% de ceux de l'enquête ménages déplacements (p -value < 0,01%). Compte-tenu du décalage temporel entre les deux enquêtes, cette différence peut s'expliquer par une augmentation de la motorisation ces dernières années. Cependant, l'importance de l'écart laisse supposer que les données de l'enquête ménages déplacements en face-à-face surestiment en partie la motorisation.

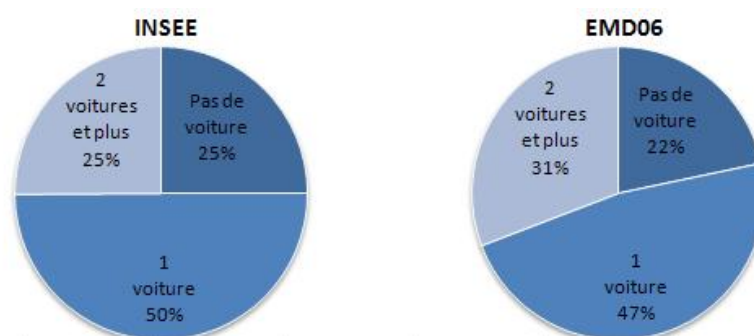


FIG. 16 – Répartition des ménages par enquête et par nombre de voitures
Source : EMD Lyon 2006 et INSEE - RGP 1999

⁹⁸Enquête Ménages Déplacements réalisée en face-à-face à Lyon en 2006.

II.2 Les hommes, les jeunes et les non diplômés : des profils peu concernés par l'enquête en face-à-face

La comparaison des données individuelles de l'enquête ménages déplacements en face-à-face avec celles de l'INSEE donne des résultats intéressants⁹⁹ (annexe IX).

Les femmes semblent légèrement sur-représentées dans l'enquête en face-à-face (52,5%, vs. 52,8%; p-value < 0,01%), car dans les faits elles acceptent plus volontiers de répondre aux enquêtes que les hommes. L'ensemble des membres du foyer étant interrogés dans les deux enquêtes, nous pouvons supposer qu'il existe un biais de sélection dans l'enquête ménages standard (sous-représentation des ménages d'une seule personne ou monoparentaux dont le représentant principal est un homme). Nous notons également une sous-représentation des jeunes adultes (20-29 ans) dans l'enquête standard (14,8%, vs. 18,9%; p-value < 0,01%). Il est probable que les personnes jeunes aient davantage de contraintes en termes de disponibilité que les plus âgées (retraités), et qu'il leur soit plus difficile de recevoir un enquêteur en soirée à leur domicile. Concernant l'occupation des répondants, les retraités sont sur-représentés dans l'enquête en face-à-face (24,1%, vs. 18,1%; p-value < 0,01%), principalement au détriment des chômeurs et autres inactifs. Si on s'intéresse à la catégorie socioprofessionnelle des répondants, ce sont les cadres et les professions intellectuelles supérieures qui ont davantage participé à l'enquête ménages en face-à-face (22,4%, vs. 13,9%; p-value < 0,01%), contrairement aux personnes sans emploi. Ici encore, le décalage temporel peut en partie expliquer une différence de répartition des professions et catégories socioprofessionnelles, mais cette différence nous semble trop forte pour être complètement expliquée par cette seule variable. Enfin, c'est pour le niveau d'études que les différences sont les plus marquées. Dans l'enquête en face-à-face, les individus sont sur-diplômés par rapport à la population de référence, puisque 33,5% ont fait des études supérieures (vs. 19,9% dans le recensement; p-value < 0,01%). D'une manière générale les personnes ayant un plus haut niveau d'étude perçoivent probablement davantage l'intérêt d'une étude de mobilité de grande ampleur. Les entretiens de l'enquête ménages déplacements se sont déroulés sur fond de 'violences urbaines' en décembre 2005, et les enquêteurs ont eu certaines difficultés à interroger les ménages des quartiers difficiles. Par ailleurs, le recensement général de la population a un caractère obligatoire, contrairement à l'enquête ménages déplacements. Certaines catégories de la population peuvent donc refuser d'y répondre, ou déclarer ne pas être disponibles.

II.3 Les enseignements

L'échantillon face-à-face est construit pour être représentatif pour chaque zone de tirage, puisqu'une stratification géographique a été menée. Cependant, la population des répondants corrigée de la stratification géographique

⁹⁹Seules les différences pour lesquelles le test bilatéral de moyennes ou des proportions est statistiquement significatif (p-value inférieure à 0,05%) sont interprétées

n'est pas complètement représentative de l'ensemble de la population, selon les données de l'INSEE. Les ménages de l'enquête face-à-face sont de taille plus modeste, et ont davantage de voitures à leur disposition. A titre individuel, les femmes, les retraités, les cadres et professions intellectuelles supérieures et les personnes ayant fait des études supérieures sont sur-représentés dans l'enquête ménages face-à-face. A contrario, les hommes, les jeunes (20-29 ans), et les personnes n'ayant pas fait d'étude supérieures semblent sous-représentées par rapport à l'enquête réalisée par l'INSEE.

Les résultats soulignent l'intérêt qu'il y aurait à redresser les données de l'enquête ménages déplacements sur plusieurs variables socio-économiques, dans la mesure où l'on sait que ces variables sont corrélées à des pratiques de mobilité distinctes. Ce travail se heurte toutefois à la disponibilité des données de l'INSEE. Le dernier recensement date de 1999, et les données FILOCOM permettent d'actualiser le nombre de personnes par zone de résidence, mais pas les distributions selon les différentes variables socio-économiques ¹⁰⁰.

Nous pouvons conclure des résultats de cette section que l'échantillon des non-répondants à l'enquête face-à-face n'est pas représentatif de l'ensemble de la population étudiée. L'enquête web s'adresse aux non-répondants de l'enquête en face-à-face. Dans les sections III et IV, nous cherchons à savoir si l'enquête web permet de corriger certains biais de représentativité de l'enquête face-à-face.

III Les ménages internautes : quelles spécificités ?

L'objectif de cette partie est d'identifier les différences entre les caractéristiques socio-économiques des ménages des deux échantillons web et face-à-face. Nous nous intéressons notamment à la taille, la localisation et le niveau de revenu des ménages, ainsi qu'à leur équipement en moyens de communication et de motorisation.

Dans l'enquête web, une seule personne du ménage répondait à l'ensemble du questionnaire. Nous avons donc le même nombre d'observations dans les bases 'ménages' et 'personnes'. En revanche, dans l'enquête en face-à-face restreinte au SCOT de Lyon, nous comptons 6 428 ménages et 13 271 personnes de plus de 10 ans répondant au questionnaire. A ce stade de l'analyse, le redressement semble particulièrement délicat. Il nous paraît donc plus pertinent de travailler sur des échantillons non redressés.

III.1 Des différences pas toujours significatives selon le lieu de résidence du ménage

Les deux distributions, web et face-à-face (figures 17 et 18), des ménages selon le lieu de résidence font apparaître des différences. Nous remarquons que

¹⁰⁰Dès que les données du recensement en continu seront disponibles, cette contrainte pourra être résolue.

III Les ménages internautes : quelles spécificités ?

dans l'enquête standard, la proportion de ménages résidant dans le centre de l'agglomération (zones 1 et 2 : Lyon - Villeurbanne) est plus importante que dans l'enquête web (41,3%, vs. 38,6%). Etant donné que le test bilatéral nous donne un risque supérieur au niveau de signification acceptable (31,1%) de rejeter à tort l'égalité de ces proportions, nous pouvons conclure que la différence observée entre les deux échantillons n'est pas statistiquement significative.

En revanche, nous observons des écarts significatifs pour quelques secteurs de la grande périphérie. La proportion de ménages résidant dans la zone 6 (2ème couronne ouest) est beaucoup plus importante dans l'enquête web (12,8% vs. 6,8%; p-value = 0,01%), alors que les ménages de la zone 7 (3ème couronne est) sont sous-représentés parmi la population d'internautes (8,4%, vs. 10%). Cette dernière différence n'est toutefois pas significative (p-value = 32,1%). De manière générale, les répondants web sont davantage situés à l'ouest de l'agglomération. Nous pouvons supposer que ce choix résidentiel est lié à leur niveau de revenu, et qu'il aura probablement un impact sur la mobilité des ménages. Par ailleurs, cette zone enregistre un taux de rebut important (21%, vs. 16% sur l'ensemble du périmètre) (Cf. chapitre 4 section 1.3). Il semble donc que l'enquête web permette de toucher un certain nombre de ménages impossibles à joindre après plusieurs tentatives lors de l'enquête en face-à-face.

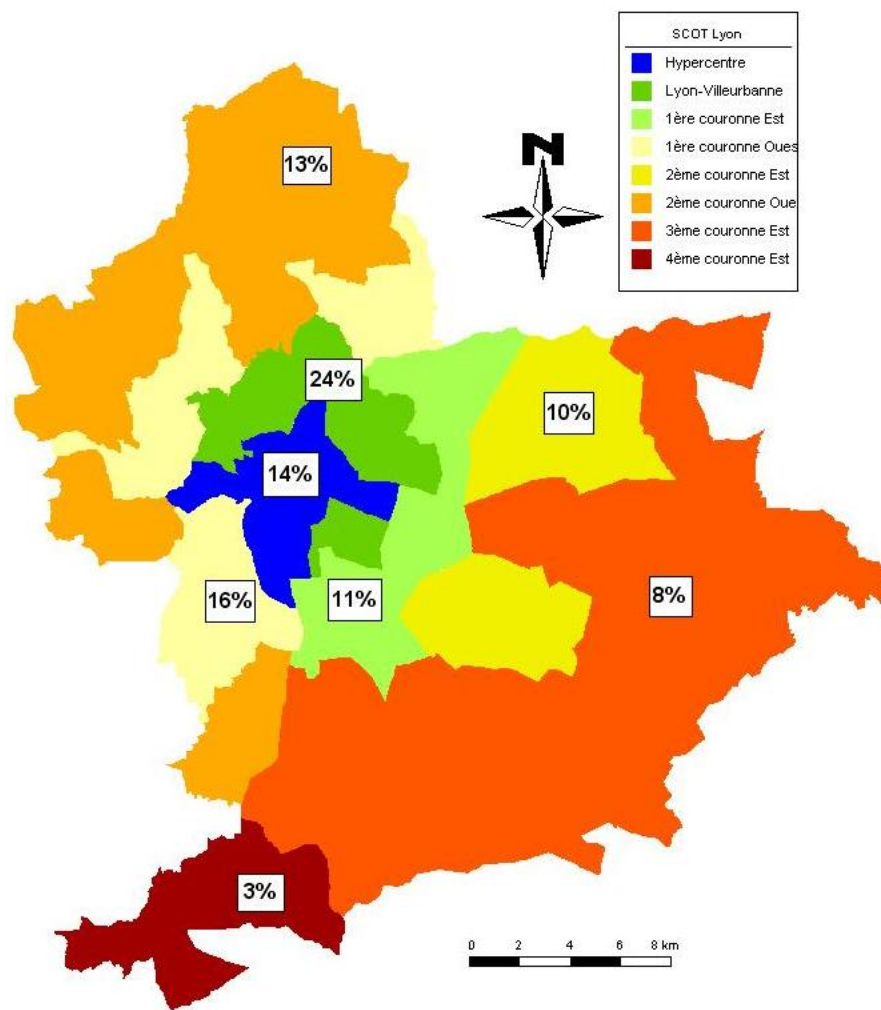


FIG. 17 – Répartition des ménages de l'enquête web par zone de résidence
Source : EMD web - Lyon 2006

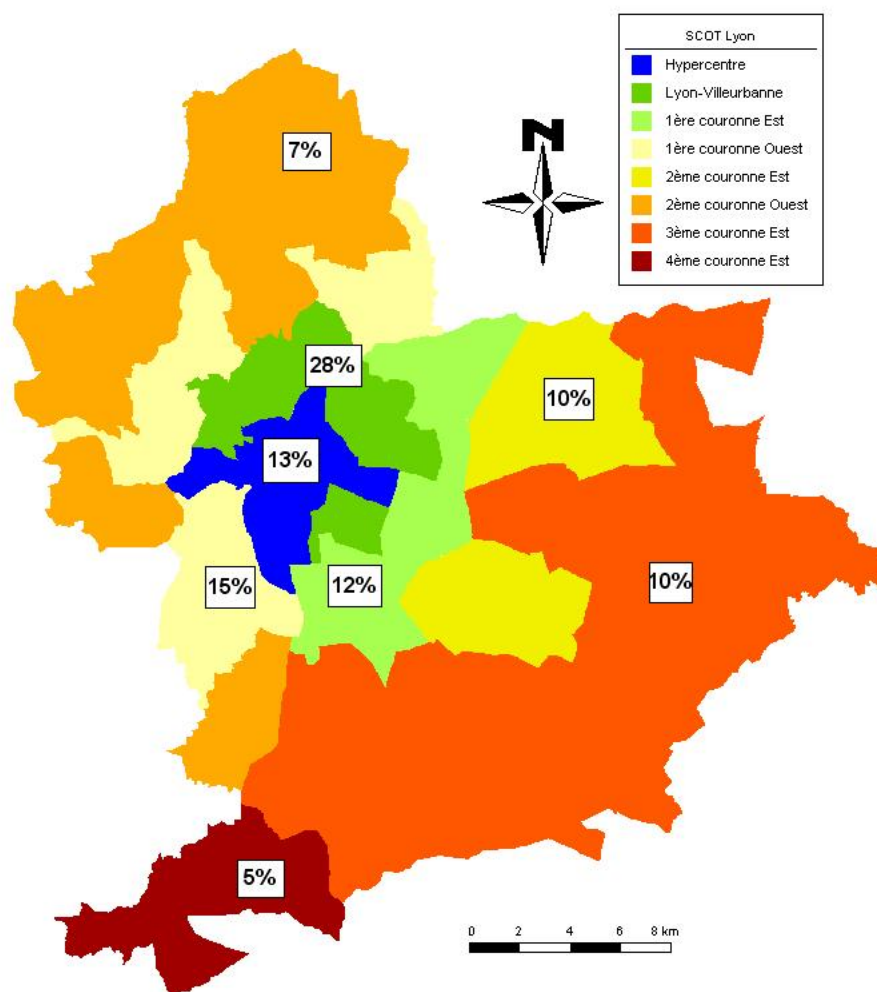


FIG. 18 – Répartition des ménages de l'enquête face-à-face par zone de résidence

Source : EMD face-à-face - Lyon 2006

III.2 Des ménages internautes de plus grande taille

La taille des ménages ayant répondu à l'enquête web semble sensiblement plus élevée que celle de la population de l'enquête face-à-face (2,60 personnes/ménage vs. 2,38 personnes/ménage ; p-value = 0,1%) (tableau 25).

	Nb personnes	Minimum	Maximum	Moyenne	Ecart-type
Enquête web	369	1	6	2,6	1,313
Enquête standard	6428	1	10	2,38	1,364

TAB. 25 – Statistiques descriptives de la variable 'Nombre de personnes'

Source : EMD web et face-à-face - Lyon 2006

Cette différence peut être liée au statut des répondants (sur-représentation des actifs chez les internautes et à la composition des ménages (par exemple

la présence d'enfants), les petits ménages étant particulièrement représentés dans le centre et les grands en périphérie (Sytral, 2007). Il semble intéressant d'affiner l'analyse en comparant la distribution de la taille des ménages entre les deux enquêtes (figure 19).

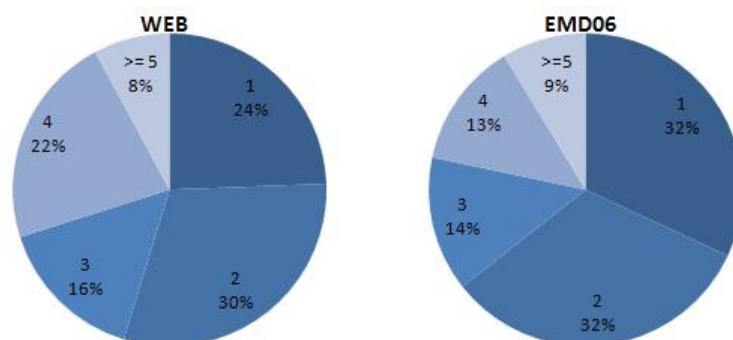


FIG. 19 – Répartition des ménages par nombre de personnes et par enquête
Source : EMD web et face-à-face - Lyon 2006

Nous remarquons que la proportion de ménages composés d'une seule personne est beaucoup plus faible dans l'enquête web (24,4%, vs. 31,9%; p-value < 0,01%). Inversement, la part des ménages composés de 4 personnes est plus importante dans l'enquête web (22%, vs. 13%; p-value < 0,01%). La composition des ménages ayant répondu sur le web diffère donc de celle des ménages répondant à l'enquête standard. Elle contient davantage de familles et moins de personnes isolées, constat qui sera à mettre en relation avec l'âge des répondants. Certaines études montrent que les ménages avec de jeunes enfants ont davantage de contraintes temporelles, limitant leur utilisation des nouvelles technologies telles que le web (Wang et Law, 2007). Dans l'enquête de mobilité menée à Lyon, les contraintes de l'enquête en face-à-face semblent toutefois plus importantes (accueil d'un enquêteur à domicile) que celles de l'enquête web pour ce type de ménages.

III.3 Des ménages internautes très motorisés

Les répondants à l'enquête web semblent davantage motorisés que ceux ayant répondu en face-à-face, puisque la moitié d'entre-eux possèdent au moins deux voitures et que moins de 10% n'en ont aucune (figure 20).

Il existe cependant une corrélation probable entre le nombre et l'âge des personnes du ménage et le nombre de voitures possédées par celui-ci. Nous avons recréé une variable 'nombre de personnes de 18 ans et plus', afin de calculer un nombre moyen de voitures par personne en âge de conduire (figure 21).

III Les ménages internautes : quelles spécificités ?

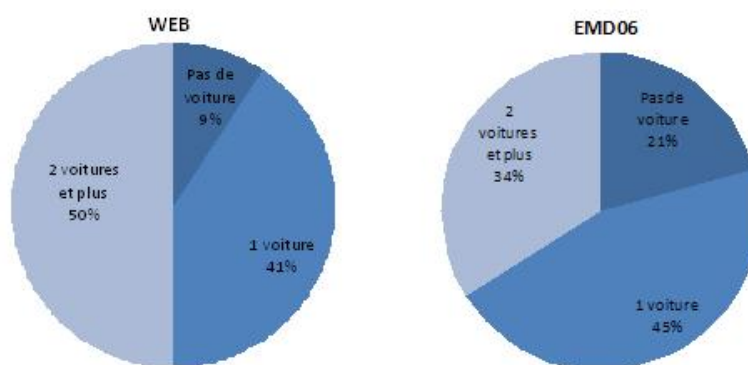


FIG. 20 – Répartition des ménages par nombre de voitures et par enquête
Source : EMD web et face-à-face - Lyon 2006

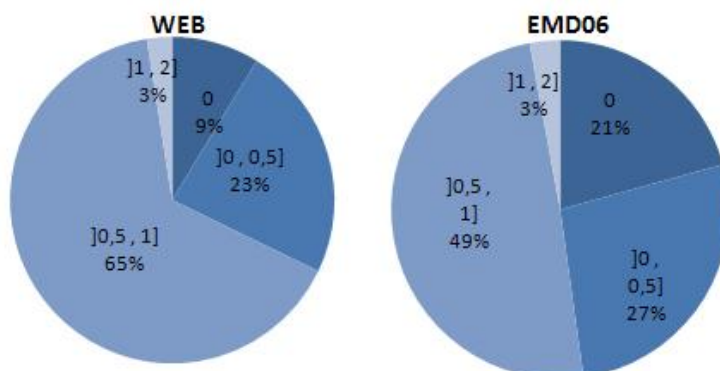


FIG. 21 – Répartition des ménages par nombre de voitures par personne de 18 ans et plus
Source : EMD web et face-à-face - Lyon 2006

Les deux distributions, web et face-à-face, des ménages selon le nombre de voitures possédées par les individus de 18 ans et plus font apparaître des différences intéressantes. La proportion de ménages pour lesquels le nombre de voitures à disposition des personnes de 18 ans et plus est supérieur à 1 est très faible et comparable entre les deux enquêtes. Nous pouvons distinguer deux cas très différents parmi les ménages possédant au moins une voiture particulière. La proportion de ménages possédant au mieux 0,5 voiture par personne en âge de conduire reste proche entre les deux enquêtes (23,5% pour le web, vs. 27,1% dans l'enquête standard), et cette différence n'est pas significative (p -value = 13,5%). En revanche, les proportions de ménages possédant entre 0,5 et 1 voiture par personne de 18 ans et plus sont très différentes entre les deux échantillons (65,3% pour le web, vs. 49,3% dans l'enquête standard ; p -value < 0,01%). L'accès à la voiture est donc plus important au sein de la population web que pour celle de l'enquête face-à-face. Ceci est en partie lié aux caractéristiques socio-économiques des répondants, les internautes ayant un niveau de revenu plus élevé (Cf. section suivante).

Le taux de motorisation est faible pour les habitants de Lyon-Villeurbanne et la moitié est de l'agglomération, mais élevé pour la couronne ouest (Sytral, 2007), où résident de nombreux répondants web. Ces derniers sont également plus jeunes, et appartiennent à une génération pour laquelle le passage du permis et la pratique de la conduite se font très tôt.

Avec la croissance du revenu par tête, les modes de transport les plus rapides tels que la voiture deviennent plus attractifs (Orfeuill, 2002b). L'accès à la voiture particulière s'est donc nettement démocratisé ces dernières années, et les catégories les plus modestes peuvent accéder un peu plus facilement à la voiture particulière, du fait de l'importance du marché de l'occasion et de la diminution des coûts d'usage (Orfeuill, 2002a). Un phénomène de rattrapage s'opère, et essentiellement auprès des populations les moins motorisées et les moins mobiles (Orfeuill et Massot, 2003). Mais à Lyon (Nicolas *et al.*, 2002) comme au niveau national (Hivert, 2000), l'accès à la voiture particulière dépend du revenu. La progression du parc automobile est tiré majoritairement par le multi-équipement des ménages (Orfeuill et Massot, 2003). Or, la dynamique de la bi-motorisation (achat d'une deuxième voiture particulière) est d'abord portée par les ménages les plus aisés.

III.4 Des revenus en moyenne plus élevés pour la population web

D'abord, on constate que la question relative aux revenus du ménage suscite un taux de non-réponse élevé. La part des ménages refusant de donner leur niveau de revenu annuel est beaucoup moins importante sur le web (1 ménage sur 4, vs. 1 ménage sur 3 ; p-value < 0,01%). Les individus sont réticents à communiquer des informations personnelles, surtout lorsqu'il s'agit de leurs revenus. Il semble plus difficile de déclarer son niveau de ressources à un enquêteur en face-à-face que sur le web, où le questionnaire est par définition auto-administré (Potoglou et Karanoglou, 2008). Toutefois, il se peut que les ménages réticents à fournir des données personnelles n'aient pas répondu à l'enquête web. Ainsi, ceux qui font la démarche de remplir le questionnaire en ligne se soumettent à l'ensemble des questions et n'hésitent pas à délivrer des réponses précises et liées à leur niveau de revenu.

Les revenus déclarés sont plus élevés dans le cas de l'enquête web, puisqu'un ménage sur 5 déclare gagner plus de 40 000 euros par an, vs. seulement un sur douze dans le cas de l'enquête en face-à-face (p-value < 0,01%). A l'opposé, les bas revenus sont sous-représentés dans l'enquête web, puisque 34,1% des ménages répondant en face-à-face déclarent percevoir moins de 20 000 euros par an, vs. seulement 21% des ménages répondant sur le web (p-value < 0,01%) (figure 22). Ces différences sont probablement corrélées au statut, au niveau de diplôme et à la profession et catégorie professionnelle des répondants, qui varient selon le mode d'enquête.

Elles peuvent également s'expliquer par la taille des ménages. Nous avons donc recréé une variable qui tient compte du nombre de personnes du ménage, et calculé un revenu moyen par unité de consommation (UC), selon les règles

III Les ménages internautes : quelles spécificités ?

établies par l'INSEE :

- le chef de ménage compte pour une unité de consommation ;
- les autres personnes du ménage de 14 ans et plus comptent pour 0,5 unités de consommation ;
- les autres personnes du ménage de moins de 14 ans comptent pour 0,3 unités de consommation.

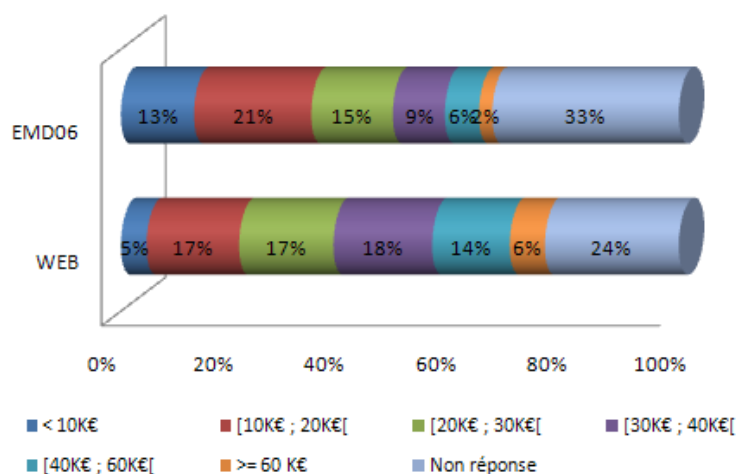


FIG. 22 – Répartition des ménages par tranche de revenus annuels nets
Source : EMD web et face-à-face - Lyon 2006

Par ailleurs, nous avons affecté à chaque ménage un revenu moyen, qui correspond au centre de la classe de revenus déclarés. Ainsi, pour la classe inférieure (moins de 10 000 euros annuels nets), nous avons retenu une valeur moyenne de 5 000 euros annuels nets. Pour la classe supérieure (plus de 60 000 euros annuels nets), nous avons pris une valeur moyenne de 70 000 euros annuels nets. Nous remarquons que les ménages ayant répondu à l'enquête web ont davantage de revenus annuels nets disponibles par unité de consommation que ceux ayant répondu en face-à-face (20,000 euros vs. 15 000 euros en moyenne ; p-value < 0,01%) (tableau 26).

Ménages	Obs.	Mini	Maxi	Moyenne	Ecart-type	1er quartile	Médiane	3ème quartile
Web	279	1 786	70 000	20 107	11 019	11 905	17 500	25 000
Standard	4 282	1 111	70 000	15 051	9 651	7 500	15 000	19 444

TAB. 26 – Statistiques descriptives de la variable 'Revenus moyens par UC'
Source : EMD web et face-à-face - Lyon 2006

Cette différence peut s'expliquer par le fait que la pénétration d'internet est encore inégale sur le territoire, et concerne davantage les ménages ayant des hauts revenus. Selon Visser et Lanzerdorf (2004), le revenu du ménage est un paramètre qui influence positivement et fortement l'utilisation des nouvelles technologies de l'information et de la communication (coût de l'investissement dans du matériel informatique, de l'abonnement internet...). Par ailleurs, ceux

qui disposent d'une connexion sur le lieu de travail ont des professions généralement supérieures, avec un bon niveau de revenu.

III.5 Des ménages web très équipés en moyens de communication

Il est intéressant de comparer les équipements de télécommunication des ménages de l'enquête en face-à-face avec ceux de l'enquête web, ces derniers étant a priori mieux équipés et plus à l'aise avec les nouvelles technologies.

III.5.1 Un équipement en téléphonie fixe au domicile comparable

La proportion de ménages possédant une ligne de téléphone fixe à domicile est proche entre les deux enquêtes (90% sur le web et 89% en face-à-face), et cette différence n'est pas statistiquement significative (p -value = 61,0%). Selon Sharp *et al.* (2004), les individus qui n'ont pas de téléphone fixe sont jeunes, urbains et souvent locataires de leur logement. Le téléphone fixe est ancien, et son taux de pénétration dans la population est élevé. Il n'y a donc pas de lien entre un tel équipement et le choix de répondre sur le web.

En revanche, les ménages ayant répondu sur le web sont moins souvent inscrits sur l'annuaire France Télécom, en comparaison de ceux de l'enquête standard (58,5%, vs. 78%; p -value < 0,01%) (figure 23). Il est probable que ces derniers possèdent des packages de téléphonie haut débit, avec des numéros spéciaux, ou soient davantage inscrits sur liste orange ou rouge pour éviter certains appels (démarcheurs, enquêteurs...) ou soient plus difficiles à joindre lors de l'enquête en face-à-face, donc sur-représentés parmi les non-répondants à l'enquête en face-à-face.

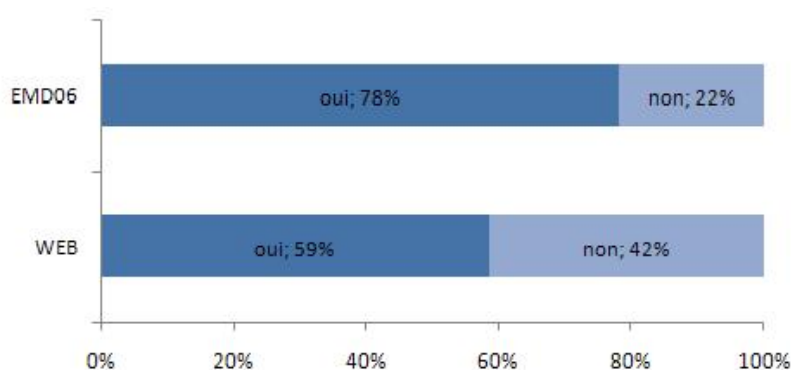


FIG. 23 – Répartition des ménages selon leur inscription sur une liste téléphonique

Source : EMD web et face-à-face - Lyon 2006

III.5.2 Des ménages web fortement connectés à internet

Près de 8 ménages sur 10 ayant répondu sur le web sont équipés d'une connexion internet au domicile, contre 1 sur 2 dans le cas de l'enquête standard

III Les ménages internautes : quelles spécificités ?

(80% ont une connexion internet au domicile, vs. 51% ; p-value < 0,01%)(figure 24). Les ménages ayant répondu sur le web sont donc mieux équipés que ceux de l'enquête standard. Cette différence est évidemment directement liée au média utilisé pour l'enquête web. Il est intéressant de s'interroger sur les 20% de ménages qui déclarent ne pas posséder de connexion internet à domicile, mais qui ont rempli le questionnaire en ligne. Deux hypothèses sont possibles : soit les individus n'ont pas souhaité répondre à la question (ce qui est peu probable, car un échappatoire sous la forme 'Ne souhaite pas répondre' était prévu), soit ils se sont connectés sur leur lieu de travail pour répondre au questionnaire.

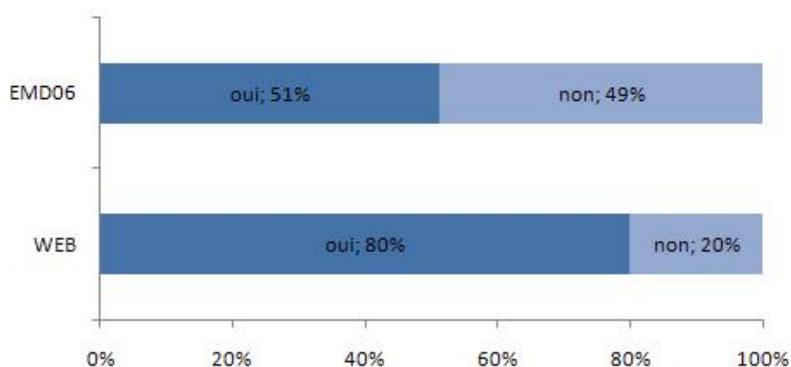


FIG. 24 – Répartition des ménages selon la possession d'une connexion internet
Source : EMD web et face-à-face - Lyon 2006

III.5.3 Des ménages web fortement équipés en téléphonie mobile

Nous nous intéressons ici à des données individuelles, puisque chaque personne était interrogée sur la possession éventuelle d'un téléphone portable.

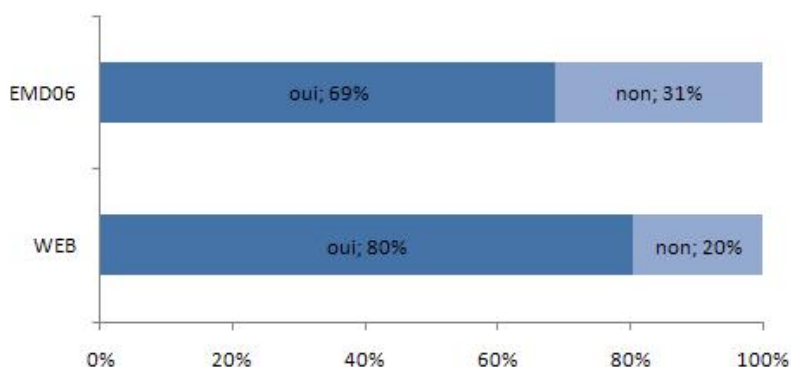


FIG. 25 – Répartition des ménages selon leur possession d'un téléphone portable
Source : EMD web et face-à-face - Lyon 2006

La proportion de répondants possédant un téléphone portable, pour motif

personnel ou professionnel est plus importante dans l'enquête web (80%, vs. 69% en face-à-face ; p-value < 0,01%) (figure 25).

Les internautes semblent mieux équipés en moyens de communication. Ceci peut être lié à la catégorie socioprofessionnelle, puisque certaines catégories possèdent davantage de téléphones pour motif professionnel, et au niveau de revenu (budget nécessaire à la mise en place d'un contrat de téléphonie mobile). Enfin, certaines études mettent en évidence une corrélation positive entre la possession d'une connexion web au domicile et celle d'un téléphone portable (Bigot et Croutte, 2007).

IV Les répondants web : quel profil ?

Dans cette section, nous comparons les caractéristiques sociodémographiques des répondants web avec ceux des enquêtés en face-à-face (genre, âge, statut...). Les données analysées concerne l'individu interrogé et non plus son ménage.

IV.1 Une population plus féminine ?

La proportion de femmes observée dans l'enquête web (56,8%) est supérieure à celle obtenue dans l'enquête standard (52,3%), mais cette différence n'est pas statistiquement significative (p-value = 10%) (figure 26).

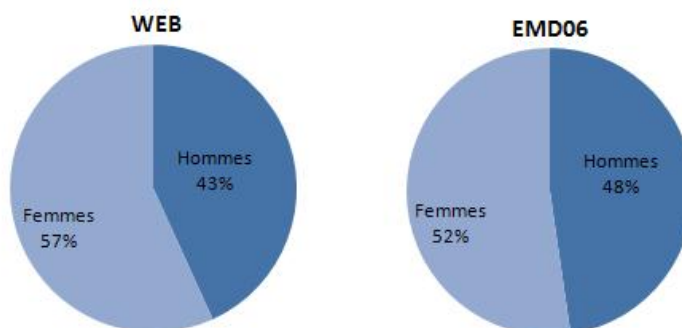


FIG. 26 – Répartition des personnes par sexe et par enquête
Source : EMD web et face-à-face - Lyon 2006

Comme le montrent des études précédentes (Visser et Lanzerdorf, 2004), le genre n'est pas très discriminant de l'utilisation d'internet, même si les hommes sont davantage adeptes des nouvelles technologies. L'effectif web un peu plus féminin rejoint cependant les résultats observés couramment dans les enquêtes, lorsqu'il n'y a pas de sélection aléatoire des répondants au sein du ménage.

IV.2 Une population dominée par les personnes en âge de travailler

Les deux distributions des personnes enquêtées selon la tranche d'âge font apparaître des différences statistiquement significatives. 70% des répondants à l'enquête web sont des adultes en âge de travailler (de 30 à 59 ans), contre 45% seulement des répondants à l'enquête face-à-face (p-value < 0,01%) (figure 27). Les jeunes ¹⁰¹ et les plus âgés semblent sous-représentés dans l'enquête web. On compte en effet 12,5% de personnes de 60 ans et plus dans l'enquête web, contre 24,1% dans l'enquête en face-à-face (p-value < 0,01%).

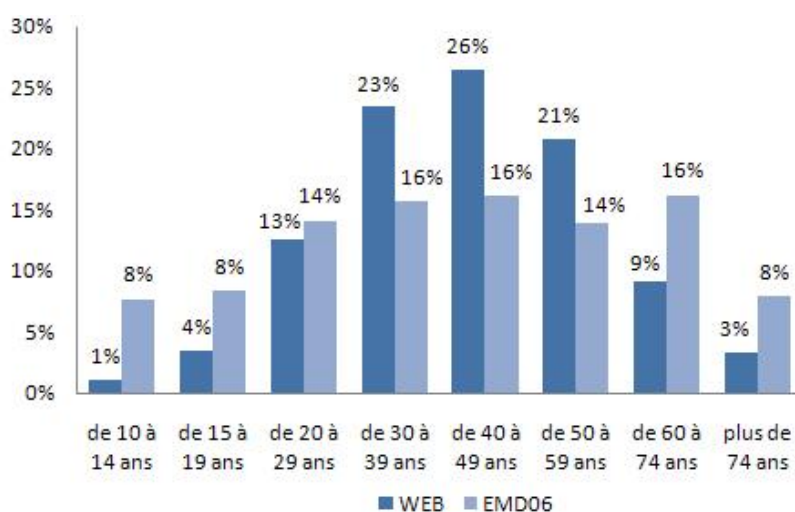


FIG. 27 – Répartition des personnes par tranche d'âge et par enquête
Source : EMD web et face-à-face - Lyon 2006

Les personnes âgées sont généralement moins familières avec internet. Elles se sentent plus à l'aise avec les modes de recueil traditionnels et préfèrent répondre en face-à-face. Il semble toutefois que les retraités répondent de plus en plus sur le web, curieux face à ce nouveau média (développement de cours d'informatique pour les seniors), ou avec l'aide de leurs petits enfants (Harvey, 2003).

Les faibles effectifs recensés dans les classes d'âge extrêmes (moins de 19 ans et plus de 74 ans) sont probablement un biais important en termes de comportements de mobilité pour la population ayant répondu à l'enquête web. Ce constat peut s'expliquer en partie par le fait que la lettre d'accompagnement précisait que seule une personne de 11 ans et plus (souvent le chef de ménage ou sont conjoint, destinataires du courrier) pouvait répondre à l'enquête en ligne. Les ménages dits 'actifs' sont davantage familiers avec les technologies de la communication, et disposent souvent d'un accès internet haut-débit sur leur lieu de travail. En revanche, ils disposent de peu de temps pour répondre aux enquêteurs à domicile.

¹⁰¹Selon Wang et Law (2007), les jeunes utilisent davantage les nouvelles technologies que les anciennes générations (mieux équipés, volonté d'explorer...).

IV.3 Des internautes majoritairement actifs

Les résultats montrent que 3/4 des répondants web sont des actifs (si nous comptabilisons les personnes à la recherche d'un emploi) contre seulement 51% des répondants en face-à-face (p-value < 0,01%) (figure 28). En revanche, les étudiants et scolaires sont fortement sous-représentés (7% pour le web, vs. 20% dans l'enquête standard ; p-value < 0,01%) et la situation est identique pour les retraités (13% pour le web, vs. 24% dans l'enquête standard ; p-value < 0,01%). Ces constats peuvent être liés à l'âge, puisque seule une personne (probablement le chef de ménage ou son conjoint) répond à l'enquête web, ce qui limite le nombre de jeunes, et que les personnes âgées sont peu familières avec internet.

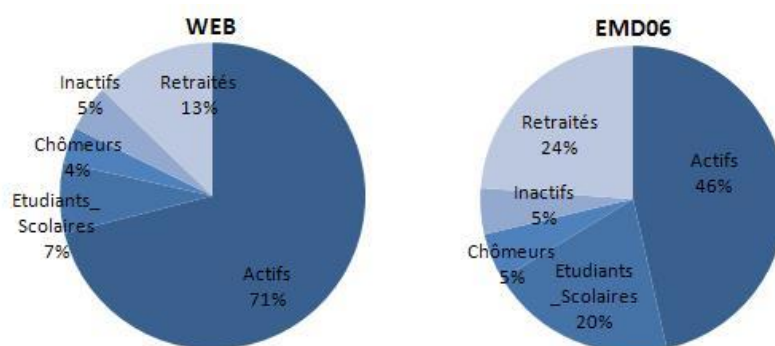


FIG. 28 – Répartition des personnes par occupation et par enquête
Source : EMD web et face-à-face - Lyon 2006

IV.4 Une sur-représentation des professions dites supérieures et des employés

Les actifs (travailleurs à temps plein ou à temps partiel) ayant répondu sur le web se répartissent de manière très inégale entre les différentes professions et catégories socioprofessionnelles (PCS). Deux profils sont sur-représentés : les cadres ou professions intellectuelles supérieures et les employés (environ 39% chacun), qui ne totalisent respectivement que 14,2% et 22,5% des effectifs dans l'enquête en face-à-face (p-value < 0,01%) (figure 29). Les autres catégories sont sous-représentées, notamment les personnes sans activité professionnelle, mais les faibles effectifs ne nous permettent pas de tirer des conclusions pertinentes. Ces résultats sont conformes avec ceux de (Yun et Trumbo, 2000), qui mettent en évidence des répondants web davantage employés et moins de professions libérales. Ils peuvent être liés au niveau de diplôme et au type d'emploi occupé. Les cadres et les employés ont essentiellement des fonctions administratives, qui nécessitent une certaine connaissance d'internet et un accès web au travail, ce qui favorise certainement une connexion web à domicile (Bates, 2001).

IV Les répondants web : quel profil ?

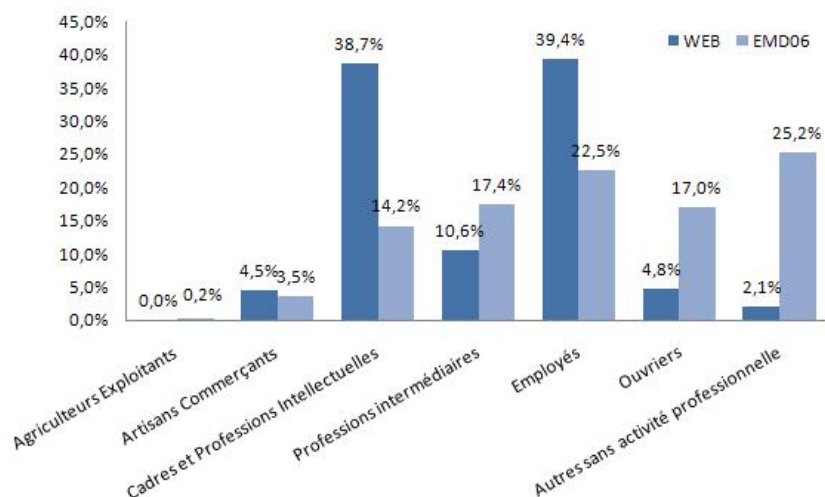


FIG. 29 – Répartition des personnes par PCS et par enquête
Source : EMD web et face-à-face - Lyon 2006

IV.5 Un niveau d'étude élevé pour les internautes

Les répondants à l'enquête web ont un haut niveau de formation, puisque près de 2/3 ont fait des études supérieures (au-delà du BAC) contre seulement 31% des répondants à l'enquête face-à-face (p -value < 0,01%) (figure 30).

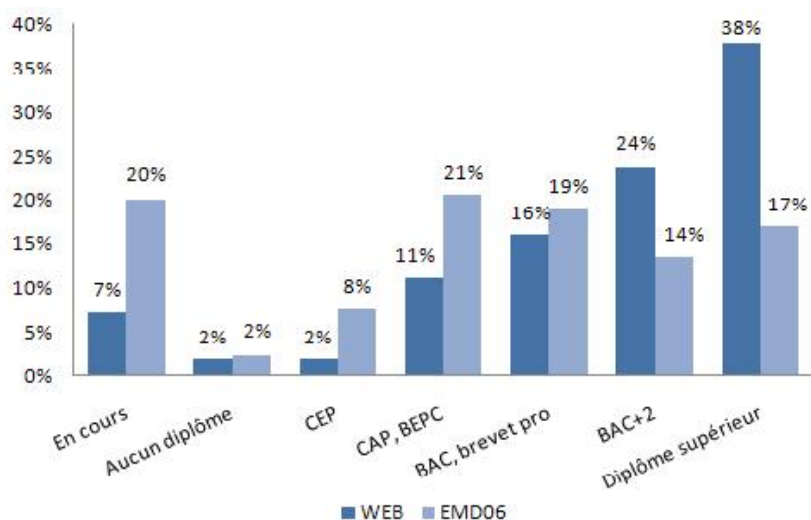


FIG. 30 – Répartition des personnes par niveau d'étude et par enquête
Source : EMD web et face-à-face - Lyon 2006

Inversement, peu d'individus sont en cours de scolarité (7% vs. 20%) (p -value < 0,01%). Le niveau d'étude est lié à la catégorie socioprofessionnelle, car davantage d'études sont requises pour les cadres et professions intellectuelles supérieures, sur-représentés dans l'enquête web. Les hauts diplômés utilisent des ordinateurs et internet plus souvent dans leur travail (Bates, 2001). Par

ailleurs, sur le web, une seule personne du ménage remplit le questionnaire. Le niveau d'études renseigné est donc probablement celui du chef de ménage ou de son conjoint, ce qui explique que la part des personnes en cours d'études reste marginale.

IV.6 Une population fortement détentrice du permis de conduire

Les personnes possédant leur permis de conduire (catégorie B) ou pratiquant la conduite accompagnée sont sur-représentées dans l'enquête web (93%, vs. 81%) (p-value < 0,01%) (figure 31). Ceci peut être lié à l'âge des répondants, puisque les personnes jeunes ne disposent pas toujours du permis de conduire (contrainte financière), et que les personnes appartenant aux classes d'âge les plus élevées n'ont pas toutes passé leur permis (surtout les femmes) (Pochet, 2003). Autour de 30 ans, 90% des hommes et 85% des femmes ont aujourd'hui leur permis, contre 88% et 64% pour l'ensemble de la population adulte (Orfeuil et Massot, 290).

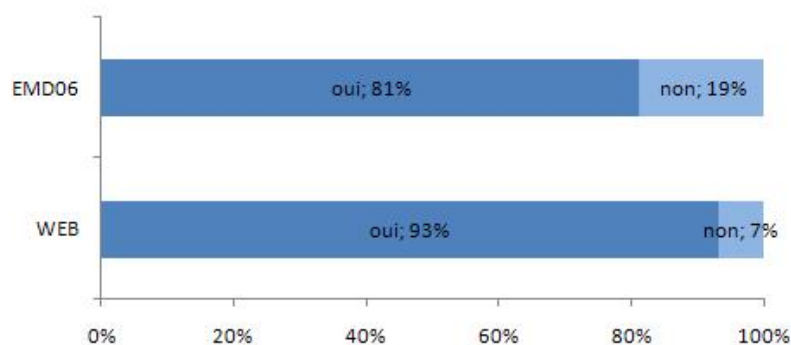


FIG. 31 – Répartition des personnes selon la possession du permis de conduire et par enquête

Source : EMD web et face-à-face - Lyon 2006

Ces résultats confirment ceux de Srinivasan et Athuru (2004), qui montrent que les individus ayant le permis de conduire sont davantage utilisateurs d'internet. Il semble donc exister un lien entre la possession d'une connexion web et le permis de conduire.

IV.7 Un lieu de travail habituel ou d'études plus fréquemment central

Concernant la localisation du lieu de travail, nous remarquons une importante différence entre les deux enquêtes, puisque l'hyper-centre (presqu'île, Part-Dieu) semble sur-représenté dans l'enquête web (26,6% vs. 16,4% ; p-value < 0,01%) (tableau 27). Cet écart s'explique probablement par le pourcentage de cadres et professions intellectuelles supérieures et d'employés plus important

dans l'enquête web. Ces emplois se situent davantage dans les sièges sociaux des entreprises et sont par nature plus présents au centre de l'agglomération. D'un autre côté, les personnes dont le lieu de travail habituel ou d'études se situe en troisième ou quatrième couronnes (zones 7, 8 et 9) apparaissent sous-représentées dans l'enquête web (5,8% vs. 15,3%; p-value < 0,01%). Attention toutefois aux faibles effectifs obtenus dans l'enquête web, qui ne permettent pas de conclure.

Lieu de travail habituel ou d'études	Enquête web		Enquête face-à-face	
	Effectif	%	Effectif	%
1 - Hypercentre	77	26,6%	1 508	16,4%
2 - Lyon Villeurbanne	88	30,3%	2 637	28,8%
3 - 1 ^{re} couronne est	33	11,4%	1 137	12,4%
4 - 1 ^{re} couronne ouest	27	9,3%	939	10,2%
5 - 2 ^{me} couronne est	24	8,3%	790	8,6%
6 - 2 ^{me} couronne ouest	16	5,5%	515	5,6%
7 - 3 ^{me} couronne est	7	2,4%	733	8 %
8 - 3 ^{me} couronne ouest	3	1 %	174	1,9%
9 - 4 ^{me} couronne est	7	2,4%	494	5,4%
10 - Externe	8	2,8%	244	2,7%
Total	290	100%	9 171	100%

TAB. 27 – Répartition des personnes par lieu de travail habituel ou d'études et par enquête

Source : EMD web et face-à-face - Lyon 2006

V Conclusion

L'analyse des caractéristiques de la population ayant répondu à l'enquête en face-à-face après redressement sur le nombre de personnes par zone de tirage fournit des résultats proches des données du recensement de 1999. Elle met toutefois en évidence certaines différences, statistiquement significatives. Les ménages de l'enquête face-à-face sont de plus petite taille et sont davantage motorisés. En ce qui concerne les individus, les femmes, les retraités, les cadres et professions intellectuelles supérieures et les personnes ayant fait des études supérieures sont sur-représentés dans l'enquête ménages face-à-face, alors que les jeunes (20-29 ans) semblent sous-représentés par rapport à l'enquête réalisée par l'INSEE. L'enquête face-à-face a été réalisée en 2006, et certaines de ces différences peuvent être en partie expliquées par le décalage temporel. Il n'en reste pas moins que compte-tenu de la stratification géographique utilisée pour l'échantillon de l'enquête en face-à-face, le sous-échantillon des non-répondants ne peut être considéré comme représentatif de l'ensemble de la population de la zone d'étude. Il est donc très peu probable que celui de l'enquête web le soit.

Cela nous a conduit à comparer les deux populations web et face-à-face non redressées. Cette comparaison nous permet de dresser un premier profil des ménages ayant répondu en ligne. Si les différences en termes de localisation résidentielle sont limitées, d'autres caractéristiques socio-économiques au ni-

veau du ménage ressortent clairement. Les ménages ayant répondu sur le web ont davantage de membres, ce qui semble cohérent avec la sous-représentation des ménages de taille modeste dans les non-répondants à l'enquête en face-à-face. On note également une sur-représentation des ménages résidant dans la 2ème couronne ouest de l'agglomération, zone où le taux de rejets à l'enquête face-à-face était particulièrement élevé. En ce qui concerne les critères socio-démographiques des individus, les répondants web sont plus souvent actifs et dans la tranche d'âge 30-50 ans, ce qui conduit à une sous-représentation des deux extrêmes de la pyramide des âges. Si les retraités sont sous-représentés dans l'échantillon des non-répondants à l'enquête en face-à-face, ce n'est pas le cas des plus jeunes, que l'enquête web ne permet apparemment pas de 'récupérer'. Ceci peut s'expliquer en partie par les consignes associées au questionnaire web, à savoir qu'un seul membre du ménage répond à l'enquête (souvent le chef de ménage ou son conjoint). Les internautes ont souvent des emplois plus qualifiés, avec un niveau d'études plus élevé et des revenus annuels moyens nets par unité de consommation supérieurs (Madre *et al.*, 2007). L'enquête web ne permet donc pas de toucher les personnes ayant un faible niveau d'éducation et qui sont sur-représentées parmi les non-répondants à l'enquête en face-à-face. Enfin, les répondants web possèdent plus souvent le permis de conduire et sont davantage motorisés, tant au niveau du nombre de véhicules par ménage qu'au niveau du nombre de véhicules moyens par personne en âge de conduire (concernant le permis de conduire, la différence s'explique essentiellement par un effet d'âge). Ils travaillent ou étudient plus souvent dans le centre de l'agglomération, et sont généralement mieux équipés en moyen de communication (connexion internet, téléphone portable). L'utilisation des nouvelles technologies est en effet plus importante chez les individus jeunes ou possédants de hauts revenus (Wang et Law, 2007). Ils semblent néanmoins plus difficiles à joindre par téléphone, puisque moins de 60% des ménages déclarent être inscrits sur la liste des abonnés France Télécom. Si certains biais de représentativité de l'enquête en face-à-face peuvent être corrigés à l'aide d'une enquête web (taille du ménage, proportion de retraités), d'autres différences se creusent (haut niveau d'éducation, bon niveau de motorisation, peu de jeunes...).

Les internautes ont donc des caractéristiques particulières. Cette population est peut-être moins disponible pour répondre à un long questionnaire en face-à-face, même si l'analyse des données de l'enquête en face-à-face comparativement aux données du recensement ne met pas en évidence une sous-représentation de ces actifs. Probablement plus sensibles aux enjeux de l'étude et familiers avec internet, ils ont accepté de se connecter pour répondre au questionnaire en ligne. Ces différences ont probablement des impacts sur la mobilité.

Chapitre 5 : Analyse comparative de la mobilité

"Par la statistique, je pourrais prouver l'existence de Dieu."
Georges Gallup (1901-1984).

Après avoir qualifié les deux échantillons (web et face-à-face) sur la base de caractéristiques socio-économiques, nous allons nous intéresser à la mobilité. L'objectif de cette partie est de mettre en évidence les différences en termes de comportements de mobilité entre l'échantillon web et face à face. Il est en effet fortement probable que les différences socio-économiques constatées entre les deux populations impactent les habitudes de déplacement des répondants. La structure des ménages (personnes seules, bi-actifs ou non, avec ou sans enfant...) constitue un paramètre essentiel pour la compréhension des déplacements quotidiens des individus (Baudelle *et al.*, 2003). Il en est de même de leur niveau de revenu et de motorisation. Par ailleurs, les nouvelles technologies telles que le téléphone portable et internet ont le plus large impact sur les scénarios de mobilité individuelle (Mokhtarian *et al.*, 2006).

La démarche est clairement analytique. L'objet concerne l'étude des pratiques de mobilité des répondants selon le média d'enquête (web ou face-à-face). Le champ est celui de la mobilité quotidienne, à l'intérieur du périmètre du SCOT de Lyon, couvert par l'enquête web. Par ailleurs, les répondants à l'enquête web devant être au moins âgés de 10 ans, nous avons appliqué cette même contrainte à l'enquête en face-à-face.

Dans une première partie, nous focalisons notre analyse sur les habitudes d'utilisation des modes de transport en semaine, avant d'étudier plus en détail les variables caractérisant la mobilité. Nous nous intéressons non seulement aux déplacements (section II), mais également aux sorties du domicile, c'est-à-dire à l'enchaînement des déplacements et activités entre un départ et un retour au domicile (section III).

I Habitudes d'utilisation des modes de transport en semaine

Les données que nous comparons dans cette section sont issues d'une partie du questionnaire recueillant les habitudes de déplacement en termes déclaratifs. Elles ne se réfèrent pas aux déplacements déclarés.

I.1 Une utilisation très fréquente de la voiture en tant que conducteur

Nous observons des différences importantes entre les deux enquêtes. Près de 7 personnes sur 10 interrogées sur le web utilisent chaque jour la voiture en tant que conducteur pour se déplacer en semaine, vs. 44,2% dans l'enquête en face-à-face (p-value < 0,01%). A contrario, 4,5% seulement des internautes n'utilisent jamais la voiture particulière pour se déplacer en semaine, vs. 32% des répondants à l'enquête standard (p-value < 0,01%) (figure 32). Les effectifs de l'enquête web sont cependant parfois faibles pour pouvoir conclure.

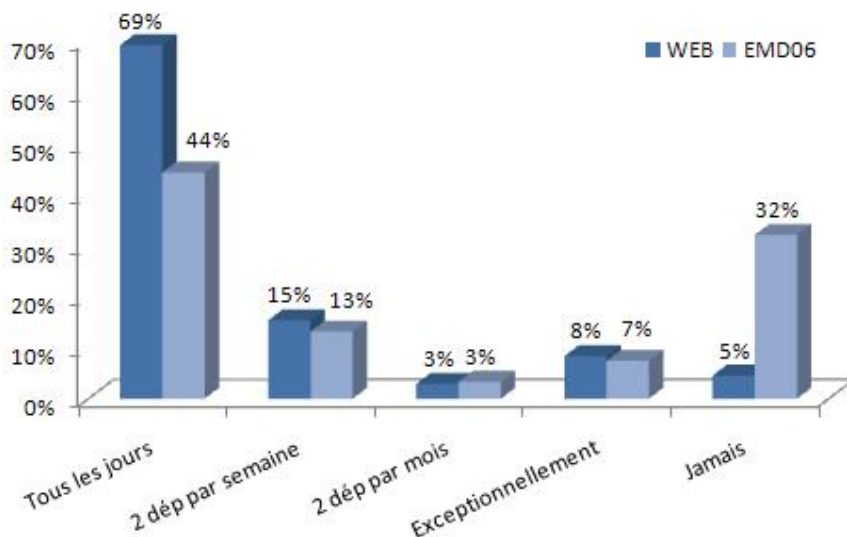


FIG. 32 – Répartition des personnes par enquête et par fréquence d'utilisation de la voiture conducteur en semaine

Source : EMD web et face-à-face - Lyon 2006

Ces résultats corroborent d'autres études, qui mettent en évidence le rôle central de l'automobile dans les déplacements quotidiens (Baudelle *et al.*, 2003). La voiture est un mode souple, indépendant, qui procure un sentiment de liberté. La praticité et la vitesse sont les atouts les plus avancés par les usagers, pour lesquels le temps compte plus que la distance parcourue. Ainsi, dès qu'il y a gain de temps, l'alternative automobile est susceptible de fonctionner (Dupuy, 1995). Les différences d'utilisation sont très probablement liées au statut des enquêtés, puisque davantage d'actifs ont répondu en ligne, et à la structure par âge. L'échantillon web regroupe en effet moins de jeunes et

de retraités, dont l'accès à la voiture particulière en tant que conducteur est souvent plus faible. La catégorie socioprofessionnelle des répondants peut également impacter les résultats, les internautes de notre enquête ayant davantage des emplois qualifiés et un niveau d'étude élevé.

I.2 Mais une utilisation moins fréquente de la voiture en tant que passager

Nous observons que 41,5% des personnes interrogées sur le web n'utilisent jamais la voiture particulière en tant que passager pour se déplacer en semaine, vs. 22,6% dans l'enquête en face-à-face (p-value < 0,01%) (figure 33). En revanche, peu de personnes utilisent chaque jour ce mode de transport pour se déplacer en semaine, quel que soit le mode d'enquête (10,0% sur le web, vs. 9,9% en face-à-face). Cette différence n'est toutefois pas significative (risque de première espèce = 95,97%) Il peut s'agir d'une part résiduelle de personnes qui ne possèdent pas leur permis, ou d'adeptes du covoiturage.

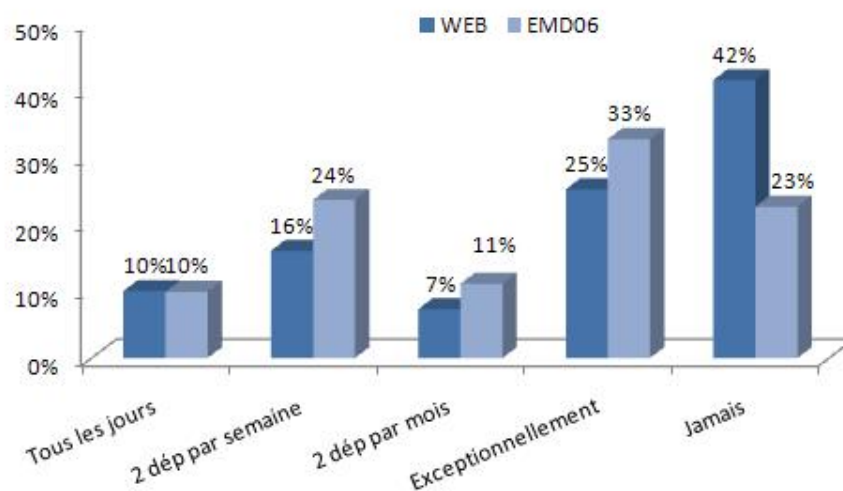


FIG. 33 – Répartition des personnes par enquête et par fréquence d'utilisation de la voiture passager en semaine

Source : EMD web et face-à-face - Lyon 2006

I.3 Un usage plus modéré des transports collectifs

Nous observons que les répondants à l'enquête en face-à-face utilisent plus fréquemment les transports collectifs en semaine (22,0%) que les répondants à l'enquête web (18,2%) (figure 34). Cette différence n'est toutefois pas significative (risque de première espèce = 8,02%). En revanche, la proportion de personnes n'utilisant jamais les transports en commun pour se déplacer en semaine est beaucoup plus importante dans l'enquête web (31,2%, vs. 24,2% pour l'enquête en face-à-face) (p-value = 0,5%).

Ces résultats peuvent s'expliquer par les différences observées en termes de motorisation. Par ailleurs, les familles nombreuses, davantage représentées

I Habitudes d'utilisation des modes de transport en semaine

dans l'enquête web, ne sont pas favorables aux transports collectifs, la présence d'enfants étant difficile lors des phases d'achat de tickets et d'embarquement.

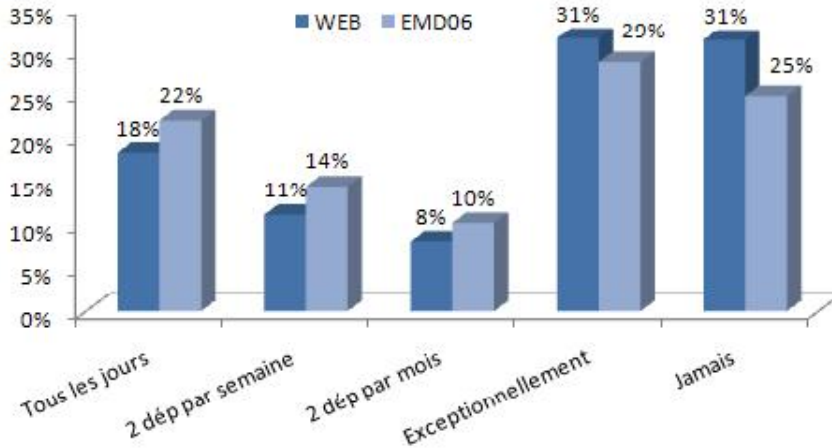


FIG. 34 – Répartition des personnes par enquête et par fréquence d'utilisation des transports collectifs en semaine

Source : EMD web et face-à-face - Lyon 2006

I.4 Une utilisation des deux-roues quasi-marginale

La proportion des répondants qui n'utilisent jamais les deux-roues pour se déplacer en semaine est plus importante dans l'enquête en face-à-face (96,3%, vs. 76,8% dans l'enquête web) (figure 35).

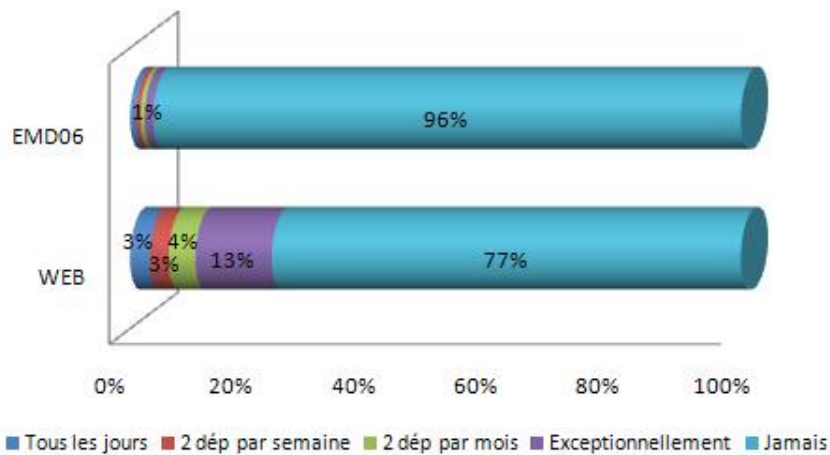


FIG. 35 – Répartition des personnes par enquête et par fréquence d'utilisation des deux-roues en semaine

Source : EMD web et face-à-face - Lyon 2006

Si cette différence est statistiquement significative (p-value < 0,01%), il est

nécessaire de prendre en compte les faibles effectifs de l'enquête web avant de tirer des conclusions. Ce constat peut s'expliquer par la forte proportion d'actifs qui ont répondu au questionnaire en ligne, à leur niveau de vie élevé et à la localisation de leur lieu de travail, situé davantage dans l'hypercentre et le centre de l'agglomération, localisation propice à l'utilisation des deux-roues. Le deux-roues moteur est surtout utilisé par les jeunes qui résident en espace périurbain et qui ne peuvent accéder aux transports collectifs pour assouvir leur soif de mobilité (Baudelle *et al.*, 2003). En revanche, chez les femmes, une certaine peur et la complexité des schémas d'activités (accompagnement des enfants, achats...) rendent plus difficile l'utilisation des deux-roues.

I.5 Conclusion sur les habitudes de déplacement

La comparaison des données de l'enquête ménages déplacements face-à-face avec celles de l'enquête web nous permet d'obtenir quelques précisions sur les habitudes des répondants en termes de fréquence d'utilisation des modes de transport. Les répondants web déclarent utiliser davantage la voiture particulière en tant que conducteur et les deux-roues pour se déplacer en semaine. A contrario, ils déclarent utiliser moins souvent les transports en commun et la voiture particulière en tant que passager comme moyens de transport quotidiens. Ces résultats sont très probablement liés aux caractéristiques socio-économiques des répondants, notamment leur niveau de revenu et de motorisation.

Le véhicule personnel est également perçu comme un prolongement de l'espace privé et les conducteurs associent une forme de reconnaissance sociale à sa propriété et à son usage, alors que le partage de modes collectifs empruntés de façon écrasante par des clients captifs est vécu sur le mode de la contrainte (Dupuy, 1995). Enfin, les transports en commun sont représentés comme contraignants, à cause de l'astreinte des usagers à des lignes et des horaires. Leur usage est dévalorisé, bien qu'ils permettent aujourd'hui l'affirmation d'une sensibilité 'écologique' et que ne pas posséder de voiture particulière devient une manière d'affirmer un style de vie.

II Caractéristiques des déplacements quotidiens

Dans cette section, nous comparons les données issues de la troisième partie du questionnaire, relatives aux déplacements effectués durant la période de référence de l'enquête.

II.1 Un nombre de déplacements plus faible

Nous constatons que la part des personnes ne s'étant pas déplacées dans l'aire d'étude durant la période de référence est plus élevée dans l'enquête web (19% vs. 11% en face-à-face ; p-value < 0,01%) (figure 36). Inversement, la part des personnes ayant effectué au moins 7 déplacements est faible parmi notre échantillon d'internautes (6% vs. 11% en face-à-face ; p-value < 0,01%).

II Caractéristiques des déplacements quotidiens

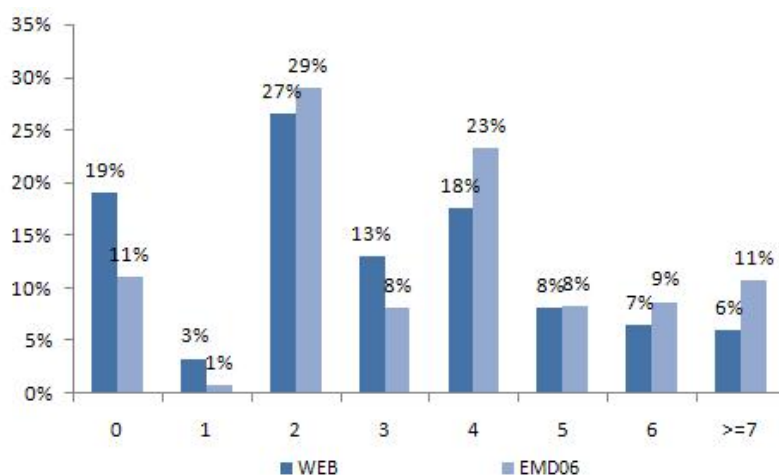


FIG. 36 – Répartition des personnes par nombre de déplacements et par enquête

Source : EMD web et face-à-face - Lyon 2006

En moyenne, nous recensons 3,63 déplacements par individu interrogé en face-à-face, contre seulement 3,00 par internaute (tableaux 28 et 29). Cette différence significative ($p\text{-value} < 0,01\%$) provient tout d'abord d'une immobilité nettement plus importante dans l'enquête web. Il est en effet plus facile de se déclarer immobile pour raccourcir l'enquête dans une enquête auto-administrée où personne ne vient relancer le répondant, qu'en présence d'un enquêteur. Par ailleurs, le manque d'ergonomie du logiciel ne favorise pas la saisie des déplacements en ligne ¹⁰². La composition des échantillons peut également expliquer les résultats observés. Les inactifs multiplient les déplacements brefs, tandis que les actifs se déplacent essentiellement pour se rendre à leur travail (Baudelle *et al.*, 2003).

Si on reprend la même analyse uniquement pour les mobiles le jour de l'enquête, l'écart subsiste mais est réduit (3,67 déplacements pour les internautes, vs. 4,08 pour les répondants en face-à-face ; $p\text{-value} < 0,01\%$). Il peut s'expliquer par le mode d'administration du questionnaire, l'enquêteur ayant pour fonction de relancer les personnes et de vérifier la pertinence du schéma de déplacements. Il est donc probable que des 'petits' déplacements aient été omis par les répondants en ligne. Soit volontairement, afin de ne pas prendre trop de temps pour remplir le questionnaire, soit involontairement, car la définition du déplacement n'est pas très évidente, notamment lors de courts arrêts au cours d'une sortie pour un motif principal. Il est également possible, comme les études des années 80 sur les schémas d'activités l'ont montré (Jones *et al.*, 1980), qu'une entrée activités et non pas déplacements dans le questionnaire aurait permis de réduire ces omissions. Mais pour assurer la cohérence avec l'enquête en face-à-face, ce choix n'a pas été privilégié.

¹⁰²Les contraintes de réalisation n'ont pas permis de choisir le prestataire et ont conduit à la transposition d'un logiciel CATI, utilisé pour les enquêtes téléphoniques assistées par ordinateur.

Analyse comparative de la mobilité

Nb de déplacements	Obs.	Minimum	Maximum	Moyenne	Ecart-type
Total	369	0	15	3,00	2,33
Personnes mobiles	299	1	15	3,67	2,02

TAB. 28 – Nombre moyen de déplacements quotidiens pour l'enquête web
Source : EMD web - Lyon 2006

Nb de déplacements	Obs.	Minimum	Maximum	Moyenne	Ecart-type
Total	13 271	0	23	3,63	2,49
Personnes mobiles	11 812	1	23	4,08	2,26

TAB. 29 – Nombre moyen de déplacements quotidiens pour l'enquête face-à-face
Source : EMD face-à-face - Lyon 2006

II.2 Modes de déplacement : une utilisation plus fréquente de la voiture au détriment de la marche

Les deux distributions, web et face-à-face, des déplacements selon les modes employés font apparaître des différences significatives. La voiture particulière en tant que conducteur semble sur-représentée dans l'enquête web (58% vs. 44% en face-à-face ; p-value < 0,01%) (figure 37). Comme nous l'avons montré précédemment, les répondants web sont davantage actifs, motorisés, et utilisent plus fréquemment la voiture particulière dans leurs déplacements quotidiens. A contrario, les déplacements à pied sont deux fois moins fréquents chez les internautes (15% vs. 29% en face-à-face ; p-value < 0,01%). Ils correspondent souvent à de petits déplacements, par nature moins bien enregistrés par les internautes.

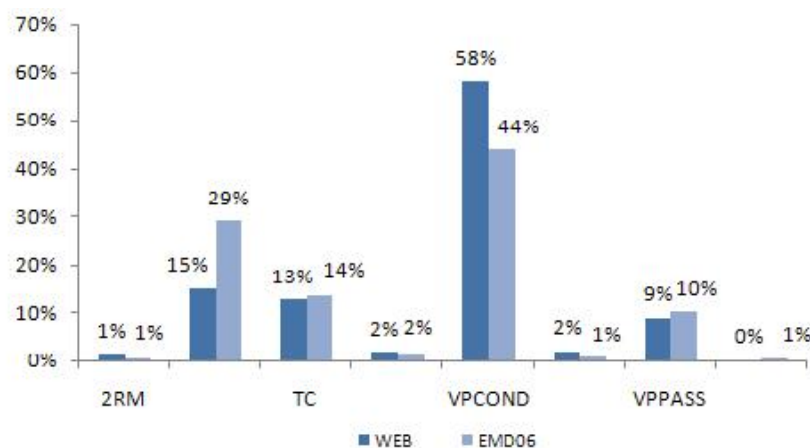


FIG. 37 – Répartition des déplacements quotidiens par mode et par enquête
Source : EMD web et face-à-face - Lyon 2006

Il est intéressant de comparer le nombre de déplacements par mode, pour voir si les différences observées entre les deux populations subsistent. Les résultats sont présentés pour l'ensemble des répondants, puis pour les mobiles,

II Caractéristiques des déplacements quotidiens

c'est-à-dire les répondants s'étant déplacés dans l'aire d'étude pendant la période de référence (tableau 30).

	Enquête web		Enquête face-à-face	
	Total	Mobiles	Total	Mobiles
Deux-roues motorisé	0,04	0,05	0,02	0,02
Marche à pied	0,44	0,56	1,05	1,18
Transports en commun	0,38	0,47	0,49	0,56
Vélo	0,06	0,07	0,06	0,06
Voiture conducteur	1,74	2,14	1,5	1,79
Voiture conducteur + Transports en commun	0,06	0,08	0,02	0,03
Voiture passager	0,27	0,33	0,37	0,41
Voiture passager + Transports en commun	0,01	0,01	0,02	0,02
Autre	0	0	0,01	0,01
Total	3,00	3,71	3,63	4,08

TAB. 30 – Nombre moyen de déplacements quotidiens par mode et par enquête
Source : EMD web et face-à-face - Lyon 2006

Ces résultats confirment les conclusions précédentes. Si on les rapproche des caractéristiques socio-économiques des deux populations, on peut poser les hypothèses suivantes. La voiture est davantage utilisée par la population web, ce qui semble logique, du fait de son niveau d'étude, de revenu et de motorisation. Les facteurs âge et sexe sont aujourd'hui encore fortement corrélés à l'usage de la voiture particulière, à cause notamment de la faible proportion de femmes âgées ayant le permis, bien que cet écart se réduise avec le renouvellement des générations (Toint *et al.*, 2004).

Si on additionne les données des modes motorisés, la mobilité des répondants web devient supérieure à celle des répondants en face-à-face (2,57 déplacements quotidiens, vs. 2,25), tout particulièrement si on exclut ceux qui ne se sont pas déplacés. Il semblerait donc que l'enquête web permette de récupérer 'correctement' les déplacements motorisés de ceux qui se déclarent mobiles, mais conduise à une forte sous-estimation des déplacements à pied, qui sont aussi de petits déplacements. Pour valider cette conclusion, il faudrait toutefois pousser l'analyse plus loin, pour tenir compte des différences socio-économiques des deux populations (Cf. chapitre 6).

Au niveau global, les déplacements individuels sont liés aux évolutions économiques, qui demandent toujours plus d'échanges. Les pratiques modales se caractérisent notamment par des choix et des contraintes en matière de programmes d'activités et de motorisation. La progression des déplacements bénéficie essentiellement à la voiture particulière, en termes de part des déplacements et de proportion de kilomètres parcourus (Pochet, 2005). Son utilisation permet de gérer des programmes d'activités complexes et de réaliser de multiples activités dans des lieux dispersés. Ainsi, très peu de conducteurs pourraient maintenir leur niveau de mobilité quotidienne, autrement qu'en voiture particulière, sans remettre en question une partie de leurs activités ou les

temps de déplacement qui y sont associés (Orfeuil et Massot, 2003).

II.3 Motifs de déplacements : un poids accru du motif 'Travail', au détriment du motif 'Etudes' principalement

Afin d'avoir des effectifs suffisants, nous segmentons les motifs de déplacement en six catégories :

- Domicile ;
- Travail (habituel ou inhabituel) ;
- Nourrice, école, collège, lycée ou université ;
- Achats et services ;
- Loisirs ;
- Accompagnement.

Le motif destination 'Travail' est sur-représenté dans l'enquête web, par rapport à l'enquête en face-à-face (22.9% vs. 15.4% ; p-value < 0,01%) (figure 38). A contrario, le motif destination 'Nourrice/Ecole/Université' semble sous-représenté parmi les internautes (0.9% vs. 6.1% en face-à-face ; p-value < 0,01%). Les effectifs constatés dans le cadre de l'enquête web sont toutefois trop faibles pour que l'on puisse conclure de manière statistiquement significative sur ce motif.

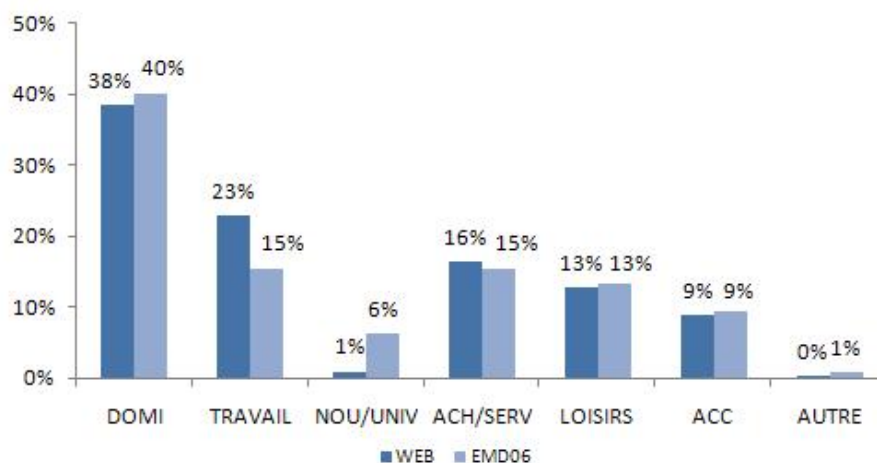


FIG. 38 – Répartition des déplacements quotidiens par motif à destination et par enquête

Source : EMD web et face-à-face - Lyon 2006

Lorsque la destination du déplacement est le domicile, il est intéressant de regarder le motif à l'origine. Nous retrouvons ici encore les différences observées pour le motif à la destination des déplacements. Le motif 'Travail' est sur-représenté dans l'enquête web (37,2%, vs. 23,4% en face-à-face ; p-value < 0,01%) et le motif 'Nourrice/ Université' est beaucoup moins recensé en ligne (1,9%, vs. 12,7% en face-à-face ; p-value < 0,01%) (figure 39).

II Caractéristiques des déplacements quotidiens

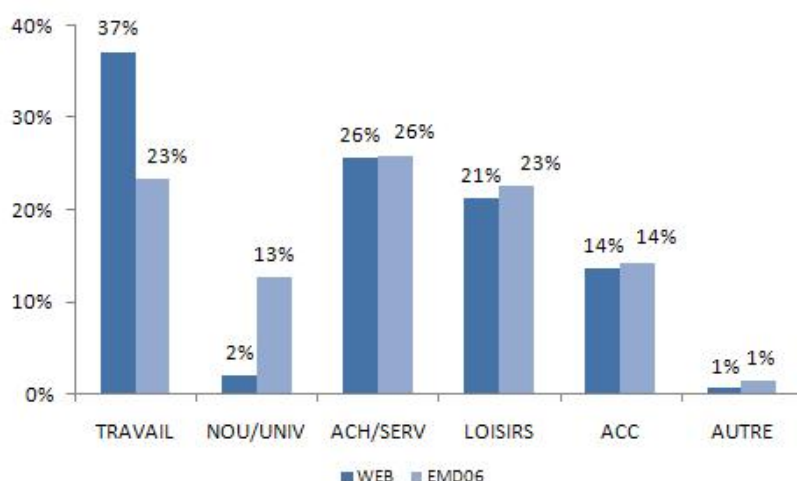


FIG. 39 – Répartition des déplacements quotidiens par enquête et par motif à l'origine, lorsque le motif à la destination est le domicile

Source : EMD web et face-à-face - Lyon 2006

L'analyse du nombre de déplacements par motif à la destination, pour l'ensemble des répondants puis pour les mobiles, c'est-à-dire les répondants s'étant déplacés dans l'aire d'étude pendant la période de référence, conduit globalement aux mêmes conclusions (tableau 31).

	Enquête web		Enquête face-à-face	
	Total	Mobiles	Total	Mobiles
Domicile	1,14	1,43	1,45	1,62
Travail	0,69	0,85	0,56	0,63
Collège-Lycée-Université	0,03	0,03	0,22	0,25
Achat-Service	0,49	0,6	0,56	0,63
Loisirs	0,38	0,47	0,48	0,54
Accompagnement	0,26	0,32	0,33	0,38
Autre	0,01	0,01	0,03	0,03
Total	3,00	3,71	3,63	4,08

TAB. 31 – Nombre moyen de déplacements quotidiens par motif à la destination et par enquête

Source : EMD web et face-à-face - Lyon 2006

Les individus sont supposés faire des déplacements dans des buts précis et ces buts consistent à s'engager dans des activités. Les activités sont donc à l'origine du processus de mobilité (Primerano *et al.*, 2008). Certains chercheurs ont traditionnellement divisé les motifs de déplacement en trois catégories : subsistance (travail et scolaire), maintenance (achats et démarches personnelles) et temps libre (loisirs) (Stopher et Metcalf, 1996). Les déplacements dits de subsistance ont une fréquence quotidienne et des localisations et horaires fixes. Ceux de maintenance ont une fréquence régulière, mais des localisations variables. Enfin, les déplacements de loisirs ont des caractéristiques variables et sont parfois inobservés.

Les déplacements liés au travail sont les plus nombreux, même si leur poids tend à diminuer ces dernières décennies (nombre de retraités en augmentation, montée du chômage, extension de la journée continue...) (Mokhtarian *et al.*, 2006).

II.4 Heure de départ des déplacements : une pointe horaire plus marquée le matin et plus tardive le soir

Nous constatons deux pics importants de départ des déplacements, le matin (entre 07h00 et 09h00 pour l'enquête web et face-à-face) et le soir (entre 17h00 et 19h00 pour l'enquête web, et entre 16h00 et 18h00 pour l'enquête face-à-face) (figure 40). Un pic plus modeste est observé entre 12h00 et 14h00 pour les deux enquêtes, alors que le nombre de déplacements reste relativement faible entre 21h00 et 07h00.

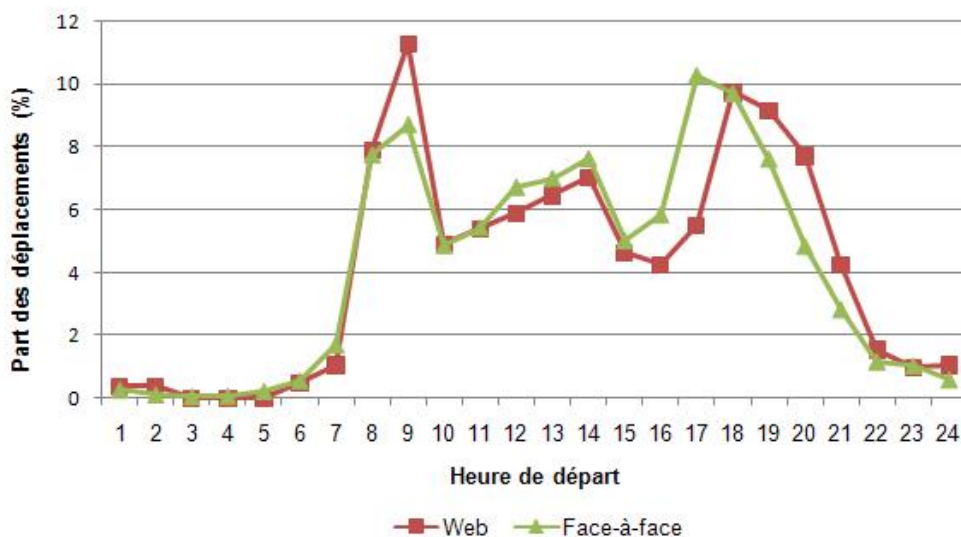


FIG. 40 – Nombre de déplacements quotidiens moyen par heure de début et par enquête

Source : EMD web et face-à-face - Lyon 2006

Les déplacements entre le domicile et le lieu de travail sont structurants (Pochet, 2005). Ils sont généralement réalisés à heures fixes, et entre des lieux identifiés. Les navettes quotidiennes sont donc le lien physique entre les milieux résidentiels et économiques (Shearmur, 2008) et les rythmes de la vie quotidienne expliquent la variation des horaires du trafic. Celle-ci résulte de comportements de personnes différentes, poursuivant des motivations différentes. La pointe du matin est simple et brève. Elle dépend surtout des scolaires et des travailleurs. La pointe du soir a même intensité et s'étale davantage dans le temps. Les personnes et les motifs concernés sont divers (Toint *et al.*, 2004). Les actifs qui disposent d'un haut niveau d'éducation ont des journées de travail plus longues, alors que les femmes et les autres personnes qui sont dépendantes

des enfants le sont moins ¹⁰³. Nous montrons dans le chapitre suivant que les internautes actifs passent davantage de temps sur leur lieu de travail que les actifs répondant en face-à-face, ce qui peut expliquer cette pointe du soir plus tardive.

II.5 Des déplacements plus longs en distance et en temps

La distance moyenne des déplacements effectués par les personnes est supérieure dans l'enquête web (4,61 km, vs. 3,78 km en face-à-face ; p-value < 0,01%) ¹⁰⁴ (tableau 32). En revanche, la dispersion des valeurs autour de la moyenne est d'un niveau comparable entre les deux enquêtes. Dans l'enquête web, nous avons une large proportion de déplacements en voiture particulière conducteur, et inversement peu de petits déplacements à pied. Nous pouvons penser que cette différence en termes de modes de déplacements explique l'écart de distance moyenne.

Distance(km)	Obs.	Minimum	Maximum	Moyenne	Ecart-type
Enquête web	1 066	0,02	40,37	4,61	4,73
Enquête standard	47 415	0,06	66,26	3,78	4,96

TAB. 32 – Distance moyenne des déplacements quotidiens calculée à vol d'oiseau par enquête

Source : EMD web et face-à-face - Lyon 2006

Par ailleurs, certaines caractéristiques sociodémographiques de répondants web peuvent expliquer ces différences. On admet généralement que les catégories les plus aisées effectuent en moyenne des déplacements de plus longue portée pour se rendre à leur travail, alors que les plus défavorisées apparaissent plus souvent captives de leur environnement immédiat (Berger, 2004). Les cadres et les professions intellectuelles supérieures ont des trajets domicile-travail plus longs que les ouvriers et les employés. En revanche, les ménages subissant le moins de contraintes familiales (une personne, un actif...) parviennent à limiter leurs distances et durées de déplacements. La distance parcourue augmente également avec le revenu du ménage, car les personnes disposant d'un haut pouvoir d'achat sont incitées à se déplacer et disposent de ressources leur permettant de le faire. Enfin, la présence d'un enfant à domicile tend à augmenter la distance quotidienne parcourue par les membres du ménage (Shearmur, 2008).

La durée moyenne des déplacements effectués est supérieure dans l'enquête web (22,96 mn, vs. 19,25 mn en face-à-face ; p-value < 0,01%) (tableau 33).

¹⁰³Les cadres du secteur privé ont augmenté leur temps de travail de trente minutes entre 1986 et 1999, alors que celui des ouvriers a diminué de six minutes et celui des employés et professions intermédiaires de dix minutes sur la même période (Dumontier et Pan Ke Shon, 1999).

¹⁰⁴Seuls les déplacements ayant leur origine et leur destination dans l'aire d'étude sont comptabilisés.

Le niveau de dispersion des valeurs autour de la moyenne est également plus important pour les internautes (33,72 mn, vs. 19,66 mn en face-à-face). Dans l'enquête web, nous avons recensé une forte proportion de déplacements en voiture particulière, de déplacement domicile-travail, mais peu de petits déplacements à pied. Les distances et durées parcourues dans l'aire d'étude sont donc en moyenne plus importantes.

Durée (mn)	Observations	Minimum	Maximum	Moyenne	Ecart-type
Enquête web	1105	0,35	570	22,96	33,72
Enquête standard	47 415	1	450	19,25	19,66

TAB. 33 – Durée moyenne des déplacements quotidiens calculée à vol d'oiseau par enquête

Source : EMD web et face-à-face - Lyon 2006

Certaines caractéristiques sociodémographiques peuvent également expliquer ces différences. Il semble exister une relation positive entre le temps de travail et le temps de transport, de longues heures de travail générant davantage de temps de déplacement (Wang et Law, 2007). Or, les internautes ont globalement des journées de travail plus longues que les répondants web.

Nous comparons ensuite les budgets temps et budgets distances des individus, c'est-à-dire les durées et distances quotidiennes moyennes engendrées par les déplacements (tableau 34). Pour cela, nous n'avons retenu dans l'analyse que les personnes mobiles la veille du jour de l'enquête et les déplacements internes au périmètre d'enquête. Les internautes consacrent davantage de temps chaque jour aux déplacements que les répondants en face-à-face (87,23 mn, vs. 78,87 mn ; p-value = 0,01%). Ce constat s'explique notamment par un nombre de kilomètres parcourus plus élevé (17,33 km, vs. 15,40 km). A ce stade de l'analyse, il est difficile de conclure sur un effet potentiel du média d'enquête, étant donné les différences socio-économiques observées entre les deux populations.

Enquête	Budget	Nb obs.	Minimum	Maximum	Moyenne	Ecart-type
Web	Distance	277	0,24	85,16	17,33	14,27
	Temps	277	5	380	87,23	54,37
Face-à-face	Distance	11 420	0,18	286,36	15,40	16,59
	Temps	11 420	2	900	78,87	59,52

TAB. 34 – Budgets temps (mn) et distances (km) quotidiens des individus par enquête

Source : EMD web et face-à-face - Lyon 2006

L'augmentation des distances parcourues est liée à l'augmentation des vitesses praticables. Les progrès de la vitesse permettent, non pas de gagner du temps, mais d'augmenter la portée spatiale des déplacements, en maintenant relativement stable le temps quotidien consacré au transport. Les études

II Caractéristiques des déplacements quotidiens

montrent en effet que le budget temps de transport quotidien est relativement stable et est compris en moyenne entre une heure et une heure trente. Il est toutefois plus important pour les actifs que pour les inactifs, et s'élève lorsque l'on progresse dans l'échelle sociale (Orfeuill, 2000).

Les données révèlent à la fois l'importance des structures familiales dans les arbitrages en termes de budget temps et le poids des appartenances socio-professionnelles.

II.6 Des échanges entre zones accrues, surtout avec le centre...

Dans ce paragraphe, nous avons segmenté le périmètre d'enquête en trois zones concentriques autour de Lyon :

- la zone dite centrale, regroupant Lyon et Villeurbanne ;
- la proche banlieue, c'est-à-dire la première couronne ouest et est ;
- la périphérie, ou deuxième et troisième couronnes ouest et est.

Les deux distributions, web et face-à-face, des déplacements selon leur zone d'origine et de destination font apparaître des différences significatives. Dans l'enquête standard, la part des déplacements intrazones est plus importante que dans l'enquête web, pour la 1ère couronne et les 2ème et 3ème couronnes (tableau 35 et 36). La différence est particulièrement accentuée pour les déplacements dont l'origine se situe dans la première couronne (42,3% des déplacements ayant leur destination dans la première couronne de l'enquête web sont internes à cette zone, vs. 66% dans l'enquête standard ; p-value < 0,01%).

Origine-Destination	Centre, Hyper-centre (1, 2)	1re couronne (3, 4)	2è-3è couronnes (5 à 9)	Total
Centre, Hyper-centre (1, 2)	384 36,1%	75 7,0%	70 6,6%	529 49,7%
1re couronne (3, 4)	78 7,3%	93 8,7%	49 4,6%	220 20,7%
2è-3è couronnes (5, 6, 7, 8, 9)	68 6,4%	58 5,5%	189 17,8%	315 29,6%
Total	530 49,8%	226 21,2%	308 28,9%	1 064 100%

TAB. 35 – Répartition des déplacements quotidiens par Origine-Destination pour l'enquête web

Source : EMD web - Lyon 2006

Par ailleurs, la part des déplacements dont l'origine ou la destination se situe dans le centre est plus importante dans l'enquête web. Ce résultat est issu du poids accru des intrazones, dans le centre, mais aussi des échanges plus importants des trois couronnes avec le centre. Cette différence dans la distribution spatiale des déplacements explique très probablement la distance

accrue des déplacements de l'enquête web, par rapport à l'enquête en face-à-face.

Origine-Destination	Centre, Hyper-centre (1, 2)	1re couronne (3, 4)	2è-3è couronnes (5 à 9)	Total
Centre, Hyper-centre (1, 2)	14 918 31,5%	2 295 4,8%	1 881 4,0%	19 094 40,3%
1re couronne (3, 4)	2 288 4,8%	7 902 16,7%	1 786 3,8%	11 976 25,3%
2è-3è couronnes (5, 6, 7, 8, 9)	1 909 4,0%	1 768 3,7%	12 668 26,7%	16 345 34,5%
Total	19 115 40,3%	11 965 25,2%	16 335 34,5%	47 415 100%

TAB. 36 – Répartition des déplacements quotidiens par Origine-Destination pour l'enquête face-à-face

Source : EMD face-à-face - Lyon 2006

II.7 ...dû à des localisations plus éclatées du domicile et du travail

Dans l'enquête face-à-face, la majorité des personnes travaillent dans la zone de leur lieu de résidence, alors que dans l'enquête web, la part du centre comme zone de travail est toujours la plus importante, quelle que soit la zone de résidence (26,3% des personnes de l'enquête standard habitant en 2ème et 3ème couronne travaillent dans l'hypercentre, vs. 44,1% des répondants à l'enquête web ; p-value < 0,01%) (tableaux 37 et 38).

Domicile-Travail	Centre, Hyper-centre (1, 2)	1re couronne (3, 4)	2è-3è couronnes (5 à 9)	Total
Centre, Hyper-centre (1, 2)	88 31,3%	16 5,7%	11 3,9%	115 40,90%
1re couronne (3, 4)	35 12,5%	26 9,3%	12 4,3%	73 26,00%
2ème-3ème couronne (5, 6, 7, 8, 9)	41 14,6%	18 6,4%	34 12,1%	93 33,10%
Total	164 58,4%	60 21,4%	57 20,3%	281 100%

TAB. 37 – Répartition des personnes par lieu de résidence et lieu habituel de travail ou d'études pour l'enquête web

Source : EMD web - Lyon 2006

Les emplois spécialisés, rares et plus prestigieux que d'autres, attirent des personnes venant de plus loin. Elles travaillent dans les centres d'affaires, surtout localisés au centre de l'agglomération et résident en banlieue, ce qui augmente leur distance domicile travail. Par ailleurs, les personnes aux revenus

III Spécificités du web en termes de sorties quotidiennes

plus modestes ont davantage des emplois locaux, moins qualifiés, qui ne nécessitent pas de longs déplacements vers le centre de l'agglomération.

Domicile-Travail	Centre, Hyper-centre (1, 2)	1re couronne (3, 4)	2è-3è couronnes (5 à 9)	Total
Centre, Hyper-centre (1, 2)	2 514 28,2%	391 4,4%	350 3,9%	3 261 36,60%
1re couronne (3, 4)	727 8,1%	1 162 13,0%	344 3,9 %	2 233 25,00%
2è-3è couronne (5, 6, 7, 8, 9)	904 10,1%	523 5,9%	2 006 22,5%	3 433 38,50%
Total	4 145 46,5%	2 076 23,3%	2 706 30,3%	8 927 100%

TAB. 38 – Répartition des personnes par lieu de résidence et lieu habituel de travail ou d'études pour l'enquête face-à-face

Source : EMD face-à-face - Lyon 2006

III Spécificités du web en termes de sorties quotidiennes

Une sortie se définit comme l'enchaînement des déplacements et des activités entre un départ et un retour au domicile. Il peut donc être intéressant pour notre analyse de comparer les caractéristiques des sorties des répondants web avec celles des répondants en face-à-face.

III.1 Des sorties moins fréquentes

Le nombre moyen de sorties effectuées quotidiennement par les répondants est inférieur dans l'enquête web (1,19 vs. 1,47 en face-à-face ; p-value < 0,01%) (tableaux 39 et 40). Le niveau de dispersion des valeurs autour de la moyenne est également plus faible dans le cas de l'enquête web (0,87 vs. 0,97 en face-à-face).

Nb de sorties	Obs.	Minimum	Maximum	Moyenne	Ecart-type
Total	369	0	5	1,19	0,87
Personnes mobiles	299	1	5	1,47	0,71

TAB. 39 – Nombre moyen de sorties quotidiennes du domicile pour l'enquête web

Source : EMD web - Lyon 2006

La répartition des répondants par nombre de sorties confirme cet écart, puisque 67,9% des répondants web ne réalisent qu'une seule sortie par jour, vs. 60,7% dans l'enquête face-à-face (p-value = 0,2%)(figure 41). Une analyse plus fine montre que le nombre de déplacements par sortie est similaire pour

Analyse comparative de la mobilité

Nb de sorties	Obs.	Minimum	Maximum	Moyenne	Ecart-type
Total	13 271	0	10	1,47	0,97
Personnes mobiles	11 812	1	10	1,65	0,87

TAB. 40 – Nombre moyen de sorties quotidiennes du domicile pour l'enquête face-à-face

Source : EMD face-à-face - Lyon 2006

les deux modes d'enquête. Il est de 2,5 déplacements par sortie en face-à-face vs. 2,47 déplacement par sortie sur le web, si nous considérons l'ensemble de la population (2,52 en face-à-face vs. 2,47 sur le web pour les mobiles). Ceci donne des indications sur la complexité des programmes d'activités. La moindre mobilité des répondants web ne semble donc pas s'expliquer par des oublis de petits déplacements au cours de sorties, mais par un nombre de sorties quotidiennes plus faible.

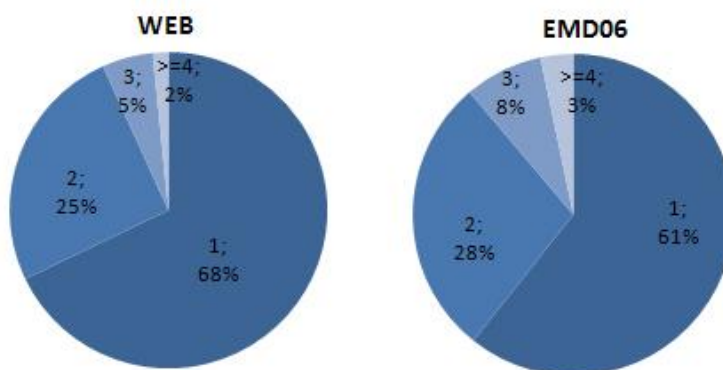


FIG. 41 – Distribution des sorties quotidiennes du domicile par enquête
Source : EMD web et face-à-face - Lyon 2006

III.2 Des sorties nettement plus longues en temps et en distance

La durée moyenne des sorties quotidiennes est supérieure dans l'enquête web (63,90 mn, vs. 50,17 mn en face-à-face ; p-value < 0,01%). Le niveau de dispersion des valeurs autour de la moyenne est également plus important dans le cas de l'enquête web (57 mn, vs. 51 mn en face-à-face) (tableau 41). Cet écart peut être lié à celui constaté pour la durée des déplacements, plus longs en moyenne pour les répondants web.

Durée (mn)	Obs.	Minimum	Maximum	Moyenne	Ecart-type
Enquête web	436	6	570	63,90	56,97
Enquête face-à-face	19 473	1	900	50,17	51,41

TAB. 41 – Durée moyenne des sorties quotidiennes du domicile par enquête
Source : EMD face-à-face - Lyon 2006

III Spécificités du web en termes de sorties quotidiennes

La distance moyenne des sorties quotidiennes est également supérieure dans l'enquête web (11,16 km, vs. 9,11 km en face-à-face ; p-value < 0,01%). Le niveau de dispersion des valeurs autour de la moyenne est en revanche relativement proche dans les deux enquêtes (11,5 km pour le web, vs. 11,9 km pour le face-à-face) (tableau 42). Cet écart peut être lié à celui observé pour les déplacements, plus longs en moyenne pour les répondants web.

Distance (km)	Obs.	Minimum	Maximum	Moyenne	Ecart type
Enquête web	439	0	75,98	11,16	11,50
Enquête face-à-face	19 473	0	183,22	9,11	11,91

TAB. 42 – Distance moyenne des sorties du domicile par enquête
Source : EMD face-à-face - Lyon 2006

III.3 Premier mode des sorties du domicile

Nous limitons l'analyse au mode utilisé pour le premier déplacement de la sortie, car l'analyse des données montre que l'unicité modale au cours de la sortie constitue la règle. Les deux distributions, web et face-à-face, des sorties du domicile selon le 1er mode employé font apparaître des différences significatives. La voiture particulière en tant que conducteur semble sur-représentée dans l'enquête web (56% vs. 42% en face-à-face ; p-value < 0,01%) (figure 42). A contrario, les déplacements à pied sont deux fois moins nombreux sur le web (15% vs. 29% en face-à-face ; p-value < 0,01%). Ces résultats confirment les conclusions relatives aux déplacements, élaborées dans les paragraphes précédents.

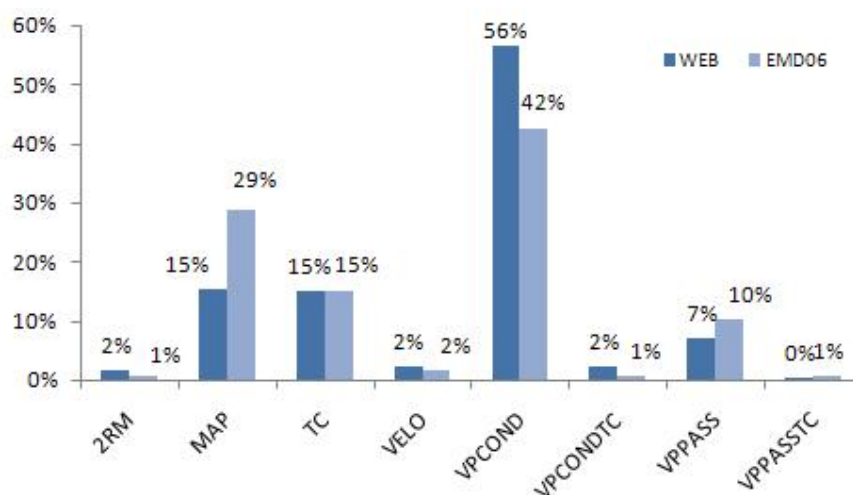


FIG. 42 – Répartition des sorties quotidiennes du domicile par 1er mode et par enquête
Source : EMD web et face-à-face - Lyon 2006

Nous comparons à présent le nombre de sorties du domicile selon le mode du premier déplacement de la sortie, pour voir si les différences observées entre les

deux populations subsistent. Les résultats sont présentés pour l'ensemble des répondants, puis pour les mobiles, c'est-à-dire les répondants s'étant déplacés dans l'aire d'étude pendant la période de référence (tableau 43).

	Enquête web		Enquête face-à-face	
	Total	Mobiles	Total	Mobiles
Deux-roues	0,02	0,02	0,01	0,01
Marche à pied	0,18	0,22	0,42	0,47
Transports en commun	0,18	0,23	0,22	0,25
Vélo	0,03	0,03	0,02	0,03
Voiture conducteur	0,67	0,84	0,63	0,7
Voiture conducteur + Transports en commun	0,03	0,03	0,01	0,01
Voiture passager	0,08	0,1	0,15	0,17
Voiture passager + Transports en commun	0	0	0,01	0,01
Autre	0	0	0	0
Total	1,19	1,47	1,47	1,65

TAB. 43 – Nombre moyen de sorties quotidiennes du domicile par 1er mode et par enquête

Source : EMD face-à-face - Lyon 2006

Les conclusions formulées à partir de la figure 42 subsistent globalement. Si l'on considère l'ensemble des internautes, l'usage de la voiture est très légèrement supérieur (un peu plus en tant que conducteur et moins en tant que passager). En revanche, si l'on considère uniquement les mobiles, l'usage de la voiture est nettement plus important parmi les internautes. Ce résultat découle probablement de la composition socio-économiques des répondants à l'enquête web (davantage d'actifs, de professions supérieures, fortement motorisés). La faiblesse de la marche subsiste, avec un usage deux fois plus faible. Les différences sont en revanche très limitées pour les autres modes.

III.4 Davantage de sorties 'Travail', au détriment des sorties 'Etudes', 'Loisirs' et 'Accompagnement'

Nous remarquons que la part des sorties du domicile ayant pour motif principal le travail est beaucoup plus importante pour les répondants de l'enquête web (46,7% vs. 29,2% dans l'enquête face-à-face ; p-value < 0,01%) (figure 43). A contrario, la part des sorties du domicile ayant pour motif l'école ou l'université est très faible dans l'enquête web (2,3% vs. 14,1% dans l'enquête en face-à-face). Si cet écart est statistiquement significatif (p-value < 0,01%), les effectifs de l'enquête web sont toutefois trop restreints pour pouvoir conclure sur ce point. Ces résultats sont liés à l'occupation et à la tranche d'âge des répondants. Sur le web, beaucoup d'actifs ont été enquêtés, ce qui explique la large part de sorties du domicile pour motif 'Travail'. Par ailleurs, nous retrouvons moins de jeunes dans l'enquête web, et moins de sorties pour motif

III Spécificités du web en termes de sorties quotidiennes

'Collège-Lycée-Université', puisque la personne qui répond au questionnaire en ligne est probablement le chef de ménage ou son conjoint, destinataires du courrier.

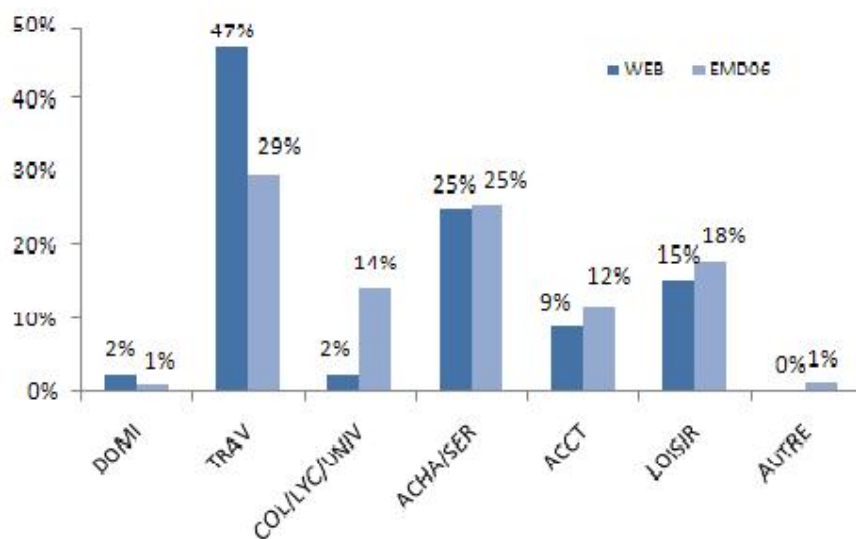


FIG. 43 – Répartition des sorties quotidiennes du domicile par motif à la destination et par enquête

Source : EMD web et face-à-face - Lyon 2006

L'analyse du nombre de sorties du domicile par motif à la destination, pour l'ensemble des répondants puis pour les mobiles, c'est-à-dire les répondants s'étant déplacés dans l'aire d'étude pendant la période de référence, conduit globalement aux mêmes conclusions (tableau 44).

	Enquête web		Enquête face-à-face	
	Total	Mobiles	Total	Mobiles
Domicile	0,03	0,03	0,01	0,01
Travail	0,55	0,70	0,44	0,49
Nourrice-Collège Lycée-Université	0,03	0,03	0,21	0,23
Achat-Service	0,3	0,36	0,37	0,42
Accompagnement	0,1	0,13	0,17	0,19
Loisirs	0,18	0,22	0,26	0,29
Autre	0	0	0,01	0,02
Total	1,19	1,47	1,47	1,65

TAB. 44 – Nombre moyen de sorties quotidiennes du domicile par motif à la destination et par enquête

Source : EMD face-à-face - Lyon 2006

IV Conclusion

Le but de ce chapitre est de mettre en évidence les différences de mobilité entre les deux échantillons et d'essayer de les expliquer par référence aux caractéristiques socio-économiques des répondants. Les niveaux de mobilité (et d'accès aux modes de transport), dépendent de la position dans le cycle de vie et du niveau de vie du ménage. Ceci légitime une approche désagrégée et la considération des caractéristiques personnelles comme éléments explicatifs des comportements de mobilité. La comparaison des données de l'enquête ménages déplacements face-à-face avec celles de l'enquête web nous permet d'obtenir quelques précisions sur la mobilité des répondants, en fonction du média utilisé pour l'enquête.

Les répondants web déclarent moins de déplacements et de sorties que les répondants à l'enquête en face-à-face. Cet écart s'explique à la fois par un nombre plus important d'immobiles (19% ne se sont pas déplacés, vs. seulement 11% dans l'enquête en face-à-face) et par une mobilité plus faible des mobiles dans l'enquête web (3,71 déplacements quotidiens, vs. 4,08 dans l'enquête en face-à-face). Les internautes utilisent davantage la voiture particulière (58% des déplacements se font en voiture dans l'enquête web, vs. 44% dans l'enquête en face-à-face), et affectionnent peu la marche à pied (respectivement 15%, vs. 29% des déplacements). Au final, l'enquête web permet de mieux 'récupérer' les déplacements motorisés des personnes mobiles (2,56 déplacements quotidiens en moyenne sur le web, vs. 2,25 en face-à-face).

Les internautes parcourent en moyenne des distances plus longues (4,61 km par déplacement, vs. 3,78 km en face-à-face) et réalisent de nombreux déplacements interzones, d'une durée plus importante (22,96 mn par déplacement, vs. 19,25 mn en face-à-face). Le budget temps de transport quotidien des répondants web est également supérieur à celui des répondants en face-à-face (87,23 mn, vs. 78,87 mn). Le principal motif des déplacements ou des sorties du domicile est le 'Travail' (23% des déplacements déclarés, vs. 15% en face-à-face), et les horaires de retour au domicile sont plus tardifs en soirée (14% des déplacements débutent entre 19h et 22h, vs. 9% en face-à-face). Enfin, leur lieu de travail habituel ou d'étude se situe davantage dans l'hypercentre de l'agglomération lyonnaise, quelle que soit leur zone de résidence (58,7%, vs. 46,5% en face-à-face). Les conclusions sont sensiblement les mêmes si on raisonne en termes de sorties du domicile.

Ces constats peuvent d'expliquer par deux facteurs principaux. D'abord, par des différences socio-économiques (permis de conduire, motorisation, revenu, niveau d'étude, âge, statut...), qui expliquent probablement une part accrue de la voiture, du motif 'Travail', des distances et durées plus longues pour les déplacements et sorties, et des horaires plus tardifs en soirée. Ensuite, par un effet du mode d'enquête, qui peut conduire à une plus forte immobilité et à une sous-déclaration du nombre de déplacements chez les répondants web. Cette sous-mobilité porte notamment sur les petits déplacements à pied et les

motifs peu contraints, et peut correspondre à de courtes sorties. Généralement, les déplacements courts ou les arrêts durant un long déplacement et les motifs non contraints sont davantage omis dans les modes auto-administrés (Bonnel, 2003). Nous chercherons dans le chapitre suivant à limiter les différences socio-économiques entre les deux échantillons, en redressant l'échantillon en face-à-face, afin de voir si les différences observées en termes de mobilité se maintiennent.

Chapitre 6 : Le redressement de l'échantillon en face-à-face : quels impacts sur la mobilité ?

"L'approche quantitative en sciences sociales a partie liée avec les statistiques."
(Castaigne *et al.*, 2006).

Au-delà des différences socio-économiques constatées entre les deux échantillons, web et face-à-face, les répondants ne semblent pas adopter les mêmes comportements de mobilité. A ce stade de l'analyse, il n'est pas possible d'imputer les différences de mobilité uniquement au choix du média. Nous savons que si les proportions de personnes interrogées dans l'enquête web, ne reflètent pas celles qui caractérisent la population de référence, au regard d'une ou plusieurs variables importantes, alors ces différences peuvent impacter les comportements de mobilité. Dans le chapitre 4, nous avons conclu que le profil des répondants web diffère substantiellement de celui des répondants en face-à-face (niveau d'éducation et d'activité, âge, revenus, motorisation...), ce qui peut expliquer en partie les écarts de mobilité constatés entre les deux échantillons. L'objectif de ce chapitre consiste à redresser l'échantillon en face-à-face, de façon à le rendre comparable à l'échantillon web au regard de caractéristiques sociodémographiques importantes. Le redressement est une opération qui consiste à affecter un poids aux observations de l'échantillon, de telle manière que l'échantillon pondéré reflète précisément la population de référence. Nous préférons redresser l'échantillon en face-à-face plutôt que celui du web, en raison du faible effectif de ce dernier

Nous revenons d'abord sur la méthode de calage sur marges, traditionnellement utilisée pour redresser les données des enquêtes ménages déplacements (section I). Puis, nous présentons les résultats de la méthode de calage sur marge sur l'échantillon en face-à-face (section II). Enfin, une fois l'échantillon face-à-face redressé sur les variables socio-économiques les plus discriminantes, nous examinons de nouveau l'incidence du mode d'enquête sur la mobilité (section III).

I Les aspects théoriques du calage sur marges

Le calage sur marges permet de redresser un échantillon provenant d'une enquête par sondage, par repondération, en utilisant une information auxiliaire disponible sur un certain nombre de variables, appelées variables de calage. Le redressement consiste à remplacer les pondérations initiales observées dans un échantillon par de nouvelles pondérations, telles que, pour chacune des variables utilisées pour le calage, les effectifs des modalités de la variable estimés dans l'échantillon après redressement soient égaux aux effectifs connus sur la population. L'intérêt de cette méthode est notamment d'assurer la cohérence entre les résultats de plusieurs enquêtes. Nous présentons les principes de cette méthode (section I.1), ainsi qu'une résolution théorique (section), en nous appuyant sur les développements de Sautory (1993).

I.1 Principes de la méthode

On considère une population $U = 1 \dots k \dots N$ d'individus, dans laquelle on a sélectionné un échantillon probabiliste s de taille n . Pour tout individu k de U , on note π_k sa probabilité d'inclusion dans s ¹⁰⁵. Soit Y une variable d'intérêt, dont on désire estimer le total sur la population :

$$Y_\pi = \sum_{k \in U} y_k \quad (14)$$

L'estimateur habituel de Y est l'estimateur de Horvitz-Thompson :

$$Y'_\pi = \sum_{k \in s} \frac{1}{\pi_k} = \sum_{k \in s} d_k * y_k \quad (15)$$

Utiliser cet estimateur sans biais de Y revient à affecter à chaque individu de l'échantillon un poids d_k égal à l'inverse de sa probabilité d'inclusion. Nous supposons connus les totaux sur la population de J variables auxiliaires ($X_1 \dots X_j \dots X_J$), disponibles sur l'échantillon s , tel que :

$$X_j = \sum_{k \in U} x_{jk} \quad (16)$$

Pour tenir compte de cette information, on va chercher à estimer le total Y' de Y , à l'aide d'un estimateur de la forme :

$$Y'_w = \sum_{k \in s} w_k * y_k \quad (17)$$

On va chercher de nouvelles pondérations, les 'poids de calage' (w_k), qui soient aussi proches que possible, au sens d'une certaine fonction de distance, des pondérations initiales (qui sont en général les 'poids de sondage', égaux aux inverses des probabilités d'inclusion (d_k), et qui assurent le calage sur les totaux des variables, c'est-à-dire qui vérifient les équations de calage :

¹⁰⁵Cette probabilité vaut n/N pour tout k , dans le cas d'un tirage aléatoire simple.

$$\forall j = 1 \dots J, \sum_{k \in s} w_k * x_{jk} = X_j \quad (18)$$

Il s'agit donc de trouver un estimateur peu différent de l'estimateur de Horvitz-Thompson, qui 'cale' l'échantillon sur les totaux des variables auxiliaires.

I.2 Résolution

On choisit une fonction de distance G, pour mesurer les distances entre les w_k et les d_k ¹⁰⁶. Le problème consiste à rechercher les poids $w_k, k \in s$ solutions de :

$$Min_{w_k} \sum_{k \in s} d_k G \frac{w_k}{d_k} \quad (19)$$

sous la contrainte de l'équation (18).

L'opération consiste à minimiser la somme pondérée des distances entre les poids de sondage d_k et les pondérations recherchées w_k , sous les contraintes du calage. Le problème se résout par l'utilisation du vecteur de multiplicateur de Lagrange associé aux contraintes $\lambda' = \lambda_1 \dots \lambda_J$. Le lagrangien vaut :

$$L = \sum_{k \in s} d_k G \frac{w_k}{d_k} - \lambda' \left(\sum_{k \in s} w_k * x_k - X \right) \quad (20)$$

Les conditions du premier ordre conduisent à :

$$w_k = d_k F(x_k \lambda) \quad (21)$$

avec F la fonction réciproque de la dérivée de G. λ est déterminé par la résolution du système non linéaire de J équations à J inconnues, résultant des équations de calage, par la méthode itérative de Newton :

$$\sum d * F(x' \lambda) x = X \quad (22)$$

Les mesures de distance sont équivalentes du point de vue des biais et de la variance (tableau 45). Elles produisent des estimateurs de même erreur quadratique, de moyenne asymptotique. Lorsque les variables auxiliaires sont toutes des variables catégorielles, pour lesquelles on connaît les effectifs des modalités dans la population, la méthode de redressement utilisée pour le calage sur marges est le raking ratio.

¹⁰⁶G est une fonction positive convexe, tel que $G(1)=G'(1)=0$.

II Application du calage sur marges à l'échantillon en face-à-face

G	$F = G^{-1}$	Type de distance	Méthode
$\frac{1}{2}(x-1)^2$	$1+u$	Khi-deux	Linéaire
$x \log(x) - x + 1$	$\exp(u)$	Entropie	Raking ratio
$\frac{1}{A}[(x-L) \log \frac{x-L}{1-L} + (U-x) \log \frac{U-x}{U-1}]$ avec $A = \frac{U-L}{(1-L)(U-1)}$ $x \in [L, U](\infty, \text{sinon})$	$\frac{L(U-1) + U(1-L)\exp(Au)}{(U-1) + (1-L)\exp(Au)}$ $x \in]L, U[$	Logistique	Raking ratio tronquée
$\frac{1}{2}(x-1)^2$ $x \in [L, U](\infty, \text{sinon})$	$1 + u \in [L, U]$	Khi-deux tronquée	Linéaire tronquée

TAB. 45 – Les fonctions de distance disponibles
Source : Sautory (1993)

II Application du calage sur marges à l'échantillon en face-à-face

Les variables servant au redressement de l'enquête sont toutes catégorielles et le redressement consiste à 'caler' les marges du tableau croisant toutes les variables de calage sur des effectifs déterminés a priori. Pour effectuer le calage sur marges, nous avons utilisé Calmar, un programme écrit dans le langage macro de SAS, qui permet de mettre en oeuvre les méthodes de calage élaborées par Deville *et al.* (1993). Calmar est un acronyme pour CALage sur MARGes : on désigne ainsi la technique de redressement qui permet d'ajuster les marges (estimées à partir d'un échantillon) d'un tableau de contingence, croisant deux (ou plus) variables catégorielles, aux marges connues dans la population. Le programme Calmar a été développé en 1990 à l'Institut National de la Statistique et des Études Économiques, où il est depuis régulièrement utilisé pour le redressement des enquêtes (Sautory, 1993)¹⁰⁷.

II.1 Sélection des variables de calage

Les variables de calage doivent être corrélées avec les indicateurs à estimer. Nous cherchons donc dans un premier temps les variables (sociodémographiques, de localisation, d'équipement...) qui impactent le plus le choix de répondre au questionnaire en ligne. À l'aide d'une régression logistique (modèle Logit), nous avons identifié sept variables importantes (tableau 46).

¹⁰⁷Le descriptif de la macro est disponible sur le site de l'INSEE, rubrique 'Nomenclatures-Définitions-Méthodes', outils statistiques.

Variables	Coef.	E.S.	Wald	ddl	Sign.	Exp(B)
Constante	-4,17	0,30	197,33	1	<0,01	0,02
Age : de 30 à 60 ans	0,75	0,24	9,59	1	<0,01	2,11
Age : moins de 30 ans	-0,08	0,26	0,11	1	0,74	0,92
Téléphone portable : non	-0,58	0,18	10,89	1	<0,01	0,56
Nombre de personnes : 1	1,41	0,23	38,02	1	<0,01	4,08
Nombre de personnes : 2	0,78	0,22	12,46	1	<0,01	2,19
Nombre de personnes : 3	0,38	0,24	2,48	1	0,12	1,46
Nombre de personnes : 4	0,60	0,23	6,86	1	0,01	1,82
connexion internet : non	-0,75	0,15	24,39	1	<0,01	0,47
diplôme : non supérieur	-0,76	0,13	31,85	1	<0,01	0,47
PCS : autre actif	-0,61	0,20	9,13	1	<0,01	0,55
PCS : cadre et prof. inter.	0,41	0,18	5,09	1	0,02	1,51
PCS : employé	0,72	0,18	16,92	1	<0,01	2,06
Possession permis : non	-0,39	0,22	3,09	1	0,08	0,68

TAB. 46 – Modèle logit du média de réponse à l'enquête (web)

Source : EMD web - Lyon 2006

Au final, on voit que deux types de variables semblent importantes pour qualifier les individus qui répondent sur le web : des variables sociodémographiques et des variables d'équipement en moyen de communication. Concernant l'équipement des ménages, ceux qui ne possèdent pas de téléphone portable ni de connexion internet à leur domicile sont peu représentés dans l'échantillon web. Il en est de même pour les individus qui ne possèdent pas leur permis de conduire. A contrario, les actifs (30-60 ans), les diplômés du supérieur et les 'petits' ménages semblent séduits par ce mode d'enquête novateur.

II.2 Détermination des poids de calage

Une fois sélectionnées les variables principales qui caractérisent les répondants en ligne, nous calculons les pourcentages des différentes modalités dans les échantillons face-à-face et web (population de référence). Les résultats sont présentés en annexe X. La macro 'CALMAR' redresse ensuite l'échantillon en face-à-face, de sorte que ces pourcentages soient égaux dans les deux populations, web et face-à-face (tableau 47).

II Application du calage sur marges à l'échantillon en face-à-face

Variables	Modalités	Enquête web	Enquête face-à-face
Téléphone portable	Oui	87,1%	87,5%
	Non	12,9%	12,5%
Connexion internet	Oui	79,9%	80,0%
	Non	20,1%	20,0%
Possession permis	Oui	91,6%	91,6%
	Non	8,4%	8,4%
Diplôme	Supérieur	67,2%	67,8%
	Non supérieur	32,8%	32,2%
Age	< 30 ans	17,1%	17,0%
	de 30 à 60 ans	73,4%	73,7%
	> 60 ans	9,5%	9,3%
Nombre de personnes	1 personne	24,4%	25,7%
	2 personnes	30,4%	29,6%
	3 personnes	15,4%	15,2%
	4 personnes	22,0%	21,2%
	5+ personnes	7,9%	8,4%
PCS	Employé	27,4%	27,5%
	Cadre et prof. inter.	28,5%	29,0%
	Autre emploi	14,6%	14,7%
	Sans emploi	29,5%	28,7%

TAB. 47 – Effectifs des modalités des variables de calage
Source : EMD web - Lyon 2006

II.3 Des différences socio-économiques disparaissent...

Suite au redressement de l'échantillon face-à-face, les deux populations ne diffèrent plus au regard des variables utilisées pour le redressement. Mais il est intéressant de noter que, suite à ce traitement, des différences initialement constatées sur d'autres variables disparaissent également, comme le montrent les valeurs de la statistique du chi-deux présentées dans l'annexe XI.

Les distributions des individus par genre et par enquête ne font pas apparaître de différence significative (p -value = 84,7%). Le redressement portant notamment sur la catégorie socioprofessionnelle et le niveau d'étude, il n'existe également plus de différence dans la distribution des revenus des deux enquêtes (p -value = 14,3%), constat qui confirme l'existence d'une corrélation forte entre ces trois variables. De même, la distribution des répondants selon les zones du lieu de travail est identique entre les deux populations (p -value = 5,8%). A l'origine, les répondants web occupaient essentiellement des emplois de cadres et d'employés, davantage situés dans le centre de l'agglomération. En redressant l'échantillon web sur la variable 'catégorie socio-professionnelle', nous uniformisons également la répartition des lieux de travail entre les deux échantillons. Enfin, il n'y a plus de différence concernant l'occupation des répondants (p -value = 10,3%). Dans l'enquête web, nous avons 70% d'actifs (majoritairement entre 30 et 60 ans) et peu d'individus dans les classes d'âge extrêmes (scolaires, retraités). Un redressement selon l'âge permet d'effacer ces différences.

Le redressement effectué sur des variables socio-économiques a également impacté certaines variables liées à la mobilité, comme l'équipement automobile des ménages. La motorisation est liée au niveau de vie (donc au revenu), mais également à la composition du ménage. Ces deux variables ayant contribué au redressement de l'échantillon face-à-face, les distributions web et face-à-face du nombre moyen de voitures particulière des ménages auxquels appartiennent les individus sont à présent identiques entre les deux populations (p-value = 71,6%). La conclusion est la même si nous nous intéressons au nombre moyen de voitures disponibles par personne de 18 ans et plus.

Par ailleurs, les pratiques modales individuelles découlent en grande partie des stratégies de localisation résidentielle que les ménages appliquent au fur et à mesure qu'ils construisent leur système de lieux personnels (lieux fréquentés de manière récurrente dans leurs activités quotidiennes). Le choix de la localisation est donc important, car il détermine l'étendue géographique du système de lieux personnels, ainsi que le budget temps consacré aux déplacements les plus fréquents (Flamm, 2003). L'absence de différence en termes de localisation résidentielle des ménages aura probablement un impact sur la mobilité des répondants.

II.4 ...alors que d'autres subsistent

II.4.1 Une présence moindre dans l'annuaire des abonnés France Télécom

En ce qui concerne l'équipement des ménages en téléphonie fixe, les différences sont toujours importantes entre les deux enquêtes (figure 44).

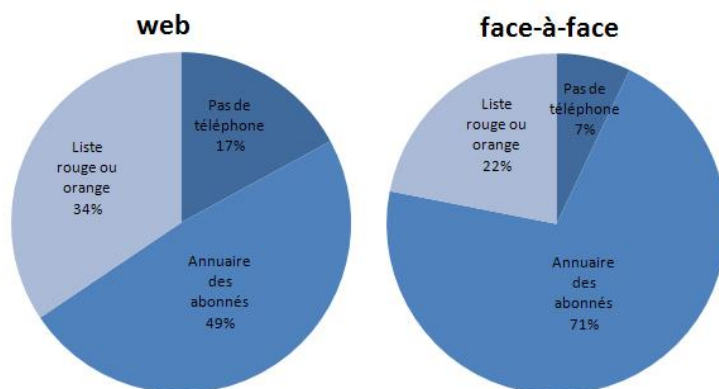


FIG. 44 – Répartition des personnes par enquête et par équipement en téléphone fixe après redressement

Source : EMD web et face-à-face redressée - Lyon 2006

Nous avons créé une variable qui tient compte à la fois de la possession d'un téléphone fixe et de l'inscription le cas échéant sur l'annuaire des abonnés ou sur une liste plus restrictive. Deux constats émergent du tableau. D'abord, La proportion de ménages possédant une ligne de téléphone fixe à domicile est moins élevée dans l'enquête web (82,9%, vs. 92,9% en face-à-face ; p-value

II Application du calage sur marges à l'échantillon en face-à-face

< 0,01%). Ensuite, les répondants web équipés d'une ligne fixe sont davantage inscrits sur une liste rouge ou orange que ceux de l'enquête face-à-face (41,5%, vs. 23,5% ; p-value < 0,01%).

II.4.2 Une propension plus élevée à déclarer leurs revenus

Bien que la distribution des revenus ne diffère plus entre les deux populations, le taux de non-réponse partielle à cette question reste identique : une personne sur quatre dans l'enquête web et une sur trois dans l'enquête face-à-face refusent de déclarer leur niveau de revenu (p-value = 0,3%) (figure 45)

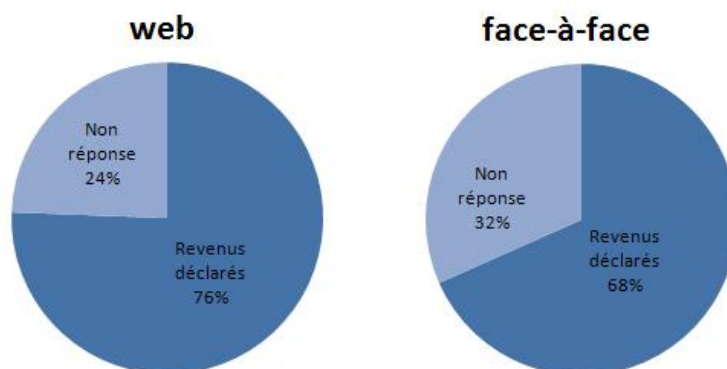


FIG. 45 – Répartition des personnes par enquête et par déclaration du revenu après redressement

Source : EMD web et face-à-face redressée - Lyon 2006

II.4.3 Une plus forte proportion de déplacements le vendredi

De même, la différence en termes de jour de référence utilisé pour le recueil des déplacements persiste après le redressement de l'échantillon face-à-face.

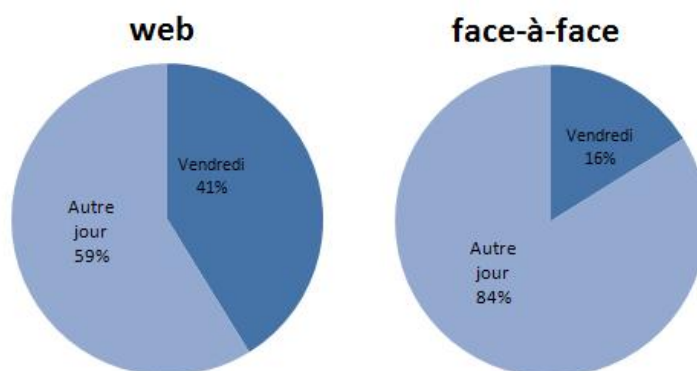


FIG. 46 – Répartition des personnes par enquête et par jour de référence des déplacements après redressement

Source : EMD web et face-à-face redressé - Lyon 2006

Ceci s'explique par la différence de méthodologie qui existe entre les deux enquêtes. Pour le face-à-face, des quotas ont permis d'assurer une répartition

homogène des interviews selon les jours de la semaine. Mais dans le cas du web, le jour de référence se définit comme le jour précédent la première connexion de l'internaute. Nombreux sont ceux qui se sont connectés le week-end, pour prendre le temps de répondre à l'enquête. Le vendredi reste donc sur-représenté dans l'enquête web (41%, vs. 16% en face-à-face ; p-value < 0,01%) (figure 46).

II.4.4 Conclusion sur le redressement de l'échantillon en face-à-face

Les écarts de profil entre les répondants web et face-à-face sont donc fortement réduits. D'une part, grâce aux sept variables sociodémographiques utilisées pour le redressement, d'autre part grâce aux corrélations de ces variables avec d'autres. Ainsi, suite au redressement, des différences initialement constatées sur d'autres variables disparaissent également (revenu, motorisation, genre, activité). Il existe toutefois des facteurs que le redressement ne semble pas avoir impacté, comme la propension à déclarer son niveau de revenu, l'équipement en téléphone fixe, l'inscription sur liste rouge et le jour de référence des déplacements. Ces variables diffèrent fortement selon les populations web et face-à-face, mais ne semblent pas directement liées au nombre de déplacements quotidiens effectués par les individus. Elles ne nécessitent donc pas un redressement, avant l'analyse de la mobilité.

III Des comportements de mobilité spécifiques

Suite au redressement de l'échantillon face-à-face, les différences en termes de caractéristiques socio-économiques sont considérablement réduites entre les deux échantillons, web et face-à-face. L'objectif de cette section est d'évaluer l'impact du redressement sur les comportements de mobilité. Nous comparons ces échantillons au regard des variables caractérisant les déplacements, pour voir dans quelle mesure les différences initialement constatées évoluent.

III.1 Utilisation des modes de transport

Nous revenons d'abord sur les habitudes en termes de mode de transport des répondants. Ces questions sont essentielles, car l'enquête ménages déplacements ne porte que sur une seule journée. Or, une journée de la vie d'une personne n'est pas représentative de ses pratiques, puisque beaucoup de modes n'ont qu'un usage occasionnel (Orfeuill, 2002a).

III.1.1 Une utilisation très fréquente de la voiture en tant que conducteur

Nous observons des différences importantes entre les deux enquêtes : près de 7 personnes sur 10 interrogées sur le web utilisent chaque jour la voiture en tant que conducteur pour se déplacer en semaine, vs. 59,4% en face-à-face (p-value = 0,1%). A contrario, 4,5% n'utilisent jamais la voiture particulière pour leurs déplacements (vs. 14,6% dans l'enquête en face-à-face (p-value <

III Des comportements de mobilité spécifiques

0,01%) (figure 47). Cette différence est toutefois moins marquée qu'avant le redressement (44,2% sur le web et 32% en face-à-face). Les effectifs de l'enquête web sont cependant parfois faibles pour pouvoir conclure.

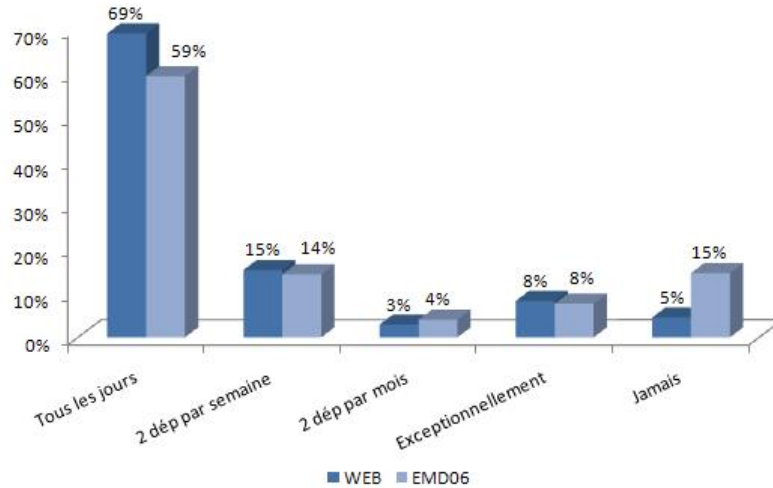


FIG. 47 – Répartition des personnes par enquête et par fréquence d'utilisation de la voiture conducteur en semaine après redressement

Source : EMD web et face-à-face redressée - Lyon 2006

Les disparités d'usage de la voiture particulière entre les deux populations après redressement traduisent probablement des choix de modes de vie différents (Orfeuil, 2000).

III.1.2 Mais une utilisation moins fréquente de la voiture en tant que passager

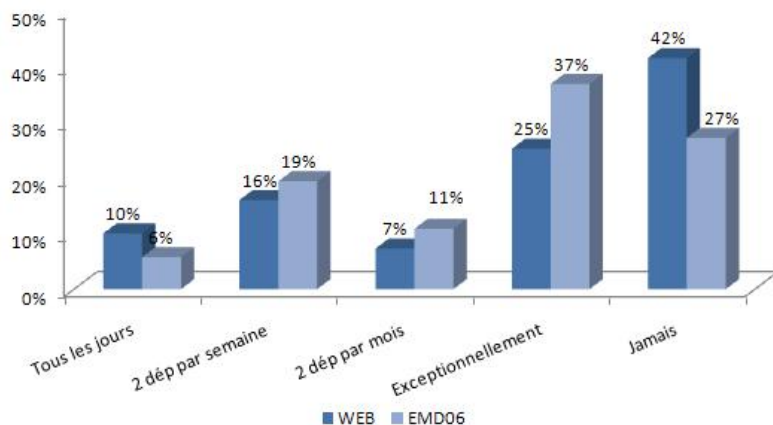


FIG. 48 – Répartition des personnes par enquête et par fréquence d'utilisation de la voiture passager en semaine après redressement

Source : EMD web et face-à-face redressée - Lyon 2006

41,5% des personnes interrogées sur le web n'utilisent jamais la voiture particulière en tant que passager pour se déplacer en semaine, vs. 27,2% dans l'enquête en face-à-face (p-value < 0,01%) (figure 48). Cette proportion est plus importante qu'avant le redressement de l'échantillon face-à-face (22,6%). En revanche, moins de personnes utilisent chaque jour ce mode de transport pour se déplacer en semaine, quel que soit le mode d'enquête (10,0% sur le web, vs. 5,7% en face-à-face ; p-value < 0,01%) (9,9% avant le redressement). Il peut s'agir d'une part résiduelle de personnes qui ne possèdent pas leur permis, ou qui sont adeptes du covoiturage.

III.1.3 Un usage plus modéré des transports collectifs

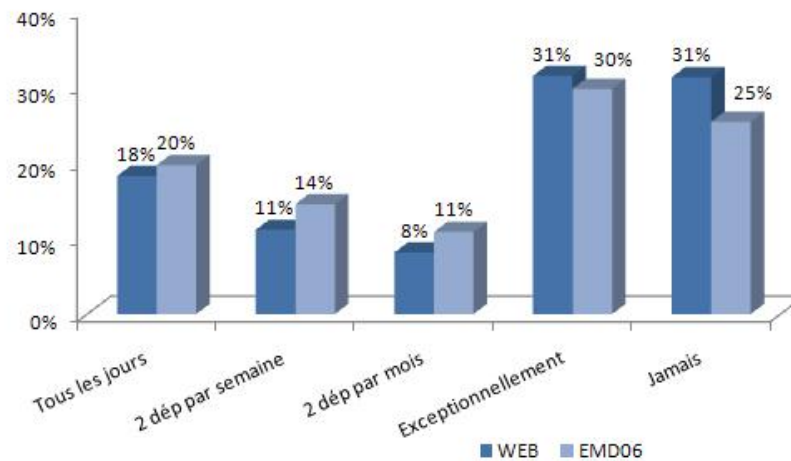


FIG. 49 – Répartition des personnes par enquête et par fréquence d'utilisation des transports collectifs en semaine après redressement

Source : EMD web et face-à-face redressée - Lyon 2006

Les répondants à l'enquête en face-à-face utilisent plus fréquemment les transports collectifs pour leurs déplacements quotidiens en semaine (19,6%) que les répondants à l'enquête web (18,2%) (figure 49). Cette différence n'est toutefois pas significative (risque de première espèce = 8,02%). En revanche, la proportion de personnes n'utilisant jamais les transports en commun pour se déplacer en semaine (31,2% dans l'enquête web et 25,4% pour l'enquête en face-à-face) diffère de façon significative entre les deux enquêtes (p-value = 1,2%). De manière générale, peu de changements sont constatés après le redressement.

III.1.4 Une utilisation des deux-roues peu fréquente, mais supérieure à celle de l'enquête en face-à-face

La proportion des répondants qui n'utilisent jamais les deux-roues pour se déplacer en semaine est plus importante dans l'enquête en face-à-face (95,8%,

III Des comportements de mobilité spécifiques

vs. 76,8% dans l'enquête web ; p-value < 0,01%) (figure 50). Si cette différence est statistiquement significative, attention toutefois aux faibles effectifs de l'enquête web. Après redressement, davantage de personnes utilisent les deux-roues, bien que cette part soit encore marginale.

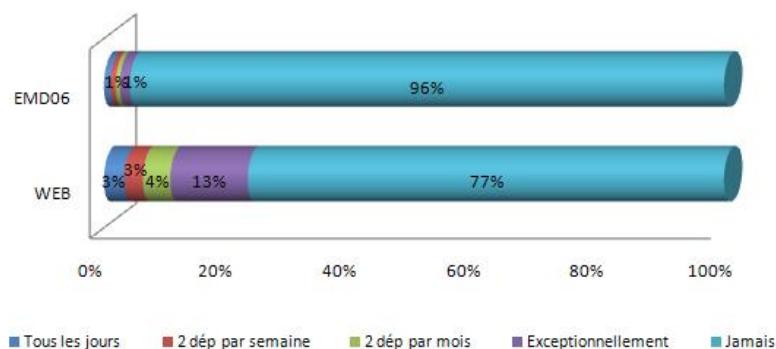


FIG. 50 – Répartition des personnes par enquête et par fréquence d'utilisation des deux-roues en semaine après redressement

Source : EMD web et face-à-face redressée - Lyon 2006

III.2 Des internautes moins mobiles

En comparant les deux enquêtes après redressement, il est intéressant de constater que la part des répondants ne s'étant pas déplacés dans l'aire d'étude durant la période de référence est beaucoup plus élevée dans l'enquête web (19% vs. 7,4% dans l'enquête en face-à-face ; p-value < 0,1%) (figure 51).

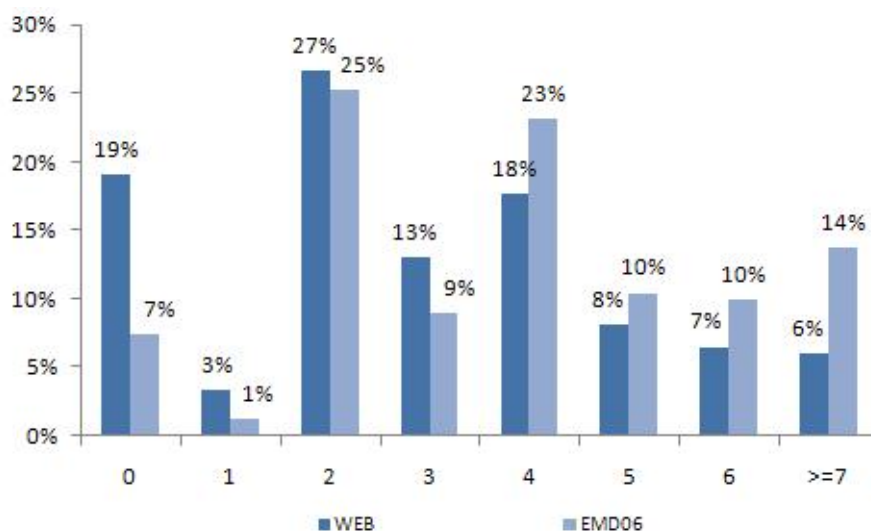


FIG. 51 – Répartition des individus par nombre de déplacements et par enquête après redressement

Source : EMD web et face-à-face redressée - Lyon 2006

Cet écart est plus important qu'avant le redressement, où la part des immobiles dans l'échantillon face-à-face était de 11%. Inversement, la part des individus ayant effectué au moins 7 déplacements est faible parmi notre échantillon d'internautes (6% vs. 13,7% dans l'enquête en face-à-face ; p-value < 0,1%). Cet écart semble statistiquement significatif, mais les faibles effectifs de l'enquête web ne nous permettent pas de conclure.

Nous recensons en moyenne 4,02 déplacements par personne interrogée en face-à-face, contre seulement 3,00 par internaute (tableau 48). Cette différence significative (p-value < 0,01%) peut s'expliquer par le mode d'administration du questionnaire, l'enquêteur ayant pour fonction de relancer les personnes et de vérifier la pertinence du schéma de déplacements. Ainsi, il est probable que des 'petits' déplacements aient été omis par les actifs répondant en ligne. D'autres ont pu juger certains déplacements inutiles pour l'étude et ne pas les saisir. L'écart s'est creusé par rapport à la population totale (3,00 vs. 3,63 déplacements), les répondants de l'échantillon web étant majoritairement constitué d'actifs, qui se déplacent davantage.

Enquête	Ech.	Obs.	Minimum	Maximum	Moyenne	Ecart-type
Web	Total	369	0	15	3,00	2,32
	Mobiles	299	1	5	3,67	2,02
Face-à-face	Total	13 271	0	23	4,04	2,60
	Mobiles	12 291	1	10	4,53	2,39

TAB. 48 – Nombre moyen de déplacements par enquête après redressement
Source : EMD web et face-à-face redressée - Lyon 2006

La différence entre les deux échantillons est plus faible lorsque nous considérons uniquement les personnes mobiles (3,67, vs. 4,34 déplacements), la proportion de personnes n'ayant pas déclaré de déplacements sur la période de référence étant beaucoup plus importante dans l'enquête web. L'écart s'est creusé par rapport à la population totale (3,71, vs. 4,08 déplacements).

III.3 Une mobilité plus faible à pied

Les déplacements à pied sont largement sous-représentés dans l'enquête web (15,1% vs. 23,7% en face-à-face ; p-value < 0,1%) (figure 52). Ce sont de petits déplacements, probablement moins bien enregistrés par les internautes. Inversement, la part des déplacements en modes motorisés est supérieure chez les internautes (69%, vs. 63,2% en face-à-face ; p-value < 0,1%). Cet écart s'explique essentiellement par les déplacements effectués en voiture passager (9,1% sur le web vs. 6,8% en face-à-face ; p-value = 0,1%), la proportion de déplacements en voiture particulière en tant que conducteur étant peu différente entre les deux enquêtes (59,9% sur le web vs. 56,4% en face-à-face ; p-value = 9,8%). Après le redressement, on observe pour l'enquête en face-à-face un rattrapage des déplacements en mode mécanisé (de 44% à 55% pour la voiture conducteur et de 6% à 10% pour la voiture passager), au détriment de la marche à pied (24% vs. 29%) et des transports en commun (11% vs. 14%).

III Des comportements de mobilité spécifiques

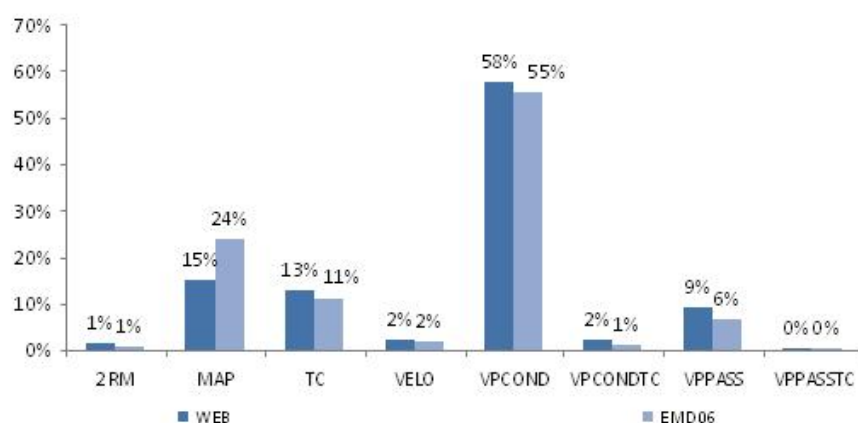


FIG. 52 – Répartition des déplacements des individus par mode et par enquête après redressement

Source : EMD web et face-à-face redressée - Lyon 2006

Il est également intéressant de comparer le nombre de déplacements par mode, pour voir si les différences observées entre les deux populations subsistent. Les résultats sont présentés pour l'ensemble des actifs, puis pour les actifs mobiles, c'est-à-dire ceux s'étant déplacés dans l'aire d'étude pendant la période de référence (tableau 49). La proportion d'immobiles étant nettement supérieure parmi les internautes, la différence de mobilité se réduit lorsque l'analyse se limite aux mobiles. Elle subsiste toutefois de manière très marquée pour la marche à pied. Après redressement, les internautes se déplacent deux fois moins à pied que les individus de l'échantillon face-à-face. On observe également une utilisation un peu moins soutenue de la voiture en tant que conducteur. Pour les autres modes, les résultats sont très proches, même si les effectifs sont parfois très limités.

	Enquête web		Enquête face-à-face	
	Total	Mobiles	Total	Mobiles
Deux-roues	0,04	0,05	0,02	0,02
Marche à pied	0,44	0,56	0,98	1,10
Transports en commun	0,38	0,47	0,45	0,50
Vélo	0,06	0,07	0,07	0,07
Voiture conducteur	1,74	2,14	2,21	2,49
Voiture conducteur + Transports en commun	0,06	0,08	0,03	0,04
Voiture passager	0,27	0,33	0,27	0,30
Voiture passager + Transports en commun	0,01	0,01	0,01	0,01
Autre	0	0	0	0
Total	3,00	3,71	4,04	4,53

TAB. 49 – Nombre moyen de déplacements par mode et par enquête après redressement

Source : EMD web et face-à-face redressée - Lyon 2006

Après redressement, le nombre de déplacements effectués en voiture particulière augmente substantiellement, au détriment de ceux effectués à pied, en transport en commun ou en tant que passager. Cet écart peut s'expliquer par le niveau d'éducation et d'emploi plus élevé des répondants et par leur accès plus large au permis de conduire.

III.4 Motifs de déplacements : moins d'accompagnements et de loisirs

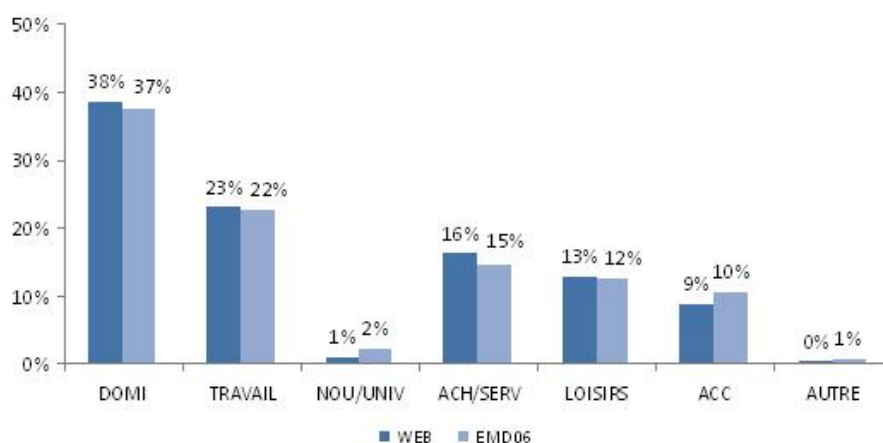


FIG. 53 – Répartition des déplacements des individus par motif à la destination et par enquête après redressement

Source : EMD web et face-à-face redressée - Lyon 2006

Les deux distributions sont relativement proches, bien que non équivalentes (risque de première espèce = 3,8%) (figure 53).

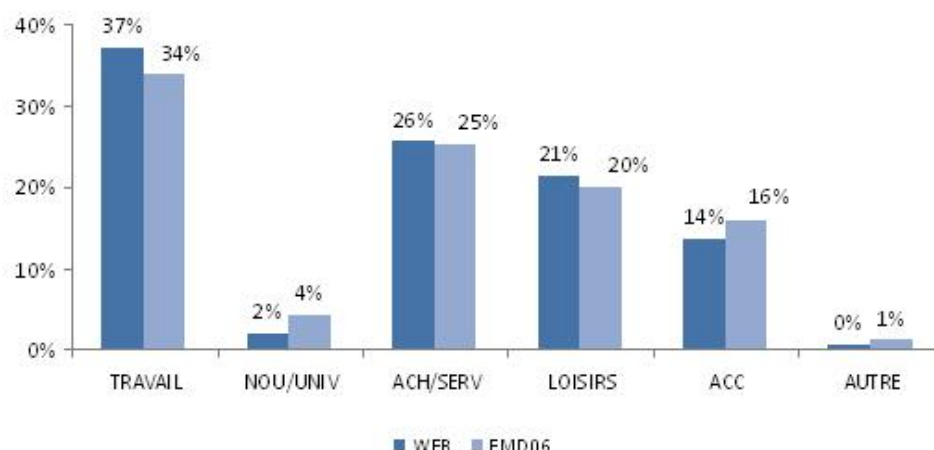


FIG. 54 – Répartition des déplacements des individus par motif à l'origine et par enquête après redressement lorsque le motif à la destination est le domicile

Source : EMD web et face-à-face redressée - Lyon 2006

III Des comportements de mobilité spécifiques

La conclusion est plus radicale si on regarde le motif à l'origine des déplacements qui ont pour motif à la destination le domicile, puisqu'il n'y a pas de différence significative entre les deux échantillons (p -value = 9,2%)(figure 54).

Si on analyse les données, non plus en répartition par motif, mais en nombre de déplacements par motif, on arrive à une conclusion un peu différente (tableau 50). La mobilité des mobiles est un peu plus faible pour les internautes, par rapport aux personnes interrogées en face-à-face. L'écart concerne tous les motifs, même s'il est plus prononcé pour les loisirs et l'accompagnement. Il concerne donc surtout de petits déplacements ou des déplacements que les internautes considèrent probablement comme moins importants.

	Enquête web		Enquête face-à-face	
	Total	Mobiles	Total	Mobiles
Domicile	1,14	1,43	1,53	1,71
Travail	0,69	0,85	0,87	0,98
Collège-Lycée-Université	0,03	0,03	0,09	0,10
Achat-Service	0,49	0,60	0,60	0,68
Accompagnement	0,26	0,32	0,41	0,46
Loisirs	0,38	0,47	0,51	0,57
Autre	0,01	0,01	0,03	0,03
Total	3,00	3,71	4,04	4,53

TAB. 50 – Nombre moyen de déplacements par motif à la destination et par enquête après redressement

Source : EMD web et face-à-face redressée - Lyon 2006

On observe un rééquilibrage après redressement, entre les déplacements pour motif travail (en augmentation : +0,31 déplacements pour l'ensemble de la population et +0,35 pour les mobiles) et ceux pour motif étude (en nette diminution : -0,13 déplacements pour l'ensemble de la population et -0,15 pour les mobiles), qui s'explique essentiellement par le redressement de l'âge des individus (plus d'actifs, moins de jeunes). Les déplacements pour motif accompagnement sont plus élevés après redressement (+0,08 déplacements pour l'ensemble de la population et les mobiles). Cela peut s'expliquer en partie par l'augmentation de la taille des ménages, donc la présence d'enfants, et par la proportion de personnes possédant le permis de conduire.

III.5 Pointe horaire du matin plus marquée et du soir plus tardive

Globalement les déplacements sont plus tardifs dans l'enquête web. Ceux débutant entre 14h00 et 17h00 sont sous-représentés dans l'enquête web (14,4%, vs. 18,4% en face-à-face ; p -value < 0,01%) (figure 55). Ce constat peut s'expliquer par le redressement selon le type d'emploi ¹⁰⁸. Les déplacements débutant

¹⁰⁸Avant le redressement, 21% des déplacements de l'enquête en face-à-face débutaient entre 14h00 et 17h00.

entre 19h00 et 22h00 sont en revanche sur-représentés parmi la population d'internautes (13,5% vs. 10,6% en face-à-face ; p-value = 0,3%).

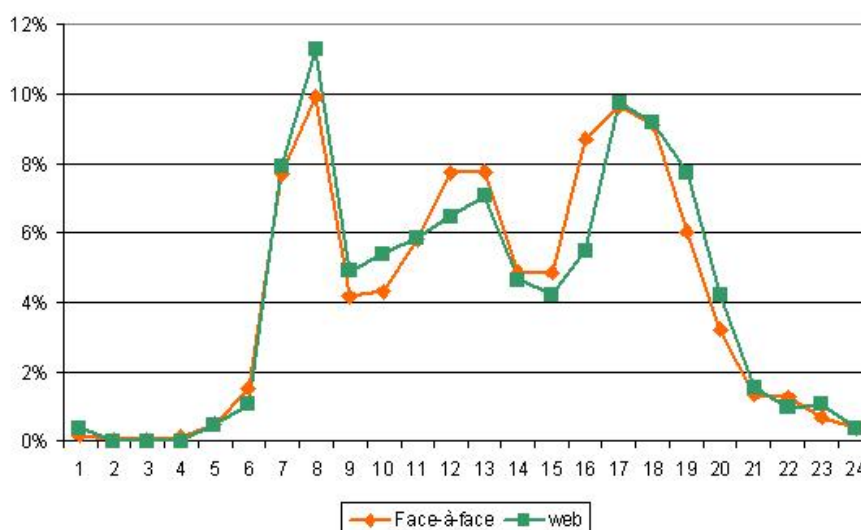


FIG. 55 – Répartition des déplacements des individus par heure de début et par enquête après redressement

Source : EMD web et face-à-face redressée - Lyon 2006

III.6 Des déplacements plus longs en distance et en temps

La distance moyenne des déplacements effectués par les individus après redressement est légèrement supérieure dans l'enquête web (4,61 km, vs. 4,08 km en face-à-face ; p-value < 0,001%)¹⁰⁹ (tableau 51). La dispersion des valeurs autour de la moyenne est en revanche plus importante dans l'enquête en face-à-face. L'écart s'est toutefois réduit après le redressement de l'enquête en face-à-face (4,61 km, vs. 3,78 km), car les déplacements domicile-travail générés par les actifs augmentent la moyenne des déplacements quotidiens.

De nombreuses caractéristiques personnelles influent sur la distance parcourue par les actifs pour se rendre à leur travail et le temps qu'ils y consacrent. La profession exercée est l'une des variables les plus discriminantes (Baudelle *et al.*, 2003). Nous avons redressé l'échantillon face-à-face sur la catégorie socio-professionnelle, mais il peut subsister des différences en termes de profession exercée.

Distance (km)	Obs.	Minimum	Maximum	Moyenne	Ecart type
Enquête web	1 066	0,02	40,37	4,61	4,73
Enquête face-à-face	52 354	0,06	66,32	4,08	5,24

TAB. 51 – Distance moyenne des déplacements par enquête après redressement
Source : EMD web et face-à-face redressée - Lyon 2006

¹⁰⁹Seuls les déplacements ayant leur origine et leur destination dans l'aire d'étude sont comptabilisés.

III Des comportements de mobilité spécifiques

La durée moyenne des déplacements est supérieure dans l'enquête web (22,96 mn, vs. 19,20 mn en face-à-face; p-value < 0,01%). Le niveau de dispersion des valeurs autour de la moyenne est en revanche similaire entre les deux enquêtes (tableau 52). Cet écart reste du même ordre que celui observé avant redressement (22,96 mn, vs. 19,25 mn).

Durée (mn)	Obs.	Minimum	Maximum	Moyenne	Écart type
Enquête web	1 066	0,35	250	22,96	21,15
Enquête face-à-face	52 354	1	450	19,20	20,22

TAB. 52 – Durée moyenne des déplacements par enquête après redressement
Source : EMD web et face-à-face redressée - Lyon 2006

L'analyse des budgets temps (tableau 53) met en évidence une durée quotidienne consacrée aux déplacements quasi identique entre les deux échantillons après redressement (environ 87 mn). En revanche, la distance quotidienne parcourue semble être légèrement plus faible chez les internautes (17,33 km, vs. 18,45 km).

	Obs.	Budget distance (km)	Budget temps (mn)
Enquête web	277	17,33	87,23
Enquête face-à-face	11 621	18,45	86,88

TAB. 53 – Budgets temps et distances par enquête et par personne après redressement

Source : EMD web et face-à-face redressée - Lyon 2006

III.7 Des journées plus longues pour les actifs

La durée quotidienne passée sur le lieu de travail est plus importante pour les répondants web (20 mn en moyenne) après le redressement de l'échantillon face-à-face. Par ailleurs, le nombre moyen de sorties pour motif travail des personnes actives est légèrement inférieur dans l'enquête web (1,12 vs. 1,17) (tableau 54).

	Obs.	Nb de sorties 'Travail'	Budget temps de travail
Enquête web	179	1,12	8h15mn
Enquête face-à-face	7 366	1,17	7h55mn

TAB. 54 – Nombre de sorties pour motif 'Travail' et budget temps de travail moyen par actif et par enquête après redressement

Source : EMD web et face-à-face redressée - Lyon 2006

Ces résultats appuient les conclusions précédentes sur la plus forte présence hors domicile des internautes et leur nombre plus limité de déplacements et de sorties. Ils peuvent s'expliquer par les différences de type d'emplois qui peuvent subsister entre les échantillons après le redressement. En effet, nous avons redressé l'échantillon face-à-face sur les critères 'cadres et professions

intellectuelles supérieures' et 'employés'. Mais ces catégories recouvrent des professions très diverses. Les opportunités de retours au domicile pour le déjeuner sont réduites et peu compatibles avec le niveau de responsabilité des actifs internautes. Enfin, le temps supplémentaire passé sur le lieu de travail par les internautes pèse sur la réalisation de déplacements en fin de soirée.

III.8 Un nombre de sorties du domicile plus faible

Le nombre de sorties du domicile est sensiblement plus faible pour les répondants de l'enquête web (62,9% ne réalisent qu'une seule sortie, vs. 54,6% dans l'enquête face-à-face ; p-value = 0,5%) (figure 56).

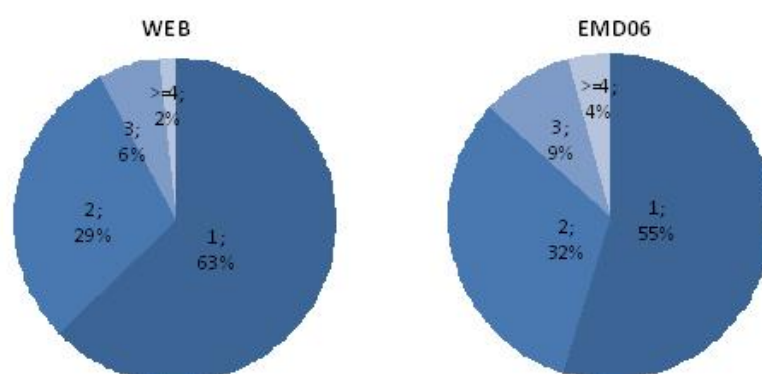


FIG. 56 – Répartition des individus par nombre de sorties du domicile et par enquête après redressement

Source : EMD web et face-à-face redressée - Lyon 2006

Le nombre moyen de sorties effectuées après redressement de l'échantillon face-à-face est nettement inférieur dans l'enquête web (1,19 vs. 1,53). Le niveau de dispersion des valeurs autour de la moyenne est également plus faible pour les internautes (0,87 vs. 0,97) (tableau 55). Si nous considérons l'ensemble des individus, l'écart est plus important que celui observé avant le redressement de l'échantillon en face-à-face (1,19 vs. 1,47), car les actifs sont plus mobiles. En revanche, pour les mobiles, le redressement n'affecte pas le nombre moyen de sorties. Deux hypothèses sont plausibles. Soit les internautes ont oublié de saisir certaines sorties du domicile, soit la différence est imputable au type d'emploi.

		Obs.	Minimum	Maximum	Moyenne	Ecart-type
Enquête web	Total	369	0	5	1,19	0,87
	Mobiles	299	1	5	1,47	0,72
Enquête face-à-face	Total	3 271	0	10	1,53	0,97
	Mobiles	12 291	1	10	1,65	0,90

TAB. 55 – Nombre moyen de sorties par enquête après redressement

Source : EMD web et face-à-face redressée - Lyon 2006

Une analyse plus fine montre que le nombre de déplacements par sortie n'est plus identique entre les deux échantillons, après le redressement (2,47 déplacements par sortie sur le web, vs. 2,63 déplacements par sortie en face-à-face). Les résultats sont identiques, suivant que nous considérons l'ensemble de la population ou uniquement les mobiles.

III.9 Des sorties plus longues en durée et en distance

La durée moyenne des sorties est supérieure dans l'enquête web (63,90 mn, vs. 54,94 mn dans l'enquête face-à-face ; p-value < 0,001%). Le niveau de dispersion des valeurs autour de la moyenne est en revanche similaire dans les deux enquêtes (environ 57 mn) (tableau 56). L'écart s'est réduit après le redressement de l'échantillon en face-à-face (63,90 mn, vs. 50,17 mn), les personnes actives avec un haut niveau d'études étant davantage représentées.

Durée (mn)	Obs.	Minimum	Maximum	Moyenne	Ecart type
Enquête web	436	6	570	63,90	56,97
Enquête face-à-face	20 282	1	900	54,94	56,22

TAB. 56 – Durée moyenne des sorties par enquête après redressement
Source : EMD web et face-à-face redressée - Lyon 2006

La distance moyenne des sorties effectuées par les actifs est supérieure dans l'enquête web (11,16 km, vs. 10,51 km en face-à-face ; p-value < 0,01%). Le niveau de dispersion des valeurs autour de la moyenne est en revanche plus important pour l'enquête en face-à-face (11,50 km pour le web, vs. 13,02 km pour le face-à-face) (tableau 57). L'écart est plus important après le redressement de l'enquête face-à-face (10,51 km, vs. 9,11 km), les déplacements pour motif 'Travail', plus nombreux pouvant impacter la distance moyenne quotidienne parcourue.

Distance (km)	Obs.	Minimum	Maximum	Moyenne	Ecart type
Enquête web	439	0,02	75,98	11,16	11,50
Enquête face-à-face	20 282	0,06	183,22	10,51	13,02

TAB. 57 – Distance moyenne des sorties par enquête après redressement
Source : EMD web et face-à-face redressée - Lyon 2006

III.10 Premier mode des sorties du domicile : une sous-estimation des sorties à pied chez les internautes

Les sorties du domicile à pied sont largement sous-représentées dans l'enquête web (15,3% vs. 22,3% en face-à-face ; p-value < 0,01%) (figure 57).

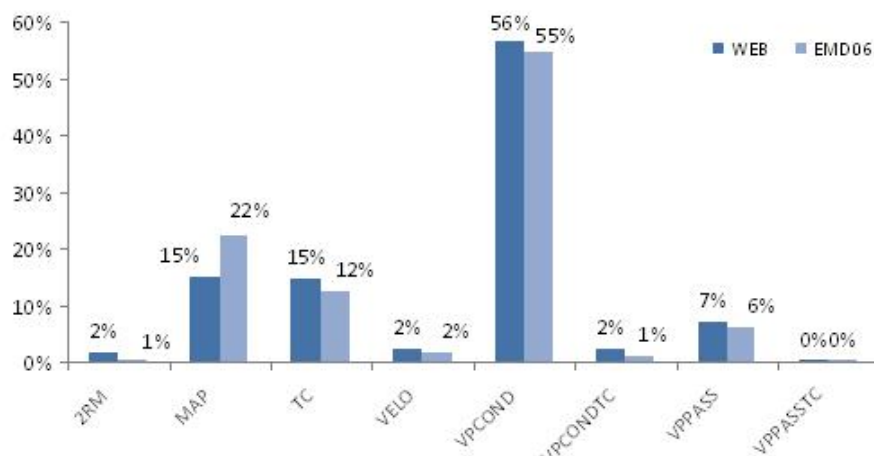


FIG. 57 – Répartition des sorties du domicile par 1er mode et par enquête après redressement

Source : EMD web et face-à-face redressée - Lyon 2006

Les déplacements à pied sont de petits déplacements, probablement moins bien enregistrés par les internautes. En revanche, les modes motorisés semblent davantage représentés parmi les internautes (66,1% vs. 62,8% en face-à-face), bien que cette différence ne soit pas significative (p-value = 15,7%). Après le redressement de l'échantillon face-à-face, nous observons un rattrapage des sorties en voiture conducteur, au détriment de celles réalisées à pied et en transports collectifs. Ce rattrapage s'explique par un meilleur accès au permis de conduire et un niveau d'études et de revenus supérieurs après le redressement.

Il est également intéressant de comparer le nombre de sorties du domicile par 1er mode, pour voir si les différences observées entre les deux populations subsistent. Les résultats sont présentés pour l'ensemble des actifs, puis pour les actifs mobiles, c'est-à-dire ceux s'étant déplacés dans l'aire d'étude pendant la période de référence (tableau 58).

L'écart entre le nombre de sorties des deux échantillons, calculé sur les seuls mobiles, se réduit. Il s'explique par un nombre de sorties beaucoup plus faible pour la marche à pied, mais également un peu plus faible pour la voiture en tant que conducteur. En revanche, pour les autres modes, les mobilités sont très proches, même si elles sont faibles pour les deux modes d'enquête. Il est probable que les sorties à pied qui sont souvent courtes en distance et fréquemment en temps soient sous-estimées dans l'enquête web. Il est plus délicat de conclure quant aux sorties voiture dans la mesure où il peut aussi bien y avoir sous-estimation que des différences de comportements liés aux différences socio-économiques, mais aussi à des retours plus tardifs le soir et à

III Des comportements de mobilité spécifiques

des distances domicile-travail plus longues, qui peuvent limiter la mobilité des personnes concernées.

	Enquête web		Enquête face-à-face	
	Total	Mobiles	Total	Mobiles
Deux-roues	0,02	0,02	0,01	0,01
Marche à pied	0,18	0,22	0,34	0,37
Transports en commun	0,18	0,22	0,19	0,21
Vélo	0,03	0,03	0,03	0,03
Voiture conducteur	0,67	0,83	0,84	0,90
Voiture conducteur + Transports en commun	0,03	0,03	0,02	0,02
Voiture passager	0,08	0,10	0,10	0,10
Voiture passager + Transports en commun	0	0	0,01	0,01
Autre	0	0	0	0
Total	1,19	1,47	1,53	1,65

TAB. 58 – Nombre moyen de sorties du domicile par 1er mode et par enquête après redressement

Source : EMD web et face-à-face redressée - Lyon 2006

III.11 Motif principal des sorties du domicile : une prépondérance du motif 'travail'

La part des sorties du domicile ayant pour motif principal le travail est plus importante pour les répondants de l'enquête web (46,7% vs. 44,1% dans l'enquête face-à-face) (figure 58), bien que la différence ne soit pas significative (p-value = 28,5%).

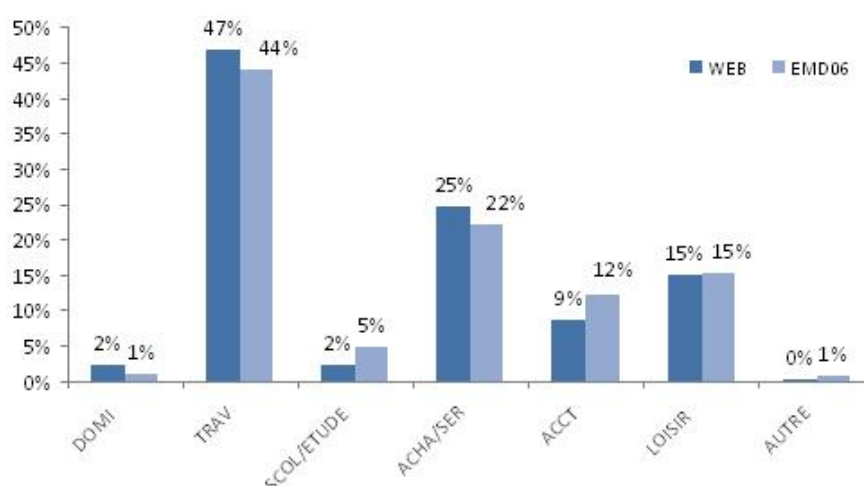


FIG. 58 – Répartition des sorties du domicile par motif principal et par enquête après redressement

Source : EMD web et face-à-face redressée - Lyon 2006

A contrario, la part des sorties du domicile ayant pour motif l'accompagnement est plus faible dans l'enquête web (8,7% vs. 12,2% en face-à-face; p-value = 2,5%). Les effectifs de l'enquête web sont toutefois relativement restreints, pour pouvoir conclure sur ce point. Les écarts se sont resserrés après le redressement de l'échantillon en face-à-face, puisque l'on observe davantage de déplacements pour motif 'travail' et moins de déplacements pour motif 'Scolaire-Etudes'. Ceci s'explique principalement par le redressement effectué sur la variable âge.

L'analyse du nombre de sorties du domicile par motif à la destination (tableau 3.38), pour l'ensemble des actifs puis pour les actifs mobiles, c'est-à-dire ceux s'étant déplacés dans l'aire d'étude pendant la période de référence, conduit globalement aux mêmes conclusions (tableau 59).

	Enquête web		Enquête face-à-face	
	Total	Mobiles	Total	Mobiles
Domicile	0,03	0,03	0,02	0,02
Travail	0,56	0,69	0,67	0,73
Scolaire-Etudes	0,03	0,03	0,07	0,08
Achat-Service	0,30	0,36	0,34	0,36
Accompagnement	0,10	0,13	0,19	0,20
Loisirs	0,18	0,22	0,23	0,25
Autre	0	0	0,01	0,01
Total	1,19	1,47	1,53	1,65

TAB. 59 – Nombre moyen de sorties par motif à la destination et par enquête après redressement

Source : EMD web et face-à-face redressée - Lyon 2006

IV Conclusion

En redressant l'échantillon face à face à partir des caractéristiques socio-économiques les plus discriminantes entre les deux échantillons web et face-à-face, on cherche à obtenir des populations dont les caractéristiques soient relativement proches et à conserver des effectifs suffisants pour avoir des résultats statistiquement significatifs. Nous avons montré que ces populations sont effectivement proches mais qu'il subsiste certaines différences, notamment en termes de jour de référence des déplacements, de présence dans l'annuaire, et de propension à déclarer les revenus du ménage. Ces différences sont encore susceptibles d'expliquer des différences de mobilité, mais leur incidence devrait être limitée (ces variables ne ressortent pas comme explicatives dans l'analyse de la mobilité).

Après le redressement, les différences en termes de nombre de déplacements ou de sorties subsistent et ont même tendance à s'amplifier un peu. Cet écart s'explique à la fois par une plus forte immobilité des internautes et par une moindre déclaration de déplacements et de sorties. L'utilisation de la voiture n'est plus supérieure dans l'enquête web par rapport à l'enquête en face à face, elle devient même un peu plus faible. En revanche, la marche à pied est toujours

autant délaissée et l'usage des autres modes similaire dans les deux enquêtes. Enfin, le poids accru du travail s'estompe, les internautes ayant une mobilité en termes de déplacements ou de sorties proches de celle des répondants à l'enquête standard. L'écart de mobilité se porte donc sur les autres motifs et tout particulièrement sur les accompagnements et les loisirs. Les déplacements et les sorties restent beaucoup plus longues en termes de temps et de distance, très probablement du fait du poids très réduit des motifs accompagnements et loisirs et du très faible nombre de sorties marche à pied.

Deux interprétations de ces résultats sont possibles. La première conduit à une sous-estimation de la mobilité imputable au média web utilisé. Cette interprétation peut être étayée par le fait que la sous-estimation concerne surtout des sorties courtes tant en temps qu'en distance principalement faite à pied pour des motifs comme les loisirs ou l'accompagnement qui peuvent être perçus comme moins importantes par les internautes. Le manque d'ergonomie du questionnaire en ligne et la lourdeur de la saisie des déplacements peuvent alors inciter à cette omission. On peut toutefois objecter que les différences socio-économiques ne sont peut être pas complètement occultées par le redressement de l'échantillon face-à-face. En premier lieu, les catégories de PCS ou de revenu sont trop agrégées (il est possible que les PCS++ et les 'très' hauts revenus soient davantage présents dans la population web) et le redressement ne corrige donc peut être pas toutes les différences. Ensuite, la pointe du soir reste plus tardive pour l'échantillon web par rapport à l'échantillon face-à-face redressé. Cela traduit une présence accrue sur le lieu de travail qui limite peut être la réalisation d'autres activités moins contraintes. Ces différences pourraient expliquer au moins en partie les différences de mobilité, notamment de marche à pied. Si cette hypothèse est avérée, l'enquête web permettrait d'enquêter des individus qui échappent à l'enquête face-à-face. Pour corroborer cette interprétation, on peut s'appuyer sur l'hypothèse de Zahavi de constance des budgets temps de déplacements (Zahavi, 1979). Les internautes ont un budget temps de déplacements équivalent à celui des répondants en face-à-face redressé. L'application de la conjecture de Zahavi ferait qu'ils ne pourraient effectuer davantage d'activités et donc de déplacements sans augmenter leur budget temps de déplacement. Ce n'est évidemment pas une démonstration tant la conjecture de Zahavi a été critiquée ¹¹⁰, tout au plus cette hypothèse n'est pas contradictoire avec nos observations.

L'analyse des données ne permet pas de conclure formellement entre ces différentes hypothèses. Il nous semble que chacune d'elles explique très probablement une partie du différentiel de mobilité observé. La combinaison de modes d'enquête permet donc probablement d'enquêter des personnes répondant peu en face-à-face, mais elle pose aussi le problème de la comparabilité des données. Si cette question a retenu l'attention dans certaines disciplines, elle ne semble pas avoir été encore largement développée dans le domaine des

¹¹⁰We conclude (with prior researchers) that travel time expenditures are not constant except, perhaps, at the most aggregate level' (Mokhtarian et Chen, 2004).

transports. Pour aller plus loin dans l'analyse, il est nécessaire de recourir à des techniques économétriques empruntées au domaine des variables qualitatives. Le modèle de sélection de l'échantillon, dont il faut estimer les paramètres à l'aide de la procédure en deux étapes élaborée par Heckman (1979) nous semble une piste prometteuse car elle permet par exemple d'isoler l'effet des différences sociodémographiques de celui du mode d'enquête sur la mesure de la mobilité quotidienne.

Conclusion de la partie II

Nous avons vu que les internautes ont des caractéristiques sociodémographiques différentes du reste de la population. Ainsi, ils possèdent un niveau d'études et de revenus élevés, sont plutôt actifs, fortement motorisés et équipés en moyens de communication (téléphone portable, connexion internet à domicile...), qui subsistent même si nous redressons l'échantillon face-à-face pour le rendre comparable à l'échantillon web. En posant l'hypothèse que ces caractéristiques impactent les comportements de mobilité, l'utilisation du web comme média d'enquête serait un moyen d'augmenter la qualité des données.

Les internautes se déplacent moins que les individus répondant en face à face ; cette moindre mobilité concerne tout à la fois les déplacements et les sorties et subsiste même lorsqu'on limite les différences socio-économiques entre les deux échantillons. Cet écart s'explique par un double effet : une immobilité plus grande des internautes et une moindre déclaration de déplacements. Une analyse par mode ou motif montre que le différentiel s'explique principalement par une mobilité marche à pied et en voiture particulière plus faible, et par une participation aux activités d'accompagnements ou de loisirs nettement moins élevée. Ces données sont cohérentes avec l'hypothèse d'une sous-déclaration imputable au média web dans la mesure où l'on sait que les risques d'omission de déclaration de déplacements concernent surtout les déplacements ou sorties courts en temps ou en distance et des motifs moins contraints. On peut toutefois objecter que les différences socio-économiques peuvent expliquer au moins en partie les différences de mobilité.

L'analyse des données ne permet pas de conclure formellement entre ces hypothèses. Il nous semble que chacune d'elles explique très probablement une partie du différentiel de mobilité observé. La combinaison de modes d'enquête pose ainsi le problème de la comparabilité des données. Si cette question a retenu l'attention dans certaines disciplines, elle ne semble pas avoir été encore largement développée dans le domaine des transports. Pour aller plus loin dans l'analyse, il est nécessaire de recourir à des techniques économétriques permettant par exemple d'isoler et de quantifier l'effet des différences socio-économiques de celui du mode d'enquête sur la mobilité quotidienne.

Troisième partie

La comparabilité des données : enjeux et méthodes de correction

Introduction de la partie III

Si le fait de proposer des médias différents permet d'augmenter le taux de réponse global, la comparabilité des données reste un exercice difficile. La comparaison du nombre moyen de déplacements quotidiens déclaré par les répondants montre que les internautes se déplacent moins que les individus interrogés en face à face. L'hypothèse d'une sous-déclaration imputable au média web est tentante, mais il est également possible que les différences socio-économiques observées entre les deux échantillons expliquent au moins en partie cet écart de mobilité. Les internautes ont un niveau d'étude et un revenu nettement supérieurs, conduisant à une très forte motorisation. Ils sont plus souvent cadres et employés et travaillent davantage dans le centre, ce qui conduit à des durées hors domicile plus importantes, réduisant les possibilités de participation à des activités moins contraintes et l'usage de la marche. Le risque est de confondre le phénomène étudié (la variation de la mobilité individuelle) avec le processus de sélection des répondants.

Un des objectifs de cette partie est de montrer qu'il est possible de distinguer l'effet dû au mode d'enquête de celui lié aux différences socio-économiques observées entre les échantillons de répondants, et de quantifier l'impact du mode d'enquête sur la mobilité déclarée. Le modèle économétrique envisagé pour cette analyse est le modèle de sélection de l'échantillon, dont nous estimons les paramètres à l'aide de la procédure en deux étapes, élaborée par James Heckman et d'autres à la fin des années 1970. Le second objectif est de mieux comprendre les déterminants de la mobilité, en distinguant ceux qui sont à l'origine de la décision de se déplacer de ceux qui impactent le nombre de déplacements réalisé par les personnes mobiles. Afin d'étudier séparément ces deux effets, nous utilisons un modèle à obstacle (ou modèle Hurdle).

Dans le chapitre 7, nous mettons en évidence l'impact du mode d'enquête (web ou face-à-face) sur la mesure des comportements de mobilité. Nous proposons une formalisation du problème du biais de sélection de l'échantillon, et développons un modèle explicatif des comportements de mobilité permettant de s'affranchir de ce biais. Le chapitre 8 est consacré à une analyse comparative des déterminants de l'immobilité, d'une part, des facteurs qui influencent la mobilité des personnes mobiles, d'autre part. Nous mettons ainsi en évidence l'intérêt des modèles à obstacle (Hurdle) pour modéliser les différentes facettes d'un phénomène complexe : la mobilité quotidienne.

Chapitre 7 : Impact du mode d'enquête sur la mesure des comportements de mobilité

"Il faut utiliser les modèles, non y croire."
Henri Theil (1924-2000).

Dans le chapitre précédent, nous avons montré que la qualité des réponses, et plus précisément le niveau de mobilité quotidienne, varie entre les deux enquêtes web et face-à-face. Le recueil d'informations via différentes sources peut générer des résultats parfois peu comparables et le danger, lorsque l'on fusionne des bases de données, est de générer un biais de sélection des individus, qui compromet la pertinence des modèles explicatifs des comportements. Ce biais de sélection est l'objet d'une importante littérature, théorique et empirique (Winship et Mare, 1992), mais les applications aux enquêtes transport sont rares à ce jour. Les données de l'enquête ménages déplacements de Lyon nous permettent de mettre en évidence le problème de sélection, les non-répondants à l'enquête standard en face-à-face choisissant de remplir ou pas le questionnaire sur le web.

Ces différences constatées entre les deux enquêtes peuvent provenir d'un effet média ou du fait que tous les répondants n'ont pas la possibilité de répondre en ligne, les effets de mode et de sélection étant confondus (De Leeuw, 2005). Il est alors nécessaire de rechercher, toutes choses égales par ailleurs, quel est l'effet marginal du web sur les résultats. Autrement dit, identifier les propriétés propres au web en tant que mode d'enquête. Selon Lozar Manfreda et Vehovar (2002b), Willke *et al.* (1999) et Wydra (1999), les effets de mode surviennent surtout lorsque le web est comparé à des enquêtes administrées en face-à-face. Il est donc essentiel d'établir une procédure pour pouvoir comparer les résultats des deux enquêtes, le choix du média ayant une influence non négligeable sur la mobilité individuelle (Van Evert et Moritz, 2000; Moritz et Brög, 1999).

L'objet de ce chapitre est d'évaluer plus précisément l'effet du mode d'enquête sur le nombre de déplacements quotidiens déclaré par les répondants. L'évaluation nécessite de tester l'existence d'un biais de sélection de l'échantillon, puis le mesurer et le neutraliser, le cas échéant. La section I présente le

problème de l'évaluation, en le transposant à l'analyse comparative du mode d'enquête sur la mobilité quotidienne. Il s'agit de mettre en évidence les enjeux du biais de sélection susceptible de se produire dans les protocoles d'enquête mixte. Suivent les développements théoriques relatifs au biais de sélection, plus spécifiquement ceux concernant la méthode en deux étapes d'Heckman (1979) et celle des variables instrumentales (Heckman et Navarro-Lozano, 2004). L'intérêt et les limites respectives de ces deux méthodes sont expliqués (section II). Puis, nous recensons les variables disponibles pour l'élaboration d'un modèle explicatif de la mobilité et estimons un premier modèle qui ne tient pas compte du biais de sélection (section III). Nous développons ensuite un modèle économétrique permettant de s'affranchir de l'effet de sélection des individus, et l'appliquons aux données de l'enquête ménages déplacements de Lyon. Nous justifions la spécification du modèle retenu et les traitements réalisés pour améliorer la robustesse des résultats. Nous présentons l'ensemble des régressions et interprétons les paramètres d'intérêt du modèle, en particulier la significativité et le signe de l'inverse du ratio de Mills (section IV). Enfin, nous isolons et quantifions la part du différentiel de mobilité liée au mode d'enquête et celle venant du biais de sélection (section IV). Seuls les individus mobiles (qui se sont déplacés durant la période de référence), de 18 ans et plus sont concernés dans ce chapitre ¹¹¹.

I La formalisation du problème

Dans l'enquête ménages déplacements de Lyon réalisée en 2006, les réponses des enquêtes web et face-à-face ne sont pas comparables. L'analyse des données montre que les répondants web déclarent en moyenne moins de déplacements quotidiens que les répondants à l'enquête en face-à-face. Ce constat peut notamment s'expliquer par les différences socio-économiques, qui caractérisent les deux échantillons. En redressant l'enquête en face-à-face, on a cherché à isoler une population dont les caractéristiques socio-économiques sont proches de celles de l'enquête web et dont les effectifs sont suffisants pour avoir des résultats statistiquement significatifs. Les différences en termes de nombre de déplacements subsistent et ont même tendance à s'amplifier un peu par rapport à l'ensemble de la population. Cet écart s'explique à la fois par une plus forte immobilité des internautes actifs et par une moindre déclaration de déplacements et de sorties. Deux interprétations sont possibles de ces résultats. La première conduit à une sous-estimation de la mobilité imputable au média web. Plusieurs facteurs différencient les modes et peuvent être à l'origine de la diversité des réponses ¹¹², comme ceux qui influencent la transmission de l'information, et les effets de l'interviewer. Les différences observées dans les

¹¹¹Dans l'enquête web, une seule personne remplit le questionnaire. Il s'agit le plus souvent du destinataire de la lettre avis, c'est-à-dire le chef de ménage ou son conjoint, ce qui explique que les individus de moins de 18 ans soient fortement sous-représentés.

¹¹²Ces différences concernent principalement les questions sensibles, d'attitude, demandant un effort de mémoire important au répondant, ayant beaucoup de réponses possibles et les répondants peu familiers avec internet (Lozar Manfreda et Vehovar, 2002a).

réponses des enquêtes web et face-à-face peuvent également venir du mode d'administration du questionnaire (auto-administré, vs. administré), de l'informatisation du questionnaire, ou être plus spécifiques (Willke *et al.*, 1999). Par exemple, le manque d'ergonomie du questionnaire web peut rendre pénible la saisie des déplacements¹¹³, sans compter les biais relatifs à l'aisance des répondants face à un ordinateur. Cette première interprétation peut être étayée par le fait que la sous-estimation concerne surtout des déplacements courts, tant en temps qu'en distance, principalement effectués à pied, pour des motifs comme les loisirs ou l'accompagnement, qui peuvent être perçus comme moins importants par les internautes (Bonnel et Le Nir, 1998). Avec une meilleure ergonomie, il serait possible de réduire cette omission, mais des contraintes techniques n'ont pas permis d'exploiter toutes les possibilités interactives du média web.

La seconde interprétation concerne les caractéristiques socio-économiques des internautes. Il peut en effet exister des raisons pour que les individus qui répondent sur le web ne présentent pas un niveau de mobilité équivalent à ceux qui ont répondu en face-à-face, et que ces raisons ne soient pas liées au mode de réponse. Selon la méthodologie utilisée, les individus choisissent de répondre ou non en face-à-face, et dans la négative de remplir ou pas le questionnaire en ligne. La présence des répondants dans un groupe est probablement déterminée par des facteurs extérieurs, qui peuvent impacter la variable d'intérêt du modèle étudié. Dit autrement, il est fort probable que des caractéristiques socio-économiques, pas toujours observables, influencent le choix des individus de recevoir un enquêteur à domicile ou de répondre, le cas échéant, sur le web et impactent leurs comportements de mobilité (Berk, 1983; Ressource System Group, 2002). Les groupes de répondants peuvent différer sur des aspects systématiques et on sait que lorsque des observations sont exclues d'un échantillon de manière non aléatoire, il y a un risque de biais de sélection. Il est fort probable que seuls les individus intéressés par le thème de l'enquête et ayant un accès privé à l'ordinateur (et à une connexion internet) répondent à l'enquête web (Stanton, 1998; Morrel-Samuels, 2003). Le choix de ce média d'enquête dépend donc en grande partie des facteurs socio-économiques et de l'aisance avec l'ordinateur¹¹⁴.

Si le mode retenu pour remplir le questionnaire est lié à la mobilité, alors les internautes pourraient déclarer un nombre de déplacements plus faible, même s'ils ne répondraient pas sur le web, mais en face-à-face. La simple comparaison de la mobilité des répondants web et face-à-face, sans correction du biais de sélection, biaise l'effet réel du média. Ignorer l'existence du biais de sélection, en particulier pour les répondants web, peut donc avoir des conséquences sur la validité d'un modèle explicatif des comportements de mobilité.

Le choix du mode peut modifier les réponses aux questions formulées iden-

¹¹³Les enquêtes assistées par ordinateur génèrent globalement moins d'erreurs, mais également moins d'activités enregistrées.

¹¹⁴Plus le répondant est jeune, a fait des hautes études et utilise un e-mail, plus il a de chance de répondre sur le web (Romano et Himmelmann, 2002).

tiquement entre deux enquêtes, et le web comme nouveau mode de collecte introduit des effets qui sont encore peu connus (Dillman et Christian, 2005). Il est alors nécessaire de scinder la différence de mobilité observée entre les répondants web et face-à-face en deux parties : l'effet de sélection et l'effet du mode d'enquête ¹¹⁵.

II Le biais de sélection de l'échantillon

On sait depuis les années 50 que l'estimation d'une équation sur un sous-échantillon obtenu de façon sélective dans la population peut conduire à des biais (Roy, 1951). Cependant, les premiers développements économétriques des conséquences de cette sélection des individus datent de 1974, avec les travaux d'Heckman (1979). L'exemple typique est celui d'une équation de salaire estimée sur les seules femmes actives, alors même que le comportement d'activité relève d'un arbitrage dans lequel le salaire que la personne peut obtenir sur le marché intervient. Depuis, de nombreux articles ont mis en évidence l'importance du biais de sélection dans les enquêtes réalisées en sciences humaines et sociales (Maddala, 1986). On notera par exemple le modèle de migration aux USA analysé par Nakosteen et Zimmer (1980), ou celui du taux d'activité féminin de Mroz (1987), bien que l'utilisation la plus fréquente des modèles d'auto-sélection concerne l'évaluation d'un traitement ou d'une formation.

Dillman (1978) a mis en évidence le biais de sélection qui résulte de protocoles d'enquêtes où plusieurs modes de recueil de données sont disponibles ¹¹⁶. Les sciences sociales contiennent par ailleurs plusieurs présentations formelles du biais de sélection (Heckman, 1979; Goldberger, 1981). Le recours aux développements économétriques semble intéressant ici pour isoler l'effet des différences sociodémographiques de celui du mode d'enquête sur la mobilité quotidienne d'une part, pour quantifier cet effet du mode d'enquête sur le comportement de mobilité des répondants d'autre part. Le modèle traditionnellement utilisé pour mettre en évidence le problème d'auto-sélection des répondants est le modèle Tobit II ¹¹⁷, appelé plus couramment modèle de sélection ¹¹⁸ de l'échantillon.

II.1 L'endogénéité du mode d'enquête au niveau de mobilité

Nous présentons des méthodes permettant de formaliser l'endogénéité du mode d'enquête au niveau de mobilité.

¹¹⁵Si la période d'enquête n'est pas la même, alors les effets de temps se mélangent aux effets de mode (De Leeuw, 2005).

¹¹⁶Le biais de sélection peut être soit positif soit négatif, c'est-à-dire que si on n'en tient pas compte, l'effet estimé du mode d'enquête peut être soit supérieur soit inférieur à son véritable effet.

¹¹⁷Cette classification est due à Amemiya (1984).

¹¹⁸Un modèle de sélection est un modèle dans lequel la variable dépendante y n'est pas toujours observée. Le critère de sélection ne porte pas directement sur la valeur de y , mais est défini par une équation auxiliaire.

II.1.1 Un effet direct du mode d'enquête non observable

La décision de répondre ou non à l'enquêteur au domicile, puis, le cas échéant, de se soumettre ou pas à l'enquête en ligne, est prise par chaque individu contacté pour répondre à l'étude. Le fait de répondre en ligne est une variable aléatoire notée I_i , qui prend la valeur 1 si l'individu i a répondu au questionnaire web et la valeur 0 s'il a répondu à l'enquête en face-à-face. Le nombre de déplacements réalisé par l'individu i (Y_i) représente la variable d'intérêt du modèle (c'est à partir du niveau de mobilité individuelle déclaré que nous souhaitons évaluer l'effet du mode d'enquête). Il s'agit d'une variable générale, qui tient compte des caractéristiques positives (comme l'anonymat des réponses) et négatives (par exemple, la lourdeur dans la saisie des réponses) du média web.

Nous sommes amenés à distinguer deux variables latentes, Y_{0i} et Y_{1i} , qui correspondent au nombre de déplacements déclaré par un individu i , conditionnellement au média de réponse utilisé (respectivement $I_i = 0$ ou $I_i = 1$). Le problème de l'évaluation provient du fait qu'il n'est jamais possible d'observer simultanément, pour un individu i , le nombre de déplacements déclaré en face-à-face et sur le web. Si l'individu i a répondu sur le web, alors le nombre de déplacements saisi sur le web (Y_{1i}) est observé, alors que le nombre de déplacements déclaré en face-à-face (Y_{0i}) est inconnu. Dans ce cas, la valeur non observée (Y_{0i}) est qualifiée de résultat contrefactuel. Symétriquement, pour un individu i ayant répondu en face-à-face, le nombre de déplacements recueilli par l'enquêteur (Y_{0i}) est observé, alors que le nombre de déplacements qu'aurait déclaré cet individu sur le web (Y_{1i}) est inconnu. C'est la valeur non observée Y_{1i} , qui correspond dans ce cas au résultat contrefactuel.

Nous pouvons dire, à partir du raisonnement précédent, que la variable de résultat observée (Y_i) est liée aux variables latentes de résultat Y_{0i} et Y_{1i} , par l'équation :

$$Y_i = I_i * Y_{1i} + (1 - I_i) * Y_{0i} \quad (23)$$

L'effet causal du mode d'enquête sur le niveau de mobilité déclaré correspond à l'écart entre les variables Y_{1i} et Y_{0i} . Il représente la différence entre le nombre de déplacements saisi par l'internaute et celui que ce même individu aurait déclaré à un enquêteur présent à son domicile. Le problème fondamental de l'évaluation est que cette différence n'est pas directement observable. Une des solutions pour déterminer cet effet consiste à poser des hypothèses sur la loi jointe du quadruplet $(Y_{1i}, Y_{0i}, Z_i, X_i)$, avec X_i les variables explicatives du niveau de mobilité individuelle pour l'individu i et Z_i les variables explicatives du fait de répondre en ligne pour l'individu i . L'estimation de l'effet du média web sur la mobilité correspond en effet à l'espérance de la différence des deux variables latentes : $E(Y_{1i} - Y_{0i})$. L'effet moyen du média web sur le nombre de déplacements déclaré s'écrit :

$$\Delta_{I1} = E(Y_{1i} - Y_{0i} | I_i = 1) \quad (24)$$

Et l'effet marginal du média web sur le nombre de déplacements déclaré s'écrit :

$$ME_I = E(Y_{1i} - Y_{0i} | X_i, Z_i) \quad (25)$$

II.1.2 Application du modèle de Rubin

Par analogie avec le modèle de Rubin (1974), on peut considérer que le mode d'enquête (web) s'apparente à un traitement à évaluer. Il existe un problème d'auto-sélection des individus, lorsqu'un traitement exerce un effet différent sur la population des traités de l'effet qu'il n'aurait exercé sur la population des non traités si celle-ci avait été traitée. Cet effet provient des différences existant entre les populations de traités et de non traités, c'est-à-dire entre les répondants web et face-à-face. Si le nombre de déplacements déclaré par un individu i est indépendant de la probabilité pour ce même individu de répondre en ligne (si $Y_{1i}, Y_{0i} \perp\!\!\!\perp I_i$), alors :

$$\Delta_{I1} = E(Y_{1i} | I_i = 1) - E(Y_{0i} | I_i = 0) \quad (26)$$

$$\Delta_{I1} = E(Y_i | I_i = 1) - E(Y_i | I_i = 0) \quad (27)$$

A contrario, nous pouvons supposer que l'effet du média pour les individus répondant en ligne ($E(Y_{1i} | I_i = 1)$) est différent de l'effet qu'engendrerait ce média pour un individu répondant en face-à-face ($E(Y_{1i} | I_i = 0)$). Une première source d'endogénéité du mode d'enquête au niveau de mobilité individuelle peut provenir de variables omises, qui sont corrélées à la fois au nombre de déplacements quotidiens et à la propension individuelle à refuser une interview en face-à-face, puis à répondre sur le web. Par exemple, si le chef de ménage a un emploi qualifié, alors ses horaires de travail ne permettent pas à un enquêteur d'interroger l'ensemble des personnes au domicile, mais sa familiarité avec l'utilisation des nouvelles technologies (notamment pour son activité professionnelle) l'encourage à répondre en ligne. Il s'agit de variables cachées, dont on a de bonnes raisons de penser qu'elles existent, mais qu'on ne peut pas mettre en évidence. L'incapacité à isoler ces effets produit des estimations non convergentes du coefficient de la variable mode de réponse, dans le modèle de régression explicatif du nombre de déplacements déclaré.

II.2 Les méthodes d'estimation disponibles

Afin de corriger ces biais, il faut développer un modèle économétrique plus spécifique. Le principe est d'utiliser les informations dont on dispose sur les individus non traités pour construire pour chaque individu traité un contre-factuel, c'est-à-dire une estimation de ce qu'aurait été sa situation s'il n'avait pas été traité (Brodaty *et al.*, 2007). Les méthodes d'évaluation procèdent toujours en plusieurs étapes. Par une fonction de sélection, on estime pour chaque individu la probabilité de recevoir un traitement. Puis, on régresse la variable d'intérêt, en intégrant dans le modèle des estimateurs corrigés. Deux grands

types de variables sont considérés dans l'équation de sélection : les variables observables et les variables inobservables.

II.2.1 La sélection sur variables observables

Les méthodes qui s'appuient sur les hypothèses d'indépendance conditionnelle à des caractéristiques observables sont appelées les méthodes de sélection par appariement. Il s'agit de contrôler les performances d'un traitement par la probabilité de le recevoir, conditionnellement à des variables observables. Le principe de cette méthode est d'utiliser les informations dont on dispose sur les individus non traités (répondants web) pour construire pour chaque individu traité (répondants face-à-face) un contrefactuel, c'est-à-dire une estimation de ce qu'aurait été sa situation s'il n'avait pas été traité. L'hypothèse de sélection sur variables observables, selon laquelle, conditionnellement aux variables observables Z_i (variables explicatives du fait de répondre en ligne pour un individu i), le fait pour un individu de répondre sur le web est rendue indépendante, est très forte. Nous avons :

$$(Y_{0i}, Y_{1i}) \perp\!\!\!\perp I_1 \mid P(Z_i) \quad (28)$$

II.2.2 La prise en compte de variables inobservables

Une autre solution pour résoudre le problème de sélection est de recourir à une modélisation jointe des résultats potentiels (Y_{0i}, Y_{1i}) et de l'affectation au traitement (I_i) , en faisant l'hypothèse que ces trois variables dépendent de termes d'erreur inobservables, potentiellement corrélés entre eux. Cette méthode suppose, contrairement à la méthode par appariement, que les résultats (le nombre de déplacements déclarés) et les choix (répondre ou pas sur le web), soient additivement dissociables. Cette méthode est par définition plus globale et produit ainsi des estimations plus robustes que la méthode par appariement, car lorsque des variables inobservables existent, elles biaisent les méthodes d'appariement sur variables observables (Heckman et Navarro-Lozano, 2004). Cette hypothèse de modélisation du biais de sélection apparaît beaucoup plus pertinente pour évaluer l'effet du média web, car les critères déterminants le choix d'un mode de réponse ne sont pas tous connus de l'analyste. La méthode de Heckman (1979) à deux étapes et la méthode des variables instrumentales reposent toutes deux sur une modélisation du lien entre les variables inobservables et les variables observables.

II.2.3 La méthode des variables instrumentales

La méthode des variables instrumentales, abordée notamment par (Heckman et Robb, 1985), permet de mettre en évidence le biais de sélection. Elle consiste à rechercher une ou plusieurs variables qui influent sur le choix du mode d'enquête, mais pas sur le nombre de déplacements déclaré. Ces variables, non corrélées au terme d'erreur de l'équation du modèle de mobilité

quotidienne, peuvent servir à l'estimation, sans qu'il y ait un biais. Plus formellement, la méthode des variables instrumentales cherche à estimer le nombre de déplacements quotidiens déclarés, sachant les variables explicatives de la mobilité X_{1i} et X_{2i} et les variables de sélection Z_i . Soit :

$$E(Y_{1i} | X_{1i}, Z_i) \text{ et } P(I_i = 1 | X_{1i}, Z_i)$$

Cette méthode diffère de l'estimation en deux étapes d'Heckman, qui suppose une condition supplémentaire de dépendance du nombre de déplacements déclarés à l'utilisation du média web pour la saisie des réponses :

$$E(Y_{1i} | X_{1i}, Z_i, I_i)$$

Soit le modèle de régression linéaire explicatif de la mobilité quotidienne :

$$Y_i = \sum_k \beta_k X_{ki} + \delta I_i + u_i \quad (29)$$

avec Y_i le nombre moyen de déplacements réalisé par l'individu i , X_i un vecteur de variables indépendantes ou explicatives de la mobilité et I_i le fait de répondre à l'enquête en ligne. Le fait de répondre en ligne peut être expliqué par le modèle suivant :

$$I_i = 1, \text{ssi } I_i^* > 0$$

$$I_i = 0, \text{ssi } I_i^* \leq 0$$

avec :

$$I_i^* = \sum_m \delta_m Z_{mi} + \epsilon_i \quad (30)$$

La variable I_i^* est une variable latente et Z_{mi} un vecteur de variables explicatives, dites de conditionnement. Nous supposons que ϵ_i suit une loi normale et est liée à u_i par une loi normale bivariée. Pour corriger le biais de sélection par la méthode des variables instrumentales, nous remplaçons dans la fonction d'intérêt la variable I_i par l'estimation de la propension à répondre sur le web (I_i^*), effectuée à l'aide d'un modèle probit. Soit :

$$Y_i = \sum_k \beta'_k X_{ki} + \delta' I_i^* + u'_i \quad (31)$$

avec u'_i les résidus corrigés du biais de sélection de la régression linéaire. Les estimations des paramètres par la méthode des moindres carrés ordinaires sont convergentes.

Il faut noter que si l'effet des variables inobservables sur la propension à répondre en ligne est strictement équivalent pour les répondants web et les répondants face-à-face, alors la propension à répondre en ligne est un instrument valide, et on peut utiliser la méthode des variables instrumentales. Si en revanche l'effet des variables inobservables sur la propension à répondre en ligne est différent pour les répondants web et les répondants face-à-face, mais que

le fait de répondre en ligne est rendu indépendant des variables inobservables par des variables observables, alors la méthode des variables instrumentales s'apparente à une méthode de traitement du biais de sélection par appariement.

La méthode des variables instrumentales s'effectue en une seule étape, et ne nécessite pas de modélisation détaillée du processus de participation à l'enquête web. En revanche, si les estimations sont sans biais (lorsque le modèle est bien défini), cette méthode ne donne pas une estimation de l'ampleur du biais de sélection. Par ailleurs, la recherche des variables instrumentales exige un examen approfondi du processus de sélection des individus et les caractéristiques sociodémographiques traditionnellement recueillies dans les enquêtes sont rarement suffisantes, car souvent liées aux résultats du modèle estimé. L'identification de variables instrumentales pertinentes est donc une tâche difficile. C'est pourquoi nous préférons utiliser la méthode en deux étapes (section II.2.2).

II.3 La méthode en deux étapes

Cette procédure consiste à estimer, dans une première équation, la probabilité pour un individu i de répondre en ligne puis, dans une seconde étape, les paramètres de l'équation d'intérêt, en s'affranchissant de l'endogénéité éventuelle du mode d'enquête à la mobilité.

II.3.1 Développements théoriques

En utilisant la méthode d'estimation en deux étapes développée par Heckman en 1979, le modèle explicatif de la mobilité peut se formaliser, pour chaque individu i .

-

Pour les répondants web :

$$Y_{1i} = \sum_k \beta_{1k} X_{1ki} + u_{1i} \quad (32)$$

Pour les répondants en face-à-face :

$$Y_{2i} = \sum_k \beta_{2k} X_{2ki} + u_{2i} \quad (33)$$

Avec Y_{1i} et Y_{2i} le nombre moyen de déplacements réalisé par l'individu i , X_{1i} et X_{2i} deux vecteurs de variables indépendantes ou explicatives de la mobilité et u_{1i} et u_{2i} deux termes d'erreur supposés normaux, qui tiennent compte des forces non observées qui pourraient influencer sur la mesure des résultats. En réalité, nous estimerons deux modèles, un sur le sous-échantillon des répondants en face-à-face, et un sur le sous-échantillon des répondants sur le web ¹¹⁹. Le modèle ainsi défini n'impose pas que les coefficients des variables explicatives

¹¹⁹Notons que seulement un des paramètres Y_{1i} et Y_{2i} est observé, selon que l'individu choisisse de répondre en face-à-face ou sur le web.

X_{1ki} et X_{2ki} soient identiques pour les répondants web et les répondants en face-à-face.

Soit la fonction de sélection traduisant la probabilité pour un individu i de répondre sur le web :

$$I_i^* = \sum_m \delta_m Z_{mi} + \epsilon_i \quad (34)$$

$$I_i = 1, \text{ssi } I_i^* > 0$$

$$I_i = 0, \text{ssi } I_i^* \leq 0$$

Avec I_i^* la variable de sélection, Z_m un ensemble de variables déterminantes du choix du web et ϵ les termes d'erreur supposés normaux. La mobilité observée pour tout individu i se définit comme :

$$Y_i = Y_{1i}, \text{ssi } I_i = 1$$

$$Y_i = Y_{2i}, \text{ssi } I_i = 0$$

Nous posons l'hypothèse restrictive qu'il existe des variables de Z_i qui ne sont pas dans X_i . Par ailleurs, nous supposons que les termes d'erreur des équations 32, 33 et 34 (u_1 , u_2 , et ϵ) suivent une loi normale bivariée de moyennes nulles et de corrélations ρ_1 et ρ_2 . La mobilité n'est observable que si les individus répondent en face-à-face ou sur le web. Des perturbations aléatoires vont donc affecter simultanément les variables endogènes des équations de sélection et d'intérêt, et les termes d'erreur des deux équations peuvent être corrélés :

$$(\epsilon, u_{1i}) \sim N(0, 0, \sigma_\epsilon, \sigma_{u1}, \rho_1)$$

$$(\epsilon, u_{2i}) \sim N(0, 0, \sigma_\epsilon, \sigma_{u2}, \rho_2)$$

L'hypothèse centrale est que la distribution normale de ϵ_i est liée à u_{1i} et u_{2i} par une loi normale bivariée et normalement distribuée, où ρ_1 et ρ_2 sont les coefficients de corrélation entre les résidus. Nous faisons l'hypothèse que les éléments inobservés ϵ_i et (u_{1i}, u_{2i}) sont indépendants des variables explicatives X_{1i} et X_{2i} et des variables de conditionnement Z_i .

Prenons tous les individus avec (X_i, Z_i) donné. Formellement, la régression de Y_i sur X_i dans l'échantillon tronqué ¹²⁰ est :

$$E(Y_{1i} | I_i = 1) = E(Y_{1i} | X_{1i}, Z_i, I_i = 1) \quad (35)$$

$$E(Y_{1i} | I_i = 1) = \sum_k \beta_{1k} X_{1ki} + E(u_{1i} | Z_i, I_i = 1) \quad (36)$$

¹²⁰Un échantillon est tronqué lorsque les observations sont faites seulement pour certains individus, constituant un sous-ensemble de la population observée (Tobin, 1958).

$$E(Y_{2i} | I_i = 0) = E(Y_{2i} | X_{2i}, Z_i, I_i = 0) \quad (37)$$

$$E(Y_{1i} | I_i = 0) = \sum_k \beta_{2k} X_{2ki} + E(u_{2i} | Z_i, I_i = 0) \quad (38)$$

Si le choix du mode de réponse des individus est systématiquement corrélé avec leur mobilité, alors les espérances conditionnelles de la mobilité et des termes d'erreur ne sont pas égales à leurs espérances :

$$E(Y_{1i} | I_i = 1) \neq E(Y_{1i}) \Leftrightarrow E(u_{1i} | I_i = 1) \neq E(u_{1i}) = 0$$

$$E(Y_{2i} | I_i = 0) \neq E(Y_{2i}) \Leftrightarrow E(u_{2i} | I_i = 0) \neq E(u_{2i}) = 0$$

Nous imposons une normalisation sur la variance de ϵ . Soit $\sigma_\epsilon = 1$ ¹²¹. Cette condition se justifie par le fait que nous n'observons pas la valeur de la variable latente I_i^* , mais seulement son signe (Davidson et McKinnon, 1993). Nous observons donc le résultat du choix individuel de répondre ou non à l'enquête sur le web, mais pas la propension d'un individu à répondre en ligne (I_i^*). Sous l'hypothèse de normalité, nous pouvons écrire¹²² :

$$u_{1i} = \rho_1 \sigma_{u1i} \epsilon_i$$

$$u_{2i} = \rho_2 \sigma_{u2i} \epsilon_i$$

σ_{u1i} représente la covariance entre les résidus u_{1i} de la régression linéaire et les résidus ϵ_i ($I_i = 1$), σ_{u2i} représente la covariance entre les résidus u_{2i} de la régression linéaire et les résidus ϵ_i ($I_i = 0$).

En remplaçant dans l'expression 36, nous obtenons :

$$E(Y_{1i} | I_i = 1) = \sum_k \beta_{1k} X_{1ki} + E(u_{1i} | Z_i, I_i = 1) \quad (39)$$

$$E(Y_{1i} | I_i = 1) = \sum_k \beta_{1k} X_{1ki} + E(u_{1i} | Z_i, I_i^* > 0) \quad (40)$$

$$E(Y_{1i} | I_i = 1) = \sum_k \beta_{1k} X_{1ki} + E(u_{1i} | \epsilon_i > - \sum_m \delta_m Z_{mi}) \quad (41)$$

$$E(Y_{1i} | I_i = 1) = \sum_k \beta_{1k} X_{1ki} + E(\rho_1 \sigma_{u1i} \epsilon_i | \epsilon_i > - \sum_m \delta_m Z_{mi}) \quad (42)$$

$$E(Y_{1i} | I_i = 1) = \sum_k \beta_{1k} X_{1ki} + \rho_1 \sigma_{u1i} E(\epsilon_i | \epsilon_i > - \sum_m \delta_m Z_{mi}) \quad (43)$$

En remplaçant dans l'expression 38, nous obtenons :

$$E(Y_{2i} | I_i = 0) = \sum_k \beta_{2k} X_{2ki} + E(u_{2i} | Z_i, I_i^* \leq 0) \quad (44)$$

$$E(Y_{2i} | I_i = 0) = \sum_k \beta_{2k} X_{2ki} + \rho_2 \sigma_{u2i} E(\epsilon_i | \epsilon_i \leq - \sum_m \delta_m Z_{mi}) \quad (45)$$

¹²¹ Compte-tenu de la nature des données, seul le signe de I_i^* est observable, et non sa valeur, ce qui empêche l'estimation de la variance de l'équation 34 (Cameron et Triverdi, 2005).

¹²² En effet, $cov(u, \epsilon) = \rho \sigma_u \sigma_\epsilon = \rho \sigma_u$ et $cov(\rho \sigma_u \epsilon, \epsilon) = \rho \sigma_u V(\epsilon) = \rho \sigma_u$

La troncature sur ϵ entraîne donc une troncature sur Y_1 et Y_2 si, u_1 et ϵ , d'une part, et u_2 et ϵ , d'autre part, sont corrélés (ρ_1 et $\rho_2 \neq 0$). Soit ϕ la fonction de densité et Φ la fonction de répartition de la loi normale. Le détail du calcul de l'espérance d'une loi normale tronquée en s est présenté en annexe XII.

Par définition, nous avons :

$$E(\epsilon_i \mid \epsilon_i > -\sum_m \delta_m Z_{mi}) = \frac{\phi(-\sum_m \delta_m Z_{mi})}{1 - \Phi(-\sum_m \delta_m Z_{mi})} \quad (46)$$

$$E(\epsilon_i \mid \epsilon_i \leq -\sum_m \delta_m Z_{mi}) = -\frac{\phi(-\sum_m \delta_m Z_{mi})}{\Phi(-\sum_m \delta_m Z_{mi})} \quad (47)$$

En remplaçant dans les expressions 43 et 45, nous obtenons ¹²³ :

$$E(Y_{1i} \mid I_i = 1) = \sum_k \beta_{1k} X_{1ki} + \rho_1 \sigma_{u1i} \frac{\phi(\sum_m \delta_m Z_{mi})}{\Phi(\sum_m \delta_m Z_{mi})} \quad (48)$$

$$E(Y_{1i} \mid I_i = 1) = \sum_k \beta_{1k} X_{1ki} + \rho_1 \sigma_{u1i} \lambda_{1i} \quad (49)$$

$$E(Y_{2i} \mid I_i = 0) = \sum_k \beta_{2k} X_{2ki} + \rho_2 \sigma_{u2i} \frac{-\phi(\sum_m \delta_m Z_{mi})}{1 - \Phi(\sum_m \delta_m Z_{mi})} \quad (50)$$

$$E(Y_{2i} \mid I_i = 0) = \sum_k \beta_{2k} X_{2ki} + \rho_2 \sigma_{u2i} \lambda_{2i} \quad (51)$$

Il s'agit en fait de remplacer ϵ_i par son espérance conditionnelle aux valeurs de δZ_i , c'est-à-dire par $\frac{\phi(\delta Z_i)}{\Phi(\delta Z_i)}$. Les ratios $\lambda_{1i} = \frac{\phi}{\Phi}$ et $\lambda_{2i} = \frac{-\phi}{1-\Phi}$ sont appelés inverse du ratio de Mills ¹²⁴. Pour chaque observation, nous calculons l'inverse du ratio de Mills, qui correspond à l'espérance conditionnelle des résidus ϵ_i à $I_i = k$, avec $k = 0, 1$. Cette variable est d'une manière générale la principale source de biais des estimations des coefficients du modèle de régression. La source d'endogénéité du mode d'enquête au niveau de mobilité provient ici de variables omises, qui sont corrélées à la probabilité de choisir le web comme

¹²³Rappelons que : $\phi(-\sum_m \delta_m Z_{mi}) = \phi(\sum_m \delta_m Z_{mi})$ et que : $1 - \Phi(-\sum_m \delta_m Z_{mi}) = \Phi(\sum_m \delta_m Z_{mi})$

¹²⁴Ce ratio est appelé également Lambda d'Heckman, car noté $\lambda(x|\beta)$ par l'auteur (1979).

média d'enquête et au nombre de déplacements saisis. Le biais de sélection correspond donc à un biais de valeur manquante. En effet, si on estime les expressions 32 et 33 par la méthode des moindres carrés ordinaires, on omet deux variables :

$$\frac{\phi\left(\sum_m \delta_m Z_{mi}\right)}{\Phi\left(\sum_m \delta_m Z_{mi}\right)} = \lambda_{1i} \quad (52)$$

$$\frac{-\phi\left(\sum_m \delta_m Z_{mi}\right)}{1 - \Phi\left(\sum_m \delta_m Z_{mi}\right)} = \lambda_{2i} \quad (54)$$

Dans ce cas, on peut s'attendre à ce que le modèle soit biaisé, puisque les estimations de β_{1k} et β_{2k} sont non convergentes. Il est d'ailleurs probable que l'ampleur, le signe et la significativité des coefficients diffèrent lorsqu'ils sont estimés par la méthode en deux étapes. Ces différences dépendent des coefficients $\rho_1\sigma_{u1}$ et $\rho_2\sigma_{u2}$ et des coefficients des variables concernées estimés dans le modèle de sélection (Hoffman et Link, 1984).

La méthode en deux étapes permet donc d'estimer les fonctions de mobilité des échantillons web et face-à-face en s'affranchissant du biais de sélection des individus, grâce à l'intégration de l'inverse du ratio de Mills dans les expressions 32 et 33. Soit le modèle :

Pour les répondants web :

$$Y_{1i} = \sum_k \beta_{1k} X_{1ki} + \rho_1 \sigma_{u1} \lambda_{1i} + e_{1i} \quad (55)$$

Pour les répondants en face-à-face :

$$Y_{2i} = \sum_k \beta_{2k} X_{2ki} + \rho_2 \sigma_{u2} \lambda_{2i} + e_{2i} \quad (56)$$

II.4 Test de l'existence significative d'un biais de sélection

L'existence d'un biais de sélection est testée par l'hypothèse que le coefficient estimé de l'inverse du ratio de Mills est nul dans chaque groupe (web et face-à-face). Les hypothèses sont les suivantes :

Pour l'échantillon web :

$$H_0 : \rho_1 \sigma_{u1} = 0$$

$$H1 : \rho_1\sigma_{u1} \neq 0$$

Pour l'échantillon face-à-face :

$$H0 : \rho_2\sigma_{u2} = 0$$

$$H1 : \rho_2\sigma_{u2} \neq 0$$

Si la t-value est inférieure à la valeur critique, il n'est pas possible de rejeter l'hypothèse nulle d'absence de corrélation entre les termes d'erreur des équations de sélection et d'intérêt. Dans ce cas, il n'y a pas de biais de sélection significatif dans le modèle, et on peut appliquer la méthode des moindres carrés ordinaires pour estimer directement les coefficients β_{1k} et β_{2k} . Cela peut s'expliquer par le fait que les facteurs non observés conduisant à une sélection non aléatoire des répondants selon le mode d'enquête ne sont pas quantitativement importants dans l'évaluation du modèle explicatif du nombre de déplacements quotidiens, ou bien que ces facteurs sont quantitativement importants mais ne sont pas en relation directe avec les résultats du modèle étudié.

Dans le cas contraire, l'effet de sélection est significatif et le choix du mode apparaît, sous ces hypothèses, endogène au niveau de mobilité. La méthode en deux étapes permet d'obtenir des estimations non biaisées des coefficients β_{1k} et β_{2k} . Il est alors possible d'obtenir une estimation de l'ampleur du biais de sélection et d'estimer l'effet véritable du mode d'enquête sur le nombre de déplacements déclaré quotidiennement par les répondants.

II.4.1 Sens et intensité du biais de sélection

Il est difficile d'estimer le signe et l'importance de la corrélation entre les deux termes d'erreur. L'interprétation des coefficients $\rho_1\sigma_{u1}$ et $\rho_2\sigma_{u2}$ est donc complexe, mais intéressante. Les inverses du ratio de Mills (λ_1 ou λ_2) sont par définition positifs. En revanche, $\rho_1\sigma_{u1}$ et $\rho_2\sigma_{u2}$ peuvent prendre tous les signes. Si $\rho_1\sigma_{u1}$ est négatif, alors les termes d'erreurs des équations de sélection et d'intérêt sont négativement corrélés. Il existe des facteurs inobservés qui font que les individus ne peuvent répondre en face-à-face, mais remplissent le questionnaire en ligne, et qui impactent négativement la mobilité. Cela signifie que le nombre de déplacements déclaré pourrait être en moyenne significativement plus élevé si ces individus avaient répondu en face-à-face. Par analogie, si $\rho_2\sigma_{u2}$ est négatif, le nombre de déplacements déclarés par les individus en face-à-face aurait été en moyenne significativement inférieur si ces derniers avaient choisi le web comme mode d'enquête. L'interprétation est inversée si $\rho_1\sigma_{u1}$ et $\rho_2\sigma_{u2}$ sont positifs.

Notons qu'il est possible d'obtenir des estimateurs efficaces du coefficient de corrélation ρ (entre les résidus u_{1i} et ϵ_i d'une part, u_{2i} et ϵ_i d'autre part), ainsi que de la covariance σ , dont le produit est le paramètre de l'inverse du ratio de Mills. Même si l'hypothèse $\rho = 0$ ne peut être acceptée, les estimateurs par la méthode des moindres carrés (dans la deuxième étape) restent non biaisés. L'efficacité des estimations des paramètres dépend de l'hypothèse de la distribution normale et bivariée des résidus des deux étapes.

II.4.2 Conclusion : Les limites de la méthode en deux étapes

La première étape de la méthode consiste à estimer l'équation de sélection à l'aide d'un modèle probit, pour obtenir des estimations des δ_m . Pour chaque observation sélectionnée, le modèle calcule la valeur λ_{1i} ou λ_{2i} (inverse du ratio de Mills). Dans une seconde étape, on estime les paramètres β_{1k} et $\rho_1\sigma_{u1}$, par une régression des moindres carrés ordinaires de Y_1 sur X_{1k} et λ_1 , et les paramètres β_{2k} et $\rho_2\sigma_{u2}$, par, une régression des moindres carrés ordinaires de Y_2 sur X_{2k} et λ_2 . Les équations du modèle de mobilité Y_1 et Y_2 contiennent donc non seulement le vecteur de variables explicatives (respectivement X_{1k} et X_{2k}), mais aussi une nouvelle variable construite, ou inverse du ratio de Mills λ_1 et λ_2 .

Le paramètre qui fait que le modèle de sélection proposé par Heckman diffère d'un modèle probit suivi d'un modèle de régression linéaire est l'existence d'un coefficient de corrélation (ou covariance) entre les termes d'erreur des équations de sélection et d'intérêt (Verbeek, 2004). Il est alors nécessaire de supposer que des facteurs inobservables jouent à la fois sur le niveau de mobilité des individus et sur le média utilisé pour répondre à l'enquête.

Si la méthode en deux étapes d'Heckman est largement utilisée pour traiter du biais de sélection, certaines critiques ont été formulées à son égard. Le modèle d'Heckman souffre de plusieurs difficultés. La première concerne l'hypothèse de normalité (Lee, 1982). En effet, les estimations des paramètres semblent très sensibles aux distributions des termes d'erreurs des équations de sélection et d'intérêt du modèle. La littérature propose des approches alternatives, fondées sur des estimateurs non paramétriques, permettant de s'affranchir de l'hypothèse de normalité. Les résultats obtenus diffèrent cependant peu du modèle paramétrique de Heckman (Greene, 2002), et les hypothèses moins fortes du modèle génèrent des résultats moins robustes (Winship et Mare, 1992).

Par ailleurs, on peut expliquer une différence entre les estimations des paramètres par la méthode en deux étapes et celles obtenues par la méthode des moindres carrés ordinaires, par une forte colinéarité entre les régresseurs et les variables manquantes λ_1 et λ_2 . La méthode en deux étapes est donc un compromis entre un biais de sélection et une erreur due à l'introduction d'un régresseur fortement corrélé aux variables explicatives du modèle (Stolzenberg et Relles, 1997). Afin de respecter l'indépendance des variables de conditionnement Z_m et des résidus de seconde étape e_1 et e_2 , il est préférable de trouver des variables de conditionnement suffisamment indépendantes du niveau de mobilité ¹²⁵.

Le ratio de Mills est très sensible à la colinéarité pouvant exister entre les deux équations et les variables explicatives de l'équation d'intérêt X_{ki} sont souvent les mêmes que celle de l'équation de sélection Z_{mi} . Ceci peut conduire

¹²⁵Lorsque le vecteur Z est entièrement contenu dans X_1 et X_2 , l'identification de σ et ρ repose entièrement sur la non-linéarité de l'inverse du ratio de Mills. La contrainte de non-linéarité peut être relâchée, en trouvant des variables explicatives au sein de Z qui soient distinctes des variables des vecteurs X_1 et X_2 .

III Modèle explicatif ne tenant pas compte du biais de sélection

à des procédures de restriction des variables explicatives, la solution idéale étant de différencier totalement les variables indépendantes de l'équation de sélection de celles de l'équation de niveau.

Un autre inconvénient est que l'inverse du ratio de Mills est traité comme une variable explicative dans la régression de la seconde étape, alors qu'il correspond à une partie des résidus. Il n'est donc pas évident d'interpréter le signe du paramètre de l'inverse du ratio de Mills, en particulier lorsque la régression est réalisée sur l'ensemble des individus (web et face-à-face). La méthode de Heckman reste néanmoins recommandée pour tester l'existence d'un biais de sélection (Davidson et McKinnon, 1993).

Enfin, il est utile de rappeler que la sélection des répondants peut être fondée sur les facteurs observables ou inobservables. C'est la richesse des données d'enquêtes qui détermine quels facteurs sont observés et lesquels ne le sont pas. Ainsi, plus les données disponibles sont riches et plus la part du biais de sélection imputable à des facteurs inobservables est réduite. Bien qu'il existe toujours des facteurs inobservés qui influent sur la sélection pour le mode d'enquête en ligne, cela n'implique pas nécessairement que la simple comparaison des répondants web et des répondants face-à-face mettra en évidence un biais de sélection. Dans notre étude, il y a un biais de sélection si et seulement si les facteurs inobservés qui influent sur la participation ou non à l'enquête web impactent également le nombre de déplacements déclarés par les individus interrogés.

III Modèle explicatif ne tenant pas compte du biais de sélection

Dans cette section, nous cherchons à identifier, par enquête, les principaux facteurs susceptibles d'expliquer les différences de mobilité entre les individus, sans tenir compte de l'existence éventuelle d'un biais de sélection. Nous définissons la mobilité quotidienne comme le nombre de déplacements moyen effectué par les individus durant une journée. Pour mener notre analyse, nous utilisons un modèle de régression multiple. Un modèle est un moyen de représenter une réalité, c'est-à-dire d'en donner une représentation simplifiée et éventuellement de simuler ce qui peut survenir si un facteur est modifié.

Nous présentons d'abord les variables disponibles pour l'analyse (section III.1), avant de procéder à une régression linéaire multivariée, permettant d'analyser l'influence des facteurs retenus sur la mobilité quotidienne (section III.2).

III.1 Les variables disponibles pour l'analyse de la mobilité

Dans les modèles de régression multiple, on cherche à déterminer la valeur d'une variable numérique, dite « à expliquer » ou « dépendante », par les

valeurs de plusieurs autres variables, dites « explicatives ». Le choix des variables explicatives se fait a priori, en fonction des éléments bibliographiques et empiriques publiés sur le sujet et selon le bon sens de l'analyste.

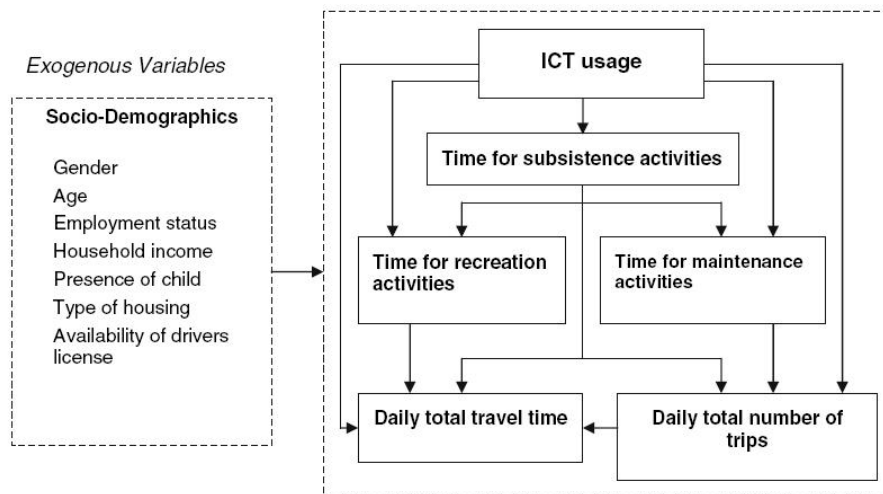


FIG. 59 – Hypothèse de relations entre l'usage des nouvelles technologies, les caractéristiques socio-économiques et le comportement de mobilité des répondants

Source : Wang et Law (2007)

La complexité du phénomène à expliquer, la mobilité quotidienne, conduit à prendre en compte de nombreuses dimensions susceptibles d'influer sur le nombre de déplacements déclaré, comme le montre à titre d'exemple la figure 59. La progression de l'individu dans le cycle de vie dépend en grande partie de son âge, et s'accompagne d'un certain nombre de caractéristiques socio-économiques (passage du permis, entrée dans la vie active, accès à la motorisation, arrivée des enfants, changement de lieu de résidence ...) qui vont impacter plus ou moins son niveau de mobilité. Face au grand nombre de variables disponibles dans l'enquête déplacements, le choix des variables pertinentes à inclure dans le modèle est difficile. Les premières analyses ont mis en évidence l'importance de dix sept d'entre-elles. Quelques statistiques descriptives sont présentées en annexe XIII.

III.1.1 Variables sociodémographiques

Sexe : homme ou femme.

Age : âge de chaque personne.

Nombre d'enfants du ménage : nombre d'enfants de moins de 18 ans présents dans le ménage.

Nombre de personnes du ménage : nombre de personnes qui composent le ménage.

Revenus déclarés : variable dichotomique, indiquant si l'individu a déclaré le niveau de revenus annuels nets de son ménage.

III Modèle explicatif ne tenant pas compte du biais de sélection

Activité : variable combinant le statut du répondant et le lieu de travail des actifs (travailleurs à temps plein ou à temps partiel, apprentis, personnes en formations et en stage). Quatre modalités sont disponibles : les non actifs, les actifs travaillant à Lyon ou à Villeurbanne, les actifs dont le lieu de travail se situe dans la reste de l'agglomération et les actifs n'ayant pas communiqué la zone de leur lieu de travail.

Distance du domicile au centre de l'agglomération : distance entre la zone de résidence du ménage et la zone centrale de l'agglomération (par hypothèse, la zone 10203, Bellecour), exprimée en m.

Densité de la zone de résidence : densité de la zone de résidence du ménage, exprimée en nombre d'habitants par kilomètre carré.

Il semble a priori aisé de mettre en évidence l'effet de certains facteurs tels que le revenu du ménage ou la catégorie socioprofessionnelle du répondant sur la mobilité individuelle. Cependant la variable 'Profession et catégorie socioprofessionnelle' est difficile à exploiter dans notre modèle, du fait des effectifs très inégaux et parfois faibles de certaines catégories. Il n'est également pas envisageable d'introduire la variable 'Revenu' dans le modèle, la non-réponse partielle étant très élevée (1 personne sur 3 en face-à-face et 1/4 sur le web). Comme le précisent Bonnafous et Puel (1983), un facteur regroupe les deux variables évoquées ci-dessus et montre un effet non négligeable sur la mobilité. Il s'agit du niveau d'étude ou de diplôme, lié positivement au nombre de déplacements.

Niveau d'étude : niveau d'étude de la personne, en trois modalités : en cours d'études, études non supérieures (allant de pas d'études, jusqu'au BAC) et études supérieures (BAC+2 et plus).

Diplôme : dernier diplôme obtenu par la personne, en deux modalités : non supérieur (allant jusqu'au BAC) et supérieur (BAC+2 et plus).

III.1.2 Variables caractéristiques de l'équipement en télécommunication

Téléphone portable : possession d'un téléphone portable par la personne, à titre personnel ou professionnel.

Liste de téléphone : la variable est scindée en trois modalités, selon que l'individu ne possède pas de téléphone fixe, est inscrit sur la liste rouge ou orange ou est inscrit sur l'annuaire des abonnés.

Connexion internet : possession d'une connexion internet au domicile du ménage.

III.1.3 Variables caractéristiques de la mobilité

Le nombre de voitures du ménage et la possession du permis constituent une bonne indication de la situation des répondants.

Nombre de voitures du ménage : nombre de voitures particulières possédées par le ménage, rapporté au nombre de personnes en âge de conduire (18 ans et plus).

Possession du permis : possession du permis de conduire ou pratique de la conduite accompagnée.

Vendredi : variable qui prend la valeur "1" si le jour de référence pour le recueil des déplacements est le vendredi, et la valeur "0" sinon

III.1.4 Variables caractéristiques du choix du mode de réponse

Mode de réponse : variable qui prend la valeur "1" si le mode de recueil des données est le web, et la valeur "0" pour le face-à-face.

III.2 Analyse de la mobilité par un modèle de régression linéaire multiple

Considérons l'équation (57), qui permet d'examiner l'effet du mode d'enquête sur le nombre de déplacements quotidiens moyen d'un individu, à l'aide d'une régression linéaire multiple, de la forme :

$$Y_i = \sum_k \beta_k X_{ki} + \alpha I_i + u_i \quad (57)$$

Avec Y_i le nombre moyen de déplacements réalisé par les individus (variable dépendante), X_i un vecteur de variables explicatives et I_i une variable muette indiquant si l'individu a répondu sur internet. Nous formulons dans un premier temps des hypothèses sur la forme des variables explicatives à inclure dans le modèle de régression. Puis, nous traitons le problème de la colinéarité de certaines d'entre-elles, avant d'analyser en détail les résultats des modèles. Nous utiliserons d'abord la base de données dans son intégralité, avant de limiter notre analyse à chacun des deux sous-échantillons (web et face-à-face). Les internautes ont en effet des caractéristiques particulières (Cf. chapitres 5 à 7), qui justifient le développement d'un modèle explicatif plus spécifique.

III.2.1 Hypothèses sur la forme des variables explicatives

Nous savons que le niveau de mobilité est plus faible chez les individus mobiles ayant répondu sur le web (3,78 vs. 4,19 déplacements en face-à-face ; p-value < 0,001%). Suite aux analyses exploratoires menées sur les échantillons web et face-à-face, nous pouvons formuler des hypothèses concernant l'impact des variables explicatives retenues sur le nombre de déplacements quotidiens moyen des individus.

Effet du genre : des femmes plus mobiles que les hommes A priori, les femmes se déplacent davantage que les hommes, car elles réalisent quotidiennement un plus grand nombre d'activités. En plus de leur activité professionnelle, elles assurent souvent l'accompagnement des enfants (école, garderie, activités

III Modèle explicatif ne tenant pas compte du biais de sélection

de loisirs...), ainsi que les achats et démarches administratives du ménage. Cette relation se vérifie dans les deux échantillons web et face-à-face (tableau 60), même si la différence est plus accentuée parmi les internautes.

Modalités	Nb de déplacements moyen	
	Enquête face-à-face	Enquête web
Homme	4,08	3,30
Femme	4,28	4,16
P-value	< 0,001%	< 0,001%

TAB. 60 – Nombre de déplacements moyen par genre et par enquête

Effet de l'âge : une relation non linéaire L'âge est lié au cycle de vie des individus, donc à la composition du ménage.

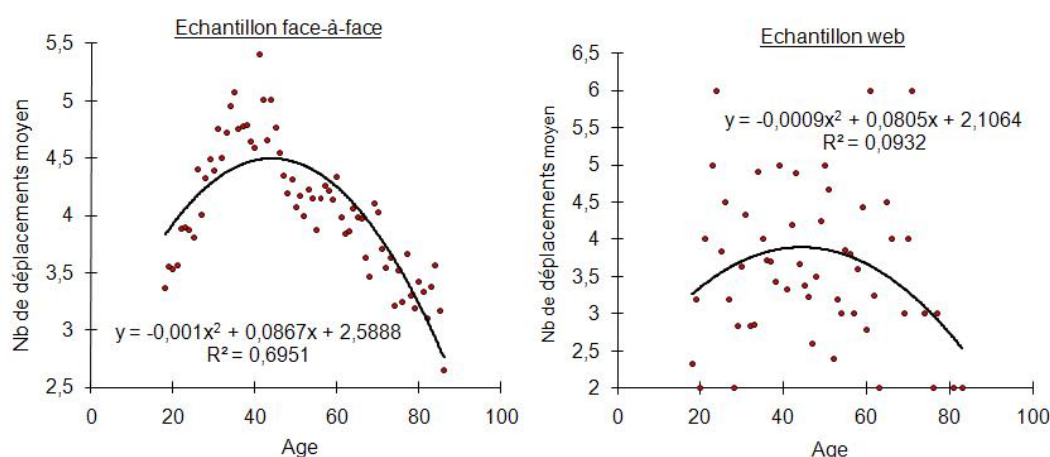


FIG. 60 – Relation entre l'âge des répondants et le nombre de déplacements moyen par enquête

Source : EMD Lyon (2006)

Si nous retenons les personnes âgées de 86 ans maximum dans l'échantillon face-à-face (pour avoir des effectifs suffisants) nous pouvons conclure que le nombre de déplacements augmente avec l'âge jusqu'à un maximum (situé entre 40 et 50 ans), puis décroît par la suite. Nous retrouvons alors une interprétation traditionnelle qui suppose qu'avec l'entrée dans la vie active et l'arrivée des enfants la mobilité s'accroît, mais qu'à partir d'un certain seuil, l'avancée en âge se traduit par une sédentarisation des individus. Cette relation est forte pour l'échantillon en face-à-face ($R^2 = 0,70$). Pour l'échantillon web, les effectifs par âge sont toutefois peu élevés et le nuage de point indique une relation moins significative entre l'âge et le nombre de déplacements ($R^2 = 0,09$) (figure 60).

Effet de la structure du ménage : un impact fort sur la mobilité Nous savons que la mobilité est influencée par la taille du ménage, mais l'impact du nombre de personnes du ménage sur la mobilité est ambigu.

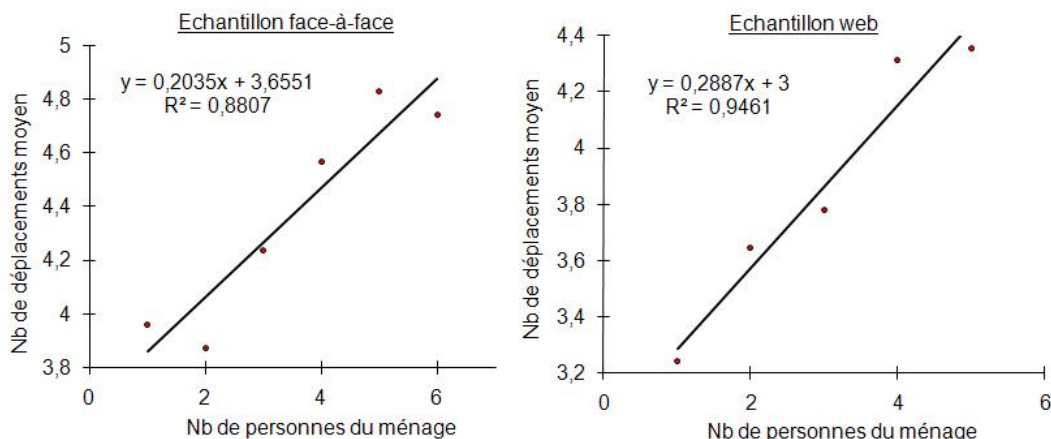


FIG. 61 – Relation entre le nombre de personnes du ménage et le nombre de déplacements moyen par enquête

Source : EMD Lyon (2006)

Si nous retenons les ménages de 1 à 6 personnes en face-à-face et de 1 à 5 personnes sur le web (pour les familles plus nombreuses, les effectifs sont trop faibles pour être considérés), alors le nombre de déplacements augmente avec le nombre de personnes du ménage. La relation est significative, quel que soit l'échantillon ($R^2 = 0,88$ en face-à-face et $R^2 = 0,95$ sur le web) (figure 61). L'augmentation de la taille du ménage s'accompagne donc d'un accroissement des besoins de mobilité. Cette hypothèse se retrouve dans les travaux de (Hanson, 1982), où l'augmentation de la taille des ménages va de pair avec l'augmentation du nombre de déplacements ¹²⁶. Nous pouvons toutefois poser l'hypothèse que les déplacements contraints sont mieux répartis lorsque la taille du ménage est importante.

Les programmes d'activités quotidiens déterminent les stratégies de mobilité des individus. Ces activités sont souvent obligatoires, se déroulent en des temps et lieux plus ou moins contraints, et sont déterminées en coordination avec les autres membres du ménage (accompagnement des enfants par exemple) (Kaufmann *et al.*, 2005; Orfeuil, 2000). Les enfants ont beaucoup d'activités qui ne sont pas liées au travail ou à l'école et qui nécessitent la participation d'un adulte de ménage. Les habitudes de déplacements des enfants influencent donc celles des parents et autres adultes du ménage. Par ailleurs, les enfants dont les parents sont actifs et ont un haut niveau de revenus ou d'éducation, ont davantage de chance de participer à des activités extérieures (Mohammadian et Bekhor, 2008).

Les analyses exploratoires laissent penser que le nombre de déplacements

¹²⁶Selon Pouyanne (2004), au lieu de se contenter d'une influence directe de la composition familiale sur la mobilité, on peut adopter l'hypothèse que la composition familiale détermine un type de localisation (zones périphériques et peu denses de l'agglomération), qui vient influencer sur les comportements de mobilité. Par ailleurs, les retraités, dont le niveau de mobilité est plus faible que celui observé pour l'ensemble de la population, sont sur-représentés dans les ménages de une à deux personnes. La corrélation entre la taille du ménage et la mobilité n'est donc pas forcément transformable en un lien causal direct.

III Modèle explicatif ne tenant pas compte du biais de sélection

augmente avec le nombre d'enfants du ménage jusqu'à un certain seuil (entre 2 et 4 enfants), puis décroît ensuite ¹²⁷. Cette relation est significative quel que soit l'échantillon considéré ($R^2 = 0,95$ en face-à-face et $R^2 = 0,81$ sur le web) (figure 62).

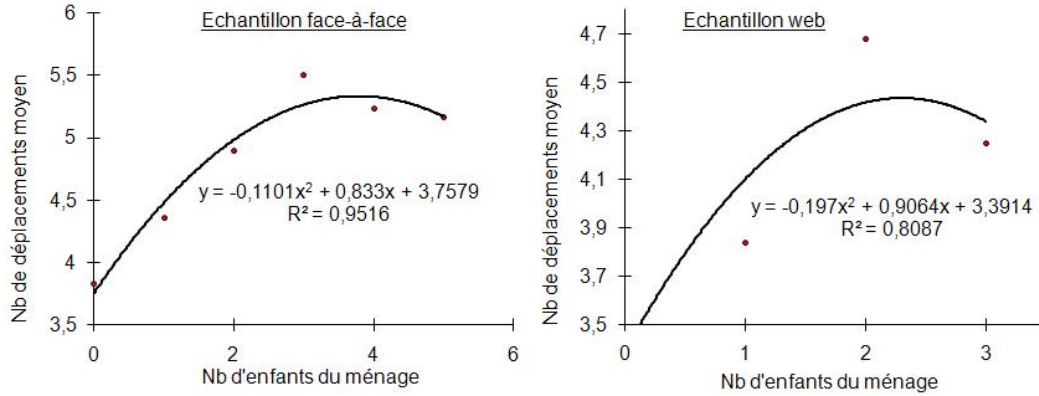


FIG. 62 – Relation entre le nombre d'enfants du ménage et le nombre de déplacements moyen par enquête

Source : EMD Lyon (2006)

Effet du niveau d'éducation : le niveau d'étude ou le diplôme acquis influencent positivement la mobilité Le niveau d'étude est lié au revenu et au statut et semble corrélé positivement avec la mobilité. Il en est de même pour la variable 'Diplômes' ¹²⁸ (tableau 61). Cette différence est significative quel que soit le mode d'enquête (p-value < 0,001%).

Modalités	Nb de déplacements moyen	
	Enquête face-à-face	Enquête web
Niveau d'étude : supérieur	4,53	3,95
Niveau d'étude : non supérieur	4,07	3,23
Niveau d'étude : en cours	3,54	4,43
P-value	< 0,001%	< 0,001%
Diplôme : supérieur	4,39	4,01
Diplôme : non supérieur	4,04	3,20
P-value	< 0,001%	< 0,001%

TAB. 61 – Nombre de déplacements moyen par niveau d'éducation et par enquête

Suivant les modèles utilisés, il conviendra d'utiliser l'une ou l'autre de ces variables. Il faut toutefois noter que la variable 'niveau d'étude' est partielle-

¹²⁷Nous avons sélectionné les ménages de l'enquête en face-à-face de 5 enfants maximum, afin de garder des effectifs suffisants.

¹²⁸La variable 'Diplômes' est plus synthétique que la variable 'niveau d'étude', puisque les individus en cours d'études se voient affecter la modalité 'supérieur' ou 'non supérieur', en fonction du type de formation suivie.

ment reliée à l'âge, les personnes en cours d'études étant majoritairement des jeunes (moyenne d'âge de 20,7 ans en face-à-face et de 22,2 ans sur le web).

Effet du niveau d'activité : des inactifs davantage mobiles A priori, le niveau d'activité impacte positivement le temps passé sur le lieu de travail et le temps de déplacement quotidien total. En revanche, le temps disponible pour les activités dites de maintenance (achats, démarches ...) et de divertissement (loisirs) est plus restreint ¹²⁹. De nombreuses études mettent cependant en évidence une mobilité supérieure pour les actifs, ces derniers cumulant les déplacements liés au travail avec ceux nécessaires au déroulement de la vie quotidienne. L'analyse des données montre que les actifs internautes se déplacent moins que les inactifs, mais la relation est inversée dans le cas de l'enquête en face-à-face. Ces différences sont significatives, quel que soit le mode d'enquête (p-value < 0,001%) (tableau 62). Elles peuvent s'expliquer en partie par la structure des échantillons. Dans l'enquête face-à-face, les retraités, peu mobiles, représentent une large proportion des inactifs. Ceci n'est pas vrai dans l'échantillon web, peu de répondants étant à la retraite. Les inactifs internautes sont donc en âge d'être "actifs" (présence d'enfants...) et ont un niveau de mobilité supérieur.

Modalités	Nb de déplacements moyen	
	Enquête face-à-face	Enquête web
Actifs	4,37	3,67
Inactifs	3,96	4,14
P-value	< 0,001%	< 0,001%

TAB. 62 – Nombre de déplacements moyen par niveau d'activité et par enquête

Effet de la localisation résidentielle : des disparités selon le mode d'enquête Les localisations résidentielles possibles pour un type de famille donné obéissent à un système de contraintes (accès aux transports collectifs, budget temps de transport quotidien, motorisation...). Il y a donc une forte dépendance entre la localisation résidentielle des ménages et leur forme de mobilité, relativement aux activités possibles (Orfeuil, 2002a; Ettema *et al.*, 1996) ¹³⁰. L'éloignement géographique du domicile par rapport au centre de l'agglomération joue a priori un rôle positif sur la mobilité, puisque la dispersion des activités en périphérie occasionne davantage d'accompagnements. Cependant, certains auteurs avancent que la mobilité est globalement supérieure au centre, en raison de l'importance de la marche à pied, et du rôle des transports en commun (Bonafous et Puel, 1983).

¹²⁹Les activités liées à l'accompagnement ne devraient pas souffrir du niveau d'activité, les actifs devant souvent gérer le transport de leurs enfants en plus de leur activité professionnelle, bien que les accompagnements peuvent être inclus dans des sorties complexes.

¹³⁰L'impact de l'éloignement du lieu de résidence sur la mobilité peut néanmoins être modéré par l'accès à l'automobile. Ettema *et al.* (2007) précisent qu'en cas de bi-motorisation, les adultes sont moins dépendants et peuvent participer à davantage d'activités hors du domicile sans procéder à des arbitrages à l'intérieur du ménage.

III Modèle explicatif ne tenant pas compte du biais de sélection

Cette relation contradictoire entre la distance du domicile au centre de l'agglomération et le nombre de déplacements déclaré semble se vérifier dans les échantillons web et face-à-face. Dans l'échantillon en face-à-face, la mobilité des individus diminue, jusqu'à un éloignement de la zone de résidence de 10km du centre de l'agglomération, et augmente ensuite ($R^2 = 0,72$). Nous constatons le contraire dans l'échantillon web ($R^2 = 0,90$). Il est intéressant de noter que, dans les deux cas, cette relation n'est pas linéaire, et s'inverse pour un seuil de 10km (figure 63).

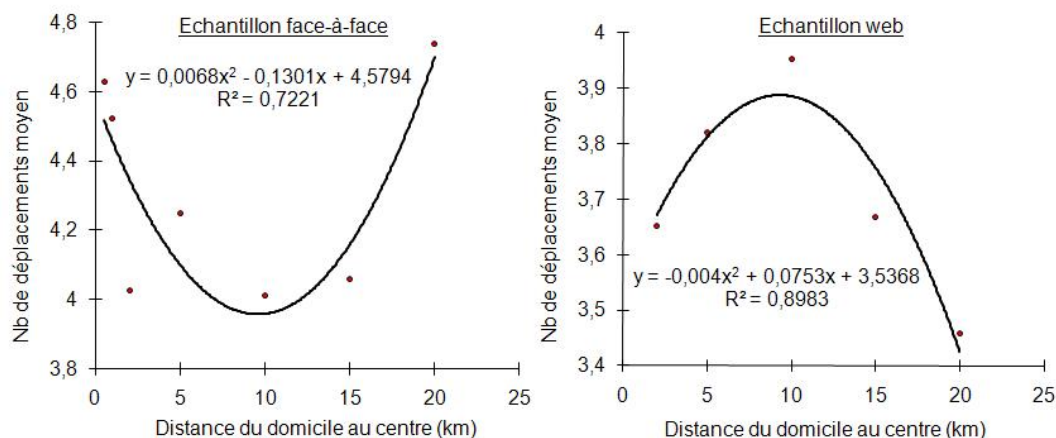


FIG. 63 – Relation entre la distance du domicile au centre de l'agglomération (km) et le nombre de déplacements moyen par enquête

Source : EMD Lyon (2006)

De façon analogue, la densité de la zone d'habitation conditionne la richesse des services et la qualité des transports publics qui peuvent être associés. Si certaines études montrent que le nombre de déplacements a tendance à augmenter quand la densité est faible et à diminuer lorsqu'elle est plus forte, il est cependant probable que les déplacements courts soient sous-comptabilisés dans les zones les plus peuplées, surtout s'ils font partie d'une chaîne de déplacements complexe (Banister, 1997).

Effet de l'équipement du ménage : le rôle non négligeable du téléphone portable Concernant les équipements du ménage, la possession d'un téléphone portable semble liée positivement à la mobilité. Cet équipement permet en effet de rester en relation avec de nombreux contacts personnels, et les personnes avec un haut niveau d'interaction sociale ont tendance à se déplacer davantage. Par ailleurs, dans la sphère professionnelle, le téléphone portable est souvent confié aux salariés ayant à se déplacer dans le cadre de leur mission. Cette relation semble se vérifier de façon significative dans les deux échantillons (tableau 63).

Impact du mode sur la mesure des comportements de mobilité

Modalités	Nb de déplacements moyen	
	Enquête face-à-face	Enquête web
Téléphone portable : oui	4,31	3,83
Téléphone portable : non	3,84	3,75
P-value	< 0,001%	< 0,001%

TAB. 63 – Nombre de déplacements moyen par possession d’un téléphone portable et par enquête

Selon Hjorthol (2002), l’accès et l’utilisation de l’ordinateur à la maison n’a pas, a priori, d’effet direct sur la mobilité. Les personnes équipées d’une connexion web à domicile font moins de déplacements pour le travail, mais reprogramment leurs activités et conservent au final le même nombre de déplacements quotidiens. Nous ne prenons pas non plus en compte les variables ‘liste de téléphone’ et ‘connexion internet’ dans le modèle explicatif de la mobilité. Nous verrons dans la section suivante que ces variables influencent davantage le choix du web comme média de réponse que le niveau de mobilité individuelle.

Effet de la motorisation : un accès à l’automobile qui encourage la mobilité Hanson (1982) montre que le niveau de motorisation individuel (nombre de voitures du ménage par personne de 18 ans et plus) et le fait de posséder le permis de conduire sont corrélés positivement à la mobilité. Les personnes qui ne possèdent pas le permis ou qui n’ont pas de véhicule à disposition sont en effet plus contraintes dans leurs déplacements, même si d’autres facteurs peuvent intervenir. Ces relations se vérifient dans les deux échantillons. L’impact de la possession du permis sur la mobilité semble toutefois moins important dans l’échantillon web (tableau 64).

Modalités	Nb de déplacements moyen	
	Enquête face-à-face	Enquête web
Possession permis : oui	4,30	3,81
Possession permis : non	3,47	3,34
P-value	< 0,001%	< 0,001%

TAB. 64 – Nombre de déplacements moyen par possession du permis de conduire et par enquête

En ce qui concerne le nombre de voitures à disposition, la relation est très forte dans l’échantillon face-à-face ($R^2 = 0,72$), mais moins prononcée pour l’échantillon web ($R^2 = 0,32$), probablement à cause des plus faibles effectifs (figure 64). La variable ‘Nombre de voitures par personne de 18 ans et plus’ est traditionnellement liée au niveau de revenu du ménage. Cette variable est donc intéressante, puisque si le revenu est un déterminant majeur de la mobilité (Pouyanne, 2004), le fort taux de non-réponse partielle dans les questionnaire web et face-à-face ne nous permet pas de la prendre en considération.

III Modèle explicatif ne tenant pas compte du biais de sélection

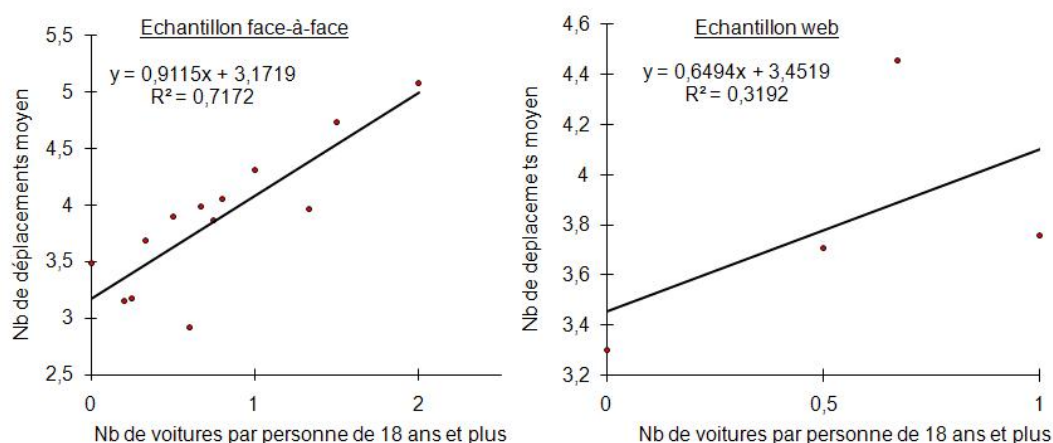


FIG. 64 – Relation entre le nombre de voitures du ménage par personne de 18 ans et plus et le nombre de déplacements moyen par enquête

Source : EMD Lyon (2006)

Le jour des déplacements et le fait de déclarer le niveau de revenu du ménage : deux facteurs explicatifs de la mobilité Nous pouvons supposer que la réduction du temps de travail profite aux autres activités et génère davantage de déplacements pour motifs achats et loisirs. Nos premiers tris confirment cette hypothèse, et montrent un accroissement des déplacements le vendredi. Cette relation est très significative dans l'enquête en face-à-face, mais semble s'inverser dans l'enquête web (tableau 65). Ceci peut s'expliquer par le fait que les internautes qui choisissent de saisir leurs réponses le week-end ou le lundi peuvent avoir oublié leur emploi du temps du vendredi, jour de référence des déplacements.

Modalités	Nb de déplacements moyen	
	Enquête face-à-face	Enquête web
Vendredi : oui	4,40	3,68
Vendredi : non	4,15	3,84
P-value	< 0,001%	< 0,001%

TAB. 65 – Nombre de déplacements moyen effectué le vendredi par enquête

Par ailleurs, nous observons que la mobilité est plus importante si la personne choisit de communiquer le niveau de revenus annuels nets du ménage (p-value < 0,001%). Il est probable que ces répondants doivent faire face à moins de freins pour délivrer des réponses précises concernant leurs déplacements (bien que le fait de dévoiler le niveau de salaire ne soit pas une variable directement liée au niveau de mobilité). Nous vérifions cette relation dans les deux échantillons (tableau 66), les personnes acceptant de déclarer leurs revenus étant plus mobiles que les autres.

Modalités	Nb de déplacements moyen	
	Enquête face-à-face	Enquête web
Revenus déclarés : oui	4,32	3,83
Revenus déclarés : non	3,94	3,61
P-value	< 0,001%	< 0,001%

TAB. 66 – Nombre de déplacements moyen effectué par enquête, selon que l’individu déclare ou non son niveau de revenus

III.2.2 Le problème de la colinéarité des variables explicatives

L’existence d’une colinéarité ¹³¹ entre les variables explicatives d’un modèle de régression linéaire est un problème fréquent. Il est donc nécessaire de s’assurer que les variables explicatives d’une régression multiple ne se répètent pas entre elles, au risque d’aboutir à un modèle instable (des coefficients peuvent avoir un signe opposé à celui attendu), et donc peu utile (Greene, 2002). Pour que celui-ci soit robuste, une des solutions consiste à étudier la matrice des corrélations. Il s’agit plus simplement de la matrice des coefficients de corrélation, calculés sur plusieurs variables prises deux à deux. Si cette dernière indique une très forte liaison entre deux variables explicatives (corrélation positive ou négative), c’est que l’une d’elles est de trop et ne peut être intégrée dans le modèle. Le choix qui consiste à réduire le nombre de variables explicatives d’un modèle pour éviter de rencontrer des problèmes de multicollinéarité n’est cependant pas toujours judicieux, et peut amener à mal spécifier le modèle ¹³². L’omission de variables explicatives pertinentes peut en effet biaiser les résultats, puisqu’en surestimant la variance, il est possible de conclure à tort à la significativité de certains coefficients, qui ne le seraient pas si les variables omises avaient été intégrées (Greene, 2002).

Le coefficient de corrélation de Pearson Dans le cadre d’une régression linéaire simple, on ‘résume’ un nuage de points par une droite, dite de régression. On mesure la qualité de la régression simple par le coefficient de corrélation linéaire de Pearson, rapport de la covariance entre x et y au produit des écarts-types empiriques. Soit :

$$r = \frac{\sigma_{xy}}{\sigma_x \sigma_y} \tag{58}$$

Ce coefficient est compris entre -1 et 1. Un signe négatif indique que y varie en sens inverse de x et plus le coefficient est proche de 0, plus les deux variables sont linéairement indépendantes. Le niveau de précision du risque d’erreur sur l’estimation de la corrélation dépend du nombre de degrés de liberté du modèle (Saporta, 2006).

¹³¹Les vecteurs u et v sont colinéaires s’il existe un réel k tel que u = kv.

¹³²Une solution consiste à construire des variables synthétiques, qui agrègent plusieurs variables fortement corrélées.

Quelques exemples de colinéarité des variables explicatives La multicollinéarité est toujours plus ou moins présente dans les jeux de données et la plupart des variables issues de l'enquête ménages déplacements que nous avons retenues pour l'analyse de la mobilité sont de près ou de loin corrélées entre-elles. Par exemple, l'éloignement du lieu de résidence en périphérie dépend d'un ensemble de variables liées au cycle de vie du ménage (effet de l'âge, du revenu, de la taille du ménage, de sa motorisation...). Pouyanne (2004), constate par ailleurs que la corrélation positive entre le niveau de diplôme et le nombre de déplacements est le fruit de l'effet localisation des diplômés du supérieur dans des zones denses (pour profiter des aménités ou par volonté de se rapprocher des fonctions dites 'supérieures', davantage présentes dans l'hypercentre), où la mobilité est supérieure.

Parmi les variables explicatives sélectionnées dans notre modèle, deux associations semblent particulièrement délicates : la densité de la zone de résidence et son éloignement du centre de l'agglomération (coefficient de Pearson = -0,594; p-value < 0,001%), d'une part, le nombre de personnes du ménage et le nombre d'enfants du ménage (coefficient de Pearson = 0,784; p-value < 0,001%), d'autre part.

Il existe a priori une colinéarité entre la densité de la zone de résidence et son éloignement par rapport au centre de l'agglomération. La forme de la relation reste ambiguë et ne peut être considérée comme linéaire (le coefficient de détermination n'est pas très important : 0,35 en face-à-face et 0,41 sur le web). Selon les figures 65 et 66, l'élasticité entre la densité de la zone de résidence et son éloignement du centre de l'agglomération n'est pas constante. Quelle que soit l'enquête, la densité de la zone de résidence décroît fortement pour un éloignement du centre de l'agglomération compris entre 0 et 3 kilomètres, puis semble se stabiliser au-delà.

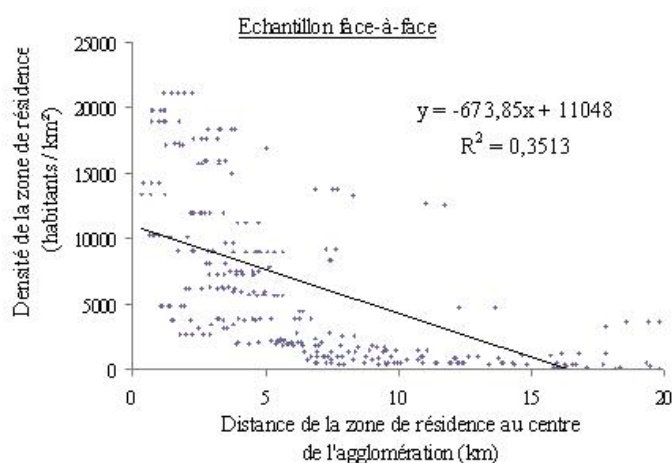


FIG. 65 – Relation entre la distance du domicile au centre de l'agglomération et la densité de la zone de résidence pour l'échantillon face-à-face

Source : EMD Lyon (2006)

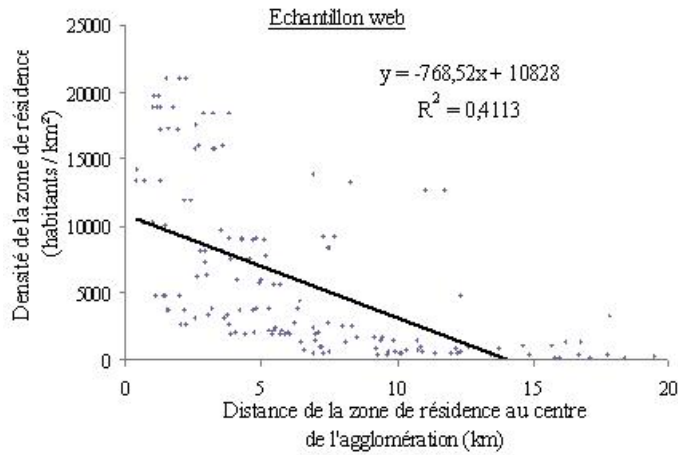


FIG. 66 – Relation entre la distance du domicile au centre de l’agglomération et la densité de la zone de résidence pour l’échantillon web

Source : EMD Lyon (2006)

Par ailleurs, comme le montre la figure 67, les variables 'Nombre de personnes du ménage' et 'Nombre d'enfants de moins de 18 ans du ménage' sont fortement corrélées (R^2 proche de 1 pour les deux échantillons). Cette corrélation peut poser un vrai problème de multicollinéarité dans le cadre des modèles de régression statistique. Suivant les échantillons concernés et les analyses réalisées, nous utiliserons donc l’une ou l’autre de ces variables.

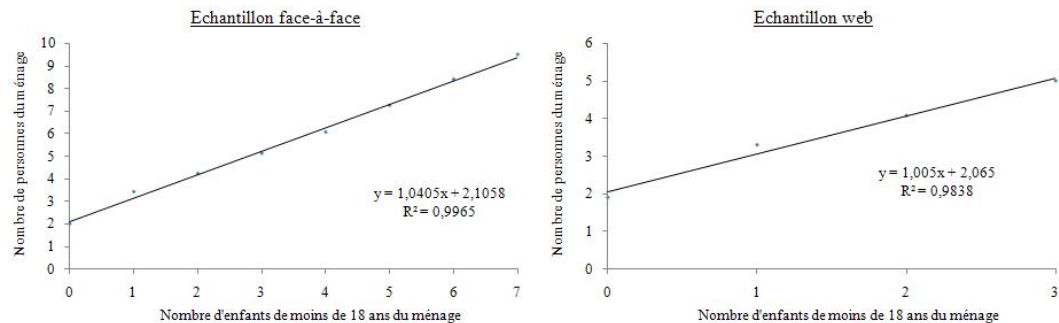


FIG. 67 – Relation entre le nombre d’enfants et le nombre de personnes du ménage

Source : EMD Lyon (2006)

III.2.3 Présentation des résultats du modèle de régression linéaire multiple

Nous présentons les variables explicatives retenues par le modèle, les paramètres estimés, leur écart-type et leur niveau de significativité, ainsi que la valeur du t de Student.

Ensemble des répondants (face-à-face et web) L’introduction des variables explicatives désignées ci-dessus dans le modèle de régression multiple

III Modèle explicatif ne tenant pas compte du biais de sélection

appliqué à l'ensemble de l'échantillon (web et face-à-face) donne les résultats présentés dans le tableau 67.

Toutes les variables présentées sont significatives à 1% et les coefficients ont un signe conforme à nos prévisions (section 3.2.1). Nous remarquons l'influence particulièrement forte de l'âge, du nombre de voitures à disposition, du niveau d'activité, du nombre d'enfants du ménage et du mode d'enquête. Le coefficient de la variable 'Mode' est négatif (-0,66), ce qui signifie qu'en moyenne un internaute déclare effectuer quotidiennement 0,7 déplacements de moins qu'un répondant en face-à-face possédant les mêmes caractéristiques socio-économiques.

Echantillon total	Coeff.	T-value	Pr(> t)	Sign.
Constante	2,45e+00	10,91	< 2e-16	***
Sexe : homme	-2,55e-01	-5,60	2,24e-08	***
Age	4,81e-02	5,49	4,05e-08	***
Age ²	-5,79e-04	-6,55	5,85e-11	***
Possession permis : oui	3,76e-01	5,01	5,45e-07	***
Vendredi : oui	2,01e-01	3,26	<0,01	**
Nb d'enfants / ménage	5,06e-01	8,98	< 2e-16	***
(Nb d'enfants / ménage) ²	-3,47e-02	-2,04	0,04	*
Nb de voiture / personne	6,05e-01	8,85	< 2e-16	***
Téléphone portable : oui	1,36e-01	2,39	0,02	*
Revenu déclaré : oui	3,29e-01	6,95	3,95e-12	***
niveau d'étude : en cours	-8,26e-01	-7,03	2,23e-12	***
niveau d'étude : non supérieur	-1,65e-01	-3,19	<0,01	**
Mode : web	-6,59e-01	-4,61	4,00e-06	***
Activité : non actif	4,86e-01	7,67	1,85e-14	***
Distance domicile / centre	-1,15e-01	-7,06	1,81e-12	***
(Distance domicile / centre) ²	6,01e-03	7,49	7,60e-14	***

Signif. : 0 '***' 0,001 '**' 0,01 '*' 0,05 '.' 0,1 ' ' 1

TAB. 67 – Régression linéaire appliquée à l'ensemble de l'échantillon (web et face-à-face)

Critères	Valeur
R ²	9,7%
F-stat	67,78
ddl	9946
P-value	< 2,2e-16

TAB. 68 – Significativité de la régression linéaire appliquée à l'ensemble de l'échantillon (web et face-à-face)

Seule 10% de la variance de la variable 'Nombre de déplacements' est expliquée par l'ensemble des variables explicatives qui interviennent dans le modèle (tableau 68). Il est donc fort probable que le niveau de mobilité individuelle soit influencé par d'autres variables, non connues de l'analyste. Cette faible qualité s'explique également par le caractère désagrégé de notre modèle et

Impact du mode sur la mesure des comportements de mobilité

surtout par une forte variabilité intra personnelle selon le jour d'enquête. Ce modèle est cependant significatif (p-value < 0,001%).

Echantillon face-à-face Le modèle appliqué au sous-échantillon de l'enquête face-à-face est quasiment le même que le précédent, comme le montre le tableau 69. Les signes des coefficients sont identiques et leurs valeurs absolues sont très proches. Cette similitude entre les deux modèles était prévisible, puisque les observations recueillies en face-à-face représentent 97% du total de notre échantillon. Nous avons pris soin d'enlever la variable 'mode', qui n'a plus de sens ici, tous les individus ayant répondu en face-à-face.

Echantillon face-à-face	Coeff.	T-value	Pr(> t)	Sign.
Constante	2,45e+00	10,78	<2e-16	***
Sexe : homme	-2,43e-01	-5,25	1,57e-07	***
Age	4,89e-02	5,52	3,46e-08	***
Age ²	-5,73e-04	-6,42	1,44e-10	***
Possession permis : oui	3,71e-01	4,90	9,82e-07	***
Vendredi : oui	2,13e-01	3,38	<0,01	***
Nb d'enfants / ménage	5,04e-01	8,81	<2e-16	***
(Nb d'enfants / ménage) ²	-3,39e-02	-1,97	0,05	*
Nb de voiture / personne	6,12e-01	8,81	<2e-16	***
Téléphone portable : oui	1,37e-01	2,37	0,02	*
Revenu déclaré : oui	3,32e-01	6,92	4,73e-12	***
niveau d'étude : en cours	-8,28e-01	-2,89	<0,01	**
Activité : non actif	4,74e-01	7,39	1,64e-13	***
Distance domicile / centre	-1,19e-01	-7,15	9,50e-13	***
(Distance domicile / centre) ²	6,22e-03	7,63	2,60e-14	***

Signif. : 0 '***', 0,001 '**', 0,01 '*', 0,05 '.', 0,1 ' ', 1

TAB. 69 – Régression linéaire appliquée à l'échantillon face-à-face

De même que précédemment, environ 10% de la variance de la variable 'Nombre de déplacements' est expliquée par l'ensemble des variables explicatives qui interviennent dans le modèle (tableau 70). Le niveau de significativité du modèle est en revanche très élevé (p-value < 2,2e-16).

Critères	Valeur
R2	9,7%
F-stat	70,56
ddl	9694
P-value	< 2,2e-16

TAB. 70 – Significativité de la régression linéaire appliquée à l'échantillon face-à-face

Echantillon web Beaucoup moins de variables ont des coefficients significatifs dans le modèle de régression appliqué à l'échantillon web (tableau 71).

III Modèle explicatif ne tenant pas compte du biais de sélection

Ceci s'explique essentiellement par les faibles effectifs, puisque seulement 369 individus ont rempli entièrement le questionnaire sur internet.

Comme nous l'avions supposé dans la section 3.2.1, la taille du ménage, l'inactivité, le niveau de formation et plus encore le nombre de voitures à disposition impactent positivement la mobilité des internautes. En revanche, le fait d'être un homme ou de résider loin du centre-ville pèse sur le nombre de déplacements déclaré. Le test de la variable 'Nombre de personnes' montre l'existence d'un lien très significatif entre la taille du ménage et la mobilité (p-value < 0,001%). Le niveau de formation et d'activité du répondant, ainsi que l'éloignement de son lieu de résidence sont des facteurs moins importants (significativité des coefficients plus faible : p-value < 10%). Etant donné les faibles effectifs de l'échantillon web et le risque d'erreur acceptable, nous les conservons tout de même dans notre modèle.

Echantillon web	Coeff.	T-value	Pr(> t)	Sign.
Constante	2,51	5,71	3,03e-08	***
Sexe : homme	-0,79	-3,30	<0,01	**
Nb de voiture / personne	0,74	2,02	0,04	*
Nb de personnes / ménage	0,34	3,59	<0,01	***
Diplômes : supérieur	0,46	1,65	0,10	.
Activité : non actif	0,48	1,75	0,08	.
Distance domicile / centre	-0,05	-1,73	0,08	.
Signif. : 0 '***' 0,001 '**' 0,01 '*' 0,05 '.' 0,1 ' ' 1				

TAB. 71 – Régression linéaire appliquée à l'échantillon web

Toutes les variables présentées sont significatives à 10% et le coefficient de détermination ajusté du modèle est légèrement meilleur (proche de 11%). Le modèle est toujours très significatif (p-value = 3,07e-06).

Critères	Valeur
R2	10,63%
F-stat	6,33
ddl	263
P-value	3,07e-06

TAB. 72 – Significativité de la régression linéaire appliquée à l'échantillon web

III.2.4 Premières conclusions

Nous avons analysé de manière empirique la relation entre la mobilité, définie par le nombre de déplacements déclaré, et un certain nombre de facteurs explicatifs (sociodémographiques, équipement du ménage, motorisation et mode d'enquête), conformément à l'équation (57). Nous pouvons cependant nous demander si le coefficient α de la variable 'Mode' mesure l'impact réel du mode d'enquête sur la mobilité quotidienne. La réponse est positive, si l'individu qui choisit de répondre sur le web déclare un nombre de déplacements identique à celui qu'il aurait déclaré en face-à-face. Cependant, la variable I ne

peut pas être considérée comme exogène dans ce modèle. Les individus contactés choisissent de répondre ou pas en face-à-face, et éventuellement acceptent de saisir leurs réponses en ligne, selon certaines caractéristiques, qui peuvent également avoir un impact sur leur niveau de mobilité. Par exemple, les ménages qui répondent en ligne sont ceux pour lesquels il n'a pas été possible de fixer un rendez-vous avec un enquêteur à domicile. On peut supposer qu'il s'agit de ménages peu disponibles (contraints par leur activité professionnelle par exemple), dont le manque de disponibilité impacte négativement la mobilité, et qui disposent d'une connexion internet pour accéder à l'enquête en ligne. En excluant systématiquement des individus de la même manière (individus n'ayant pas accès au web par exemple), on introduit le besoin d'un régresseur additionnel que la méthode des moindres carrés ordinaires ignore (Kmenta, 1971). L'auto-sélection des répondants doit donc être corrigée, afin d'obtenir des estimations non biaisées des coefficients des variables du modèle explicatif de la mobilité.

IV Modèle explicatif incluant le biais de sélection

Nous avons détaillé en section II les fondements économétriques des modèles de régression multivariés qui tiennent compte du biais de sélection. Dans cette section, nous analysons plus précisément les résultats des estimations du modèle d'Heckman en deux étapes, menées sur l'échantillon des répondants à l'enquête ménages déplacements de Lyon (face-à-face et web). Dans la première étape du modèle d'Heckman, nous estimons pour chaque répondant la probabilité de ne pas participer à l'interview en face-à-face et d'utiliser le web pour remplir le questionnaire, à l'aide d'un modèle probit (section III.1). Puis, dans une seconde étape, on estime l'équation de mobilité par la méthode des moindres carrés ordinaires, en incorporant l'espérance conditionnelle aux variables de conditionnement des résidus de la première étape (section III.3). Enfin, nous chercherons à formuler un modèle stable, qui tienne compte des éventuelles interactions entre le mode d'enquête et les variables socio-économiques (section III.4)

IV.1 Première étape : équation de sélection

Nous présentons d'abord les fondements théoriques des modèles probit (section 4.1.1). Puis, nous commentons les résultats des estimations effectuées sur notre échantillon (section 4.1.2), avant de mener une analyse de sensibilité (section 4.1.3).

IV.1.1 Le modèle probit

A l'aide d'un modèle probit, nous cherchons à modéliser le comportement d'individus, qui choisissent ou pas de répondre en face-à-face et, dans la négative, de remplir ou non le questionnaire web. La variable à expliquer y_i est

une variable binaire, qui peut prendre deux valeurs : 0 si l'individu répond en face-à-face ou 1 s'il répond sur le web. Soit U_{1i} l'utilité de l'individu i s'il refuse de recevoir un enquêteur à domicile et choisit de répondre en ligne, et U_{0i} son utilité s'il choisit de répondre en face-à-face. Ces niveaux d'utilité ne sont pas directement observés, mais dépendent de caractéristiques socio-économiques et d'équipements en moyens de communication des ménages. Supposons que les niveaux d'utilité U_{1i} et U_{0i} sont expliqués par un modèle linéaire, dont les variables explicatives sont x_{1i} et x_{2i} , avec ϵ_{1i} et ϵ_{0i} des aléas, tel que :

$$U_{1i} = \alpha_1 + \alpha_{11}x_{1i} + \alpha_{21}x_{2i} + \epsilon_{1i} \quad (59)$$

$$U_{0i} = \alpha_0 + \alpha_{10}x_{1i} + \alpha_{20}x_{2i} + \epsilon_{0i} \quad (60)$$

L'écart entre U_{0i} et U_{1i} est une variable latente, en fonction de laquelle la décision de répondre ou pas sur le web est prise. La règle de décision est la suivante : l'individu choisit de répondre sur le web si U_{1i} est supérieure à U_{0i} . Soit P_i la probabilité pour un répondant de saisir ses réponses en ligne :

$$P_i = P(y_i = 1) = P(U_{0i} < U_{1i}) \quad (61)$$

$$P_i = P(\epsilon_{0i} - \epsilon_{1i}) \leq (\alpha_1 - \alpha_0) + x_{1i}(\alpha_{11} - \alpha_{10}) + x_{2i}(\alpha_{21} - \alpha_{20}) \quad (62)$$

$$P_i = P(\epsilon_i \leq x_i' \beta) = F(x_i' \beta) \quad (63)$$

avec F la fonction de répartition, ϵ_i un aléa égal à $\epsilon_{0i} - \epsilon_{1i}$, x_i' le vecteur des variables explicatives et β le vecteur des paramètres à estimer, tel que :

$$\beta = \begin{pmatrix} \alpha_1 - \alpha_0 \\ \alpha_{11} - \alpha_{10} \\ \alpha_{21} - \alpha_{20} \end{pmatrix} = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{pmatrix}$$

Dans les modèles désagrégés, il est nécessaire de faire des hypothèses sur les termes résiduels. Les modèles probabilistes les plus usuels sont les modèles Logit et Probit. Le modèle Logit impose des hypothèses très restrictives sur la répartition des ϵ , mais sa formulation mathématique est simple. A l'inverse, le modèle Probit ne nécessite aucune hypothèse sur la répartition des termes d'erreur, mais reste un modèle complexe lorsque le nombre d'alternatives est supérieur à deux ¹³³.

Le modèle de choix est binaire, puisque nous cherchons à modéliser la probabilité qu'un individu réponde sur le web. La décision est représentée par la variable y_i , qui ne peut prendre que deux valeurs. Soit :

¹³³Dans le modèle probit, les résidus suivent une distribution normale, alors que dans le modèle logit, ils suivent une distribution de Gumbel. Ces distributions sont proches dans leur forme, et les modèles probit et logit donnent généralement des résultats similaires. Cependant, les valeurs estimées des paramètres ne sont pas directement comparables, puisque les variances des lois logistiques et normales ne sont pas identiques.

$$P_i = P(y_i = 1) = F(x'_i\beta) \quad (64)$$

La fonction F est une fonction de répartition, et on note f sa fonction de densité. Dans un modèle probit, on suppose que la fonction de répartition est une loi normale¹³⁴. On note Φ la fonction de répartition et ϕ la fonction de densité de la loi normale centrée réduite. La probabilité P_i devient :

$$P_i = \int_{-\infty}^{x'_i\beta} \frac{1}{2\pi} e^{-\frac{t^2}{2}} dt = \Phi(x'_i\beta) \quad (65)$$

Les valeurs des coefficients β sont celles qui maximisent la probabilité qu'un événement survienne (refus du face-à-face puis choix du web pour répondre au questionnaire). L'estimation est faite par la méthode du maximum de vraisemblance, dont le principe est assez simple. Pour chaque individu, il est possible de calculer la probabilité de répondre par un média donné (web ou face-à-face), et de connaître les paramètres de la fonction qui renvoie ces probabilités. Pour une observation donnée, la vraisemblance est la probabilité que le modèle renvoie le bon résultat. La méthode du maximum de vraisemblance consiste à trouver pour un modèle donné, la valeur des paramètres qui a le plus de chances de conduire à la situation observée. On va chercher la probabilité d'obtenir le bon résultat pour l'ensemble de notre échantillon, en formant le produit des probabilités individuelles. Il faut ensuite estimer les coefficients de la fonction qui maximise cette vraisemblance. La fonction de vraisemblance est donnée par :

$$L = P(Y_1 = y_1, \dots, Y_n = y_n) \quad (66)$$

Les variables aléatoires Y_i étant distribuées de manière indépendantes, cette fonction peut s'écrire :

$$L = \prod_{i=1}^n (P_i)^{y_i} (1 - P_i)^{1-y_i} \quad (67)$$

Car y_i prend la valeur 1 avec une probabilité P_i et la valeur 0 avec une probabilité $1 - P_i$. L'expression logarithmique de la fonction de vraisemblance est donnée par :

$$\text{Log}(L) = \sum_{i=1}^n y_i * \text{Log}(P_i) + (1 - y_i) * \text{Log}(1 - P_i) \quad (68)$$

$$\text{Log}(L) = \sum_{i=1}^n y_i * \text{Log}[F(x'_i\beta)] + (1 - y_i) * \text{Log}[1 - F(x'_i\beta)] \quad (69)$$

Le vecteur des paramètres β est obtenu en maximisant la fonction $\text{Log}(L)$ par rapport à β . Il est possible d'appliquer les tests standard de contraintes sur les paramètres, et de tester la significativité du modèle avec le test du ratio

¹³⁴Dans le modèle logit, on suppose que la fonction de répartition est une loi logistique

IV Modèle explicatif incluant le biais de sélection

de vraisemblance ¹³⁵. La mesure de la qualité de l'ajustement se fait au niveau global, par une analyse du R^2 ¹³⁶ et au niveau de chaque coefficient par une comparaison entre les choix prédits et les choix réels.

IV.1.2 Estimation du modèle de sélection des répondants

Deux types de variables sont introduites dans notre modèle en qualité de variables explicatives : des caractéristiques sociodémographiques (âge, occupation, nombre de personnes du ménage, diplôme, lieu de travail ...), et des variables concernant l'équipement en télécommunication des ménages (connexion internet, téléphone portable, inscription sur l'annuaire des abonnés). Le modèle probit explicatif du choix du web, appliqué aux variables sélectionnées ci-dessus, donne les résultats présentés dans le tableau 73.

	Coefficient	Ecart-type	Pr(> z)	Signif.
Constante	-3,71	3,47e-01	<2e-16	***
Age	0,05	1,44e-02	<0,001	***
Age ²	-4,78e-04	1,62e-04	<0,01	**
Connexion internet : oui	0,49	8,65e-02	1,03e-08	***
Téléphone portable : oui	0,26	9,19e-02	<0,01	**
Liste téléphone : oui	0,37	6,60e-02	1,68e-08	***
Liste téléphone : pas de téléphone	0,25	1,18e-01	0,04	*
Nb de personnes / ménage	-0,10	2,49e-02	5,10e-05	***
Lieu de travail : non précisé	-0,99	3,64e-01	<0,01	**
Lieu de travail : périphérie	-0,24	7,39e-02	<0,01	**
Lieu de travail : inactif	-0,24	8,80e-02	<0,01	**
Diplôme : supérieur	0,41	6,67e-02	8,16e-10	***
Densité de la zone de résidence	-2,17e-05	6,68e-06	<0,01	**
Revenu déclaré : oui	0,40	7,54e-02	1,04e-07	***
Vendredi : oui	0,47	6,66e-02	1,89e-12	***
Distance domicile / centre	-1,90e-05	7,90e-03	0,01	*

Signif. : 0 '***', 0,001 '**', 0,01 '*', 0,05 '.', 0,1 ' ', 1

TAB. 73 – Equation de sélection - Modèle probit

Les variables retenues dans le modèle sont significatives à 5% (probabilité < 5% de rejeter à tort l'hypothèse 'le coefficient est nul'). L'étude de leur signe donne une idée de l'influence des divers facteurs sur le choix de répondre par le web : un coefficient positif augmente la probabilité de répondre en ligne, mais un coefficient négatif diminue la probabilité de répondre sur internet ¹³⁷. L'ensemble des signes des coefficients sont ceux que nous espérons.

¹³⁵ $LR = -2(\text{LogL}(\beta_{MV}^*) - \text{LogL}(\beta_{MV}))$, avec $\text{LogL}(\beta_{MV}^*)$ la valeur de la fonction LogL pour un modèle non contraint (contenant l'ensemble des variables explicatives) et $\text{LogL}(\beta_{MV})$ la valeur de la fonction LogL pour un modèle contraint (suppression d'une ou plusieurs variables explicatives)

¹³⁶Suggéré par Mc Fadden

¹³⁷Nous travaillons en écart par rapport à une personne de référence, ou à une modalité de référence en ce qui concerne les variables nominales.

La possession d'une connexion internet au domicile, d'un téléphone portable, l'inscription sur la liste rouge ou orange ou l'absence de ligne de téléphone fixe au domicile ou le fait de déclarer son niveau de revenus augmentent la probabilité de répondre sur le web. Bien que l'accès au site hébergeant le questionnaire puisse se faire en dehors du domicile, les ménages possédant une connexion internet ont une probabilité plus élevée de répondre sur le web. Les internautes sont également mieux équipés que la moyenne en moyens de communication, ou appartiennent à une catégorie socioprofessionnelle plus élevée (possession d'un téléphone portable personnel ou professionnel). La variable 'liste de téléphone' est scindée en trois modalités, selon que l'individu ne possède pas de téléphone fixe, soit inscrit sur la liste rouge ou orange ou soit inscrit sur l'annuaire des abonnés. Les packages de téléphonie haut débit, couplés à l'abonnement internet, permettent de ne plus être abonné à France Télécom, donc de ne plus figurer sur l'annuaire. Il est probable que les internautes affectionnent les nouvelles technologies, et certains ne disposent plus de ligne fixe, mais d'un téléphone mobile uniquement. Fortement sollicités par les démarches commerciales, il est probable que les internautes préfèrent s'exclure de l'annuaire des abonnés en s'inscrivant sur la liste rouge ou orange.

La probabilité de répondre en ligne augmente également avec le niveau de diplôme et l'âge. Inversement, elle diminue avec l'éloignement géographique du lieu de travail des actifs (par rapport au centre de l'agglomération), le fait d'être inactif et le nombre de personnes du ménage. Les personnes diplômées sont plus familiarisées avec internet. Elles ont probablement utilisé cet outil durant leurs études et l'utilisent encore au domicile ou sur leur lieu de travail. L'âge est lié au cycle de vie des individus. Il est probable qu'avec la progression dans la vie professionnelle, les revenus et l'équipement en moyens de communication s'améliorent. En revanche, internet est une technologie relativement récente et non maîtrisée par tous ; certaines études montrent que l'usage d'internet diminue avec l'âge. L'introduction d'un terme quadratique permet de prendre en compte un impact non linéaire de l'âge. Par ailleurs, la localisation des entreprises est liée aux types d'emplois. Les bureaux, où travaillent des personnes qualifiées, qui disposent plus facilement d'un accès à internet sur leur lieu de travail, sont davantage présents dans le centre de l'agglomération. Les inactifs ont en revanche du temps à consacrer à un enquêteur à domicile, et sont donc moins tentés par un questionnaire web.

Du fait de la colinéarité entre les deux variables 'Densité de la zone de résidence' et 'Distance du domicile par rapport au centre', leur introduction simultanée pourrait rendre les coefficients instables. Cependant, la corrélation existant entre ces variables n'affecte pas le résultat, puisque si on exclut la variable 'Distance du domicile par rapport au centre', le coefficient attaché à la variable 'Densité de la zone de résidence' reste significatif et l'impact marginal est identique. Le résultat n'est donc pas dû à des problèmes de multicollinéarité, puisque l'omission d'une variable caractérisant le répondant ne le modifie pas.

Critères	Valeur
R2	15,2%
AIC	1994,2
ddl	9941

TAB. 74 – Indicateurs de la qualité du modèle de sélection

Le calcul du coefficient de détermination ajusté du modèle (R^2)¹³⁸ nous renvoie la valeur 0,152 (tableau 74). Ce résultat peut s'expliquer par l'absence d'une variable explicative importante dans le modèle de sélection. Si cette variable influe également sur le comportement de mobilité, alors il est impossible de modéliser le nombre de déplacements quotidiens par une simple régression linéaire. Le modèle probit utilisé ci-dessus apporte une solution, en calculant pour chaque individu un facteur de correction, appelé inverse du ratio de Mills.

IV.1.3 Analyse de sensibilité

L'estimation d'un modèle probit permet d'évaluer l'influence des différentes variables explicatives sur la survenue de l'événement à expliquer¹³⁹. Afin d'étudier le choix du mode de réponse à l'enquête, nous allons nous intéresser à quelques profils types d'individus. L'objectif est de savoir, pour chaque profil, quelle est la variation de probabilité de l'événement $Y_i=1$ (réponse en ligne), en cas de variation d'une des variables exogènes. Nous considérons pour ces simulations que le jour de référence des déplacements n'est pas le vendredi.

- Profil 1 : individu d'âge moyen (46 ans) n'ayant pas fait d'études supérieures, appartenant à un ménage de deux personnes, travaillant au centre de l'agglomération, ne possédant pas de connexion internet ni de ligne fixe, mais un téléphone portable, n'ayant pas déclaré son niveau de revenu annuel et résidant à 2 kilomètres du centre de l'agglomération, dans une ville de densité moyenne.
- Profil 2 : individu d'âge moyen (46 ans) ayant fait des études supérieures, appartenant à un ménage de quatre personnes, travaillant au centre de l'agglomération, possédant une connexion internet et un téléphone portable, inscrit sur la liste rouge ou orange ayant déclaré son niveau de revenu annuel et résidant à 5 kilomètres du centre de l'agglomération dans une ville de densité moyenne.

Les probabilités de répondre sur le web ou en face-à-face de ces différents profils sont présentées dans le tableau 75.

¹³⁸Le coefficient de détermination est un indicateur qui permet de juger la qualité d'une régression linéaire, simple ou multiple. D'une valeur comprise entre 0 et 1, il mesure l'adéquation entre le modèle et les données observées. Son défaut étant de croître avec le nombre de variables explicatives, on s'intéresse davantage au coefficient de détermination ajusté.

¹³⁹Les effets marginaux ne sont plus égaux aux valeurs des paramètres, comme c'est le cas dans les modèles linéaires.

TAB. 75 – Probabilités de choix du mode de réponse

	Web	Face-à-face
Profil 1	9%	91%
Profil 2	43%	57%

Les personnes correspondant au profil 2 utilisent davantage le web pour répondre au questionnaire que celles correspondant au profil 1. Ces dernières n'ont pas de connexion internet au domicile et n'ont pas effectué d'études supérieures, ce qui réduit leur degré probable d'aisance avec la navigation sur le web, ainsi que leur faculté à se connecter une vingtaine de minutes pour remplir un questionnaire en ligne.

Si l'on change le profil 1, pour considérer les individus célibataires, alors la probabilité de répondre en ligne atteint 11% : la taille du ménage impacte négativement la probabilité de répondre sur internet. Il en est de même pour le niveau d'étude, puisque le fait de prendre les personnes n'ayant pas effectué d'études supérieures dans le profil 2 donne une probabilité de répondre en ligne de 28%. Mais ce sont les variables d'équipement en moyens de communication qui ont l'effet le plus important. Ainsi, si les individus du profil 2 ne disposent plus d'une connexion internet à domicile, la part du web chute à 25%.

L'âge est une variable quantitative continue, dont la valeur moyenne dans l'échantillon est égale à 46 ans. Il est intéressant de connaître l'influence marginale de cette variable sur chaque profil, c'est-à-dire : de combien augmente la probabilité de répondre sur le web si l'âge augmente de dix ans ? Le coefficient de l'âge est sensiblement positif. En revanche, son terme quadratique, qui permet de rendre compte de la relation non linéaire de l'âge avec la probabilité de répondre sur le web est négatif. Une croissance de l'âge jusqu'à 51 ans augmente la probabilité de répondre au questionnaire en ligne, mais l'effet s'inverse au-delà de cette limite. Ainsi, une diminution de l'âge de 1 an diminue la probabilité de répondre sur le web de 1 point en ce qui concerne le profil 1 ($P=8\%$) et de 2 points en ce qui concerne le profil 2 ($P=41\%$). Nous allons à présent nous intéresser à la seconde étape du modèle d'Heckman.

IV.2 Deuxième étape : équation d'intérêt

La deuxième étape consiste à expliquer les différences de comportements en termes de mobilité, au moyen d'un modèle spécifique qui comprend :

- une variable dépendante (le nombre moyen de déplacements réalisé par les individus) ;
- plusieurs variables indépendantes ou explicatives (les facteurs observés censés avoir un effet sur la mobilité) ;
- l'inverse du rapport de Mills (variable obtenue dans la première étape) ;
- et un terme d'erreur (pour tenir compte des forces non observées qui pourraient influencer sur la mesure des résultats).

L'estimation des coefficients des variables explicatives et de la variable relative au biais de sélection se fait par une régression des moindres carrés. Nous ne retenons ici que les variables qui impactent directement la mobilité des individus. Les variables sociodémographiques prises en compte dans l'équation de sélection ne sont donc réintroduites dans la seconde étape que si elles semblent jouer un rôle significatif sur le nombre de déplacements déclaré. Dans ce cas, l'effet marginal des régresseurs sur la mobilité a deux composantes. Il y a un effet direct sur la moyenne de Y_1 et Y_2 , capté par β_{1k} et β_{2k} et un effet indirect dû à leur présence dans λ_1 et λ_2 . La compensation de ces deux effets permet de mettre en évidence l'impact marginal d'une variation des variables explicatives pour un mode d'enquête donné.

En réalité, nous estimerons deux modèles, un sur le sous-échantillon des répondants en face-à-face (section 4.2.1) et un sur le sous-échantillon des répondants sur le web (section 4.2.2), afin de tester l'existence et le sens d'un biais de sélection dans chaque sous-échantillon de répondants.

IV.2.1 Analyse de la mobilité pour l'échantillon en face-à-face

Le modèle restreint au sous-échantillon des individus ayant répondu en face-à-face nous donne des résultats intéressants (tableau 76), puisque l'ensemble des coefficients des variables explicatives prennent les signes attendus.

Le fait d'être un homme et d'appartenir à un ménage de taille élevée impacte négativement la propension à se déplacer. À l'inverse, la mobilité semble être une fonction croissante de l'éloignement du domicile par rapport au centre de l'agglomération, du nombre d'enfants et de la volonté de déclarer son niveau de revenus annuels nets. Ces résultats peuvent s'expliquer notamment par les déplacements pour motif 'Accompagnement', plus nombreux pour les habitants de la périphérie (dispersion des activités), les ménages avec enfants (vie scolaire et associative) et les femmes. La moindre réticence des personnes qui déclarent leurs revenus à communiquer l'ensemble des activités effectuées la veille de l'interview impacte favorablement la mobilité. Le nombre de déplacements quotidiens moyen augmente également avec la possession du permis de conduire et le nombre moyen de voitures par personne du ménage en âge de conduire. L'effet de la motorisation ou de la possession du permis de conduire est important, puisque le fait de ne pas pouvoir se déplacer en voiture particulière limite les occasions de déplacement. Enfin, l'introduction de termes quadratiques traduit un impact non linéaire de l'âge et du nombre d'enfants sur la mobilité. Le nombre de déplacements croît jusqu'à l'âge de 40 ans, puis décroît par la suite.

Le coefficient associé à l'inverse du ratio de Mills est significatif. Il existe donc un biais de sélection des individus. Le signe de ce coefficient est négatif, ce qui signifie qu'en moyenne le choix du web comme mode d'enquête par ces répondants aurait un impact négatif sur le nombre de déplacements déclaré.

Impact du mode sur la mesure des comportements de mobilité

Echantillon face-à-face	Rég. sans correction			Rég. avec correction		
	Coeff.	Pr(> z)	Sign.	Coeff.	Pr(> z)	Sign.
Constante	2,01	<2e-16	***	2,64	<2e-16	***
Sexe : homme	-0,24	2,14e-07	***	-0,24	2,41e-07	***
Age	0,07	1,38e-15	***	0,06	1,73e-11	***
Age ²	-7,30e-04	<2e-16	***	-6,40e-04	2,44e-12	***
Possession permis : oui	0,47	4,61e-10	***	0,45	4,23e-09	***
Nb d'enfants / ménage	0,71	<2e-16	***	0,71	<2e-16	***
(Nb d'enfants / ménage) ²	-0,01	0,0267	*	-0,01	0,05	*
Nb de voitures / personne	0,57	<2e-16	***	0,53	2,00e-14	***
Nb de personnes / ménage	-0,19	5,13e-12	***	-0,18	4,78e-11	***
Revenu déclaré : oui	0,32	1,33e-11	***	0,27	2,90e-07	***
Activité : non actif	0,29	6,85e-07	***	0,30	2,30e-07	***
Distance domicile / centre	6,70e-06	0,1478		9,50e-06	0,04	*
Mills	NA	NA		-0,18	<0,01	**

Signif. : 0 '***', 0,001 '**', 0,01 '*', 0,05 '.', 0,1 ' ', 1

TAB. 76 – Analyse de la mobilité (Echantillon face-à-face)

Environ 9% de la variance de la variable 'Nombre de déplacements' est expliquée par l'ensemble des variables explicatives qui interviennent dans le modèle (tableau 77). Le niveau de significativité du modèle est en revanche très élevé (p-value < 2,2e-16).

Critères	Valeur
R2	9,2%
F-stat	81,63
ddl	9697
P-value	< 2,2e-16

TAB. 77 – Significativité de la régression utilisée dans la seconde étape (échantillon face-à-face)

IV.2.2 Analyse de la mobilité pour l'échantillon web

Les estimations des coefficients du modèle explicatif de la mobilité appliqué à l'échantillon des répondants web sont disponibles dans le tableau 78. Peu de coefficients sont significatifs, ce qui s'explique notamment par les différences d'effectifs entre les deux échantillons de répondants. La comparaison des deux modèles estimés (web et face-à-face) montre que l'ordre de grandeur des coefficients est le même, mais que celui de leurs écarts-types varie fortement ¹⁴⁰, puisque 13 271 individus ont été interrogés en face-à-face contre seulement 369 sur le web (soit un rapport de 1 à 36). L'ordre de grandeur des écarts-types des coefficients estimés varie dans un rapport de 1 à 6 entre les deux échantillons web et face-à-face. Les valeurs de la statistique de test sont donc beaucoup

¹⁴⁰La variance des coefficients estimés est : $V(\beta) = \frac{1}{n} * \frac{s^2}{V(x)}$, avec n le nombre d'observations, s² la variance de l'échantillon et V(x) la variance de la population.

IV Modèle explicatif incluant le biais de sélection

plus faibles dans le cas de l'échantillon web et ne dépassent que rarement le seuil critique de 1,96 permettant de conclure à la significativité statistique des coefficients (au risque $\alpha = 5\%$).

Les estimations non corrigées et corrigées des coefficients diffèrent beaucoup. Les coefficients non corrigés du biais de sélection auraient pu inclure des faux positifs ou des faux négatifs, mais ce n'est pas le cas ici puisqu'après correction les coefficients conservent leur signe. L'ajout de la variable 'Mills' permet d'identifier l'impact réel des facteurs socio-économiques sur la mobilité des répondants web. Le fait d'être un homme impacte négativement la propension à se déplacer, les femmes effectuant globalement plus de déplacements. En revanche, contrairement à ce que l'on observe dans l'échantillon en face-à-face, certains coefficients ne prennent pas le signe attendu. Ainsi, le fait de résider en périphérie diminue la propension à se déplacer. Les actifs cadres sont fortement représentés dans l'échantillon web. Ils habitent souvent en périphérie, disposent d'un haut niveau de formation et déclarent des revenus élevés, mais leur emploi est chronophage. Ils ont peu de temps libre en semaine pour effectuer des activités non contraintes, ce qui limite les possibilités de déplacements. Par ailleurs, la mobilité semble croître, ici, avec le nombre de personnes du ménage.

Le coefficient de l'inverse du ratio de Mills est significatif. Il y a donc des variables qui influent sur le choix de remplir le questionnaire en ligne et la mobilité des répondants, et un biais d'endogénéité du mode sur la mobilité. Le recours à la méthode d'estimation en deux étapes est justifié, puisque le choix du web apparaît, dans ces conditions, endogène au niveau de mobilité déclaré. Par ailleurs, le signe négatif de l'inverse du ratio de Mills signifie que la mobilité pourrait être en moyenne significativement plus élevée si ces répondants n'avaient pas répondu sur le web.

Echantillon web	Rég. sans correction			Rég. avec correction		
	Coeff.	Pr(> z)	Sign.	Coeff.	Pr(> z)	Sign.
Constante	0,89	0,50		3,04	0,08	.
Sexe : homme	-0,85	7.22e-04	***	-0,87	<0,001	**
Age	0,08	0,21		0,04	0,54	
Age ²	-8,57e-04	0,22		-4,38e-04	0,55	
Possession permis : oui	0,26	0,64		0,45	0,44	
Nb d'enfants / ménage	0,40	0,42		0,47	0,36	
(Nb d'enfants / ménage) ²	-0,12	0,53		-0,15	0,43	
Nb de voitures / personne	0,68	0,08	.	0,69	0,08	.
Nb de personnes / ménage	0,27	0,08	.	0,30	0,06	.
Revenu déclaré : oui	0,36	0,23		0,17	0,65	
Activité : non actif	0,74	0,03	*	0,82	0,02	*
Distance domicile / centre	-5,70e-05	0,03	*	-5,20e-05	0,06	.
Mills	NA	NA		-0,73	1,46e-03	**
Signif. : 0 '***', 0,001 '**', 0,01 '*', 0,05 '.' , 0,1 ' ' 1						

TAB. 78 – Analyse de la mobilité (Echantillon web)

Globalement le modèle de régression est meilleur, puisque nous avons 4% de variance expliquée en plus ($R^2 = 15\%$, vs. 11% sans l'introduction du biais de sélection) (tableau 79).

Critères	Valeur
R^2	15,2%
F-stat	3,50
ddl	234
P-value	8,78e-05

TAB. 79 – Significativité de la régression utilisée dans la seconde étape (échantillon web)

IV.2.3 Interprétation des coefficients des inverses du ratio de Mills

Nous avons vu que les coefficients estimés $\rho_1\sigma_{u1}$ et $\rho_2\sigma_{u2}$ peuvent prendre tous les signes, en fonction des signes de ρ_1 et ρ_2 , c'est-à-dire du signe de la corrélation entre les résidus de l'équation de sélection et ceux de l'équation d'intérêt (respectivement celle appliquée à l'échantillon web et celle appliquée à l'échantillon face-à-face). Les résidus correspondent à des variables non observées et par conséquent non prises en compte dans le modèle, qui peuvent avoir un effet sur la variable à expliquer. Par exemple, le fait de travailler à temps partiel n'est pas une variable explicative du modèle de mobilité. Pourtant, cette caractéristique implique une moindre présence au travail, ce qui peut avoir un effet sur la disponibilité des individus (capacité à recevoir un enquêteur à domicile) et leur niveau de mobilité.

Les coefficients de l'inverse du ratio de Mills correspondent au produit de ρ_1 par σ_{u1} pour l'échantillon web ($= -0,731$) et de ρ_2 par σ_{u2} pour l'échantillon en face-à-face ($= -0,180$). La procédure d'estimation en deux étapes d'Heckman ne nous permet pas de distinguer ces valeurs. Il est cependant possible d'estimer les écarts-types des résidus de l'équation d'intérêt. Soit $\sigma_{u1} = 1,88$ et $\sigma_{u2} = 2,23$. Ensuite, l'estimation du paramètre ρ permet d'évaluer la corrélation entre l'inverse du ratio de Mills et le niveau de mobilité, c'est-à-dire la force de l'endogénéité du mode d'enquête au nombre de déplacements déclarés :

$$\rho_1 = -0,73/1,88 = -0,39$$

$$\rho_2 = -0,18/2,23 = -0,08$$

A partir de ces résultats, nous pouvons conclure que le nombre de déplacements déclaré par les répondants web apparaît fortement corrélé à l'inverse du ratio de Mills ($\rho_1 = -0,39$). La corrélation est plus faible dans le groupe des répondants en face-à-face ($\rho_2 = -0,08$). Il y a donc des facteurs qui incitent les individus à ne pas recevoir un enquêteur à domicile, puis à accepter de répondre en ligne, et qui influencent leur niveau de mobilité.

IV.3 Test de stabilité du modèle

Il s'agit maintenant de tester si les estimations des coefficients générés par le modèle en deux étapes sont stables et de quantifier un éventuel impact du mode d'enquête sur la mobilité.

IV.3.1 Comparaison d'un modèle contraint et non contraint

Nous comparons deux modèles de régression multiple explicatifs du nombre moyen de déplacements déclaré : un modèle contraint et un modèle non contraint. Le modèle contraint ne considère comme facteurs explicatifs que les variables retenues dans le modèle en deux étapes ci-dessus. Le modèle non contraint inclut également l'ensemble des interactions entre les variables explicatives et le mode d'enquête. Nous cherchons à savoir si les interactions entre les variables explicatives et le mode d'enquête, retirées dans le modèle contraint, ont un pouvoir explicatif significatif sur le nombre de déplacements quotidiens moyen des individus. Dans ce cas, il est impossible de conclure à l'existence d'un effet stable du mode d'enquête sur les réponses des enquêtés.

IV.3.2 Test du rapport des vraisemblances

Afin de pouvoir conclure sur la significativité des coefficients et comparer les modèles, il faut utiliser des tests statistiques. Les principaux sont le test de Wald et le test de rapport des vraisemblances, qui donnent souvent des résultats proches. Ces tests suivent le même principe : il s'agit de comparer l'information apportée par un modèle non contraint, contenant un certain nombre de variables explicatives, et celle apportée par un modèle contraint (contenant un sous-ensemble des variables explicatives issues du modèle non contraint). Si la différence entre les deux modèles n'est pas significative, alors les variables explicatives retirées dans le modèle contraint n'ont aucun impact sur la variable à expliquer. Les hypothèses sont les suivantes :

H_0 = le supplément d'information apporté par les variables explicatives du modèle non contraint n'est pas significatif au seuil fixé (généralement $\alpha = 5\%$), ce qui se traduit par la nullité de tous les coefficients concernés.

H_1 = Au moins une variable explicative ajoutée dans le modèle non contraint a une influence significative sur le phénomène étudié.

Dans le test de rapport de vraisemblance, on rejette l'hypothèse nulle si la vraisemblance sous l'hypothèse alternative est significativement supérieure à la vraisemblance sous l'hypothèse nulle. La statistique du test est la suivante :

$$\Lambda = [-2 * \text{Log}(\text{vraisemblance}_{mc})] - [-2 * \text{Log}(\text{vraisemblance}_{mnc})]$$

avec mc le modèle contraint et mnc le modèle non contraint. Elle suit une loi du Chi-deux à (n-1) degrés de liberté, avec n le nombre total de variables explicatives.

IV.3.3 Application aux données de l'enquête

Nous raisonnons ici sur la totalité de l'échantillon (web et face-à-face), l'introduction de la variable 'Mode' dans le modèle permettant de distinguer les répondants. Soit les hypothèses :

H_0 : les deux modèles contraint et non contraint sont équivalents, et les interactions entre les variables explicatives de la mobilité et le mode d'enquête ne sont pas significatives.

H_1 : les deux modèles contraint et non contraint ne sont pas équivalents, et les interactions entre les variables explicatives de la mobilité et le mode d'enquête sont significatives.

Les estimations du modèle de mobilité non contraint appliqué à l'ensemble de l'échantillon sont présentées dans le tableau 80, ceux du modèle contraint (par les variables retenues dans le modèle à deux étapes) dans le tableau 82.

Modèle non contraint	Coefficients	Pr(> z)	Signif.
Constante	2,64	<2e-16	***
Sexe : homme	-0,240	2,19e-07	***
Age	0,06	1,47e-11	***
Age ²	-6.36e-04	2,05e-12	***
Possession permis : oui	0,44	3,74e-09	***
Nb d'enfants / ménage	0,71	<2e-16	***
(Nb d'enfants / ménage) ²	-0,03	0,046	*
Nb de voitures / personne	0,53	1,63e-14	***
Nb de personnes / ménage	0,18	4,09e-11	***
Revenu déclaré : oui	0,27	2,63e-07	***
Activité : non actif	0,30	2,09e-07	***
Distance domicile / centre	9.50e-03	0,04	*
Mills	-0,18	<0.01	**
Mode	0,40	0,84	
(Sexe : homme) * mode	-0,63	0,04	*
(Age) * mode	-0,02	0,82	
(Age ²) * mode	1.98e-04	0,82	
(Possession permis : oui) * mode	0,01	0,99	
(Nb d'enfants / ménage) * mode	-0,24	0,69	
((Nb d'enfants / ménage) ²) * mode	-0,12	0,61	
(Nb de voitures / personne) * mode	0,16	0,74	
(Nb de personnes / ménage) * mode	0,48	0,01	**
(Revenu déclaré : oui) * mode	-0,10	0,81	
(Activité : non actif) * mode	0,52	0,23	
(Distance domicile / centre) * mode	0,06	0,06	.
(Mills) * mode	-0,55	0,18	

Signif. : 0 '***', 0,001 '**', 0,01 '*', 0,05 '.', 0,1 ' ', 1

TAB. 80 – Modèle non contraint

IV *Modèle explicatif incluant le biais de sélection*

Critères	Valeur
R ²	9,33%
F-stat	40,89
ddl	9931
P-value	<2,2e-16

TAB. 81 – Significativité du modèle non contraint

Modèle contraint	Coefficients	Pr(> z)	Signif.
Constante	2,54	<2e-16	***
Sexe : homme	-0,25	4,19e-08	***
Age	0,06	7,72e-12	***
Age ²	-6,41e-04	1,02e-12	***
Possession permis : oui	0,45	1,84e-09	***
Nb d'enfants / ménage	0,71	<2e-16	***
(Nb d'enfants / ménage) ²	-0,04	0,04	*
Nb de voitures / personne	0,53	1,61e-14	***
Nb de personnes / ménage	-0,17	3,38e-10	***
Revenu déclaré : oui	0,26	3,62e-07	***
Activité : non actif	0,32	3,20e-08	***
Distance domicile / centre	8,44e-06	0,07	.
Mills	-0,15	4,85e-03	**

Signif. : 0 '***' 0,001 '**' 0,01 '*' 0,05 '.' 0,1 ' ' 1

TAB. 82 – Modèle contraint

Critères	Valeur
R ²	8,99%
F-stat	81,84
ddl	9944
P-value	<2,2e-16

TAB. 83 – Significativité du modèle contraint

Le test de rapport de vraisemblance appliqué aux deux modèles, contraint et non contraint, renvoie une probabilité de rejeter à tort l'hypothèse d'absence d'interaction égale à 3,12e-04. La p-value est inférieure à la valeur critique de 5%. Nous pouvons donc rejeter l'hypothèse nulle d'équivalence des modèles contraint et non contraint et conclure qu'il existe une interaction significative entre les variables explicatives de la mobilité et le mode d'enquête considéré, qui permet d'expliquer le nombre de déplacements effectué par les répondants. Trois variables semblent interagir significativement avec le mode de réponse : le sexe, le nombre de personnes composant le ménage et la distance entre le centre de l'agglomération et le lieu de résidence du ménage. En revanche, le coefficient de la variable 'Mills' n'est pas significativement différent entre les deux échantillons.

IV.3.4 Formulation d'un modèle stable

Les résultats du modèle de régression appliqué aux variables explicatives et interactions significatives sont présentés dans le tableau 84. On applique ensuite le test de vraisemblance aux modèles non contraint (contenant l'ensemble des interactions entre les variables explicatives de la mobilité et le mode d'enquête) et non contraint simplifié (qui ne laisse que les trois interactions significatives comme variables explicatives).

La probabilité de rejeter à tort l'hypothèse selon laquelle ces modèles sont équivalents est de 42,83%. Les interactions entre les variables et le mode d'enquête supprimées n'ont aucun pouvoir explicatif significatif de la mobilité quotidienne et nous conservons la formulation du modèle non contraint simplifié ci-dessous.

Modèle stable	Coefficients	Pr(> z)	Signif.
Constante	2,65	<2e-16	***
Sexe : homme	-0,24	2,08e-07	***
Age	0,06	1,03e-11	***
Age ²	-6,36e-04	1,53e-12	***
Possession permis : oui	0,45	2,63e-09	***
Nb d'enfants / ménage	0,69	<2e-16	***
(Nb d'enfants / ménage) ²	-0,03	0,051	.
Nb de voitures / personne	0,53	8,07e-15	***
Nb de personnes / ménage	-0,17	1,93e-10	***
Revenu déclaré : oui	0,27	2,84e-07	***
Activité : non actif	0,32	5,13e-08	***
Distance domicile / centre	9,54e-06	0,042	*
Mills	-0,19	6,15e-04	***
Mode	-0,50	0,18	.
(Sexe : homme) * mode	-0,55	0,06	.
(Nb de personnes / ménage) * mode	0,19	0,10	.
(Distance domicile / centre) * mode	-5,9e-05	0,05	.

Signif. : 0 '***' 0,001 '**' 0,01 '*' 0,05 '.' 0,1 ' ' 1

TAB. 84 – Modèle stable

Critères	Valeur
R ²	9,20%
F-stat	63,33
ddl	9940
P-value	<2,2e-16

TAB. 85 – Significativité du modèle stable

V Quantification de l'impact du mode d'enquête sur la mobilité

Ce paragraphe illustre de quelle manière les apports des techniques économétriques permettent de comprendre les différences de mobilité observées entre deux échantillons. Nous y détaillons l'impact des variables explicatives sur le nombre de déplacements déclaré, après correction du biais de sélection. Notre regard porte essentiellement sur les variables qui interagissent significativement avec le mode de réponse choisi. L'échantillon web étant de taille modeste au regard de l'échantillon face-à-face, nous conservons dans l'analyse les interactions significatives au seuil d'erreur de 10%, exception faite du coefficient de la variable mode (p-value = 18%).

V.1 Les différentes composantes de l'influence du mode d'enquête sur la mobilité

Nous allons détailler deux composantes de l'influence du mode d'enquête sur la mobilité. D'abord l'influence directe du média, ensuite l'impact des interactions entre le mode et certaines variables explicatives.

V.1.1 Impact direct du mode d'enquête

Le mode de recueil de données, web ou face-à-face, impacte directement le niveau de mobilité. Si le questionnaire est rempli en ligne, le nombre de déplacements décroît de 0,5, ce qui confirme nos analyses exploratoires.

V.1.2 Impact des interactions entre le mode et certaines variables explicatives

Par ailleurs, trois variables semblent interagir avec le mode : le sexe, le nombre de personnes du ménage et la distance entre le domicile et le centre de l'agglomération. Pour ces deux dernières, la relation bivariée avec le mode d'enquête peut être formalisée par la figure 68.

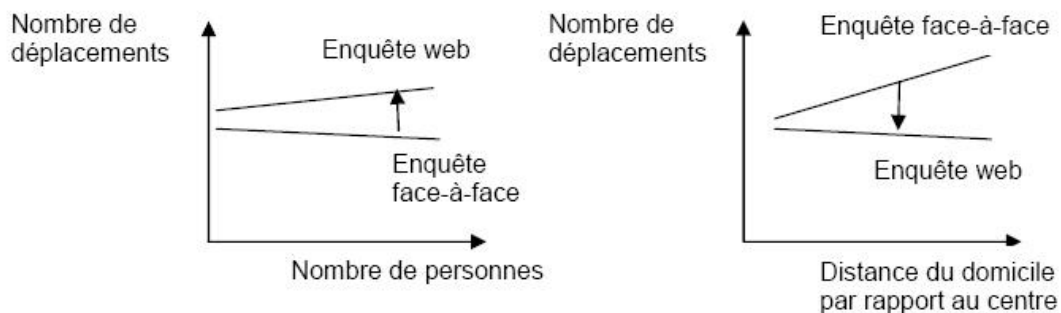


FIG. 68 – Relation bivariée entre le nombre de personnes, l'éloignement du domicile par rapport au centre et le nombre de déplacements des individus

La distance de la zone de résidence au centre de l'agglomération Le coefficient de la variable 'distance du domicile au centre * mode' est légèrement négatif ($-5,90e-05$). Ainsi, la mobilité des répondants web diminue avec l'éloignement de leur lieu de résidence du centre de l'agglomération : l'impact est de $9,54e-06$ déplacements pour les répondants en face-à-face, vs. $-4,95e-05$ déplacements pour les répondants web. Les répondants web occupent davantage d'emplois de cadres et d'employés, situés dans le centre de l'agglomération. Avec l'éloignement du lieu de résidence, la distance domicile-travail augmente. Le temps disponible pour les activités moins contraintes est donc réduit, ce qui impacte négativement leur mobilité. Nous remarquons que ce coefficient est très faible par rapport aux autres. Ceci s'explique par l'unité choisie (m) pour calculer la distance entre le domicile et le centre de l'agglomération.

Le genre Le fait d'être un homme impacte négativement la mobilité des répondants web : l'impact est de $-0,24$ déplacements par jour pour les répondants en face-à-face, vs. $-0,79$ pour les répondants web. Comme nous l'avons décrit précédemment, les femmes se déplacent davantage que les hommes, car elles doivent gérer des activités familiales en plus de leur activité professionnelle. Cet effet est amplifié en ce qui concerne les répondants web. Nous pouvons penser que leur niveau d'emploi, plus exigeant (beaucoup de cadres), leur laisse encore moins de latitude pour leurs déplacements. Rappelons ici que $3/4$ des répondants web sont des actifs.

La taille du ménage A contrario, la mobilité des répondants web augmente avec le nombre de personnes du ménage : l'impact sur le nombre de déplacements quotidiens est de $-0,17$ par personne pour les répondants en face-à-face, vs. $0,02$ pour les répondants web. Ces derniers disposent d'un revenu annuel supérieur. Il est donc possible que le nombre de personnes accroisse les besoins et les déplacements (pour motif achat, loisir...) des membres du ménage.

V.2 Les profils types de répondants

Les coefficients des variables :

- Mode ;
- (Sexe : homme) * mode ;
- (Nb de personnes / ménage) * mode ;
- (Distance domicile / centre) * mode.

permettent de quantifier, pour chaque répondant, l'impact du mode d'enquête sur le comportement de mobilité.

Soit un homme de 35 ans, actif et possédant le permis de conduire, qui réside à 1 km du centre ville. On fait l'hypothèse qu'il appartient à un ménage composé de 4 personnes (dont 1 enfant de moins de 18 ans), et de 2 voitures et qu'il n'a pas communiqué ses revenus lors de l'enquête. Ce profil de répondant, déclare sur le web 3,99 déplacements, vs. 4,58 en face-à-face.

Si on considère à présent un homme actif de 22 ans, sans enfant, vivant en couple à 500 mètres du centre de l'agglomération, qui a son permis, 1 seule voiture à disposition et qui a déclaré ses revenus, le nombre de déplacements est égal à 3,97 sur le web et 3,82 en face-à-face. La différence de mobilité est donc fortement atténuée.

A contrario, une femme active de 53 ans ayant son permis, appartenant à un ménage de 5 personnes (4 adultes et un enfant de moins de 18 ans) résidant à 5 km du centre de l'agglomération, qui possède 3 voitures et qui n'a pas déclaré ses revenus déclare, toutes choses égales par ailleurs, 4,44 déplacements sur le web et 4,30 déplacements en face-à-face.

D'une manière générale, lorsque les valeurs des variables qui interagissent positivement avec le mode d'enquête (la variable 'mode' prend la valeur 1 si l'individu répond sur le web et la valeur 0 s'il répond en face-à-face) sont élevées, la différence de mobilité s'atténue entre les répondants web et face-à-face. A contrario, lorsque les valeurs des variables qui interagissent négativement avec le mode d'enquête web sont élevées, cette différence de mobilité s'accroît.

VI Conclusion

Le problème de l'évaluation de l'effet du mode d'enquête sur la mobilité provient du fait qu'il n'est pas possible d'observer simultanément la mobilité déclarée sur le web et en face-à-face, pour un répondant donné. A l'aide de techniques économétriques, nous avons montré qu'il est possible de tester l'existence d'un biais de sélection des répondants lors d'un protocole d'enquête mixte, de le quantifier, puis d'isoler l'impact des différences socio-économiques propres aux enquêtés de l'effet média. La méthode que nous avons utilisée se base sur la procédure d'estimation en deux étapes, empruntée à Heckman. Dans la première étape, on estime par un modèle probit la probabilité pour un individu de répondre en ligne en fonction de ses caractéristiques socio-économiques et des équipements du ménage. Puis, dans une seconde étape, on estime par la méthode des moindres carrés ordinaires (MCO) l'équation d'intérêt, en incorporant l'espérance conditionnelle aux variables de conditionnement des résidus de la première étape. Les coefficients estimés des variables explicatives du modèle ne sont alors plus biaisés et reflètent l'impact des caractéristiques de l'individu et de son ménage sur la mobilité, indépendamment du choix du mode de recueil de données.

Les analyses montrent que la différence de mobilité observée entre les échantillons web et face-à-face ne peut être totalement imputée à l'effet du mode d'enquête, mais s'explique par deux effets : l'effet lié aux différences socio-économiques entre les populations et l'effet lié au mode de recueil de données. Le biais de sélection est statistiquement significatif, quel que soit l'échantillon considéré, c'est-à-dire qu'il existe des variables qui influent à la fois sur le choix de remplir le questionnaire en ligne et sur la mobilité des répondants. Le signe du coefficient du ratio de Mills indique que la mobilité des internautes pourrait être en moyenne significativement plus élevée si ces derniers avaient répondu

en face à face. Inversement, le nombre de déplacements déclaré par les individus soumis à l'enquête en face-à-face pourrait être plus faible si ces derniers avaient saisi leurs réponses sur le web. L'effet mode est également significatif, mettant en évidence l'effet du média web sur le nombre de déplacements déclaré. L'analyse précise que l'effet du web n'est pas uniforme sur la population des internautes. Les variables observées qui interagissent directement avec le mode d'enquête sont : le sexe, le nombre de personnes du ménage, ainsi que la distance entre le domicile et le centre de l'agglomération.

L'état actuel de diffusion du web au sein de la population française ne permet pas la réalisation d'une enquête web exclusive auprès de l'ensemble de la population. L'échantillon ainsi constitué ne serait pas représentatif et les résultats biaisés, ainsi que l'illustre notre exemple. L'utilisation du web dans les protocoles d'enquêtes mixtes est en revanche intéressante, mais impose d'analyser le biais de sélection des individus. La question est évidemment de savoir si ce biais est suffisamment important pour devoir être corrigé. Ceci dépend de la part de la population exclue par un mode de collecte de données et de la précision des données attendue. Dans cette expérience, le relativement faible taux de pénétration du web dans la population et les exigences croissantes des modèles de planification ne permettent pas d'occulter le biais de sélection de l'échantillon.

Des développements complémentaires vont permettre d'approfondir l'analyse comparative de la mobilité selon le mode d'enquête. Nous avons montré dans les chapitres précédents que la sous-mobilité des répondants web s'explique en partie par une immobilité plus importante. Un moyen de mieux appréhender les facteurs à l'origine de cette sous-mobilité est d'utiliser un modèle explicatif de l'immobilité selon le mode d'enquête. Dans le chapitre suivant, nous mobilisons des techniques qui ont l'avantage de séparer les facteurs explicatifs de la décision de se déplacer de ceux qui influent sur le niveau de mobilité.

Chapitre 8 : Intérêt du modèle "Hurdle" pour la compréhension des comportements de mobilité

"Le Temps est l'image mobile de l'éternité immobile."
Platon (-428,-347).

Dans le chapitre précédent, nous avons cherché à expliquer le nombre de déplacements moyen quotidiennement effectués par les habitants de l'agglomération lyonnaise. Cependant, il s'avère qu'une part non négligeable de l'échantillon a déclaré être immobile le jour de référence de l'enquête. Dans le cadre de deux estimations économétriques, nous allons nous intéresser aux facteurs impactant la décision de se déplacer, d'une part, le niveau de mobilité, d'autre part. Il s'agit de donner une estimation de la mobilité quotidienne moyenne des individus, ainsi que la fonction reliant cette dernière aux caractéristiques socio-économiques, à la motorisation et à l'équipement des ménages en moyens de communication en considérant que, pour une part non négligeable de la population enquêtée, la mobilité quotidienne est nulle (aucun déplacement déclaré). Notre démarche va notamment consister à mieux caractériser les déterminants de l'immobilité.

La variable dépendante de nos modèles, le nombre de déplacements quotidiens déclaré, est quantitative mais ne peut prendre que des valeurs entières. Le modèle développé fera donc référence aux techniques employées pour la régression sur les variables de comptage. Par ailleurs, l'explication du niveau de mobilité ne se fera que pour les individus mobiles, c'est-à-dire ayant effectué au moins un déplacement dans le périmètre de l'enquête durant la période de référence. La probabilité pour un individu de se situer dans l'intervalle des personnes mobiles sera prise en compte avec des notions similaires à celles utilisées dans les modèles qualitatifs. Selon Thomas (2000), il existe trois façons de considérer les variables qualitatives en statistique : les incorporer comme variables explicatives dans un modèle de régression, étudier leur corrélation ou les traiter comme des variables dépendantes. Pour ce dernier cas, il faut distinguer les modèles à variables qualitatives binaires, comme les modèles Logit et Probit, des modèles à variable dépendante limitée, où la variable expliquée n'est observée que sur un intervalle, comme les modèles Tobit. La particula-

rité des modèles qualitatifs réside dans le fait que les techniques d'inférences par la méthode des moindres carrés ne sont pas adaptées. Nous appliquons à la place la méthode du maximum de vraisemblance. De nombreux problèmes rencontrés lors de la régression linéaire (multicolinéarité, hétéroscédasticité. . .) sont toujours présents, mais peu de travaux traitent des effets du non respect de certaines hypothèses sur la qualité de l'estimation des modèles à variable dépendante limitée (Maddala, 1986).

Afin d'examiner les déterminants de la mobilité chez les répondants web et face-à-face, nous utilisons un modèle de Poisson à obstacle, appelé plus couramment dans la littérature "Poisson Hurdle model". Il est alors possible de distinguer les facteurs incitant les personnes à se déplacer au moins une fois durant la période de référence de l'enquête de ceux influençant le niveau de mobilité des personnes mobiles. L'objectif est de répondre séparément à ces deux questions : comment expliquer que certains individus se déplacent alors que d'autres restent immobiles durant la période de référence ? Une fois la décision de mobilité prise, qu'est-ce qui pousse certaines personnes à se déplacer davantage que d'autres durant la même période ? De nombreux exemples relatifs aux sciences sociales sont détaillés dans la littérature, comme les travaux de Bohara et Krieg (1996) concernant le taux de migration aux USA, ceux de King (1989) sur l'implication des nations dans les conflits internationaux ou ceux de Bounie *et al.* (2006) sur les dépenses effectuées en France par carte de crédit.

I Pour poser le problème

Le tableau 86 donne la distribution des individus par enquête (web et face-à-face) selon le nombre de déplacements quotidiens déclaré. Nous constatons que, globalement, les répondants en face-à-face se déplacent davantage que les répondants web. Par ailleurs, nous observons une forte proportion de personnes immobiles, en particulier dans l'enquête web (environ une personne sur cinq ne se déplace pas durant la période de référence). Dans ce cas, quels sont les déterminants de la mobilité quotidienne ?

Enquête	Obs.	0	1	2	3	4	5	6	7+	Total
Face-à-face	Ind.	1459	110	3854	1081	3096	1106	1151	1414	13271
	%	11%	1%	29%	8%	23%	8%	9%	11%	100%
Web	Ind.	70	12	98	48	65	30	24	22	369
	%	19%	3%	27%	13%	18%	8%	7%	6%	100%

TAB. 86 – Distribution du nombre de déplacements par personne et par enquête

La faible proportion de personnes n'effectuant qu'un seul déplacement, quelle que soit l'enquête, laisse penser qu'une fois la décision de se déplacer prise, les répondants effectuent plusieurs déplacements. Le nombre de déplacements déclaré par les personnes mobiles est toutefois plus important dans l'échantillon en face-à-face. Dès lors, comment peut-on expliquer la fréquence

des déplacements entre les répondants ? Enfin, et plus généralement, les facteurs explicatifs de la décision de mobilité et de la fréquence des déplacements sont-ils identiques, ou bien peut-on isoler, des facteurs spécifiques pour chaque type de comportement ?

Notre analyse tente de mettre en évidence les facteurs motivant les individus à effectuer un premier déplacement, mais également à comprendre pourquoi certains se déplacent davantage que d'autres. Nous pensons que la décision de se déplacer distingue fortement les individus. En effet, ceux qui ont besoin de se déplacer en semaine pour effectuer des activités pour leur compte (travail, loisirs ...) ou celui d'autrui (accompagnement ...) ont probablement des caractéristiques différentes de ceux qui déclarent rester à leur lieu de résidence. Plusieurs facteurs peuvent expliquer l'immobilité des répondants, comme la volonté de ne pas communiquer l'information, une maladie ou une absence d'activité le jour de référence des déplacements, ou le mode d'enquête ¹⁴¹. Par ailleurs, le niveau de mobilité des personnes mobiles est très variable, et nous avons vu dans le chapitre précédent que certains facteurs sociodémographiques influencent positivement le nombre de déplacements. A contrario, d'autres caractéristiques semblent modérer cette mobilité. Ces observations nous amènent à supposer que le modèle statistique qui détermine la probabilité de se déplacer diffère de celui qui détermine la fréquence des déplacements. Dans les modèles standard applicables aux données de comptage, comme le modèle de Poisson, ces deux procédés sont forcés d'être identiques. Ces modèles ne permettent pas une structure conditionnelle et donnent des estimateurs biaisés (Grogger et Carson, 1991). C'est pourquoi, à la place d'appliquer le modèle de Poisson, nous estimons un modèle qui permette de différencier le processus de choix binaire (mobile / immobile) de l'estimation de fréquence de la variable d'intérêt (nombre de déplacements déclaré par personne mobile).

A cause de cette différence supposée, notre modèle est scindé en deux parties. Nous estimons d'abord la probabilité individuelle de se déplacer durant la période de référence. Puis, conditionnellement à cette décision, nous estimons dans un second temps la fréquence des déplacements déclarée. Il est possible de distinguer le modèle économétrique donnant le résultat binaire (se déplace / ne se déplace pas) de celui qui détermine le niveau de mobilité, à l'aide d'une procédure connue sous le nom de 'Hurdle model' ou modèle à obstacle (Mullahy, 1986). Il n'est cependant pas possible de distinguer, avec les données disponibles, l'immobilité réelle (individus ne s'étant pas déplacés) du refus de déclaration (non déclaration des déplacements effectués). Nous estimons la décision de faire un déplacement initial et d'effectuer des déplacements additionnels à l'aide des mêmes variables explicatives dans les deux modèles, ce qui nous permet de comparer directement les coefficients estimés dans les deux équations. Si les modèles génèrent des résultats significativement différents, alors nous pourrions conclure que modéliser simplement la décision de mobilité omet d'expliquer la motivation de nombreuses personnes de réaliser des déplacements additionnels, ou qu'évaluer simplement le niveau de mobilité

¹⁴¹Le caractère auto-administré du web peut encourager certains individus à se déclarer immobiles, pour éviter d'avoir à saisir leurs déplacements.

de permet pas de prendre en compte les facteurs à l'origine de l'immobilité d'un certain nombre de répondants.

II Formalisation économétrique

La méthodologie adaptée est basée sur une spécification de Poisson à obstacle (Hurdle Poisson)¹⁴². D'abord, la décision de se déplacer est modélisée par un résultat binaire, à savoir si un individu a effectué un déplacement dans le périmètre d'enquête durant la période de référence ou pas. Lorsque la décision de se déplacer est prise, générant la réalisation d'un déplacement, cela s'appelle 'crossing a Hurdle'. Le nombre de déplacements effectués par l'individu est supposé suivre un modèle de Poisson tronqué en zéro (Zeileis *et al.*, 2008).

II.1 La loi de Poisson

La distribution de Poisson est adaptée aux variables entières (Zorn, 1996). En statistique, une variable aléatoire Y suit une loi de Poisson¹⁴³ de paramètre réel positif θ , notée $Y \sim P(\theta)$, si et seulement si elle suit, pour tout entier naturel k , une loi de probabilité définie par :

$$P(Y = k) = \frac{e^{-\theta} \theta^k}{k!} \quad (70)$$

θ est le paramètre unique de la loi de Poisson. Il représente à la fois la moyenne et la variance de la distribution de la variable d'intérêt. Soit :

$$E(Y) = \theta \quad (71)$$

$$V(Y) = \theta \quad (72)$$

L'espérance conditionnelle de la distribution d'une variable aléatoire Y , étant donné un ensemble de facteurs explicatifs x_k est donnée par :

$$E(Y | x_k) = e^{(\theta_k x_k)} \quad (73)$$

où $\beta = (\beta_0 \dots \beta_k)$ est un vecteur de paramètres inconnus à estimer et $x = (1 \dots x_k)$ un vecteur de variables explicatives. Nous notons :

$$\theta_i \equiv e^{(\beta x_i)} \quad (74)$$

Pour chaque observation i tirée au hasard dans la population, la probabilité de y_i conditionnelle à l'ensemble de facteurs x_{ki} est donnée par la relation :

¹⁴² 'The idea underlying the Hurdle formulations is that a binomial probability model governs the binary outcome of whether a count variable has zero or a positive realization. If the realization is positive, the 'Hurdle' is crossed and the conditional distribution of the positives is governed by a truncated-at-zero count data model' (Mullahy, 1986).

¹⁴³ Ou loi des événements rares.

$$P(Y = y_i | x_{ki}) = \frac{e^{-\theta_i} \theta_i^{y_i}}{y_i!} \quad (75)$$

avec $y = 0, 1, 2, \dots$

On a alors :

$$\ln L[P(y_i)] = -\theta_i + y_i \ln \theta_i - \ln y_i! \quad (76)$$

$$\ln L[P(y_i)] = -e^{\beta x_i} + y_i(\beta x_i) - \ln y_i! \quad (77)$$

Si nous supposons que les observations concernant différents individus i sont mutuellement indépendantes, il est possible d'estimer simplement les paramètres β_k par la méthode du maximum de vraisemblance. La fonction de vraisemblance est en effet égale à la somme du logarithme des probabilités, exprimée en fonction de β :

$$\ln(L) = -\sum_{i=1}^n e^{\beta x_i} + \sum_{i=1}^n y_i(\beta x_i) - \sum_{i=1}^n \ln y_i! \quad (78)$$

On a désormais $(k+1)$ paramètres à estimer et $(k+1)$ conditions du premier ordre, qui peuvent s'écrire sous la forme du vecteur :

$$\frac{\partial \ln(L)}{\partial \beta_k} = \sum_{i=1}^n (y_i - \theta_i) x_i = 0 \quad (79)$$

Si la distribution de Poisson est appropriée, et en posant l'hypothèse que la répartition des y_i est aléatoire dans l'échantillon, nous obtenons une estimation non biaisée de β .

II.2 Le modèle de Poisson censuré

Les données utilisées en microéconomie sont souvent caractérisées par une censure de la variable dépendante. Un intervalle de valeurs possibles pour cette variable est alors transformé en une simple valeur c et la variable dépendante est égale à c pour une part significative des observations. Les méthodes de régression classiques ne permettent pas de tenir compte de la différence qualitative qui existe entre les valeurs limites (égales à c) et les valeurs continues de la variable dépendante.

Dans notre modèle, la variable à expliquer y_i est une variable dite de comptage, observable seulement sur un certain intervalle. L'objectif de l'analyse est de connaître les déterminants de la variable d'intérêt y_i , sans éliminer de l'échantillon les individus pour lesquels cette variable est nulle. Dans la littérature, les modèles appropriés pour ce type d'analyse sont les modèles Tobit

¹⁴⁴, utilisés notamment lorsqu'un grand nombre d'observations de la variable à expliquer sont nulles. Si nous conservons dans l'échantillon les observations concernant les individus immobiles, nous appliquons le modèle tobit censuré, et les valeurs observées sont censurées à zéro. Si nous choisissons de supprimer ces observations, alors il faut appliquer le modèle tobit tronqué. On dit qu'un modèle est tronqué si les variables explicatives ne sont pas observables lorsque la variable dépendante passe en-dessous (ou au-dessus) d'un certain seuil c . Ce cas peut se produire, soit si les individus pour lesquels $y_i < c$ ne sont pas interrogés, soit si les réponses aux variables explicatives x_i n'ont de sens que lorsque $y_i > c$.

On considère une variable aléatoire y_i , distribuée selon une loi de Poisson. Nous savons par hypothèse, que :

$$E(Y | x_k) = e^{(\theta_k x_k)}$$

Un échantillon de n observations de couples (y_i, x_{ki}) est constitué, mais la variable y_i n'est pas toujours observable. On dispose à la place de la variable latente y_i^* , qui est reliée à la variable y_i par la relation ¹⁴⁵ :

$$y_i^* = y_i, \text{ ssi } y_i > c \quad (80)$$

$$y_i^* = c, \text{ sinon} \quad (81)$$

où c est une constante. Si le vecteur x_k est observé pour tout i , que y_i^* soit ou non supérieure à c , alors l'échantillon est censuré et seule la variable dépendante n'est observée que sur un certain intervalle. Si nous éliminons les observations telles que $y_i^* \leq c_i$, alors l'échantillon est tronqué.

Si l'échantillon est censuré en zéro, alors la constante c est nulle et le modèle peut s'écrire :

$$P(y_i = y_i^*) = \frac{e^{-\theta_i} \theta_i^{y_i}}{y_i!} \quad (82)$$

avec $\theta_i = e^{\beta_k x_{ki}}$, x_i un vecteur de variables explicatives et β un vecteur de paramètres.

Si $y_i = 0$, alors nous pouvons poser :

$$P(y_i = 0) = \frac{e^{-\theta_i} \theta_i^0}{0!} \quad (83)$$

$$P(y_i = 0) = e^{-\theta_i} \quad (84)$$

¹⁴⁴En référence à Tobin qui l'a proposé la première fois pour analyser l'achat de biens durables par les ménages (Tobin, 1958). De nombreux exemples d'applications du modèle Tobit sont présents dans la littérature, comme ceux décrivant le nombre d'heures de travail ou les salaires des personnes selon leur statut (Amemiya, 1984).

¹⁴⁵Le point de censure peut théoriquement être vers le haut ou vers le bas, selon le problème à modéliser.

En revanche, si $y_i > 0$, alors nous pouvons écrire :

$$P(y_i > 0) = 1 - P(y_i = 0) \quad (85)$$

$$P(y_i > 0) = 1 - e^{-\theta_i} \quad (86)$$

Par ailleurs, nous pouvons calculer la probabilité de y_i , conditionnellement à la réalisation de l'événement $y_i > 0$. Soit :

$$P(y_i | y_i > 0) = \frac{\frac{e^{-\theta_i} \theta_i^{y_i}}{y_i!}}{1 - e^{-\theta_i}} \quad (87)$$

II.3 Modèle Hurdle

Certaines données empiriques contiennent plus de valeurs nulles qu'il est autorisé pour appliquer le modèle de Poisson. La catégorie de modèle capable de prendre en compte cette propriété est appelé modèle Hurdle. Il contient deux composantes. Un modèle tronqué en zéro, de type Poisson, est employé pour modéliser les valeurs positives de la variable d'intérêt et une composante de type 'Hurdle' permet de distinguer les valeurs nulles des valeurs positives de la variable d'intérêt. Pour cette dernière, un modèle binomial ou une distribution censurée en zéro de type Poisson peut être employée. Plus formellement, le modèle Hurdle combine un modèle de données de comptage tronqué à gauche en $y = 1$ et un modèle Hurdle en zéro censuré à droite en $y = 1$.

Soit un modèle de Poisson censuré en zéro, dans lequel la probabilité de survenue de l'événement y est influencé par un vecteur de variables explicatives x_1 , à l'aide de la fonction :

$$\theta_i \equiv e^{(\beta_1 x_{1i})} \quad (88)$$

et un modèle de Poisson tronqué en zéro, dans lequel la fréquence de survenue de l'événement y est influencée par un vecteur de variables explicatives x_2 , à l'aide de la fonction :

$$\lambda_i \equiv e^{(\beta_2 x_{2i})} \quad (89)$$

La question de la méthode à utiliser pour calculer les valeurs prévues des paramètres β_1 et β_2 reste entière. S'agissant d'un modèle non linéaire, nous allons utiliser la méthode du maximum de vraisemblance pour évaluer les paramètres du modèle. Pour une observation donnée, la vraisemblance est la probabilité que le modèle renvoie le bon résultat. On cherche alors la probabilité d'obtenir le bon résultat pour l'ensemble de l'échantillon en formant le produit des probabilités individuelles. La méthode consiste à estimer les coefficients de la fonction qui maximise cette vraisemblance, au moyen d'un algorithme numérique. Cette spécification présente l'avantage que les composantes tronquées et censurées peuvent être estimées séparément. La fonction de vraisemblance jointe pour le modèle Hurdle de Poisson est :

$$L = \prod_{y_i=0} [P(y_i = 0)] \prod_{y_i>0} [1 - P(y_i = 0)][P(y_i | y_i > 0)] \quad (90)$$

Le premier terme représente le résultat binaire caractérisant la réalisation de y_i et le dernier terme correspond au processus de Poisson, conditionnellement à la réalisation de y_i . Soit :

$$L = \prod_{y_i=0} [e^{-\theta_i}] \prod_{y_i>0} [1 - e^{-\theta_i}] \left[\frac{e^{-\lambda_i} \lambda_i^{y_i}}{y_i!} \right] \quad (91)$$

$$\ln(L) = - \sum_{y_i=0} \theta_i + \sum_{y_i>0} \ln(1 - e^{-\theta_i}) + \sum_{y_i>0} \ln \left[\frac{e^{-\lambda_i} \lambda_i^{y_i}}{y_i!} \right] \quad (92)$$

Il est possible d'utiliser le même vecteur de variables explicatives pour estimer la réalisation d'une valeur positive de la variable d'intérêt (x_1) et la fréquence de survenue de cette même variable (x_2). Ceci permet de comparer directement les deux modèles. Soit :

$$L = \prod_{y_i=0} [e^{-\theta_i}] \prod_{y_i>0} [1 - e^{-\theta_i}] \left[\frac{e^{-\theta_i} \theta_i^{y_i}}{y_i!} \right] \quad (93)$$

$$L = \prod_{y_i=0} [e^{-\theta_i}] \prod_{y_i>0} \left[\frac{e^{-\theta_i} \theta_i^{y_i}}{y_i!} \right] \quad (94)$$

$$\ln(L) = - \sum_{y_i=0} \theta_i + \sum_{y_i>0} \ln \left[\frac{e^{-\theta_i} \theta_i^{y_i}}{y_i!} \right] \quad (95)$$

II.4 Evaluation de la pertinence du modèle Hurdle

Afin d'évaluer la pertinence de l'utilisation d'un modèle 'Hurdle', nous calculons une statistique qui combine la valeur de Log-vraisemblance du modèle Hurdle et celle du modèle de Poisson simple :

$$stat = 2 * [\log Lik_{Hurdle} - \log Lik_{Poisson}] \quad (96)$$

Pour tester si le modèle Hurdle est statistiquement significatif, on compare cette statistique à une valeur lue dans la table du Khi-deux. Les hypothèses sont :

H_0 = l'écart entre la valeur de Log-vraisemblance du modèle Hurdle et celle du modèle de Poisson simple n'est pas significatif. Ceci signifie que le modèle Hurdle n'apporte pas un intérêt majeur dans la compréhension du phénomène étudié.

H_1 = l'écart entre la valeur de Log-vraisemblance du modèle Hurdle et celle du modèle de Poisson simple est significatif. Le modèle Hurdle permet alors de mieux rendre compte de l'influence des facteurs explicatifs sur la variable étudiée.

III Application à l'équation de mobilité quotidienne

Pour illustrer nos propos, nous utilisons les données issues de l'enquête ménages déplacements réalisée à Lyon en 2006. L'analyse est menée sur l'échantillon web, d'une part, et face-à-face, d'autre part. Ces deux populations n'étant pas homogènes en termes de caractéristiques socio-économiques et d'habitude de mobilité, il est probable que l'impact des facteurs ne soit pas le même. Seules les personnes de 18 ans et plus sont considérées, par souci de cohérence avec le chapitre précédent et de comparaison entre les deux échantillons. Une seule personne du ménage remplit le questionnaire en-ligne, et nous avons montré qu'il s'agit surtout du chef de ménage ou de son conjoint. A contrario, dans l'enquête en face-à-face, l'ensemble des membres du ménage sont interrogés, dont les enfants. Ne pas prendre en compte les moins de 18 ans revient à exclure les scolaires de l'analyse, largement sous représentés dans l'échantillon web, et dont le niveau de mobilité est traditionnellement spécifique.

III.1 Les variables disponibles pour l'analyse

III.1.1 Définition de l'immobilité

Dans une enquête déplacements, une personne immobile est une personne qui a déclaré n'avoir effectué aucun déplacement. Il existe certaines raisons au fait que des individus choisissent de ne pas se déplacer durant la période de référence de l'enquête, et un répondant immobile est un problème pour l'analyste. Cette personne n'a-t-elle réellement pas quitté son domicile, a-t-elle oublié les déplacements effectués ou utilise-t-elle cet échappatoire pour réduire le niveau de pénibilité associé à l'enquête ? Les enquêtes ménages déplacements excluent certaines personnes. On peut se demander combien d'individus résidant à leur domicile sont incapables de se déplacer. Seule une infime proportion de la population enquêtée semble concernée (environ 1%). D'autres personnes sont dans une incapacité temporaire à sortir de chez elles (1%), devaient rester chez elles (1%) ou n'ont pas souhaité sortir à cause de la météo (Madre *et al.*, 2007).

Le nombre d'immobiles peut varier fortement selon les enquêtes, une forte proportion pouvant cependant correspondre à des 'soft refusals'. Il s'agit pour les répondants d'éviter de passer du temps à livrer leurs réponses, en déclarant dès le début de l'interview ne pas avoir effectué de déplacement durant la période de référence. *'Something in the interaction between the survey, the*

survey protocol, and the fieldwork firm invites the respondent . . . to use the soft refusal'(Madre *et al.*, 2007) ¹⁴⁶. Le mode d'enquête peut avoir une influence. D'une façon générale, les enquêtes auto-administrées laissent apparaître davantage de personnes immobiles, l'enquêteur n'étant pas présent pour relancer les répondants ¹⁴⁷. Le web permet de répondre à distance, dans des conditions pas toujours propices à la concentration du répondant. La qualité de la rédaction des questions et le professionnalisme des enquêteurs permettent sans nul doute de réduire les 'refus mous'. Mais les sociétés d'études sont plus souvent tournées vers le taux de réponse global que vers la part d'immobiles dans les enquêtes (Madre *et al.*, 2007). Il est probable que moins un répondant aura passé de temps en déplacements, plus il aura tendance à adopter cette attitude, bien que ce 'refus mou' touche toutes les classes de la population, même les plus mobiles. Il peut donc faire perdre des individus qui se sont beaucoup déplacés, surtout dans une interview auto-administrée (Armoogum *et al.*, 2005). Stopher *et al.* (2004b) rapporte que plus les individus sont sollicités pour participer à l'interview, plus ils risquent de se déclarer immobiles. La part d'immobiles est importante, car elle impacte directement le nombre moyen de déplacements et d'activités déclaré (Madre *et al.*, 2003).

Si les décisions de sous-déclaration sont faites de façon aléatoire par le répondant, et ne sont pas liées au nombre de déplacements ou d'activités pratiqués, alors les résultats modélisés ne sont pas biaisés (Han et Polak, 2003). Madre *et al.* (2003) précisent que le 'refus mou' est un facteur aléatoire qui permet d'utiliser un modèle binomial de détermination de l'immobilité.

III.1.2 Description des variables

Nous avons choisi un ensemble de variables caractérisant l'individu et son ménage, pouvant expliquer potentiellement le niveau de mobilité et plus particulièrement la décision de mobilité. Nous avons mis en évidence dans le chapitre précédent que trois types de variables influent sur la mobilité. Il y a d'une part des variables socio-économiques, telles que l'occupation de l'individu, son âge, le nombre d'enfants du ménage ou la localisation du lieu de résidence. Mais des variables concernant l'équipement du ménage en moyen de communication ou liées à la motorisation ont également un impact non négligeable sur le nombre de déplacements déclaré. Nous avons retenu dix variables dans l'analyse :

Sexe : homme ou femme.

Age : âge de chaque personne.

Nombre d'enfants du ménage : nombre d'enfants de moins de 18 ans présents dans le ménage.

Nombre de personnes du ménage : nombre de personnes qui composent le ménage.

¹⁴⁶Le taux d'immobiles 'acceptable' dans les enquêtes se situe entre 18% et 12% dans les enquêtes transport portant sur un jour de semaine (Madre *et al.*, 2007).

¹⁴⁷Madre *et al.* (2007) met en évidence un taux d'immobilité de l'ordre de 20% pour les enquêtes postales, contre 13% à 15% pour les enquêtes téléphoniques et face-à-face.

Revenus déclarés : variable dichotomique, indiquant si l'individu a déclaré le niveau de revenu annuel net de son ménage.

Activité : variable renseignant le statut du répondant. Deux modalités sont disponibles : les actifs et les non actifs.

Téléphone portable : possession d'un téléphone portable par la personne, à titre personnel ou professionnel.

Nombre de voitures du ménage : nombre de voitures particulières possédées par le ménage, rapporté au nombre de personnes en âge de conduire (18 ans et plus).

Possession du permis : possession du permis de conduire ou pratique de la conduite accompagnée.

Vendredi : variable qui prend la valeur '1' si le jour de référence pour le recueil des déplacements est le vendredi, et la valeur '0' sinon.

III.1.3 Statistiques descriptives

Les tableaux suivants donnent quelques statistiques descriptives sur les variables retenues : nombre d'observations, valeurs minimum et maximum, moyenne et écart-type pour les variables continues (tableaux 87 et 88), modalités et effectifs, pour les variables nominales (tableaux 89 et 90).

Variabes	Obs.	Moyenne	Ecart-type	Minimum	Maximum
Age	11 577	47,34	18,11	18	98
Nb d'enfants du ménage	11 577	0,65	1.02	0	7
Nb de voitures du ménage	11 577	0,66	0.39	0	4
Nb de personnes du ménage	11 577	2,78	1.42	1	10

TAB. 87 – Statistiques descriptives des variables continues (échantillon face-à-face)

Variabes	Obs.	Moyenne	Ecart-type	Minimum	Maximum
Age	360	44,06	13,98	18	86
Nb d'enfants du ménage	361	0,62	0,91	0	3
Nb de personnes du ménage	361	2,58	1.28	1	6

TAB. 88 – Statistiques descriptives des variables continues (échantillon web)

III Application à l'équation de mobilité quotidienne

Variables	Observations	Modalités	Effectifs
Sexe	11 577	Homme	5 424
		Femme	6 153
Possession du permis	11 577	Oui	9 694
		Non	1 883
Activité	11 577	Actif	6 067
		Non actif	5 510
Revenus déclarés	11 577	Oui	7 521
		Non	4 056
Téléphone portable	11 577	Oui	8 204
		Non	3 373

TAB. 89 – Statistiques descriptives des variables nominales (échantillon face-à-face)

Variables	Observations	Modalités	Effectifs
Activité	361	Actif	260
		Non actif	101
Revenus déclarés	361	Oui	275
		Non	86
Vendredi	361	Oui	150
		Non	211

TAB. 90 – Statistiques descriptives des variables nominales (échantillon web)

III.1.4 Hypothèses sur l'influence des différents facteurs

Nous étudions dans ce paragraphe l'impact des variables explicatives retenues. Les résultats diffèrent parfois de ceux du chapitre précédent, car nous considérons à présent l'ensemble des répondants, et plus seulement les mobiles.

Impact du genre : des hommes plus mobiles que les femmes (échantillon face-à-face) En ce qui concerne le genre, plusieurs enquêtes transport réalisées en France et en Belgique mettent en évidence un taux d'immobiles plus important chez les femmes (Armoogum *et al.*, 2005). Par ailleurs, l'analyse descriptive de l'échantillon face-à-face montre que les hommes déclarent en moyenne davantage de déplacements que les femmes (figure 69 ¹⁴⁸), ce qui vient contredire les résultats du chapitre précédent concernant uniquement les personnes mobiles. Ceci s'explique par un plus fort taux d'immobilité chez les femmes (13,81% vs. 10,01%). Les femmes sont moins souvent titulaires du permis de conduire que les hommes, bien que la proportion de femmes ayant

¹⁴⁸Ce diagramme appelé boîte à moustaches s'utilise pour comparer un même caractère dans deux populations de tailles différentes. Il s'agit de tracer un rectangle allant du premier quartile au troisième quartile et coupé par la médiane. On ajoute des segments aux extrémités menant jusqu'aux premier et neuvième déciles. Les points en dehors des segments correspondent aux valeurs extrêmes.

le permis soit de plus en plus importante, et ont un accès plus restreint et moins régulier à la voiture particulière (Hine, 2004). Le genre ne semble pas discriminant de la décision de mobilité, pour l'échantillon web.

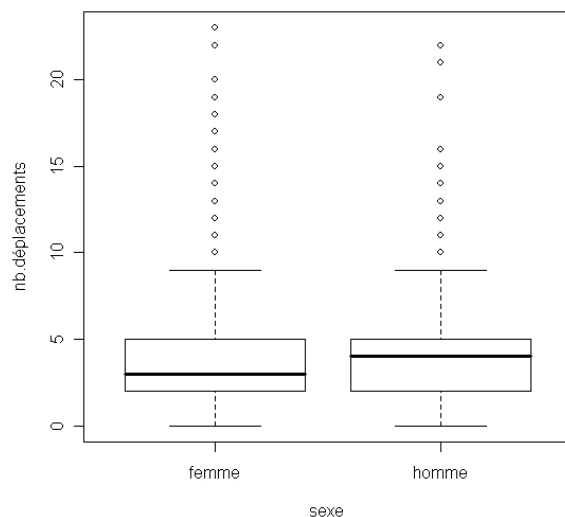


FIG. 69 – Relation entre le genre des répondants et le nombre de déplacements (enquête face-à-face)

Source : EMD Lyon (2006)

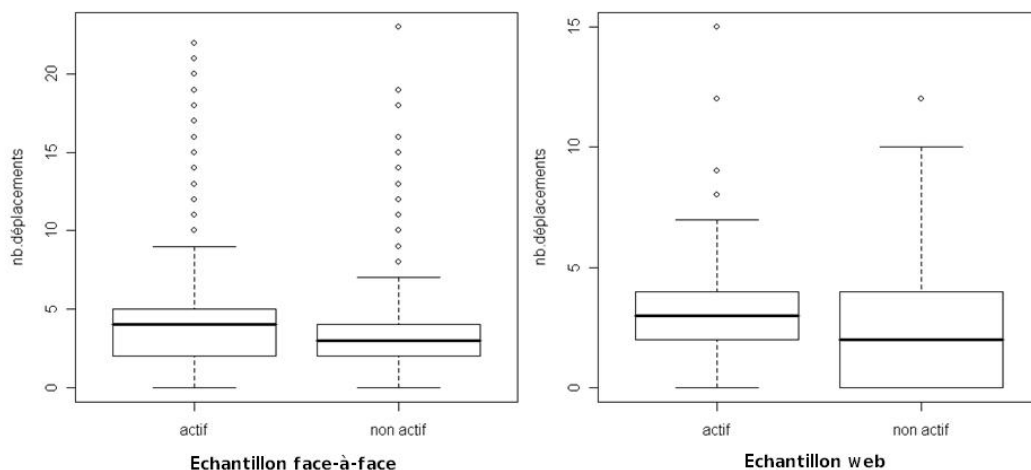


FIG. 70 – Relation entre l'activité des répondants et le nombre de déplacements par enquête

Source : EMD Lyon (2006)

Impact du statut : davantage de personnes immobiles chez les inactifs Le même raisonnement peut-être appliqué à la variable 'activité'. L'analyse descriptive montre, quel que soit l'échantillon, que le nombre moyen de

déplacements est sensiblement plus important chez les actifs (figure 70). Or, nous avons vu précédemment que dans l'échantillon web, parmi les mobiles, le niveau de mobilité était plus important chez les inactifs, ce que confirment d'autres analyses (Madre *et al.* (2003) et Hubert (2003)). De nouveau, cette différence s'explique par un taux d'immobiles soit plus important chez les inactifs (18,95% pour les inactifs et 5,75% pour les actifs en face-à-face, vs. respectivement 30,69% et 15% sur le web). Cette relation est probablement liée à l'âge.

Impact du jour de référence : des immobiles plus nombreux le vendredi Par ailleurs, le taux d'immobiles peut dépendre du jour concerné par le recensement des déplacements, et donc du jour de référence des déplacements. Si les interviews sont conduites en semaine, alors les actifs sont probablement plus difficiles à joindre. Ces derniers répondent davantage le week-end, par le web, au sujet des déplacements de la journée du vendredi. Ce décalage entre la saisie des réponses et la réalisation des déplacements pèse sur la mobilité déclarée.

Le moment de l'interview, ou de saisie des réponses, peut donc biaiser le nombre de déplacements déclaré et impacter le taux d'immobilité (Christensen, 2006). La variable 'vendredi' sera donc prise en compte pour l'analyse de l'échantillon web, en posant l'hypothèse qu'elle pèse sur la probabilité de se déplacer (28,67%, vs. 12,80% d'immobiles)¹⁴⁹.

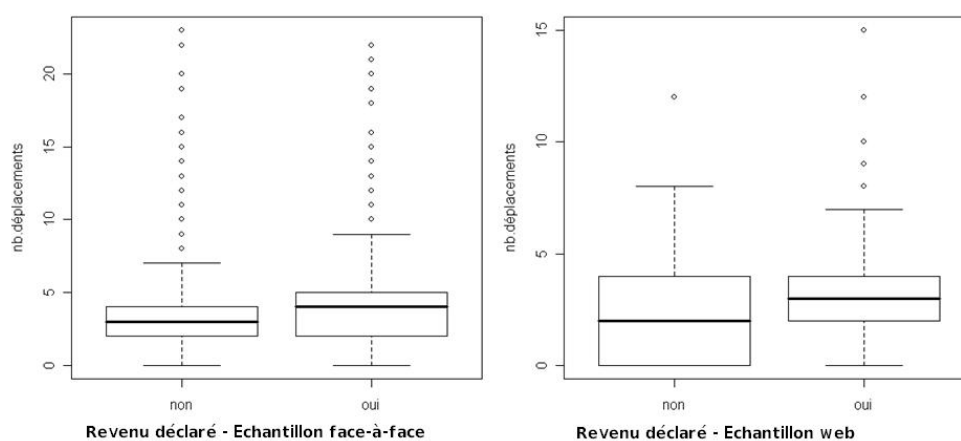


FIG. 71 – Relation entre la déclaration des revenus annuels du ménage et le nombre de déplacements par enquête

Source : EMD Lyon (2006)

Impact de la décision de déclarer les revenus : un facteur générateur de déplacements Le fait de déclarer son revenu à l'enquêteur ou sur son ordinateur peut être interprété comme une volonté de participer à l'enquête

¹⁴⁹Celle-ci est contraire à la tendance observée dans l'enquête nationale transport, où la plus faible immobilité est observée le vendredi (Armoogum *et al.*, 2005).

et de ne pas vouloir 'retenir' des informations jugées trop personnelles. Nous formulons donc l'hypothèse que cette variable est négativement liée à l'immobilité (11,02%, vs. 13,90% en face-à-face et 16%, vs. 30,24% sur le web) et au contraire propice à un nombre de déplacements déclarés élevé (figure 71).

Impact de la possession d'un téléphone portable et du permis de conduire : une influence non négligeable sur la mobilité Enfin, dans l'enquête face-à-face, deux autres variables semblent impacter le taux d'immobilité : la possession du permis de conduire et celle d'un téléphone portable¹⁵⁰. Il est probable que ces deux variables influencent positivement la mobilité (figure 72). Si le ménage ne dispose pas d'une voiture, alors le risque d'immobilité est plus important (Madre *et al.*, 2003). Posséder le permis est un atout pour se déplacer, puisque cela rend possible les déplacements en voiture particulière (9,67%, vs. 24,22% d'immobiles), et les possesseurs de téléphone portable sont souvent des personnes actives ou ayant un bon niveau d'interaction sociale, ce qui génère des activités et donc des déplacements (8,93%, vs. 19,57% d'immobiles).

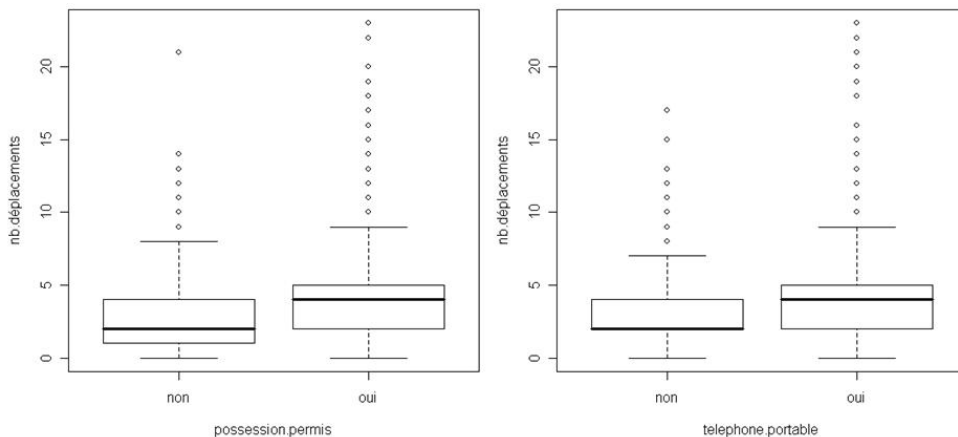


FIG. 72 – Relation entre la possession du permis de conduire ou d'un téléphone portable et le nombre de déplacements pour les répondants en face-à-face

Source : EMD Lyon (2006)

Impact de l'âge : une relation plus ambiguë Les récents comparatifs effectués sur les données d'enquêtes transport françaises (Armoogum *et al.* (2005) et Hubert (2003)) montrent un fort taux d'immobiles chez les moins de 20 ans, qui devient faible avec l'entrée dans la vie active, pour remonter ensuite. Le taux d'immobilité est donc minimal pour les classes actives et maximal pour les jeunes et les personnes âgées. L'analyse descriptive sur les échantillons web et face-à-face de l'enquête déplacements de Lyon conduit aux mêmes résultats (figure 73). L'immobilité est de manière générale plus importante chez les personnes âgées, notamment à cause de leur statut de retraité et leurs difficultés

¹⁵⁰Ces deux variables ne sont pas significatives dans l'échantillon web, puisque la plupart des internautes disposent du permis de conduire et d'un téléphone portable.

III Application à l'équation de mobilité quotidienne

à se déplacer (Madre *et al.*, 2003).

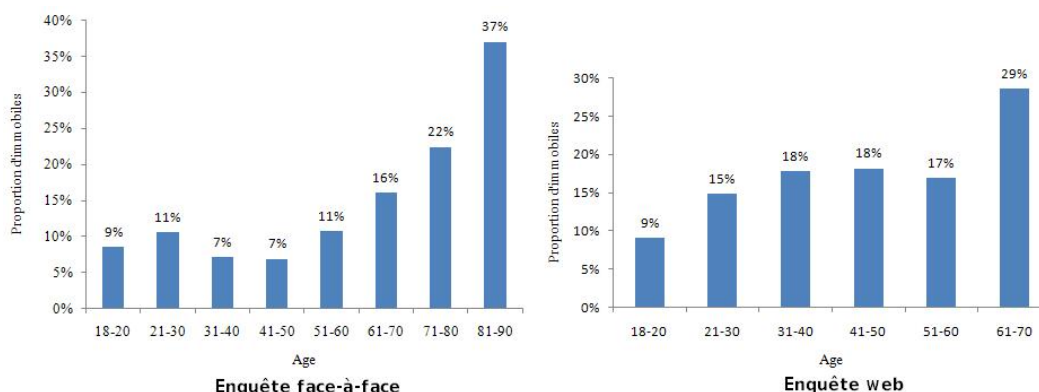


FIG. 73 – Relation entre le taux d'immobiles et l'âge des répondants par enquête

Source : EMD Lyon (2006)

Impact de la taille du ménage : deux facteurs à prendre en considération La relation entre le taux d'immobiles et le nombre de personnes n'est pas linéaire. Nous constatons plus de personnes immobiles parmi les ménages composés de une à deux personnes et ceux de cinq personnes et plus. Inversement, la part des mobiles est plus importante chez les ménages de trois à quatre personnes (figure 74).

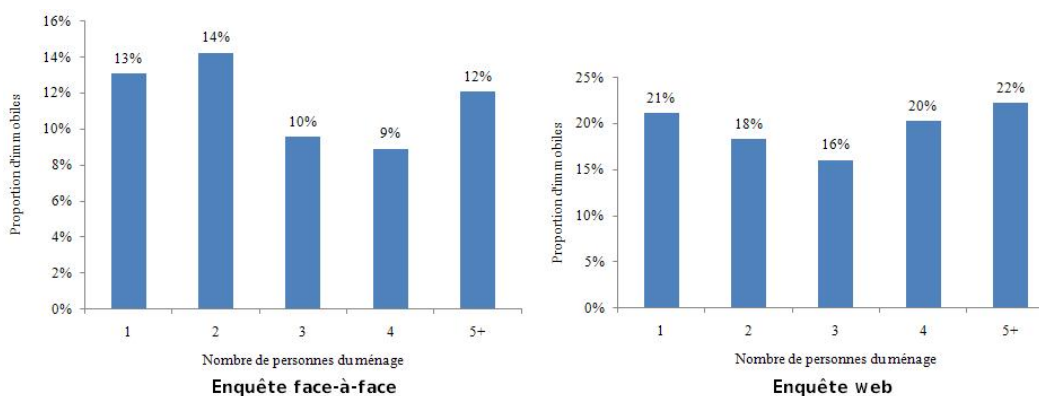


FIG. 74 – Relation entre le taux d'immobiles et le nombre de personnes du ménage par enquête

Source : EMD Lyon (2006)

Cette relation peut être mise en parallèle avec le nombre d'enfants. Les ménages composés de une à deux personnes sont majoritairement des personnes seules ou vivant en couples, sans enfant de moins de 18 ans. Ceux comprenant de 3 à 4 personnes sont pour la plupart des familles avec un à deux enfants. Comme le montre la figure 75, la relation entre le taux d'immobiles et le nombre d'enfants diffère selon le type d'enquête. Sur le web, l'immobilité semble faiblement corrélée aux nombre d'enfants présents dans le ménage (entre 18% et

20% d'immobiles). A contrario, dans l'enquête face-à-face, le taux d'immobiles est élevé chez les ménages sans enfant (14%), puis diminue jusqu'à la présence de deux enfants (7%), avant de remonter à partir de trois enfants.

Il semble donc intéressant de garder ces deux variables dans l'analyse, surtout pour rendre compte du phénomène de mobilité dans l'échantillon web.

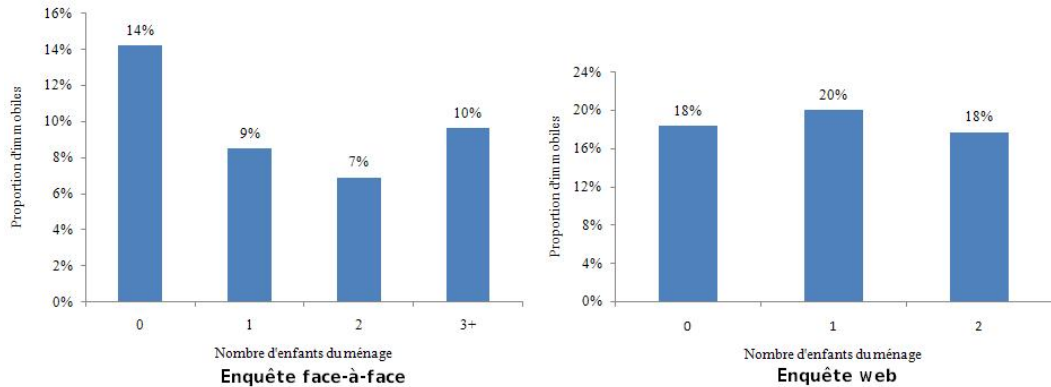


FIG. 75 – Relation entre le taux d'immobiles et le nombre d'enfants du ménage par enquête

Source : EMD Lyon (2006)

Les autres facteurs intéressants pour l'analyse de la mobilité Deux autres facteurs souvent mis en avant dans les modèles explicatifs de l'immobilité sont l'éloignement du lieu de résidence du centre de l'agglomération et le niveau de revenu du ménage. Les habitants de la périphérie sont plus immobiles que ceux du centre-ville (Madre *et al.*, 2007). Mais cette variable n'est pas significative dans nos modèles. Nous avons donc mené l'analyse sans en tenir compte. En ce qui concerne le niveau de revenu, le nombre élevé de non-réponses partielles ne permet pas de prendre en compte un nombre suffisant d'individus (25% dans l'enquête web et 30% dans l'enquête en face-à-face) dans les modèles explicatifs de l'immobilité, bien que les études montrent une forte immobilité chez les ménages déclarant un faible niveau de revenu (Madre *et al.*, 2007).

III.2 La population des répondants face-à-face

Les analyses ci-dessous autorisent la prise en compte des différentes facettes de la mobilité, en distinguant la décision de se déplacer du nombre de déplacements effectivement réalisés (tableaux 91, 92 et 93). L'ensemble des variables explicatives décrites vont permettre d'estimer la probabilité de mobilité, d'une part, et la fréquence des déplacements, d'autre part. Nous mettons plus particulièrement en évidence les effets statistiquement significatifs. Les tableaux suivants indiquent les principaux résultats de nos modèles, appliqués à l'échantillon des répondants en face-à-face. Ils sont suivis de deux indicateurs sur la qualité des modèles : le logarithme du maximum de vraisemblance et l'AIC

III Application à l'équation de mobilité quotidienne

(Akaike's information criterion) (Akaike, 1974) ¹⁵¹.

Variabes	Estimate	Std. Error	z-value	Pr(> z)	Sign.
Constante	1,094	0,031	35,033	< 2e-16	***
Sexe : homme	-0,055	0,010	-5,563	2,65e-08	***
Age	-0,003	0,001	-8,286	< 2e-16	***
Nb de voitures du ménage	0,186	0,014	13,072	< 2e-16	***
Possession du permis : oui	0,245	0,017	14,244	< 2e-16	***
Nb de personnes du ménage	-0,063	0,006	-10,476	< 2e-16	***
Activité : non actif	-0,062	0,011	-5,587	2,31e-08	***
Nb d'enfants du ménage	0,162	0,008	20,868	< 2e-16	***
Revenu déclaré : oui	0,090	0,010	8,558	< 2e-16	***
Téléphone portable : oui	0,069	0,013	5,446	5,17e-08	***

Signif. : 0 '***' 0,001 '**' 0,01 '*' 0,05 '.' 0,1 ' ' 1

TAB. 91 – Modèle de Poisson simple - Echantillon face-à-face

Variabes	Estimate	Std. Error	z-value	Pr(> z)	Sign.
Constante	1,157	0,032	35,562	< 2e-16	***
Sexe : homme	-0,073	0,010	-7,094	1,30e-12	***
Age	-0,001	0,001	-1,998	0,046	*
Nb de voitures du ménage	0,151	0,015	10,189	< 2e-16	***
Possession du permis : oui	0,152	0,018	8,334	< 2e-16	***
Nb de personnes du ménage	-0,044	0,006	-6,932	4,14e-12	***
Activité : non actif	0,018	0,011	1,577	0,115	,
Nb d'enfants du ménage	0,149	0,008	18,357	< 2e-16	***
Revenu déclaré : oui	0,082	0,011	7,457	8,82e-14	***
Téléphone portable : oui	0,040	0,013	3,019	0,003	**

Signif. : 0 '***' 0,001 '**' 0,01 '*' 0,05 '.' 0,1 ' ' 1

TAB. 92 – Modèle de Poisson tronqué - Echantillon face-à-face

¹⁵¹AIC = - 2 * log(L) + 2 * k, où L est la vraisemblance maximisée et k le nombre de paramètres dans le modèle. L'AIC représente donc un compromis entre le biais qui diminue avec le nombre de paramètres estimés et la nécessité de décrire les données avec le plus petit nombre de paramètres possible).

Intérêt du modèle "Hurdle"

Variabiles	Estimate	Std. Error	z-value	Pr(> z)	Sign.
Constante	2,925	0,183	15,982	< 2e-16	***
Sexe : homme	0,144	0,063	2,280	0,023	*
Age	-0,017	0,002	-8,417	< 2e-16	***
Nb de voitures du ménage	0,496	0,092	5,386	7,20e-08	***
Possession du permis : oui	0,537	0,079	6,762	1,36e-11	***
Nb de personnes du ménage	-0,252	0,032	-7,796	6,38e-15	***
Activité : non actif	-0,853	0,073	-11,627	< 2e-16	***
Nb d'enfants du ménage	0,238	0,047	5,032	4,85e-07	***
Revenu déclaré : oui	0,131	0,061	2,127	0,033	*
Téléphone portable : oui	0,207	0,071	2,912	0,004	**

Signif. : 0 '***' 0,001 '**' 0,01 '*' 0,05 '.' 0,1 ' ' 1

TAB. 93 – Modèle de Hurdle (binomial) - Echantillon face-à-face

Modèle	Poisson simple	Poisson Hurdle
AIC	52817	50458
logLik	-2,640e+04	-2,521e+04
df	10	20

Signif. : 0 '***' 0,001 '**' 0,01 '*' 0,05 '.' 0,1 ' ' 1

TAB. 94 – Maximum de vraisemblance - Echantillon face-à-face

La simple modélisation du nombre de déplacements par un modèle de Poisson ne permet pas de prendre en considération toutes les facettes du phénomène de mobilité. Le modèle Hurdle apporte une amélioration dans la compréhension des pratiques de mobilité des individus interrogés en face-à-face. En effet, la statistique qui combine la valeur de Log-vraisemblance du modèle Hurdle et celle du modèle de Poisson simple est significative (tableau 95).

Critères	Valeur
Statistique de test ¹⁵²	2379,55
Valeur du Chi-deux au risque $\alpha = 5\%$	31,41
Ddl	20
P-value	< 0,001

TAB. 95 – Significativité du modèle Hurdle - Echantillon face-à-face

La différence s'explique notamment par l'impact des variables 'Activité' et 'Sexe'. Le fait d'être inactif joue négativement sur la mobilité dans le modèle de Poisson simple (coefficient = -0,062; p-value < 0,001%), alors qu'il joue positivement sur la mobilité dans le modèle de Poisson tronqué (coefficient = 0,018; p-value = 11,5%). Le fait d'être un homme impacte négativement le nombre de déplacements déclarés dans le modèle de Poisson simple (coefficient = 0,055; p-value < 0,001%), alors qu'il influence positivement la décision de mobilité dans le modèle Hurdle (coefficient = 0,144, p-value = 2,3%).

III Application à l'équation de mobilité quotidienne

Une autre conclusion importante est de savoir si les causes de la mobilité initiale (décision de se déplacer ou pas durant la période de référence) sont les mêmes que celles de la fréquence des déplacements, c'est-à-dire si les variables explicatives ont la même influence dans les deux équations. Par exemple, si nous considérons la variable 'Sexe', la décision de mobilité est plus importante pour les hommes (sexe.homme = 0,144, significatif au seuil de 2,3%). En revanche, une fois la décision de se déplacer prise, ce sont les femmes qui se déplacent le plus (sexe : homme = -0,073 dans le modèle de Poisson tronqué, significatif au seuil $< 0,001\%$). La conclusion est la même en ce qui concerne l'activité des répondants. Le fait d'être inactif pénalise la décision de mobilité (activité : inactif = -0.853 dans le modèle binomial, significatif au seuil $< 0,001\%$). Mais lorsqu'une personne inactive est mobile durant la période de référence, alors elle effectue davantage de déplacements qu'une personne active aux caractéristiques socio-économiques similaires (activité : inactif = 0,018 dans le modèle de Poisson tronqué, significatif au seuil de 11,5%). Les variables liées à la motorisation ('nombre de voitures du ménage' et 'permis de conduire'), la possession d'un téléphone portable et le fait de déclarer ses revenus ont un impact positif sur la décision de se déplacer, et le nombre de déplacements déclaré, quel que soit le modèle considéré. A contrario, plus les répondants appartiennent à des ménages de grande taille et plus ils sont âgés, plus leur probabilité de se déplacer et leur niveau de mobilité sont faibles (dans les modèles de Poisson simple et tronqué).

III.3 La population des répondants web

Nous reprenons les traitements de la section précédente, en ciblant à présent les répondants web (tableaux 96, 97 et 98).

Variables	Estimate	Std. Error	z-value	Pr(> z)	Sign.
Constante	1,001	0,144	6,951	3,62e-12	***
Age	-0,005	0,002	-2,237	0,025	*
Nb de personnes du ménage	0,098	0,037	2,631	0,009	**
Activité : non actif	-0,037	0,072	-0,523	0,601	
Nb d'enfants du ménage	-0,037	0,053	-0,689	0,491	
Revenu déclaré : oui	0,249	0,078	3,188	0,001	**
Vendredi : oui	-0,231	0,064	-3,622	2,93e-04	***

Signif. : 0 '***' 0,001 '**' 0,01 '*' 0,05 '.' 0,1 ' ' 1

TAB. 96 – Modèle de Poisson simple - Echantillon web

Intérêt du modèle "Hurdle"

Variabes	Estimate	Std. Error	z-value	Pr(> z)	Sign.
Constante	1.069	0.151	7.089	1.35e-12	***
Age	-0.001	0.003	-0.425	0.671	
Nb de personnes du ménage	0.042	0.040	1.055	0.292	
Activité : non actif	0.139	0.075	1.856	0.064	.
Nb d'enfants du ménage	0.097	0.056	1.733	0.083	.
Revenu déclaré : oui	0.097	0.083	1.168	0.243	
Vendredi : oui	-0.057	0.068	-0.848	0.396	

Signif. : 0 '***' 0,001 '***' 0,01 '**' 0,05 '.' 0,1 '.' 1

TAB. 97 – Modèle de Poisson tronqué - Echantillon web

Variabes	Estimate	Std. Error	z-value	Pr(> z)	Sign.
Constante	2,240	0,682	3,284	0,001	**
Age	-0,023	0,011	-2,148	0,032	*
Nb de personnes du ménage	0,350	0,196	1,784	0,074	.
Activité : non actif	-0,752	0,318	-2,363	0,018	*
Nb d'enfants du ménage	-0,722	0,280	-2,578	0,010	**
Revenu déclaré : oui	0,782	0,312	2,510	0,012	*
Vendredi : oui	-0,975	0,289	-3,378	0,001	***

Signif. : 0 '***' 0,001 '***' 0,01 '**' 0,05 '.' 0,1 '.' 1

TAB. 98 – Modèle de Hurdle (binomial) - Echantillon web

Modèle	Poisson simple	Poisson Hurdle
AIC	1601	1473
logLik	-793,5	-722,7
df	7	14

TAB. 99 – Maximum de vraisemblance - Echantillon web

Moins de variables sont significatives dans les modèles explicatifs de la mobilité appliqués à l'échantillon des répondants web, et tout particulièrement dans le modèle de Poisson tronqué (constante, activité et nombre d'enfants du ménage). Ceci peut s'expliquer notamment par la taille réduite de l'échantillon (361 individus). Dans cet échantillon, une personne sur cinq est déclarée immobile, ce qui rend nécessaire l'identification précise des facteurs déclencheurs de la mobilité. Le modèle binomial explicatif de la décision de mobilité semble performant, puisque l'ensemble des variables sont significatives.

Ici encore, le modèle Hurdle apporte une amélioration dans la compréhension des pratiques de mobilité des individus interrogés. En effet, la statistique qui combine la valeur de Log-vraisemblance du modèle Hurdle et celle du modèle de Poisson simple est significative (tableau 100).

III Application à l'équation de mobilité quotidienne

Critères	Valeur
Statistique de test	141,53
Valeur du Chi-deux au risque $\alpha = 5\%$	23,68
Ddl	14
P-value	$< 0,001$

TAB. 100 – Significativité du modèle Hurdle - Echantillon face-à-face

La différence s'explique notamment par l'impact de la variable 'activité'. Dans le modèle de Poisson simple, cette variable n'est pas significative (p-value = 60,1%), car il y a un double effet, négatif sur la probabilité de se déplacer (coefficient = -0,752 dans le modèle de Hurdle ; p-value = 1,8%) et positif sur le niveau de mobilité (coefficient = 0,139 dans le modèle de Poisson tronqué ; p-value = 6,4%), les deux étant significatifs. La situation est identique pour la variable 'nombre d'enfants du ménage'. Le fait d'appartenir à une famille possédant de nombreux enfants semble réduire la probabilité de se déplacer (coefficient = -0,722 dans le modèle de Hurdle ; p-value = 1,0%), alors que son influence sur la mobilité est positive (coefficient = 0,097 dans le modèle de Poisson tronqué ; p-value = 8,3%).

Dans le modèle de Poisson simple, la mobilité augmente avec le nombre de personnes du ménage et le fait de déclarer son revenu (coefficients respectifs de 0,098 et 0,249 ; p-value = 9% et 1%). Ces variables jouent en fait sur la probabilité de se déplacer durant la période de référence (coefficients respectifs de 0,350 et 0,782 dans le modèle de Hurdle ; p-value = 7,4% et 1,2%), mais ne sont pas statistiquement significatives dans le modèle de Poisson tronqué explicatif du niveau de mobilité (p-value respectives de 29,2% et 24,3%).

Le jour de référence a également une influence non négligeable, puisque si les déplacements concernent la journée du vendredi, alors la probabilité de se déclarer immobile est relativement forte (coefficient dans le modèle Hurdle = -0,975 ; p-value $< 0,001\%$). De même, la probabilité de se déplacer diminue avec l'âge (coefficient dans le modèle Hurdle = -0,975 ; p-value 3,2%). Or, ces variables ne sont pas significatives dans le modèle de Poisson tronqué, explicatif du niveau de mobilité (p-value respectives = 39,6% et 67,1%). Dans le modèle de Poisson simple, cette nuance n'est pas prise en compte, le fait de se déplacer le vendredi et l'âge impactant négativement le nombre de déplacements déclaré (coefficients respectifs de -0,231 et -0,005 ; p-value $< 0,001\%$ et 2,5%).

III.4 Synthèse

Un enseignement général de ce travail est que les déterminants de la mobilité sont différents de ceux de son intensité. Nous pouvons d'abord dissocier les déterminants exclusifs (variables qui n'affectent que la décision de mobilité ou son intensité) des déterminants qui affectent à la fois la mobilité et la fréquence des déplacements, certaines variables ayant des effets positifs ou négatifs à la fois sur la probabilité de mobilité et sur son intensité.

Intérêt du modèle "Hurdle"

Variables	Echantillon face-à-face		Echantillon web	
	Décision de mobilité	Niveau de mobilité	Décision de mobilité	Niveau de mobilité
Sexe : homme	+	-	NA	NA
Age	-	-	-	-
Nb de voitures du ménage	+	+	NA	NA
Possession du permis : oui	+	+	NA	NA
Nb de personnes du ménage	-	-	+	+
Activité : non actif	-	+	-	+
Nb d'enfants du ménage	+	+	-	+
Revenu déclaré : oui	+	+	+	+
Téléphone portable : oui	+	+	NA	NA
Vendredi : oui	NA	NA	-	-

Sign. NA : variable non introduite, '+' : variable à effet positif, '-' : variable à effet négatif

TAB. 101 – Mise en perspective, par échantillon, des déterminants de la décision de mobilité et de son intensité

Ensuite, nous pouvons distinguer les facteurs qui influencent uniquement la décision de mobilité et ceux qui ne jouent que sur son intensité. Les résultats sont mis en perspective dans le tableau 101.

Comme nous l'avons vu dans le chapitre précédent, le genre n'est pas une variable qui impacte significativement la mobilité des répondants web. Il en est de même du téléphone portable, des variables de motorisation (possession du permis et nombre de voitures à disposition des personnes de 18 ans et plus). L'échantillon web étant constitué principalement de personnes actives ayant un bon niveau d'éducation et d'emploi, ces variables ne permettent pas de les distinguer. A contrario, le vendredi a un impact très important sur la mobilité des internautes, ce qui s'explique notamment par le choix du jour de connexion. En revanche, cette variable ne semble pas pertinente pour expliquer la mobilité des répondants à l'enquête en face-à-face, puisque les interviews à domicile sont réalisées de manière équivalente selon les différents jours de la semaine.

IV Conclusion

La littérature s'est souvent interrogée sur les déterminants du niveau de mobilité, sans déterminer si les facteurs qui conditionnent la mobilité sont les mêmes que ceux qui influencent son intensité. L'objectif de ce chapitre est de connaître les déterminants de la mobilité déclarée, sans éliminer de l'échantillon les individus qui sont restés immobiles durant la période de référence et de comparer les résultats obtenus sur les deux échantillons, web et face-à-face. En distinguant la décision de mobilité et la fréquence des déplacements, nous sommes capables de cerner efficacement le rôle que les différents facteurs jouent sur la mobilité quotidienne.

Tout au long de ce chapitre, nous avons montré qu'un modèle de Poisson simple n'était pas approprié pour rendre compte des facteurs qui influencent la mobilité quotidienne des individus. Nous avons donc utilisé un modèle de Poisson de type Hurdle. Les analyses économétriques montrent que les facteurs qui incitent à la mobilité ne sont pas forcément ceux qui impactent positivement le niveau de mobilité. Nos résultats confirment l'hypothèse que modéliser uniquement la décision d'effectuer un seul déplacement laisse de côté d'importantes informations concernant les multiples déplacements effectués par les répondants. Dans l'échantillon face-à-face par exemple (le plus représentatif), il ressort d'un modèle de Poisson simple que le fait d'être un homme influence négativement le niveau de mobilité. Le modèle de Poisson Hurdle met en évidence un effet plus complexe : les hommes ont une probabilité plus forte de réaliser un premier déplacement, mais une fois la décision de mobilité prise, alors leur niveau de mobilité est plus faible que celui des femmes. D'un autre côté, ne considérer qu'un modèle explicatif de la fréquence des déplacements ne permet pas de rendre compte des facteurs qui influencent la décision de mobilité durant la période de référence. L'exemple le plus significatif est l'activité, qu'il s'agisse de l'échantillon web ou face-à-face. Dans le modèle qui distingue les individus mobiles des immobiles, le fait d'être inactif incite fortement les individus à ne pas se déplacer durant la période de référence. Une fois la décision de mobilité prise, cette variable a en revanche une influence positive sur le niveau de mobilité des répondants.

Si nous comparons les échantillons web et face-à-face, l'impact des variables sociodémographiques sur le nombre de déplacements déclaré diffère parfois :

- le rôle ambigu du statut (niveau d'activité) est similaire entre les deux échantillons ;
- dans l'échantillon face-à-face, le sexe a une influence complexe sur la mobilité (impacte différemment la décision de mobilité et le niveau de mobilité), alors que sur le web c'est l'influence du nombre d'enfants qui n'est pas identique entre les modèles ;
- certains effets sont stables et similaires entre les deux échantillons, comme celui de l'âge (négatif) et le fait de déclarer ses revenus (positif) ;
- d'autres variables, comme le nombre de personnes du ménage ont une influence stable sur la mobilité, mais opposée entre les deux échantillons (négative en face-à-face et positive sur le web) ;
- le fait de se déplacer le vendredi impacte uniquement la mobilité des répondants web (négativement), car les internautes avaient le choix du jour de connexion et donc implicitement du jour de référence pour la déclaration des déplacements (ce qui n'est pas le cas dans l'échantillon face-à-face) ;
- les variables relatives à la motorisation et à la possession d'un téléphone portable influencent positivement la mobilité dans les modèles appliqués à l'échantillon face-à-face mais ne sont pas significatives dans ceux menés sur l'échantillon web (ces variables ne sont pas discriminantes pour la population d'internautes).

L'immobilité est par définition une variable difficile à évaluer, puisqu'il est possible que des participants sous estiment volontairement leur niveau de mobilité. Le grand nombre d'immobiles de l'enquête web pourrait donc provenir d'une attitude de 'refus mou', qu'adopteraient certains répondants, ainsi soupçonnés de déclarer qu'ils n'ont pas bougé de chez eux plutôt que de se plier au cadre relativement complexe du questionnaire. Ces résultats laissent penser que l'immobilité ne peut être modélisée uniquement à l'aide de variables socio-économiques, bien qu'il ne soit pas possible, à ce stade de l'analyse, de différencier les réels immobiles des 'refus mous'.

Conclusion de la partie III

La procédure d'estimation en deux étapes empruntée à Heckman permet de s'affranchir du biais d'auto-sélection des individus (volonté de répondre en face-à-face ou, au contraire, refus de l'interview à domicile et saisie des réponses en ligne). Les coefficients estimés des variables du modèle ne sont plus biaisés et reflètent l'impact des variables socio-économiques sur la mobilité, indépendamment du choix du mode de recueil de données. Dans notre exemple, le biais de sélection est statistiquement significatif. Le mode d'enquête a donc une incidence sur la mobilité qu'une simple régression linéaire ne peut mettre en évidence. L'exercice montre que les variables qui impactent la mobilité des répondants web sont le genre, le nombre de personnes du ménage, ainsi que la distance entre le domicile et le centre de l'agglomération. Certaines hypothèses ont été formulées pour tenter d'expliquer ces différences. L'intérêt de la méthode consiste également à fournir des clés pour éventuellement pouvoir redresser l'échantillon des répondants web, afin de rendre leur niveau de mobilité comparable à celui des répondants en face-à-face, bien que ce traitement soit risqué à un niveau agrégé.

Des développements complémentaires permettraient d'approfondir l'analyse comparative de la mobilité selon le mode d'enquête. D'abord, nous avons montré que la sous-mobilité des répondants web concernait certains types de motifs (non contraints) et de mode de transports (marche à pied). Une analyse des facteurs explicatifs de ces types de déplacements permettrait de corriger plus finement la mobilité quotidienne des internautes.

Par ailleurs, la sous-mobilité des répondants web s'explique par une immobilité plus importante. Il est intéressant de développer un modèle explicatif de l'immobilité selon le mode d'enquête afin de mieux appréhender les facteurs à l'origine de cette sous-mobilité. Certaines techniques ont l'avantage de séparer les facteurs explicatifs de la décision de se déplacer de ceux qui influent sur le niveau de la mobilité. C'est le cas du modèle Hurdle. Deux résultats importants ressortent de ces analyses. D'abord, certaines variables ont un effet complexe sur la mobilité, qu'une simple régression linéaire ne peut mettre en évidence. Ainsi, quel que soit le mode d'enquête, le fait d'être actif augmente la probabilité de se déplacer, mais une fois la décision de se déplacer prise, l'inactivité pèse sur le niveau de mobilité. Ensuite, des variables ont un effet similaire sur la décision de mobilité et son intensité, mais ne sont pas significative dans les deux échantillons. C'est le cas par exemple des variables liées à la motorisation ou à la possession d'un téléphone portable, qui jouent positivement sur la mo-

bilité des répondants en face-à-face. Les internautes étant fortement motorisés et équipés en moyens de communication, ces variables n'apparaissent pas significatives dans l'analyse appliquée à l'échantillon web. Inversement, le vendredi n'est pas un jour propice à la mobilité des internautes, alors que cette variable n'est pas significativement explicative de la mobilité dans l'échantillon face-à-face (les interviews sont réparties de manière uniforme entre les différents jours de la semaine).

Si le mode d'enquête semble influencer les facteurs à l'origine des comportements de mobilité et d'immobilité, il serait intéressant de développer un modèle capable de distinguer les réels immobiles des 'soft refusals', d'une part, de quantifier l'impact du mode sur l'immobilité, d'autre part.

Conclusion générale : le web, un média d'avenir pour les enquêtes ménages déplacements

"L'étude des pratiques de déplacements n'est pas une discipline académique établie, elle met en jeu des disciplines variées (anthropologie, sociologie, psychologie et psychosociologie, géographie, économie, démographie, socio-économie, ingénierie), avec des difficultés de dialogue entre les différents groupes."
Orfeuil (2002a)

L'objectif d'une enquête ménages déplacements est de comprendre les comportements de mobilité et d'obtenir des informations utiles pour alimenter les modèles de planification stratégique des transports. Afin de répondre à ces exigences, il est impératif de s'appuyer sur un support méthodologique qui permette la comparaison des résultats dans le temps et dans l'espace. La méthodologie standard Certu (1998) est éprouvée depuis plusieurs années. Elle impose des règles strictes, aussi bien dans la phase de préparation de l'enquête que durant la mise en place et le déroulement de la collecte des données. Malgré ce cadre rigoureux, de nombreuses difficultés sont apparues lors des précédentes enquêtes, en termes de mobilisation des enquêteurs et de recrutement des ménages. Ces difficultés se sont accrues lors de l'enquête réalisée à Lyon en 2006, notamment à cause de l'étendue du périmètre d'enquête et de facteurs conjoncturels qui sont venus perturber le travail des enquêteurs (grèves dans les transports en commun, climat social tendu...). A cette occasion, certaines innovations méthodologiques ont permis d'accroître la précision des données, notamment en limitant le nombre de non-réponses. Ainsi, le Laboratoire d'Economie des Transports a proposé de mener une enquête web auprès des non-répondants au questionnaire traditionnel, administré en face-à-face.

Le premier objectif de cette thèse est de réfléchir sur la possible intégration du web dans les enquêtes de mobilité et d'explorer plus largement la combinaison de plusieurs modes de collecte de données (face à face et web). Le second est de s'interroger sur la comparabilité des résultats de l'enquête en ligne avec ceux obtenus par l'enquête traditionnelle en face-à-face. Notre travail a d'abord consisté à construire un questionnaire adapté à une diffusion

sur le web, en opérant des choix par rapport à sa version papier. Ensuite, nous avons mené une analyse comparative des profils et des comportements de mobilité des individus qui ont répondu à l'enquête par l'intermédiaire d'Internet, avec ceux soumis au questionnaire standard. Enfin, nous avons développé des techniques économétriques permettant, d'une part, de tenir compte des différences socio-économiques subsistant au sein des échantillons web et face-à-face et, d'autre part, d'identifier un effet média sur les réponses fournies.

La conclusion permet de tirer les enseignements de cette expérience. Nous abordons d'abord la portée et les limites des choix méthodologiques effectués (section I). Puis, nous présentons une synthèse des principaux résultats de l'analyse comparative des échantillons web et face-à-face (section II). Dans la section III, nous revenons sur les outils économétriques mobilisés pour traiter du problème délicat de la comparabilité des données et nous nous interrogeons sur la pertinence du rapprochement des deux échantillons. Puis, nous développons quelques perspectives de recherche, en vue de la préparation des prochaines enquêtes ménages déplacements (section IV).

I Portée et limites des choix méthodologiques

Notre recherche a permis de réaliser une enquête web auprès de la population des non-répondants à l'enquête ménages déplacements réalisée à Lyon en face à face. Le but est de proposer aux ménages qui ne veulent pas recevoir un enquêteur à domicile, ou dont les membres ne peuvent pas se rendre disponibles pour répondre à l'enquête, un média alternatif et moins contraignant pour saisir leurs réponses

I.1 Un protocole d'enquête mixte : web et face-à-face

L'échantillon de l'enquête en face à face pour le territoire du SCOT de l'agglomération Lyonnaise, qui est le territoire retenu pour le test de l'enquête web, était de 11 951 ménages. 6 375 interviews ont pu être réalisées, soit un taux de réponse de 53%, en baisse assez sensible par rapport aux enquêtes précédentes de 1995, 1985 et 1976. Ces enquêtes ont elles-mêmes connu un trend négatif, suivant en cela la tendance observée au niveau international (Arostic et Burt, 1999). Même si certaines procédures permettent de réduire le taux de non-réponse et que les méthodes de redressement permettent de limiter le biais introduit par la non-réponse totale, il reste nécessaire de postuler que les comportements de mobilité des ménages qui refusent de participer à l'enquête sont identiques à ceux des ménages ayant des caractéristiques socio-économiques similaires. Or, de nombreux travaux (Ampt, 1997; Bonnel, 2003) ont montré que les comportements étaient généralement corrélés au fait de ne pas répondre ou au nombre de tentatives nécessaires pour atteindre un répondant. La non-réponse est donc susceptible de biaiser la représentativité statistique des enquêtes et notamment la mobilité, ce qui impacte forcément l'évaluation des projets de transport, d'une part, et la planification des transports, d'autre part.

C'est pour essayer de limiter ce biais de non-réponse, que le Laboratoire d'Economie des Transports a proposé de réaliser l'enquête par le web auprès des non-répondants à l'enquête ménages déplacements en face-à-face. 4 335 internautes potentiels ont été sollicités par courriers postaux, car nous ne disposions pas de leur adresse électronique et pas toujours de leur numéro de téléphone. L'enquête s'est déroulée en deux vagues successives, chacune avec deux rappels, le tout au Printemps 2006. 536 individus ont accédé au site pour tenter de répondre à l'enquête. Mais seuls 369, soit un taux de réponse de 8,5%, ont fourni une réponse suffisamment complète et exploitable. Ce taux de réponse peut sembler faible, mais il faut se rappeler que l'enquête ne concerne que des personnes qui ont déjà refusé de répondre en face à face ou qu'il n'a pas été possible de contacter. L'analyse bibliographique nous a permis de développer une méthodologie originale pour mettre en ligne l'enquête ménages déplacements. Son expérimentation sur le terrain lyonnais nous permet de faire quelques préconisations complémentaires.

I.2 Le fort potentiel du web pour les enquêtes de mobilité

Couper (2002) précise qu'il existe plusieurs façons de considérer l'apport des nouvelles technologies pour améliorer la collecte de données d'enquête : remplacement, support, efficacité, contrôle... Grâce au web, les enquêtes se démocratisent, et il est désormais possible de toucher rapidement et à faible coût un grand nombre d'individus. La réalisation d'une enquête web nécessite des moyens très limités, dès lors que la méthodologie est en place. Il est ainsi possible d'augmenter fortement l'échantillon. Ensuite, il est probable qu'elle touche une population qui est beaucoup plus réticente aux médias traditionnels. Le fait que les internautes soient beaucoup moins présents dans l'annuaire, alors qu'ils sont tout autant équipés de téléphone fixe, semble indiquer qu'ils souhaitent davantage se protéger des 'intrusions' dans le ménage que constituent les enquêtes. Enfin, la souplesse du média permet d'éviter la saisie a posteriori des réponses et le renvoi manuel des questionnaires, et autorise la validation automatique des réponses, la possibilité d'insérer des filtres complexes dans le questionnaire et le choix du moment de la réponse. L'utilisation du web pour les enquêtes ménages déplacements permet de recueillir des données de qualité, qui viennent augmenter le taux de réponse global. Il est ainsi possible d'avoir une image plus fidèle du comportement de mobilité de la population, en recueillant les données d'individus qui ne participent pas aux enquêtes traditionnelles.

Les questions de design du questionnaire, de confidentialité des données, d'échantillonnage, de mode de contact des répondants potentiels, de pilotage de l'étude et de taux de réponse sont cependant des éléments méthodologiques à prendre en considération pour réaliser des enquêtes en ligne performantes (Andrews *et al.*, 0003). En effet, si le design du questionnaire est un outil qui permet de motiver le répondant et de faciliter la saisie des données, une mauvaise utilisation peut affecter la qualité des réponses. Or, les concepteurs d'enquête se laissent souvent porter par leurs compétences informatiques pour

créer des questionnaires qui restent inaccessibles au plus grand nombre d'individus. Un arbitrage doit ainsi être fait entre le taux de réponse, la qualité des données et les coûts de l'étude.

I.3 Nouveau média, nouvelles contraintes

Une règle de base en méthodologie des enquêtes consiste à adapter le questionnaire au média utilisé. La recherche de la comparabilité des données ne doit pas conduire à figer le questionnaire, quel que soit le média utilisé ¹⁵³. Au contraire, il faut chercher à utiliser les potentialités offertes par chacun des médias, pour simplifier le travail du répondant et lui rendre la tâche la plus simple et la plus logique possible. Nous avons tenté d'appliquer ce principe lors de la rédaction du questionnaire web de l'enquête ménages, en nous inspirant non pas du questionnaire en face à face, qui est très lourd, mais de celui préconisé par le CERTU dans le cadre des enquêtes téléphoniques en villes moyennes. Ce questionnaire a été simplifié pour l'adapter à une enquête non-administrée. De nombreux filtres ont également été ajoutés à l'enquête, pour éviter des questions inutiles ou redondantes. L'analyse des résultats montre cependant qu'il aurait été préférable de retenir une entrée activité, et non pas déplacements comme dans les enquêtes téléphoniques et face à face, pour recueillir les déplacements. Il s'agit de demander à l'enquêté de décrire toutes les activités qu'il a réalisées en dehors de son domicile (les déplacements découlant des changements de localisation), et non pas tous les déplacements qu'il a effectués. De plus, dans une enquête auto-administrée, il n'y a pas d'enquêteur pour stimuler la participation à l'enquête ou pour donner des explications. Nous avons donc fourni à l'enquêté suffisamment d'éléments permettant de comprendre les enjeux de l'enquête et d'explications sur la logique du questionnaire. Ce dernier devant rester le plus concis possible, nous avons effectué un travail particulier de présentation de la logique de l'enquête, à l'aide de courtes phrases d'explication ou de transition entre les différentes parties du questionnaire.

Par ailleurs, le recueil des localisations est indispensable dans une enquête déplacements. Lors des enquêtes en face à face ce recueil s'effectue à l'aide d'un zonage, l'enquêteur disposant de plans pour effectuer la codification en temps réel. Il semble difficile d'utiliser la même méthodologie sur le web, même si techniquement il serait possible de fournir des plans détaillés à l'internaute, pour qu'il effectue une codification de ses localisations. La codification doit donc être réalisée par l'analyste à partir des informations fournies par l'enquêté. L'expérience lyonnaise a montré que, dans la majorité des cas, cette codification était possible a posteriori. Cela nécessite toutefois un travail important et l'information fournie est parfois ambiguë. Il serait donc intéressant de tester une codification en temps réel comme cela se pratique dans de nombreux pays. Il est alors possible de fournir une information de qualité en coordonnées x-y,

¹⁵³La perception des questions et l'organisation du questionnaire dépendant largement du média utilisé, un questionnaire rigoureusement identique a toutes les chances de conduire à des résultats différents selon le mode de recueil des données.

qui autorise la compatibilité avec tous les découpages et notamment celui de l'enquête en face à face.

Lors de notre expérimentation, nous avons dû avoir recours à la transposition d'un logiciel CATI (logiciel pour la réalisation d'enquête téléphonique assistée par ordinateur). Cela entraîne une certaine lourdeur et un manque d'ergonomie dans le questionnaire web. Nous préconisons donc l'utilisation d'outils adaptés à la réalisation d'enquête web comme les logiciels CAWI (logiciel pour la réalisation d'enquête web assistée par ordinateur). Il est en effet indispensable de proposer à l'internaute une ergonomie aussi proche que possible des environnements web auxquels il a l'habitude d'accéder. Nous recommandons pour les prochaines enquêtes ménages déplacements qui intégreront une composante web la conception d'un cahier des charges précisant certains éléments méthodologiques, comme ceux mentionnés ci-dessus, à l'image de ce qui a été construit pour les enquêtes en face à face par le CERTU.

Enfin, les résultats des enquêtes ménages déplacements doivent pouvoir être généralisés à l'ensemble de la population de référence. Ceci impose la construction d'un plan d'échantillonnage et la représentativité statistique des ménages sélectionnés pour répondre à l'enquête. Si le faible taux de pénétration du web limite son utilisation dans les enquêtes de mobilité à l'échelle d'une grande agglomération, ce média peut en revanche faire partie d'un protocole mixte de recueil des données, de façon à éviter le biais de couverture provenant des personnes qui ne disposent pas d'une connexion internet ou qui n'ont pas les capacités de répondre à un questionnaire en ligne (Schaefer et Dillman, 1998). Dans les protocoles d'enquêtes mixtes, le web est un mode alternatif, où l'échantillon est construit selon un recrutement par une méthode plus traditionnelle (aucune base de données officielle et fiable regroupant les adresses e-mail des internautes n'est disponible à ce jour). Les avantages financiers procurés par l'utilisation du web sont cependant moindres et des problèmes liés à la comparabilité des données peuvent apparaître. Malgré les différences observées entre les deux échantillons, que nous détaillons dans la section suivante, il nous semble que cette expérience a mis en évidence l'intérêt du web pour la réalisation d'une enquête ménages déplacements.

II Les principaux enseignements de l'approche empirique

La seconde partie du bilan concerne l'analyse de la comparabilité de la mobilité. Cette analyse est assez complexe, dans la mesure où nous devons tenir compte des différences socio-économiques entre la population des internautes (individus qui ont répondu à l'enquête web) et de l'enquête standard (individu ayant répondu à l'enquête ménages déplacements en face à face), des différences de mobilité éventuelles de ceux qui acceptent de répondre sur le web par rapport à ceux qui répondent en face à face et enfin des différences qui peuvent être imputées au média d'enquête, indépendamment des éventuelles différences de comportement. Dans la pratique, il n'est pas possible d'identifier

de manière précise et séparée chacun des trois effets. Néanmoins, l'analyse des données nous permet de faire quelques constats et hypothèses.

II.1 Le profil des répondants web

Nous pouvons tout d'abord caractériser la population des internautes comparativement à celle de l'enquête standard. Nous observons que les ménages internautes :

- ont davantage de membres (2,60 vs. 2,38 personnes par ménage) ;
- sont mieux équipés en moyens de communication, à commencer par l'accès à internet, même si 20% n'ont pas de connexion internet à domicile. En revanche, ils sont moins souvent sur l'annuaire (seuls 60% sont inscrits sur l'annuaire France Télécom), ce qui pourrait en partie expliquer la plus grande difficulté à les joindre ou leur refus, car ils souhaitent probablement ne pas être dérangés. Le web est alors une opportunité pour eux de répondre à l'enquête, sans avoir à recevoir un enquêteur à domicile ;
- sont davantage motorisés, tant au niveau du ménage qu'au niveau du nombre moyen de véhicules par individu en âge de conduire (65% des personnes de 18 ans et plus possèdent entre 0,5 et 1 voiture sur le web, vs. 49% en face-à-face) ;
- ont un revenu annuel moyen supérieur, tant au niveau du ménage que par unité de consommation.

Au niveau des individus, il apparaît clairement que :

- les actifs internautes sont très fortement sur-représentés (71% versus 46% en face à face), tout particulièrement ceux dont l'âge est compris entre 30 et 59 ans, qui constituent l'essentiel de l'échantillon web de répondants ;
- les actifs internautes sont beaucoup plus souvent cadres (39% vs. 14%) ou employés (39% vs. 22,5%) et, de ce fait, travaillent plus souvent dans le centre de l'agglomération ;
- les internautes ont un niveau d'étude nettement supérieur (deux-tiers ont fait des études après le BAC versus 30% en face à face) et possèdent davantage le permis de conduire ;
- les femmes sont un peu plus présentes dans l'échantillon web (57%, vs. 52%), même si la différence reste assez faible ;
- la localisation des internautes, en revanche, ne diffère pas sensiblement de celle de la population face à face.

II.2 Des comportements de mobilité très différents

Ces différences socio-économiques ont forcément des conséquences sur l'analyse de la mobilité. C'est pourquoi, nous avons conduit notre analyse sur l'ensemble des échantillons, puis en redressant l'échantillon face-à-face à partir des caractéristiques socio-économiques les plus discriminantes entre les populations web et face-à-face, afin d'avoir des populations plus proches. Dans ce

Conclusion générale

dernier cas, certaines différences soulignées ci-dessus disparaissent, d'autres se réduisent, et quelques unes persistent. Les différences de mobilité que nous observons peuvent donc toujours être en partie liées aux différences socio-économiques :

- le nombre de déplacements déclaré par les internautes est plus faible, que ce soit pour l'ensemble de la population (3,00 vs 3,63 en face à face) ou après redressement de l'échantillon face-à-face (3,00 vs 4,02 en face à face). Cette différence s'explique tout d'abord par une immobilité nettement plus importante dans l'enquête web (19% vs 11% en face à face). Il est en effet plus facile de se déclarer immobile pour raccourcir l'enquête dans une enquête auto-administrée, qu'en présence d'un enquêteur qui peut assurer des relances. Si on reprend la même analyse uniquement pour les mobiles le jour de l'enquête, l'écart subsiste mais est réduit (ensemble de la population 3,71 vs 4,08 et après redressement de l'échantillon face-à-face 3,71 vs 4,34) ;
- l'analyse en termes de sorties du domicile (enchaînement des déplacements entre une sortie et un retour successif au domicile) fournit des résultats très proches (1,19 vs 1,47 pour l'ensemble de la population et 1,47 vs 1,65 pour les mobiles ; 1,19 vs 1,53 pour l'ensemble de la population et 1,47 vs 1,65 si on ne considère que les personnes mobiles après redressement). Pour les mobiles, le redressement n'affecte donc pas le nombre moyen de sorties. Une analyse plus fine montre que, pour l'ensemble de la population, le nombre de déplacements par sortie est similaire pour les deux modes d'enquête, mais qu'il est légèrement plus faible pour l'enquête web, dès lors que l'on redresse l'échantillon face-à-face. Pour les répondants web, la différence en nombre de déplacements est donc le résultat d'un double phénomène : une plus forte immobilité et un plus faible nombre de sorties ;
- la voiture est davantage utilisée par les internautes, mais ce résultat semble être lié à la sur-représentation des personnes actives. Lorsque l'échantillon face-à-face est redressé, on observe un rattrapage des déplacements en modes mécanisés. Ce résultat s'explique probablement par le niveau d'éducation et d'emploi plus élevé et par un accès plus large au permis de conduire. Il subsiste lorsque l'on limite l'analyse aux mobiles le jour de l'enquête. Pour la marche, quel que soit le mode de calcul, on observe une très faible utilisation (à peu près deux fois plus faible) par la population internaute. En revanche, pour les autres modes, la mobilité est du même ordre de grandeur ;
- l'analyse des motifs montre un poids très fort du travail, qui est largement lié à la surreprésentation des actifs chez les internautes. Lorsque l'on redresse l'échantillon face-à-face et que l'on considère uniquement les personnes mobiles, la mobilité liée au travail, que ce soit en termes de déplacements ou de sorties, est très proche pour les deux populations. En revanche, la sous-estimation est forte (de l'ordre d'un tiers) pour les accompagnements et notable pour les loisirs ;
- les distances et durées des déplacements et des sorties sont nettement plus

grandes chez les internautes. Ce résultat subsiste après le redressement de l'échantillon face-à-face, même si l'écart est un peu réduit avec +5 à 12% pour la distance et +15 à 20% pour la durée, selon que l'on considère les déplacements ou les sorties. Les déplacements des internautes sont moins fréquemment internes à une zone. La distribution des lieux de résidence est proche, mais le travail est plus souvent localisé dans le centre. Ces éléments expliquent probablement en partie les distances et durées plus importantes. Il est dès lors intéressant d'analyser les budgets temps distance après redressement de l'échantillon face-à-face. L'analyse montre que les internautes ont un budget temps supérieur à celui des individus enquêtés en face à face, mais qu'en revanche leur budget distance est légèrement plus faible. Ces résultats pourraient expliquer une mobilité plus faible ;

- la distribution des déplacements dans le temps montre des pointes un peu plus marquées surtout le matin et une pointe plus tardive le soir alors que celle du matin se déroule aux mêmes horaires. L'amplitude de la présence hors domicile est donc plus importante chez les internautes.

II.3 Les interprétations possibles de ces résultats

Ces conclusions sont cohérentes avec l'hypothèse d'une sous-déclaration imputable au média web, dans la mesure où l'on sait que les risques d'omission de déclaration de déplacements concernent surtout les déplacements ou sorties courts en temps ou en distance et des motifs moins contraints (même si l'accompagnement n'est pas forcément considéré comme non contraint). On peut toutefois objecter que les différences socio-économiques peuvent expliquer au moins en partie les différences de mobilité, notamment de marche à pied. Les internautes ont un niveau d'étude et un revenu nettement supérieur, conduisant à une très forte motorisation. Ils sont plus souvent cadres et employés et travaillent plus souvent dans le centre, ce qui conduit à des durées hors domicile plus importantes, réduisant les possibilités de participation à des activités moins contraintes et l'usage de la marche. De plus, l'analyse des budgets temps et des budgets distance après redressement montre que les internautes ont un budget temps déplacements supérieur à celui des répondants en face à face et un budget distance très légèrement inférieur. Si l'on en croit l'hypothèse de Zahavi de constance des budgets temps de déplacements (Zahavi, 1979), c'est le niveau élevé du budget temps de déplacements des internautes qui viendraient limiter leur mobilité à un niveau plus faible de celui des répondants en face à face.

L'analyse des données ne permet pas de conclure formellement entre ces deux hypothèses. Il nous semble que chacune d'elles explique très probablement une partie du différentiel de mobilité observé. Pour aller plus loin dans l'analyse, il est nécessaire de développer des techniques économétriques permettant, d'une part, de tenir compte des différences socio-économiques subsistant au sein des populations d'actifs et, d'autre part, d'identifier un effet média.

III L'intérêt de l'approche économétrique

L'analyse comparative des échantillons web et face-à-face met en évidence des résultats sensiblement différents, même lorsque nous corrigeons les différences socio-économiques qui existent entre les deux échantillons. Les internautes se déplacent moins que les individus répondant en face à face. Cet écart s'explique par un double effet : une immobilité plus grande des internautes et une moindre déclaration de déplacements. Nous avons appliqué des développements économétriques afin de valider l'existence de telles différences et de pouvoir ensuite les quantifier.

III.1 L'identification du biais de sélection

Les internautes ont des caractéristiques sociodémographiques différentes du reste de la population (niveau d'étude et revenus élevés, plutôt actifs...). En posant l'hypothèse que ces différences impactent les comportements, l'utilisation du web comme média d'enquête serait un moyen d'augmenter la qualité des données. Le risque est cependant de confondre le phénomène étudié (le nombre de déplacements) avec le processus de sélection des répondants (web ou face-à-face). On peut conclure qu'il existe un problème d'auto-sélection, si le choix du web exerce un effet différent sur la mobilité des internautes qu'il n'aurait exercé sur celle des répondants en face-à-face si ces derniers avaient rempli le questionnaire en ligne. A l'aide de techniques économétriques, nous avons montré qu'il est possible de tester l'existence d'un biais de sélection des répondants, de le quantifier, puis d'isoler l'impact des différences socio-économiques propres aux enquêtés de l'effet média. La méthode que nous avons utilisée se base sur la procédure d'estimation en deux étapes empruntée à Heckman. Grâce à cette méthode, les coefficients estimés des variables du modèle explicatif de la mobilité ne sont plus biaisés et reflètent l'impact des variables socio-économiques sur le nombre de déplacements déclaré, indépendamment du choix du mode de recueil de données.

Dans notre exemple, le biais de sélection est statistiquement significatif. Le mode d'enquête a donc une incidence sur la mobilité qu'une simple régression linéaire ne peut mettre en évidence : les répondants web déclareraient davantage de déplacements s'ils répondaient en face-à-face et les individus interrogés à domicile déclareraient moins de déplacements sur le web. L'exercice montre que les variables socio-économiques qui impactent la mobilité des répondants web sont le sexe, le nombre de personnes du ménage, ainsi que la distance entre le domicile et le centre de l'agglomération. Certaines hypothèses ont été formulées pour tenter d'expliquer ces différences comportementales :

- Le fait d'être un homme impacte négativement le nombre de déplacements. Les femmes se déplacent davantage que les hommes, car elles doivent gérer des activités familiales en plus de leur activité professionnelle. Cet effet est amplifié en ce qui concerne les répondants web, dont le niveau d'emploi, plus exigeant, laisse peu de latitude pour les dépla-

- cements en soirée ;
- A contrario, le nombre de déplacements des répondants web augmente avec le nombre de personnes du ménage. Ces derniers disposent d'un revenu annuel supérieur. Il est donc probable que le nombre de personnes accroisse les besoins et donc les déplacements (pour motif achat, loisir ...) des membres du ménage ;
 - Enfin, le nombre de déplacements des internautes diminue avec l'éloignement du lieu de résidence du centre de l'agglomération. Les répondants web occupent davantage des emplois de cadres et d'employés, situés dans le centre de l'agglomération. Avec l'éloignement du lieu de résidence, la distance domicile-travail augmente et le temps disponible pour les activités moins contraintes se réduit.

III.2 La prise en compte des déterminants de l'immobilité

Dans les modèles explicatifs de la mobilité utilisés dans la section III.1, seules les personnes mobiles sont considérées. Or, un nombre important de répondants ont déclaré ne pas avoir effectué de déplacements durant la période de référence de l'étude (11% en face-à-face et 19% sur le web). Ne pas tenir compte de ces répondants nous prive d'information. A contrario, les intégrer directement dans les modèles explicatifs de la mobilité biaise les résultats.

Des développements complémentaires séparant les facteurs explicatifs de la décision de se déplacer de ceux qui influent sur le niveau de la mobilité, comme le modèle de Poisson 'Hurdle', permettent d'approfondir l'analyse comparative du nombre de déplacements effectué selon le mode d'enquête. Les analyses montrent que les facteurs qui motivent la décision de mobilité ne sont pas les mêmes que ceux qui influent sur le niveau de mobilité quotidienne. Le modèle 'Hurdle' apporte donc une amélioration dans la compréhension des pratiques de mobilité.

Dans l'échantillon face-à-face, nous concluons que :

- La probabilité de se déplacer est plus élevée pour les hommes, mais, une fois cette décision prise, ce sont les femmes qui effectuent le plus grand nombre de déplacements ;
- La probabilité de se déplacer est plus importante chez les personnes actives, mais parmi les personnes qui se déplacent, ce sont les inactifs qui sont les plus mobiles ;
- Les variables liées à la motorisation ou à l'équipement en moyens de communication des ménages impactent fortement la décision de se déplacer, mais ont un rôle modéré sur le niveau de mobilité.

En ce qui concerne l'échantillon web, l'influence des variables socio-économiques est parfois différente :

- De même que pour l'échantillon face-à-face, l'inactivité joue très négativement sur la probabilité de se déplacer, mais une fois la décision de se

déplacer prise, les inactifs se déplacent davantage ;

- La probabilité de se déplacer est moins importante le vendredi, mais augmente si l'individu a choisi de déclarer ses revenus.

Beaucoup moins de variables sont significatives lorsque nous appliquons les modèles économétriques à l'échantillon web, du fait de la taille réduite de l'échantillon. Nous pouvons cependant tirer quelques enseignements intéressants de ces analyses. En ce qui concerne les similitudes entre les populations web et face-à-face, le statut des répondants est à interpréter avec précaution. Les actifs sont moins immobiles que les inactifs, du fait que la majorité d'entre eux doivent se rendre sur leur lieu de travail, mais parmi les personnes mobiles, les inactifs ont un niveau de mobilité supérieur (davantage de temps libre pour réaliser des activités en dehors du domicile). Des différences liées au mode d'enquête apparaissent également entre les deux échantillons. Le web, par nature auto-administré, laisse une grande latitude au répondant sur le choix du moment de connexion et donc du jour de référence considéré pour le recueil des déplacements. Ceci explique en partie pourquoi la variable 'Vendredi' ressort comme significative dans les modèles appliqués à l'échantillon web, mais pas dans ceux menés sur l'échantillon face-à-face.

Ces résultats sont importants à plusieurs titres. Premièrement, sur un plan économique. Deuxièmement, d'un point de vue académique, notre travail contribue à la littérature empirique existante sur deux points : l'estimation dissociée des déterminants de la mobilité et de son intensité.

III.3 Les limites de notre cadre conceptuel

Dans nos modèles explicatifs de la mobilité quotidienne, nous avons tenu compte de trois types de variables indépendantes : des variables socio-économiques, des variables liées au ménage (motorisation et équipement en biens de communication) et une variable relative au mode de réponse utilisé. Nous supposons donc que la décision de mobilité et son intensité sont une conséquence directe des caractéristiques individuelles et de celles du ménage auquel appartient le répondant, ainsi que du média d'enquête. Il est toutefois possible que d'autres variables, non observées, influent fortement sur les habitudes de mobilité des individus. Il peut s'agir par exemple du système de transports accessible par le répondant, de ses convictions en matière d'écologie ou de contraintes en termes d'horaires de travail. Nous pouvons ainsi noter que les gens qui se déplacent sont extrêmement dissemblables (piétons, cyclistes, usagers des transports publics, automobilistes...) et que leurs différences (notamment en termes de capacité) sont souvent ignorées des modèles explicatifs de la mobilité. Le pouvoir explicatif modéré des modèles que nous avons développés peut également provenir d'une catégorisation insuffisante des variables, qui masque des différences socio-économiques. Ainsi, dans certaines catégories relativement larges (niveau d'étude, revenu...), les individus peuvent être très différents, sans que cette diversité ne soit prise en compte par le modèle.

Par ailleurs, les phénomènes de multicolinéarité potentiels entre les variables peuvent poser des problèmes d'interprétation sur la nature des liens

entre les caractéristiques de l'individu ou de son ménage et le niveau de mobilité. Ces liens sont interprétés au regard de la littérature. Il n'est cependant possible d'extraire d'une étude empirique que de simples corrélations entre variables et il est délicat de transformer ces corrélations en lien de causalité. La détermination de liens causaux est très difficile dans un système où tout est en interaction, puisqu'il est possible que la prise en considération d'une tierce variable conduise à une autre interprétation.

La richesse des données d'une enquête de comportement permet des analyses descriptives souvent utiles aux décideurs politiques. Mais si la journée est une 'tranche de vie' suffisante pour les besoins de prévision des flux, elle est de moins en moins suffisante pour la compréhension des pratiques personnelles (Orfeuill *et al.*, 2000). Les disparités de mobilité sont en partie structurées par les catégories sociodémographiques, mais résultent également pour partie du choix de mode de vie, impliquant une mobilité plus ou moins importante. '*L'organisation de l'emploi du temps, les activités et la mobilité ne jouent pas à l'échelle d'un seul individu, mais s'inscrivent dans un système complexe où chaque individu du ménage doit se coordonner avec les autres* (Orfeuill, 2002a)'. Pour bien comprendre les enjeux de la mobilité, il faudrait traiter conjointement ce qui relève des déplacements, de l'organisation des activités et des contextes familiaux et socio-économiques (André-Poyaud *et al.*, 2008). De plus, les observations ne concernent qu'une journée, et nous savons que la variance intrapersonnelle entre les jours de la semaine est élevée, en ce qui concerne la mobilité. Enfin, la taille de l'échantillon de l'enquête face-à-face est importante, et permet de retrouver une diversité des profils de répondants. Dans l'enquête web, l'effectif est plus restreint. Il est donc probable que des individus atypiques biaisent en partie certains résultats.

IV Les perspectives de recherche

La difficulté croissante pour obtenir des données d'enquêtes représentatives de la population visée et la complexité des informations nécessaires à l'alimentation de modèles de plus en plus sophistiqués ne permettent généralement plus de recueillir toutes les données au cours d'une même enquête ou selon une méthodologie unique. De plus en plus d'enquêtes reposent sur des protocoles complexes, et la mixité des modes d'enquêtes devient très populaire, et même une norme, surtout aux USA et en Europe de l'ouest (De Leeuw, 2005). Les protocoles mixtes ne sont pas nouveaux, mais des modes apparaissent et avec eux de nouvelles combinaisons.

L'enquête que nous avons réalisée est la première du genre en France et à notre connaissance dans le monde ¹⁵⁴. Comme pour toute expérimentation, nous avons identifié des pistes d'amélioration, notamment en termes d'ergonomie pour simplifier la tâche des internautes. Il nous semble souhaitable de renouveler l'expérience à l'occasion d'une autre enquête ménages déplacements

¹⁵⁴Il y a déjà eu d'autres enquêtes de mobilité réalisées sur internet, mais elles ont toujours été couplées à des enquêtes postales ou téléphoniques et non pas à des enquêtes en face à face.

en France. Si les résultats positifs se confirment, il nous semble alors envisageable de proposer une combinaison des modes d'enquête, comme cela se pratique déjà dans certains pays, le choix étant laissé à l'enquêté du média de réponse. En revanche, il est totalement exclu d'avoir recours au web de manière exclusive, compte tenu du taux de réponse, d'une part, et du taux de couverture de ce média dans la population, d'autre part, sans compter le problème de la constitution d'un échantillon représentatif, nécessaire pour généraliser les résultats à l'ensemble de la population.

L'intégration de plusieurs sources de données dans le processus d'analyse et de décision en transport devient de plus en plus complexe, du fait notamment de l'évolution rapide de l'état, de la forme et de la quantité de données disponibles. Il serait intéressant d'exploiter conjointement les résultats des enquêtes de mobilité et du recensement, ou de rapprocher des données provenant de différentes enquêtes de mobilité. Les méthodes de fusion de données et d'intégration des méthodes d'acquisition de données doivent s'adapter à ces nouveaux types d'information (Bayart *et al.*, 2009). La littérature présente les connaissances théoriques et relate des expériences sur le design des enquêtes mixtes, sans en discuter les implications en termes de comparabilité des données. Est-ce que les personnes qui répondent par un mode auraient donné les mêmes réponses s'ils avaient répondu par un autre mode ? Une fois quantifié l'impact réel du mode d'enquête sur le nombre de déplacements déclaré, il est tentant de redresser l'échantillon web, en vue d'une fusion des deux bases de données, web et face-à-face. Nous avons montré qu'en moyenne, un répondant web déclarait 0,5 déplacement de moins qu'il n'aurait déclaré, toutes choses égales par ailleurs, en face-à-face. Cette option, bien que légitime, n'est pas sans risque si nous considérons la taille relativement réduite de l'échantillon web, qui sert de base à nos estimations, et les interactions entre certaines variables socio-économiques et le mode d'enquête.

Malgré la difficulté générée par le rapprochement de bases de données issues de modes d'enquêtes différents, nous pensons que l'utilisation du face-à-face ne permet plus d'interroger un échantillon de ménages représentatif de la population générale. Le web permet d'accéder, à faible coût, à une population qui répond peu aux enquêtes et qui a des comportements de mobilité spécifiques. Suivant l'objet de leurs analyses, les chercheurs doivent alors décider quel biais est le plus important : l'erreur de mode ou la non-réponse.

**

Bibliographie

- ABDEL-ATY, M. (2003). Hybrid distribution and response techniques for an origin-destination travel survey. *ITE Journal*, Vol. 73 (No. 2), pp : 22–27.
- ABDELWAHAB, H. T. et ABDEL-ATY, M. (2003). Use of the internet to enhance transportation data collection. Institute of Transportation Engineers Annual Meeting and Exhibit, Seattle, Washington.
- ABRAHAM, S. Y., MILLER STEIGER, D. et SULLIVAN, C. (1998). Electronic and mail self-administered questionnaires : A comparative assessment of use among elite populations. American Statistical Association, Proceedings of the Section on Survey Research Methods, pp : 833-841.
- ACOCK, A. C. (1997). Working with missing data. *Family Science Review of Economics and Statistics*, Vol. 10 (No. 1), pp : 76–102.
- ADLER, T., RIMMER, L., BANDY, G. et SCHELLINGER, D. (2001). Use of respondent-interactive geocoding in the baltimore mode choice survey. *Transportation Research Record*, (No. 1719), pp : 154–158.
- ADLER, T., RIMMER, L. et CAPENTER, D. (2002). Use of internet-based household travel diary survey instrument. *Transportation Research Record*, (No. 1804), pp : 134–143.
- ADLER, T.-J. (2003). Reducing the effects of item nonresponse. Dans STOPHER, P. R. et JONES, P. M., éditeurs : *Transport Survey Quality and Innovation*, pages 529–542. Elsevier Press, Pergamon, Oxford.
- AKAIKE, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, Vol. 19, pp : 716–723.
- ALRECK, P. L. et SETTLE, R. B. (1995). *The Survey Research Handbook*. McGraw-Hill Inc., Chicago, 2nd édition.
- ALSNIH, R. (2004). New technology and travel surveys : the way forward. Working Paper No. ITS-WP-04-1, Institute of Transport Studies, University of Sydney, Australia.
- ALSNIH, R. (2005). Characteristics of web-based surveys and applications in travel research. 7th International Conference on Travel Survey Methods, Costa Rica, August. Actes sur CD-ROM.

- ALVAREZ, R.-M. et VAN BESELAEREE, C. (2003). Web-based surveys. *Internet Surveys of American Opinion (ISAO)*.
- ALVEY, W. (2004). Highlights from the session on web-based data collection. Federal Committee on Statistical Methodology, Statistical policy seminar, Web based data collection session, Washington, December.
- AMEMIYA, T. (1984). Tobit models : a survey. *Journal of Econometrics*, Vol. 24 (No. 1-2), pp : 3–61.
- AMPT, E.-M. (2000). Understanding the people we survey. International Conference Transport Surveys : Raising the standard, Grainau, Germany, May 24-30.
- AMPT, E.-M. (2003). Respondent burden. *Dans* STOPHER, P. R. et JONES, P. M., éditeurs : *Transport Survey Quality and Innovation*, pages 507–521. Elsevier Press, Pergamon, Oxford.
- AMPT, E.-S. (1997). Response rates - do they matter? *Dans* BONNEL, P., CHAPLEAU, R., LEE-GOSSELIN, M. et RAUX, C., éditeurs : *Les enquêtes de déplacements urbains : mesurer le présent, simuler le futur*, pages 115–125. Collection Programme Rhône-Alpes Recherches en Sciences Humaines, Lyon.
- ANDERSON, M., ORTEGA, J., GHOLSTON, S. et JONES, S. (2006). *Application of Advanced Technology in Transportation*, chapitre Development and Analysis of an Internet-Based Travel Survey, pages 671–676. American Society of Civil Engineers, Chicago.
- ANDRÉ-POYAUD, I., CHARDONNEL, S., CHARLEUX, L. et TAKAKA, K. (2008). La mobilité au coeur des emplois du temps des citoyens. *Dans* CHALAS, Y. et PAULHIAC, F., éditeurs : *La mobilité qui fait la ville : Actes des 3e rencontres internationales en Urbanisme de l'Institut d'Urbanisme de Grenoble*, Série Débats 66. Certu.
- ANDREWS, D., NONNECKE, B. et PREECE, J. (2003). Electronic survey methodology : a case study in reaching hard to involve internet users. *International Journal of Human-Computer Interaction*, Vol. 2 (No. 16), pp : 185–210.
- ARENTZE, T., HOFMAN, F. et TIMMERMANS, H. (2004). Multi-faceted activity-travel adjustment strategies in response to possible congestion pricing scenarios using an internet-based stated adaptation experiment. *Transport Policy*, Vol. 11 (No. 1), pp : 31–41.
- ARMOOGUM, J., AXHAUSEN, K., HUBERT, J.-P. et MADRE, J.-L. (2005). Immobility and mobility seen through trip-based versus time-use surveys. *Transport Reviews*, Vol. 28 (No. 5), pp : 641–658.

Bibliographie

- ARMOOGUM, J. J. et MADRE, J.-L. (1997). Correction de la non-réponse dans l'enquête transports 1993-94 : l'exemple des déplacements quotidiens. *Dans* BONNEL, P., CHAPLEAU, R., LEE-GOSSELIN, M. et RAUX, C., éditeurs : *Les enquêtes de déplacements urbains : mesurer le présent, simuler le futur*, pages 127–143. Collection Programme Rhône-Alpes Recherches en Sciences Humaines, Lyon.
- ATROSTIC, B.-K. et BURT, G. (1999). Household non-reponse : what we have learned and a framework for the future. Statistical Policy working paper n° 28, Federal Committee on Statistical methodology, Office of Management and Budget, Washington, pp : 153-180.
- BACHMANN, D., ELFRINK, J. et VAZZANA, G. (2000). E-mail and snail mail face off in rematch. *Marketing Research*, Vol. 11 (No. 4), pp : 10–15.
- BADAGLIACCO, J. M. (1990). Gender and race differences in computing attitudes and experience. *Social Science Computer Review*, Vol. 8 (No. 1), pp : 42–63.
- BAINES, A. D., PARTIN, M. R., DAVERN, M. et ROCKWOOD, T. H. (2007). Mixed-mode administration reduced bias and enhanced poststratification adjustments in a health behavior survey. *Journal of clinical epidemiology*, Vol. 60 (No. 12), pp : 1246–1255.
- BANISTER, D. (1997). Reducing the need to travel. *Environment and Planning B*, Vol. 24 (No. 3), pp : 437–449.
- BATES, N. (2001). Internet versus mail as a data collection methodology from a high coverage population. American Statistical Association Annual Meeting, USA, August.
- BAUELLE, G., DARRIS, G., OLLIVRO, J. et PIHAN, J. (2003). Les conséquences d'un choix résidentiel périurbain sur la mobilité : pratiques et représentations des ménages. 3ème colloque du Groupe de Travail Mobilités spatiales et fluidité sociale (GT23) : Offre urbaine et expériences de la mobilité, Stasbourg, France, 20-21 et 22 mars. <http://www.cybergeo.eu/index3430.html> (site visité le 14/02/08).
- BAYART, C., BONNEL, P. et MORENCY, C. (2009). Survey mode integration and data fusion : methods and challenges. *Dans* BONNEL, P., LEE-GOSSELIN, M., ZMUD, J. et MADRE, J.-L., éditeurs : *Transport Survey Methods : Keeping up with a Changing World*. Emerald press.
- BECKENBACH, A. (1995). Computer assisted questionning : The new survey methods in the perception of the respondents. *BMS*, Vol. 48, pp : 82–100.
- BENFIELD, J. A. et SZLEMKO, W. J. (2006). Internet-based data collection : Promises and realities. *Journal of Research Practice*, Vol. 2 (No. 2), pp : 30–51.

- BERGER, M. (2004). Mobilités résidentielles, mobilités quotidiennes : une approche des déterminants sociaux des aires de déplacement en région parisienne. Communication ESO, LADYSS, Rennes, Octobre.
- BERK, R.-A. (1983). An introduction to sample selection bias in sociological data. *American Sociologic Review*, Vol. 48 (No. 3), pp : 386–398.
- BERRY, D. M. (2004). Internet research : Privacy, ethics and alienation-an open source approach. *Internet Research*, Vol. 14 (No. 4), pp : 323–332.
- BICKART, B. et SCHMITTEIN, D. (1999). The distribution of survey contact and participation in the united states : Constructing a survey-based estimate. *Journal of Marketing Research*, Vol. 36 (No. 2), pp : 286–294.
- BIGOT, R. et CROUTTE, P. (2007). La diffusion des technologies de l'information dans la société française. CREDOC, 210 pages.
- BISHOP, I. D. (1997). Testing perceived landscape color-difference using the internet. *Landscape and Urban Planning*, Vol. 37 (No. 3-4), pp : 187–196.
- BOGEN, K. (1996). The effect of questionnaire length on response rates - a review of the literature. American Association of Public Opinion Research meeting, Research Method Section.
- BOHARA, A. K. et KRIEG, R. G. (1996). A poisson hurdle mode of migration frequency. *Journal of Regional Anamysis and Policy*, Vol. 26 (No. 1), pp : 37–45.
- BONNAFOUS, A. et PUEL, H. (1983). *Physionomies de la ville*. Série Initiation économique. Éditions ouvrières.
- BONNEL, P. (1985). *Les besoins latents de déplacement, une méthode de révélation appliquée au cas de Saint-Priest*. Thèse de docteur-ingénieur en économie des transports, Laboratoire d'Économie des Transports, Université Aix-Marseille II.
- BONNEL, P. (1995). An application of activity-based travel analysis to simulation of change in behaviour. *Transportation*, Vol. 22 (No. 1-2), pp : 73–93.
- BONNEL, P. (2003). Postal, telephone and face-to-face surveys : how comparable are they ? Dans STOPHER, P. R. et JONES, P. M., éditeurs : *Transport Survey Quality and Innovation*, pages 215–237. Elsevier Press, Pergamon, Oxford.
- BONNEL, P. (2004). Web-based survey : potential for travel surveys ? COST WATCH, Namur, 2-3 décembre, 12 pages.
- BONNEL, P. et ARMOOGUM, J. (2005). National transport surveys : What can we learn from international comparisons ? European Transport Conference, Strasbourg, Octobre, 24 pages. Actes sur CD-ROM.

Bibliographie

- BONNEL, P. et LE NIR, M. (1998). The quality of survey data : telephone versus face-to-face interviews. *Transportation*, Vol. 25 (No. 2), pp : 147–167.
- BOSNJAK, M. et TUTEN, T.-L. (2001). Classifying response behaviour in web-based surveys. *Journal of Computer-Mediated Communication*, Vol. 6 (No. 3). <http://jcmc.indiana.edu/vol6/issue3/boznjak.html> (visited on 10/24/05).
- BOUNIE, D., BOURREAU, M. et FRANÇOIS, A. (2006). Les déterminants de la détention et de l'usage de la carte de crédit : une analyse empirique sur données individuelles françaises. Télécom Paris, CREST-LEI et LAEP-Paris I. 69 pages.
- BRADLEY, M. A. (2004). Process data for understanding and modelling travel behaviour. 7th International Conference on Travel Survey Methods, Costa Rica, August.
- BRADLEY, N. (1999). Sampling for internet surveys : an examination of respondent selection for internet research. Working paper, University of Westminster, UK. <http://users.wmin.ac.uk/bradlen/papers/sam06.html> (visited on 10/18/05).
- BRAUNSBERGER, K., WYBENGA, H. et GATES, R. (2007). A comparison of reliability between telephone and web based surveys. *Journal of Business Research*, Vol. 60, pp : 758–764.
- BRÖG, W. et MEYBURG, A. (1980). Non response problems in surveys : an empirical investigation. *Transportation Research Record*, Vol. 775, pp : 34–38.
- BRÖG, W. et MEYBURG, A. H. (1982). Influence of survey methods on the results of representative travel surveys. 61st Transportation Research Board Meeting, January.
- BRODATY, T., CRÉPON, B. et FOUGÈRE, D. (2007). Les méthodes micro-économétriques d'évaluation et leurs applications aux politiques actives de l'emploi. *Économie et Prévision*, Vol. 1, pp : 93–118.
- BURKE, L. A. et JAMES, K. E. (2006). Using online surveys for primary research data collection : lessons from the field. *International Journal of Innovation and Learning*, 3 (1), 16–30.
- CAMBRIDGE SYSTEMATICS, I. (1996). Scan of recent travel surveys. Report DOT T-97-08, Federal Highway Administration, U.S. Department of Transportation. 110 pages.
- CAMERON, A. C. et TRIVERDI, P.-K. (2005). *Microeconometrics Methods and Applications*. Cambridge University Press, New York.

- CASTAIGNE, M., CORNELIS, E., HUBERT, J.-P. et TOBIN, J. (2006). Intégration des recherches quantitatives et qualitatives sur la mobilité quotidienne et les temporalités sociales (motus & quanli), rapport final du projet cp/62. FUNDP, Faculté des sciences, département de mathématiques, Namur. 142 pages.
- CERTU (1998). *L'enquête ménages déplacements « méthode standard »*. Editions du CERTU, Lyon.
- CETE (2005). Enquête ménages déplacements et moyens de communication des ménages « standard certu » de l'aire métropolitaine lyonnaise 2005-2006. Certu. Lyon, 67 pages.
- CHAPLEAU, R., ALLARD, B., TRÉPANIER, M. et MORENCY, C. (2001). Les logiciels d'enquête transport comme instruments incontournables de la planification stratégique. *Recherche Transport Sécurité*, Vol. 70, pp : 59–77.
- CHO, H. et LAROSE, R. (1999). Privacy issues in internet surveys. *Social Science Computer Review*, Vol. 17 (No. 4), pp : 421–434.
- CHRISTENSEN, L. (2004). Busy people are hard to reach. 7th International Conference on Travel Survey Methods, Costa Rica, August. Actes sur CD-ROM.
- CHRISTENSEN, L. (2006). Possible explanation for an increasing share of no-trip respondents. Dans STOPHER, P. et STECHER, C., éditeurs : *Travel Survey Methods : Quality and future Directions*. Elsevier Press, Pergamon.
- CLELAND, K. (1996). On-line research costs about half that traditional methods. *Advertising Age's Business Marketing*, Vol. 81 (No. 4), pp : 8–9.
- COBANOGLU, C., WARDE, B. et MOREO, P. J. (Vol. 2001). A comparison of mail, fax and web-based survey methods. *International Journal of Market Research*, 43 (No. 4), pp : 441–452.
- COMLEY, P. (1996). Use of the internet as a data collection method. World Association of Opinion and Marketing Research Professional (ESOMAR) Symposium, November.
- CONN, C. (2005). Using the internet for surveying : Techniques for designing, developing & delivering. Office of Academic Assessment, Northern Arizona University, USA. <http://www4.nau.edu/assessment/main/research/responserates.htm> (visited on 10/18/05).
- CONRAD, F. G., COUPER, M. P., TOURANGEAU, R. et BAKER, R. (2003). Use and non-use of clarification features in web surveys. 58th Annual Conference of the American Association of Public Opinion Research, Nashville, TN.

Bibliographie

- CONRAD, F. G. et SCHOBBER, M. F. (2000). Clarifying question meaning in a household telephone survey. *Public Opinion Quarterly*, Vol. 64 (No. 1), pp : 1–28.
- CONVERSE, P. E. (1974). Nonattitudes and american public opinion : Comments. the status of nonattitudes. *American Political Science Review*, Vol. 68, pp : 650–666.
- COOK, C., HEATH, F. et THOMPSON, R. L. (2000). A meta-analysis of response rates in web or internet based surveys. *Educational and Psychological Measurement*, Vol. 60 (No. 6), pp : 821–836.
- COUPER, M.-P. (2000). Web surveys : a review of issues and approaches. *Public Opinion Quarterly*, Vol. 65 (No. 2), pp : 230–253.
- COUPER, M.-P. (2001). The promises and perils of web surveys. Association for Survey Computing Conférence : The challenge of the Internet, May.
- COUPER, M.-P. (2002). New technologies and survey data collection : challenges and opportunities. International Conference on Improving Surveys, Copenhagen, Denmark, August.
- COUPER, M.-P., BLAIR, J. et TRIPPLET, T. (1999). A comparison of mail and e-mail for a survey of employees in federal statistical agencies. *Journal of Official Statistics*, Vol. 15 (No. 1), pp : 39–56.
- COUPER, M.-P. et NICHOLS, W. L. I. (1998). The history and development of computer assisted survey information collection methods. Dans COUPER, M. P., BAKER, R. P., BETHLEHEM, P., CLARK, C., MARTIN, J., NICHOLS, W. L. et O'REILLY, J. M., éditeurs : *Computer assisted survey information collection*, pages 1–22. John Wiley & Sons, New York.
- COUPER, M.-P. et ROWE, B. (1996). Evaluation of a computer-assisted self-interview component in a computer-assisted personal interview survey. *Public Opinion Quarterly*, Vol. 60 (No. 1), pp : 89–105.
- COUPER, M.-P., TRAUGOTT, M.-W. et LAMIAS, M.-J. (2001). Web survey design and administration. *Public Opinion Quarterly*, Vol. 64 (No. 4), pp : 464–494.
- CRAWFORD, S., MC CABE, S., COUPER, M.-P. et BOYD, C. (2002). From mail to web : improving response rates and data collection efficiencies. International Conference on Improving Surveys, Copenhagen, Denmark , August.
- DAVIDSON, R. et MCKINNON, J. (1993). *Estimation and Inference in Econometrics*. Oxford University Press, New York.
- DE HEER, W. F. et MORITZ, G. (2000). Data quality problems in travel surveys : an international overview. International Conference Transport Surveys : Raising the standard, Grainau, Germany, May 24-30.

- DE LA GODELINAIS, M. C. et MERCIER, M. A. (1992). L'équipement des ménages en 1990. *INSEE Première*, (No. 225).
- DE LEEUW, E.-D. (1992). *Data quality in mail, telephone and face-to-face surveys*. T.T. Publikaties, Amsterdam.
- DE LEEUW, E.-D. (2005). To mix or not to mix data collection modes in surveys. *Journal of Official Statistics*, Vol. 21 (No. 2), pp : 233–255.
- DE LEEUW, E.-D. et VAN DER ZOWEN, J. (1992). Data quality and mode of data collection : methodology and explanatory model. *Dans* LEBART, L., éditeur : *La qualité de l'information dans les enquêtes*, pages 11–31. Dunod, Paris.
- DE ROUVRAY, C. et COUPER, M.-P. (2002). Designing a strategy for reducing 'no opinion' responses in web-based surveys. *Social Science Computer Review*, Vol. 20 (No. 1), pp : 3–9.
- DEUTSKENS, E., JONG, A. D. et RUYTER, K. D. (2006). Comparing the generalizability of online and mail surveys and cross-national service quality research. *Marketing Letters*, Vol. 17, pp : 119–136.
- DEVILLE, J.-C., SÄRNDAL, C.-E. et SAUTORY, O. (1993). Generalized raking procedures in survey sampling. *Journal of the American Statistical Association*, Vol. 88 (No. 423), pp : 1013–1020.
- DILLMAN, D.-A. (1978). *Mail and telephone surveys : The total design method*. Wiley-Interscience, New York.
- DILLMAN, D.-A. (2000). *Mail and internet survey : the tailored design method*. John Willey & Sons, New York, 2nd édition.
- DILLMAN, D.-A. et BOWKER, D.-K. (2001). The web questionnaire challenge to survey methodologists. *Dans* ULF-DIETRICH, R. et BOSNJAK, M., éditeurs : *Dimensions of Internet Science*. Pabst Science Publishers, Lengerich, Germany.
- DILLMAN, D.-A. et CHRISTIAN, L.-M. (2005). Survey mode as a source of instability in responses across surveys. *Field Methods*, Vol. 17 (No. 1), 30–52.
- DILLMAN, D.-A., CLARK, J. R. et WEST, K. K. (1995). Influence of an invitation to answer by telephone on response to census questionnaires. *Public Opinion Quarterly*, Vol. 57, pp : 557–568.
- DILLMAN, D.-A., PHELPS, G., TORTORA, R.-D., SWIFT, K., KOHRELL, J. et BERCK, J. (2001). Response rate and measurement differences in mixed mode surveys using mail, telephone, interactive voice response and the internet. 56th Annual Conference of the American Association for Public Opinion Research (AAPOR).

Bibliographie

- DILLMAN, D.-A., REDLINE, C. et CARLEY-BAXTER, L. (1999). Influence of type of question on skip pattern compliance in self-administered questionnaires. Joint Statistical Meetings of the American Statistical Association, Indianapolis.
- DILLMAN, D.-A., SINCLAIR, M. D. et CLARK, J. R. (1993). Effects of questionnaire length, respondent-friendly design and a difficult question on response rates for occupant-addressed census mail surveys. *Public Opinion Quarterly*, Vol. 57 (No. 3), pp : 289–304.
- DILLMAN, D.-A. et TARNAI, J. (1988). Administrative issues in mixed mode surveys. Dans GROVES, R. M., BIEMER, P., LYBERG, L., MASSEY, J. T., NICHOLLS II, W. L. et WAKSBERG, J., éditeurs : *Telephone Survey Methodology*, pages 509–528. John Wiley and Sons, New York.
- DILLMAN, D.-A., TORTORA, R.-D. et BOWKER, D.-K. (1998a). Principles for constructing web surveys. SESRC Technical Report n° 98-50, Pullman, Washington.
- DILLMAN, D.-A., TORTORA, R.-D., CONRADT, J. et BOWKER, D.-K. (1998b). Influence of plain vs. fancy design on response rates for web surveys. Joint Statistical Meetings of the American Statistical Association, Dallas, Texas, August.
- DUMONTIER, F. et PAN KE SHON, J.-L. (1999). En 13 ans, moins de temps contraint et plus de loisirs. *Insee Première*, (No. 675).
- DUPUY, G. (1995). *Les territoires de l'automobile*. Anthropos. Economica, Paris.
- ETTEMA, D., SCHWANEN, T. et TIMMERMANS, H. (2007). The effect of locational factors on task and time allocation in households. *Transportation*, Vol. 34 (No. 1), pp : 89–105.
- ETTEMA, D., TIMMERMANS, H. et VAN VEGHEL, L. (1996). Effects of data collection methods in travel and activity research. Prepared for European Institute of Retailing and Services Studies, TRB Committee on Transportation Survey Methods.
- FALCONER, D. J. et HODGETT, R. A. (1999). Why executives don't respond to your survey. 10th Australasian Conference on Information Systems, Wellington, New Zealand, December.
- FLAMM, M. (2003). *Comprendre le choix modal : les déterminants des pratiques modales et des représentations individuelles des moyens de transport*. Thèse de doctorat, Faculté environnement naturel, architectural et construit, Ecole Polytechnique Fédérale de Lausanne.

- FLEMMING, G. et SONNER, M. (1999). Can internet polling work ? strategies for conducting public opinion surveys online. Annual Meeting of the American Association for Public Opinion Research, St Petersburg Beach, Florida, May.
- FOWLER, F. J. (1993). *Survey Research Methods*. Sage Publications, Newbury Park, California, 2nd édition.
- FRANCESCHINI, L. A. (2000). Navigating electronic survey methods : Three pilot studies. Annual Meeting of the Mid-South Educational Research Association, Bowling Green, USA.
- FRARY, R.-B. (1996). Hints for designing effective questionnaires. *Practical Assessment, Research & Evaluation*, Vol. 5 (No. 3). <http://pareonline.net/getvn.asp?v=5&n=3> (visited on 01/19/06).
- GADDIS, S.-E. (1998). How to design online surveys ? *Training & Development*, Vol. 52 (No. 6), pp : 67–71.
- GAY, L. R. et DIEHL, P. L. (1992). *Research Method for Business and Management*. McMillan, New York.
- GOLDBERGER, A. S. (1981). Linear regression after selection. *Journal of Econometrics*, Vol. 15 (No. 3), pp : 357–66.
- GOYDER, J. C. (1985). Face to face interviews and mail questionnaire : the net difference in response rate. *Public Opinion Quarterly*, Vol. 49 (No. 2), pp : 234–252.
- GOYDER, J. C. (1987). *The Silent Minority : Non respondents on Sample Surveys*. Polity Press, Cambridge, England.
- GREAVES, S. (1998). Applications of GIS technology in recent travel survey methpdologies. Travel Model Improvement Program, Federal Highway Administration.
- GREAVES, S. (2003). GIS and the collection of travel survey data. Dans HENSHER, D. A. et STOPHER, P. R., éditeurs : *Handbook of Transport and Spatial Information Systems*, pages 375–390.
- GREAVES, S. (2006). Simulating household travel survey data. Resource Paper for the 7th International Conference on Survey Methods in Transport, Costa Rica, August.
- GREENE, W.-H. (2002). *Econometric Analysis*. Prentice Hall International, US, 5th édition.
- GROGGER, J. T. et CARSON, R. T. (1991). Model for truncated counts. *Journal of applied econometrics*, Vol. 6, pp : 225–238.

Bibliographie

- GROVES, R. et KAHN, R. (1979). *Surveys by telephone : a national comparison with personal interviews*. Academic Press, New York.
- GROVES, R. et LEPKOWSKI, J. (1985). A mean squared error model for dual frame, mixed mode survey design. *Journal of the American Statistical Association*, Vol. 81 (No. 396), pp : 930–937.
- GROVES, R. M. (1989). *Survey Errors and Survey Costs*. Interscience. John Wiley & Sons, New York.
- GROVES, R. M., CIALDINI, R. B. et & COUPER, M. P. (1992). Understanding the decision to participate in a survey. *Public Opinion Quarterly*, Vol. 56, pp : 475–495.
- GROVES, R. M. et COUPER, M.-P. (1998). *Non-response in household interview surveys*. Interscience. John Wiley & Sons, New York.
- GUIITON, H. (1964). *Statistiques et économétrie*. Dalloz, Paris, 3ème édition.
- GUNN, H. (2002). Web-based surveys : Changing the survey process. *First Monday*, Vol. 7 (No. 12).
- HAN, X. L. et POLAK, J. W. (2003). Imputation with non-ignorable missing values : A stochastic approach. *Dans* AXHAUSEN, K. W., MADRE, J.-L., POLAK, J. W. et TOINT, P., éditeurs : *Capturing long distance travel*, pages 172–187. Research Science Press, Baldock.
- HANSON, S. (1982). The determinants of daily travel-activity patterns : relative location and socio-demographic factors. *Urban Geography*, Vol. 3, pp : 179–202.
- HARALDSEN, G., DALE, T., DALHEIM, E. et STROMME, H. (2002). Mode effects in a mail plus internet designed census. International Conference on Improving Surveys, Copenhagen, Denmark, August.
- HARPUDER, B. E. et STEC, J. A. (1999). Achieving an optimum number of callback attempts : Cost-savings vs. non-response error due to non-contacts in rdd surveys. Proceedings of the American Statistical Association, Session on Survey Research Methods, 1999.
- HARVEY, A. S. (2003). Time-space diaries : Merging tradition. *Dans* STOPHER, P. R. et JONES, P. M., éditeurs : *Transport Survey Quality and Innovation*, pages 152–180. Elsevier Press, Pergamon, Oxford.
- HECKMAN, J. (1979). Sample selection bias as a specification error. *Econometrica*, Vol. 47 (No. 1), pp : 153–161.
- HECKMAN, J. et NAVARRO-LOZANO, S. (2004). Using matching, instrumental variables, and control functions to estimate economic choice models. *The Review of Economics and Statistics*, Vol. 86, pp : 30–57.

- HECKMAN, J. et ROBB, R. (1985). Alternative methods for evaluating the impact of interventions : An overview. *Journal of Econometrics*, Vol. 30 (No. 1-2), pp : 239–267.
- HILL, R. (1998). What sample size is 'enough' in internet survey research ? *Interpersonal Computing and Technology : An electronic Journal for the 21st Century*, Vol. 6 (No. 3-4), 10 pages.
- HINE, J. (2004). Transport disadvantage and social exclusion in urban scotland. *Built environment*, Vol. 30 (No. 2), pp : 161–171.
- HIVERT, L. (2000). *Inégalité d'accès à l'automobile*. ADEME-INRETS, Paris.
- HJORTHOL, R. J. (2002). The relation between daily travel and use of home computer. *Transportation Research Part A*, Vol. 36 (No. 5), pp : 437–452.
- HOFFMAN, D. et LINK, C.-R. (1984). Selectivity bias in male wage equation : black-white comparison. *The Review of Economics and Statistics*, Vol. 66 (No. 2), pp : 320–324.
- HOJMAN, P., de DIOS ORTÚZAR, J. et RIZZI, L. (2004). Internet-based surveys to elicit the value of risk reductions. International Steering Committee for Travel Survey Conference, 12 pages.
- HUANG, H. M. (2006). Do print and web surveys provide the same results ? *Computers in Human Behavior*, Vol. 22 (No. 3), pp : 334–350.
- HUANG, H.-M. et LIAW, S.-S. (2005). Exploring users' attitudes and intentions toward the web as a survey tool. *Computers in Human Behavior*, Vol. 21 (No. 5), pp : 729–743.
- HUBERT, J.-P. (2003). Travel and time use surveys, a comparison of two belgian surveys, 25ème conférence IATUR, bruxelles, 17-19 septembre.
- HUBERT, J.-P. et TOINT, P. (2003). *La mobilité quotidienne des Belges*. Presses Universitaires de Namur, Namur, Belgique.
- HUGO, V. (1901). *Post-Scriptum de ma vie*. Calmann Lévy, Paris.
- IGBARIA, M. et PARASURMAN, S. (1989). A path analytic study of individual characteristics, computer anxiety and attitudes toward microcomputers. *Journal of Management*, Vol. 3 (No. 15), pp : 373–88.
- JEAVONS, A. (1999). Ethology and the web. observing respondent behavior in web surveys. *Marketing and Research Today*, Vol. 29 (No. 2), pp : 69–76.
- JOINSON, A. (1998). Causes and implications of disinhibited behavior on the internet. Dans GACKENBACH, J., éditeur : *Psychology and the Internet : Intrapersonal, Interpersonal, and Transpersonal Implications*. Academic Press, San Diego.

Bibliographie

- JONES, P.-M., DIX, M.-C., CLARKE, M.-I. et HEGGIE, I.-G. (1980). Understanding travel behaviour. Oxford studies of Transport. 241 pages.
- KALFS, N. et VAN EVERT, H. (2003). Nonresponse and travel surveys. Dans STOPHER, P. R. et JONES, P. M., éditeurs : *Transport Surveys, Quality and Innovation*, pages 567–585. Elsevier Press, Pergamon, Oxford.
- KAUFMANN, V. (1999). Mobilité et vie quotidienne : synthèse et questions de recherche. *2001 plus*, (No. 48).
- KAUFMANN, V. (2008). De l'espace au temps : ces mobilités hybrides qui transforment la ville. Dans CHALAS, Y. et PAULHIAC, F., éditeurs : *La mobilité qui fait la ville : Actes des 3e rencontres internationales en Urbanisme de l'Institut d'Urbanisme de Grenoble*, Série Débats 66. Certu.
- KAUFMANN, V., MONTULET, B. et LE BRETON, E. (2005). Mobilité et mobilité familiale. *Netcom, Networks and Communication Studies*, Vol. 19 (No. 3), pp : 137–139.
- KEESLING, R. (2002). Use of web-based data collection to supplement and improve mail survey response rates. American Association for Public Opinion Research Conference, St. Pete Beach, Florida. <http://www.amstat.org/sections/SRMS/proceedings/> (visited on 10/18/2005).
- KEETER, S., KOHUT, A., MILLER, A., GROVES, R. et PRESSER, S. (2000). Consequences of reducing nonresponse in a large national telephone survey. *Public Opinion Quarterly*, Vol. 64, pp : 125–148.
- KING, G. (1989). Event count mode for international relations : generalizations and applications. *International Studies Quarterly*, Vol. 33, pp : 123–147.
- KITTLESON, M. J. (1995). An assessment of the response rate via the postal service and e-mail. *Health Values*, Vol. 19 (No. 2), pp : 27–39.
- KMENTA, J. (1971). *Elements of econometrics*. McMillan, New York.
- KREITZ, M. (2001). Methods for collecting spatial data in household travel surveys. International Conference on Transport Survey Quality and Innovation, Kruger National Park, South Africa, August.
- KROSINICK, J.-A. et ALWIN, D. F. (1987). An evaluation of a cognitive theory of response order effects in survey measurement. *Public Opinion Quarterly*, Vol. 51 (No. 2), pp : 201–219.
- KROSINICK, J.-A. et ALWIN, D.-F. (1991). The reliability of survey attitude measurement : The influence of question and respondent attributes. *Sociological Methods and Research*, Vol. 20 (No. 1), pp : 139–181.

- KROSNICK, J.-A. et CHANG, L. C. (2001). A comparison of the random digit dialing telephone survey methodology with internet survey methodology as implemented by knowledge networks and harris interactive, Ohio state university.
- KROSNICK, J. A. et FABRIGAR, L. R. (1997). Designing rating scales for effective measurement in surveys. *Dans* LYBERG, L., BIEMER, P., COLLINS, M., de LEEUW, E. D., DIPPO, C., SCWARTZ, N. et TREWIN, D., éditeurs : *Survey measurement and process quality*, pages 141–165. Wiley & Sons, New York.
- KWAK, N. et RADLER, B. (2000). Comparison between mail and web surveys : response pattern, respondent profile, data quality and construct association. Annual meeting of the American Association for Public Opinion Research, Portland, Oregon.
- LANG, M. (2002). The use of web-based international surveys in information systems research. pages 187–196. European Conference on Research Methodology for Business and Management Studies, Reading, England, April.
- LAZAR, J. et PREECE, J. (1999). Designing and implementing web-based survey. *The Journal of Computer Information Systems*, Vol. 39 (No. 4), pp : 63–67.
- LEE, L.-F. (1982). Some approaches to the correction of selectivity bias. *The Review of Economic Studies*, Vol. 49 (No. 3), pp : 355–372.
- LLIEVA, J., B. S. et HEALEY, M. N. (2002). Online surveys in marketing research : Pros and cons. *International Journal of Marketing Research*, Vol. 44 (No. 3), pp : 361–378.
- LOZAR MANFREDA, K. et VEHOVAR, V. (2002a). Do mail and web surveys provide same results ? *Dans* FERLIGOJ, A. et MRVAR, A., éditeurs : *Development in social science methodology*, pages 149–169. Ljubljana.
- LOZAR MANFREDA, K. et VEHOVAR, V. (2002b). Mode effect in web surveys. Strengthening Our Community, Proceedings of the American Association for Public Research, Session on Survey Research Methods.
- LOZAR MANFREDA, K., VEOVAR, V. et BATAGELJ, Z. (2001). Web versus mail questionnaire for an institutional survey. *Dans* WESTLAKE, A., SYKES, W., MANNERS, T. et RIGG, M., éditeurs : *The Challenge of the Internet*. Association for Survey Computing.
- MADDALA, G. (1986). *Limited dependant and qualitative variables in econometrics*. Cambridge University Press, London.
- MADRE, J.-L., AXHAUSEN, K. et GASCON, M.-O. (2003). Immobility : a microdata analysis. 10th IATBR conference, Lucerne, August.

Bibliographie

- MADRE, J.-L., AXHAUSEN, K. K. et BRÖG, W. (2007). Immobility in travel diary surveys. *Transportation*, Vol. 34 (No. 1), pp : 107–128.
- MALHOTRA, N. K. (1999). *Marketing Research : An applied Orientation*. Prentice Hall, London, 3rd édition.
- MASSOT, M.-H. et ORFEUIL, J.-P. (2008). Mobilité résultante et mobilité organisatrice : les paradigmes au service de la compréhension des transformations urbaines. Dans CHALAS, Y. et PAULHIAC, F., éditeurs : *La mobilité qui fait la ville : Actes des 3e rencontres internationales en Urbanisme de l'Institut d'Urbanisme de Grenoble*, Débats 66. Certu.
- MC DANIEL, C. et GATES, R. (2004). *Marketing Research*. Hoboken, Wiley, NJ.
- MEDIAMETRIX (2000). US baby boomers and seniors are fastest growing internet demographic group.
- MEHTA, R. et SIVADAS, E. (1995). Comparing response rates and response content in mail vs. electronic mail surveys. *Journal of the Market Research Society*, Vol. 37 (No. 4), pp : 429–439.
- MERTLER, C. A. (2002). Web-based surveys : Guiding lessons for thir user. Annual Meeting of the American Educational Research Association, New Orleans, USA.
- MICHALOWSKA, A. (1999). Internet, un nouvel outil au service des études. *Marketing Magazine*, Vol. 40, pp : 34–39.
- MICHALOWSKA, A. (2000). Internet et le métier des études : évolution ou révolution? *Marketing Magazine*, Vol. 54, pp : 58–61.
- MOHAMMADIAN, A. et BEKHOR, S. (2008). Travel behavior of special population groups. *Transportation*, Vol. 35 (No. 5), pp : 579–583.
- MOKHTARIAN, P., SALOMON, I. et HANDY, S. (2006). The impacts of ict on leisure activities and travel : A conceptual exploration. *Transportation*, Vol. 33 (No. 3), pp : 263–289.
- MOKHTARIAN, P. L. et CHEN, C. (2004). Ttb or not ttb, that is the question : a review and analysis of the empirical literature on travel time (and money) budgets. *Transportation Research Part A*, Vol. 38, pp : 643–675.
- MORITZ, G. et BRÖG, W. (1999). Redesign of the dutch travel surveys : response improvement. TRB conference on personal travel : the long and short of it, Washington, June-July.
- MORREL-SAMUELS, P. (2003). Web surveys' hidden hazards. *Harvard Business Review*, Vol. 81 (No. 7), pp : 16–17.

- MORRIS, J. et ADLER, T.-J. (2003). Mixed mode survey. Dans STOPHER, P. R. et M, J. P., éditeurs : *Transport Surveys, Quality and Innovation*, pages 239–252. Elsevier Press, Pergamon, Oxford.
- MROZ, T.-A. (1987). The sensitivity of an empirical model of married women's hours of work to economic and statistical assumptions. *Econometrica*, Vol. 55 (No. 4), pp : 765–799.
- MULLAHY, J. (1986). Specification and testing of some modified count data models. *Journal of Econometrics*, Vol. 33, pp : 341–365.
- MURAKAMI, E. (2004). Survey methods. Proceedings of the National Household Travel Survey Conference, November, pp : 23-26. <http://trb.org/publications/circulars/ec071.pdf> (visited on 10/27/05).
- MURAKAMI, E. et MORRIS, J. (2003). Using technology to improve transport survey quality. Dans STOPHER, P. R. et JONES, P. M., éditeurs : *Transport Surveys, Quality and Innovation*, pages 499–506. Elsevier Press, Pergamon, Oxford.
- MYLES, R. et TIBERT, T. (1998). Internet surveys : do they work? *Institute for social research, Newsletter*, Vol. 13 (No. 1).
- NAKOSTEEN, R.-A. et ZIMMER, M.-A. (1980). Migration and income : the question of self-selection. *Southern Economic Journal*, Vol. 46 (No. 3), pp : 840–851.
- NCHRP (2006). Standardization of personal travel surveys. Report to the National Cooperative Highway Research Program on Project 08-37, Transportation Research Board, Washington, USA.
- NICHOLS, E. et SEDIVI, B. (1998). Economic data collection via the web : A census bureau case study. Annual Meetings of the American Statistical Association.
- NICOLAS, J. P., POCHET, P. et POINTBOEUF, H. (2002). Mobilité urbaine et développement durable : quels outils de mesure pour quels enjeux? *Les Cahiers Scientifiques du Transport*, (No. 41), pp : 53–76.
- NIELSEN, A.-C. (2005). Are online survey as accurate as offline surveys? American Marketing Association. <http://www.marketingpower.com/content25232.php> (visited on 10/18/2005).
- NOWACK, M. (1997). The impact of the internet on statistical organisations. *Statistical Journal of the UN Economic Commission for Europe*, Vol. 14 (No. 4), pp : 345–355.
- O'NEIL, K. M. et PENROD, S. D. (2001). Methodological variables in web-based research that may affect results : Sample type, monetary incentives,

Bibliographie

- and personal information. *Behavior Research Methods, Instruments, & Computers*, Vol. 33 (No. 2), pp : 226–233.
- OOSTERVELD, P. et WILLEMS, P. (2003). Two modalities, one answer? *Dans* FELLOWS, D. S., éditeur : *Combining Internet and CATI Surveys Effectively in Market Research*, pages 141–150. Proceedings of the ESOMAR Technovate Conference, Amsterdam.
- OPPENHEIM, A. N. (1992). *Questionnaire Design, Interviewing and Attitude Measurement*. Printer Publishers, New York.
- ORFEUIL, J.-P. (2000). L'évolution de la mobilité quotidienne : Comprendre les dynamiques, éclairer les controverses. *Synthèse INRETS*, (No. 37), 146 pages.
- ORFEUIL, J.-P. (2002a). Etat des lieux des recherches sur la mobilité quotidienne en France. *Dans* LÉVY, J. P. et DUREAU, F., éditeurs : *L'accès à la ville*, page 25. L'Harmattan, Paris.
- ORFEUIL, J.-P. (2002b). La mobilité dans le monde. Document de travail, CRETEIL. 11 pages.
- ORFEUIL, J.-P. et MASSOT, M.-H. (2003). La mobilité individuelle dans 20 ans. Séminaire au Sénat, Janvier, 9 pages.
- ORFEUIL, J.-P. et MASSOT, M.-H. (2003). Penser les mobilités de demain : Essai de clairvoyance prospective. *Revue du CERAP*, (No. 22), pp : 269–290.
- ORFEUIL, J.-P., MASSOT, M.-H. et BELLANGER, F. (2000). Eléments pour une prospective de la mobilité, Centre de recherche « espace, Transports, Environnement et Institutions », Université de Paris XII, Créteil.
- PENDYALA, R. (2003). Quality and innovation in time use and activity surveys. *Dans* R, S. P. et M, J. P., éditeurs : *Transport Survey Quality and Innovation*, pages 181–190. Elsevier Science Ltd, Oxford, UK.
- PITKOW, J. E. et RECKER, M. M. (1995). Using the web as a survey tool : results from the second www user survey. *Journal of Computer Networks and ISDN systems*, Vol. 27 (No. 6), pp : 809–822.
- POCHET, P. (2003). Mobilité et accès à la voiture chez les personnes âgées : évolutions actuelles et enjeux. *Recherche Transport Sécurité*, (No. 79), pp : 93–106.
- POCHET, P. (2005). La mobilité des personnes : principaux déterminants et grandes tendances. Journée de la population européenne, Tours, France, Juillet, 8 pages.
- POTOGLOU, D. et KARANOGLU, P. S. (2008). Comparison of phone and web based surveys for collecting household background information. International Conference on Survey Methods in Transport, Annecy, France, Mai, 23 pages.

- POUYANNE, G. (2004). *Forme urbaine et mobilité quotidienne*. Thèse de doctorat, Institut d'Economie Régionale du Sud-Ouest, Université Montesquieu-Bordeaux IV.
- PRATT, J.-H. (2003). Survey instrument design. Dans STOPHER, P. R. et JONES, P. M., éditeurs : *Transport Surveys, Quality and Innovation*, pages 137–150. Elsevier Press, Pergamon, Oxford.
- PRIMERANO, F., TAYLOR, M. A. P., PITAKSRINGKARN, L. et TISATO, P. (2008). Defining and understanding trip chaining behaviour. *Transportation*, Vol. 35, pp : 55–72.
- RESSOURCE SYSTEM GROUP, Inc. (2002). Documentation for sbir phase II final report : Computer-based intelligent travel survey system. DTRS57-00-C-10030, FHWA.
- RICHARDSON, A.-J. (2003). Behavioural mechanisms of non-response in mail-back travel surveys. *Transportation research record*, Vol. 1855, pp : 18.
- RICHARDSON, A.-J. et AMPT, E.-M. (1994). *Vital : the victorian integrated travel, activities and land-use toolkit*. Australian Road Research Board Ltd, Vermont South, Australia.
- RICHARDSON, A.-J., AMPT, E.-S. et MEYBURG, A. H. (1995). *Survey Methods for Transport Planning*. Eucalyptus Press, University of Melbourne, Australia.
- RIETVELD, P. (2002). Rounding of arrival and departure times in travel surveys : an interpretation in terms of scheduled activities. *Journal of Transportation and Statistics*, Vol. 5 (No. 1), pp : 71–82.
- ROBINSON, J. P., SHAVER, P. R. et WRIGHTSMAN, L. S. (1991). *Measures of Personality and Social Psychological Attitudes*. Academic Press, San Diego.
- ROMANO, M.-F. et HIMMELMANN, M. (2002). Determinants of web mode choice in a 'web & paper' survey in a high education population. International Conference on Improving Survey, Copenhagen, Denmark, August.
- ROOS, P. et MARCHAND, V. (2006). Le renouvellement des techniques de recueil de l'information sur la demande : l'essor des études en ligne. Centre de Ressources en Economie Gestion. 9 pages.
- ROSCOE, J. T. (1975). *Fundamental Research Statistics for the Behavioral Sciences*. Holt Rinehart & Winston, New York, 2nd édition.
- ROSE, J.-M., BLACK, I.-R., IOANNOU, C. et EFRON, A. (2004). Using microsoft excel as an alternative survey instrument. 7th International Conference on Travel Survey Methods, Costa Rica, August. Actes sur CD-ROM.
- ROY, A.-D. (1951). Some thoughts on the distribution of earnings. *Oxford Economic Papers*, Vol. 3 (No. 2), pp : 135–146.

Bibliographie

- RUBIN, D. (1974). Estimating causal effects of treatments in randomized and non-randomized studies. *Journal of Educational Psychology*, Vol. 66, pp : 688–701.
- SAMMER, G. (1997). Problems and solutions in urban travel survey. Dans BONNEL, P., CHAPLEAU, R., LEE-GOSSELIN, M. et RAUX, C., éditeurs : *Les enquêtes de déplacements urbains : mesurer le présent, simuler le futur*, pages 145–159. Collection Programme Rhône-Alpes Recherches en Sciences Humaines, Lyon.
- SAPORTA, G. (2006). *Probabilités, analyses des données et statistiques*. Editions Technip.
- SARIS, W. R. (1991). Computer assisted interviewing. *Quantitative applications in the social sciences*, (No. 80).
- SAUTORY, O. (1993). Redressement d'un échantillon par calage sur marges. Document de travail de la DSDS n°F9310, INSEE. 51 pages.
- SCHAEFER, D. et DILLMAN, D.-A. (1998). Development of a standard e-mail methodology : results of an experiment. *Public Opinion Quarterly*, Vol. 62 (No. 3), pp : 378–397.
- SCHILLEWAERT, N., LANGERAK, F. et DUHAMEL, T. (1998). Non-probability sampling for www surveys : a comparison of methods. *Journal of the Marketing Research Society*, Vol. 40 (No. 4), pp : 307–313.
- SCHILLEWAERT, N. et MEULEMEESTER, P. (2005). Comparing response distributions of offline and online data collection methods. *International Journal of Market Research*, Vol. 47 (No. 2), pp : 163–178.
- SCHLEYER, T. K. L. et FORREST, J. L. (2000). Methods for the design and administration of web-based surveys. *Journal of the American Medical Informatics Association*, Vol. 7 (No. 4), pp : 416–425.
- SCHMIDT, W. C. (1997). World wide web survey research : benefits, potential problems, and solutions. *Behavior Research Methods, Instruments and Computers*, Vol. 29 (No. 2), pp : 274–279.
- SCHONLAU, M., FRICKER, R.-D. et ELLIOTT, M.-N. (2001). Conducting research surveys via e-mail and the web. Rand Documents, 118 pages. <http://www.rand.org/publications/MR/MR1480/> (visited on 11/09/2005).
- SCHULDT, B. et TOTTEN, J. (1994). Electronic mail vs. mail survey response rates. *Marketing Research*, Vol. 6 (No. 1), pp : 36–39.
- SCHUMAN, H. et PRESSER, S. (1981). *Questions and answers in attitude surveys : experiments on question form, wording and context*. Academic Press, New York.

- SCHUMAN, H. et SCOTT, J. (1989). Response effect over time : two experiments. *Sociological Methods Research*, Vol. 17 (No. 4), pp : 398–408.
- SHARP, J., MC GUCKIN, N. et BOSE, J. (2004). Travel survey methods. Proceedings of the National Household Travel Survey Conference, November, pp : 49-55. <http://trb.org/publications/circulars/ec071.pdf> (visited on 10/27/05).
- SHASHAANI, L. (1994). Socioeconomic status, parents' sex-role stereotypes, and the gender gap in computing. *Journal of Research on Computing in Education*, Vol. 26 (No. 4), pp : 433–451.
- SHEARMUR, R. (2008). Les migrations pendulaires à montréal : pour un regard géoéconomique. Dans CHALAS, Y. et PAULHIAC, F., éditeurs : *La mobilité qui fait la ville : Actes des 3e rencontres internationales en Urbanisme de l'Institut d'Urbanisme de Grenoble*, Série Débats 66. Certu.
- SHETTLE, C. et MOONEY, G. (1999). Monetary incentives in u.s. government surveys. *Journal of Official Statistics*, Vol. 15, pp : 231–250.
- SILLS, S.-J. et SONG, C. (2002). Innovations in survey research : An application of web-based surveys. *Social Science Computer Review*, Vol. 20 (No. 1), pp : 22–30.
- SMITH, C.-B. (1997). Casting the net : Surveying an internet population. *Journal of Computer-Mediated Communication*, Vol. 3 (No. 1). <http://jcmc.indiana.edu/vol3/issue1/smith.html> (visited on 11/09/05).
- SOLOMON, D.-J. (2001). Conducting web-based surveys. *Practical Assessment, Research & Evaluation*, Vol. 7 (No. 19). <http://pareonline.net/getvn.asp?v=7&n=19> (visited on 10/27/05).
- SRINIVASAN, K. K. et ATHURU, S. R. (2004). Modeling interaction between internet communication and travel activities : Evidence from bay area California, travel survey 2000. *Transportation research record*, (No. 1894), pp : 230–240.
- STANTON, J. et ROGELBERG, S. G. (2001). Using internet/intranet web pages to collect organizational research data. *Organizational Research Methods*, Vol. 4 (No. 3), pp : 199–216.
- STANTON, J. M. (1998). An empirical assessment of data collection using the internet. *Personnel Psychology*, Vol. 51 (No. 3), pp : 709–726.
- STOLZENBERG, R.-M. et RELLES, D.-A. (1997). Tools for intuition about sample selection bias and its correction. *American Sociological Review*, Vol. 62 (No. 3), pp : 494–507.

Bibliographie

- STOPHER, P.-R. (1998). Household travel surveys : new perspectives and old problems. *Dans* GÄRLING, T., LAITILA, T. et WESTIN, K., éditeurs : *Theoretical Foundations of Travel Choice Modeling*, chapitre 17, pages 399–419. Pergamon Press, Elsevier.
- STOPHER, P.-R., COLLINS, A. et BULLOCK, P. (2004a). GPS surveys and the internet. Working Paper ITS-WP-04-18, Institute of Transport Studies, University of Sydney, Australia.
- STOPHER, P.-R., FITZGERALD, C. et XU, M. (2007). Assessing the accuracy of the sydney household travel survey with gps. *Transportation*, Vol. 34 (No. 6), pp : 723–741.
- STOPHER, P.-R. et GREAVES, S. P. (2006). Household travel surveys : Where are we going ? *Transportation Research Part A : Policy and Practice*, Vol. 41 (No. 5), pp : 367–381.
- STOPHER, P.-R. et JONES, P.-M. (2003). Summary and future directions. *Dans* STOPHER, P. R. et JONES, P. M., éditeurs : *Transport survey quality and innovation*, pages 635–646. Elsevier Press, Pergamon, Oxford.
- STOPHER, P.-R. et METCALF, H. M. A. (1996). Methods for household travel surveys. NCHRP Synthesis 236, Transportation Research Board, Washington, USA.
- STOPHER, P.-R. et STECHER, C. (2005). *Travel Survey Methods : Quality and Future Directions*. Emerald Group Publishing, Elsevier.
- STOPHER, P.-R., WILMOT, C.-G., STECHER, C. et ALSNIH, R. (2004b). Household travel surveys : proposed standards and guidelines. 7th International Conference on Travel Survey Methods, Costa Rica, August. Actes sur CD-ROM.
- SUDMAN, S. et BRADBURN, N. (1996). *Thinking about Answers : The Application of Cognitive Processes to Survey Methodology*. Jossey-Bass, San Francisco.
- SWOBODA, S. J., MUEHLBERGER, N., WEITKUNAT, R. et SCHNEEWEISS, S. (1997). Internet surveys by direct mailing : an innovative way of collecting data. *Social Science Computer Review*, Vol. 15 (No. 3).
- SYTRAL (2007). Enquête ménages déplacements 2006 de l'aire métropolitaine lyonnaise, principaux résultats. Sytral.
- THOMAS, A. (2000). *Econométrie des variables qualitatives*. Economica.
- TOBIN, J. (1958). Estimation of relationship for limited dependent variables. *Econometrica*, Vol. 26 (No. 1), pp : 24–36.

- TOINT, P., CIRILLO, C., CORNELIS, E. et HUBERT, J.-P. (2004). Les enquêtes sur les comportements de mobilité, et après? *Reflets et perspectives de la vie économique*, Vol. XLIII (No. 4), pp : 111–121.
- TOOLEY, M. (1996). Incentives and rates of return for travel surveys. *Transportation research record*, Vol. 1551, pp : 67–73.
- TOURANGEAU, R. et SMITH, T. W. (1996). Asking sensitive questions : The impact of data collection, question format, and question context. *Public Opinion Quarterly*, Vol. 60 (No. 2), pp : 275–304.
- TRUPELL, A. (2003). Use of internet tools for survey research. *Learning, and Performance Journal*, Vol. 21 (No. 1), pp : 31–37.
- TSE, A. (1998). Comparing the response rate, response speed and response quality of two methods of sending questionnaires : e-mail vs. mail. *Journal of the Market Research Society*, Vol. 40 (No. 4), pp : 353–361.
- VAN EVERT, H. et MORITZ, G. (2000). The new dutch travel survey. 9th International association for travel behaviour conference, Gold Coast, Queensland, Australia, July, 16 pages.
- VEHOVAR, V. et LOZAR MANFREDA, K. (2002). Mode effect in web surveys. American Association for Public Opinion Research Annual Conference, St. Petesburg, Florida, May.
- VERBEEK, M. (2004). *A guide to modern econometrics*. John Wiley & Sons Ltd, England, 2nd édition.
- VISSER, E. J. et LANZERDORF, M. (2004). Mobility and accessibility effects of b2c e-commerce : a literature survey. *Journal of Economic & Social Geography*, (No. 95), pp : 189–205.
- WANG, D. et LAW, F. (2007). Impacts of information and communication technologies ICT on time use and travel behavior : a structural equations analysis. *Transportation*, Vol. 34 (No. 4), pp : 513–527.
- WATT, J. H. (1999). Internet systems for evaluation research. *Dans* GAY, G. et BENNINGTON, T. L., éditeurs : *Information technologies in evaluation : social, moral, epistemological and practical implications*, pages 23–44. Jossey-Bass, San Francisco.
- WEIBLE, R. et WALLACE, J. (1998). Cyber research : The impact of the internet on data collection. *Marketing Research*, Vol. 10 (No. 3), pp : 19–26.
- WERMUTH, M., SOMMER, C. et KREITZ, M. (2001). Impact of new technologies in travel surveys. International Conference on Transport Survey Quality and Innovation, Kruger National Park, South Africa, August.

Bibliographie

- WILLKE, J., ADAMS, C. O. et GIRNUS, Z. (1999). A landmark study of the difference between mall intercept and on-line interviewing in the US. ESO-MAR Worldwide Internet Conference Net Effects.
- WILSON, J. (2004). Measuring personal travel and goods movement. Transportation Research Board Special Report n° 277, National Research Council, Washington, USA.
- WINSHIP, C. et MARE, R.-D. (1992). Models for sample selection bias. *Annual Review of Sociology*, Vol. 18, pp : 327–350.
- WITT, K. J. (1998). Best practices in interviewing via the internet. Sawtooth Software Conference, Washington, pp : 15-37.
- WITTE, J. C., AMOROSO, L. M. et HOWARD, P. E. N. (2000). Research methodology-method and representation in internet-based survey tools-mobility, community, and cultural identify in survey 2000. *Social Science Computer Review*, Vol. 18 (No. 2), pp : 179–195.
- WOLF, J., LECHL, M., THOMPSON, M. et ARCE, C. (2003). Trip rate analysis in gps-enhanced personal travel surveys. Dans STOPHER, P. R. et JONES, P. M., éditeurs : *Transport survey quality and innovation*, pages 483–498. Elsevier Press, Pergamon, Oxford.
- WONNACOTT, T. H. et WONNACOTT, R. J. (2000). *Statistique : Economie-Gestion-Sciences-Médecine*. Economica, Paris, 4ème édition.
- WRIGHT, D. L., AQUILINO, W. S. et SUPPLE, A. J. (1998). A comparison of computer assisted and paper-and-pencil self-administered questionnaires in a drug use survey. *Public Opinion Quarterly*, Vol. 62 (No. 3), pp : 331–353.
- WYDRA, D. (1999). Online tracking : A new frontier. ARFs Online Research Day : Towards Validation, Advertising Research Foundation, New York.
- YU, J. et COOPER, H. (1983). A quantitative review of research design effects on response rates to questionnaires. *Journal of Marketing Research*, No. 20, pp : 36–44.
- YUN, G.-M. et TRUMBO, C.-W. (2000). Comparative response to a survey executed by post, e-mail & web form. *Journal of Computer-Mediated Communication*, Vol. 6 (No. 1). [http ://jcmc.indiana.edu/vol6/issue1/yun.html](http://jcmc.indiana.edu/vol6/issue1/yun.html) (site visité le 24/10/2005).
- ZAHAVI, Y. (1979). The umot project. U.S. Department of Transportation and Ministry of Transport of Federal Republic Of Germany, 267 pages.
- ZEILEIS, A., KLEIBER, C. et JACKMAN, S. (2008). Regression mode for count data in r. *Journal of statistical software*, Vol. 27 (No. 8), 25 p.

- ZHANG, Y. (1999). Using the internet for survey research : A case study. *Journal of the American Society for Information Science*, Vol. 51 (No. 1), pp : 57–68.
- ZMUD, J. (2003). Designing instruments to improve response : keeping the horse before the cart. Dans STOPHER, P. R. et JONES, P. M., éditeurs : *Transport Survey Quality and Innovation*, pages 89–108. Elsevier Press, Pergamon, Oxford.
- ZMUD, J. et ARCE, C. (2000). Item nonreponse in travel surveys : Causes and solutions. International Conference Transport Surveys : Raising the standard, Grainau, Germany, May 24-30.
- ZORN, C. J. W. (1996). Zero-inflating and hurdle poisson specifications. Midwest Political Science Association, April, pp : 18-20.
- ZUKERBERG, A., NICHOLS, E. et TEDESIO, H. (1999). Designing surveys for the next millennium : Internet questionnaire design issues. American Association for Public Opinion Research Annual Conference, St. Petersburg, Florida, May.

Annexes

I Evolution du nombre d'internautes en France

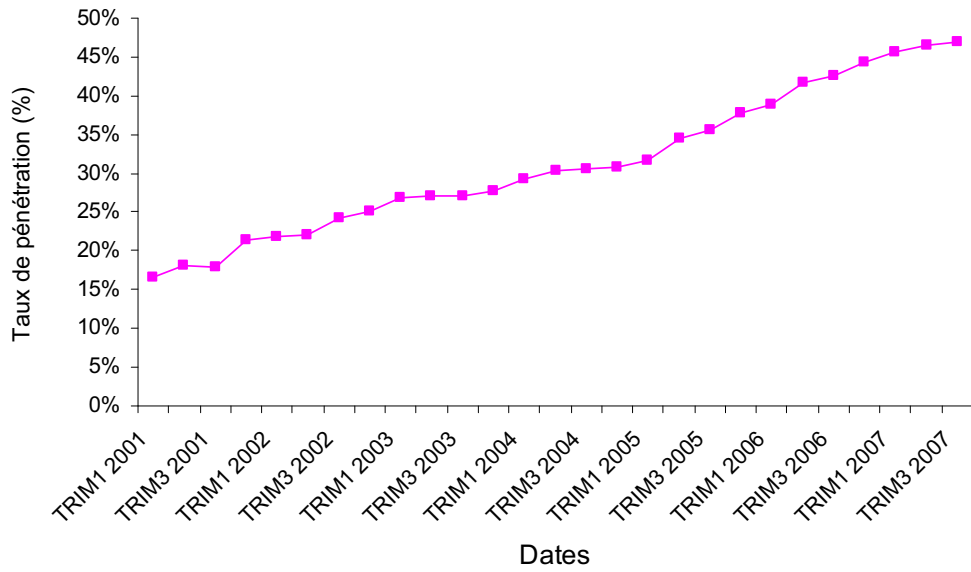
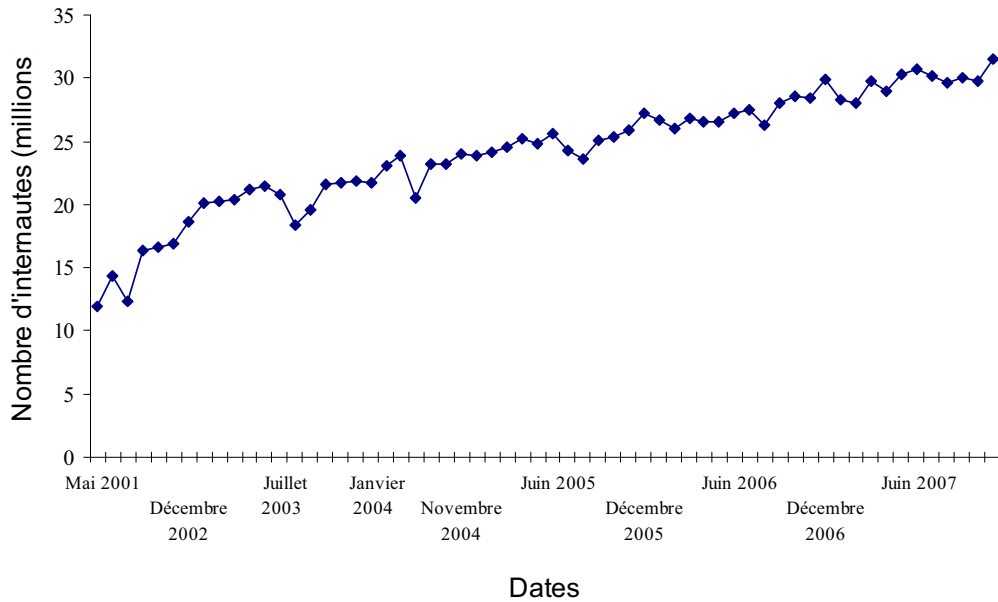


FIG. 1 – Evolution du taux de pénétration d'internet en France
Source : Médiamétrie (2007)



II Evolution des inégalités d'accès aux nouvelles technologies

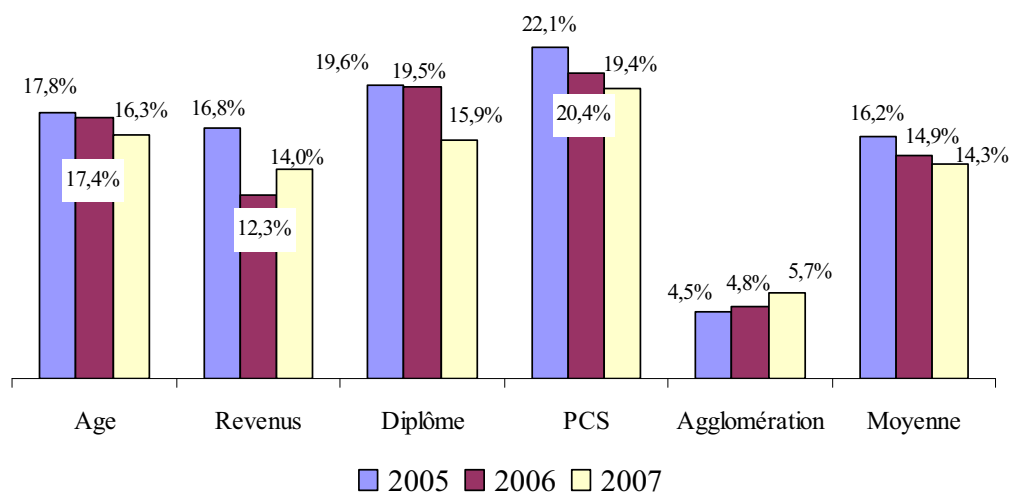


FIG. 1 – Inégalités d'équipement en ordinateur
Source : Crédoc (2007)

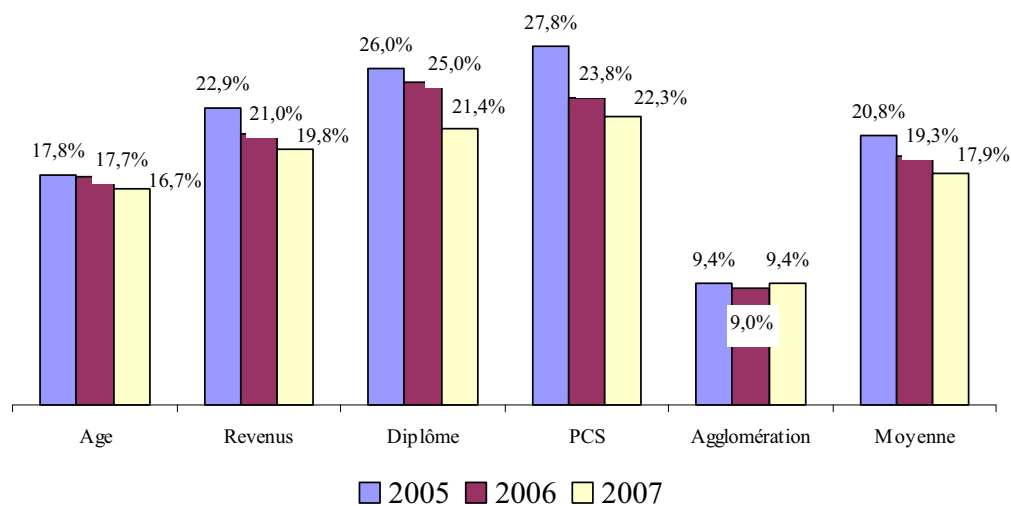


FIG. 2 – Inégalités d'équipement en internet à domicile
Source : Crédoc (2007)

III Guide d'entretien semi-directif

1) : Evaluation de la durée du questionnaire :

- minuter le questionnaire, pour avoir une estimation fiable du temps de réponse :
 - temps effectif et temps perçu par l'internaute, pour saisir l'ensemble des réponses. Le temps moyen est par hypothèse de 20mn.
 - impact de la barre de progression : cet outil a-t-il été utilisé par le répondant ? comment a-t-il perçu sa progression (positive ou négative) dans le questionnaire ?
- s'interroger sur la « lourdeur » du questionnaire et vérifier si certaines questions doivent être supprimées :
 - le questionnaire n'est-il pas trop long (redondance des questions) ?
 - par lassitude, certaines questions importantes concernant les déplacements ne sont-elles pas traitées trop rapidement ?
 - les questions posées sont-elles liées à l'étude des déplacements quotidiens (pertinence des questions...) ?

2) Evaluation du contenu du questionnaire :

- tester la compréhension des questions et la pertinence des réponses :
 - sommes-nous bien compris des personnes interrogées ?
 - quelles questions sont mal comprises, ambiguës ou obscures ?
 - les enquêtés peuvent-ils exprimer facilement ce que l'on attend d'eux, c'est-à-dire retracer le schéma des déplacements effectués la veille ?
 - dans la saisie des déplacements, les questions sont-elles suffisamment détaillées (ex : quand avez-vous quitté cet endroit ? – pas de précision de l'activité-)
- identifier l'utilisation du bon vocabulaire :
 - les termes et tournures de phrases employés sont-ils compréhensibles et adaptés à public non familier des études de transports ?
 - le concept de « déplacement » est-il correctement appréhendé par les internautes ?
- vérifier l'impact des instructions sur le bon déroulement du questionnaire :
 - les répondants ont-ils tendance à trop interpréter les questions, notamment celles concernant le type de déplacements à considérer ?
 - les instructions (aides) et des aides visuelles (graphiques) sont-elles pertinentes, visualisées et utilisées par le répondant ?
- identifier les causes possibles d'abandon :
 - repérer les points « difficiles » du questionnaire et identifier les problèmes rencontrés
 - éliminer les difficultés dans l'enregistrement des réponses
 - certaines questions n'incitent-elles pas trop à sortir du questionnaire ?
 - le répondant n'est-il pas frustré de devoir répondre obligatoirement à certaines questions (ex : utilisation de la VP) ?

3) Evaluation de la structure du questionnaire :

- vérifier que les tests textes d'accroche initiale, de conclusion et de liaison sont compréhensibles et suffisants :
 - les objectifs de l'étude sont-ils bien présentés (analyse des comportements de mobilité quotidienne) ? L'internaute dispose-t-il de suffisamment d'informations sur le sujet traité (périmètre de l'étude, destinataire des données...) ?
 - le répondant se sent-il incité à répondre (citoyenneté, importance de la participation de chacun...) et rassuré par rapport à la confidentialité des données et à la sécurité du serveur (virus) ?
 - les différentes parties sont-elles judicieusement coordonnées entre elles ? L'internaute est-il incité à progresser dans la saisie des réponses ?
- contrôler la cohérence du cheminement du questionnaire ?
 - la succession des questions est-elle logique et motivante, de façon à ne pas demander un niveau d'effort trop élevé aux enquêtés ?
 - existe-t-il des a priori dans la construction du questionnaire (filtres abusifs, répondants frustrés par l'impossibilité de certains choix...) ?
 - l'interactivité est-elle efficiente ? Est-il simple de naviguer dans le questionnaire ?
 - certaines questions doivent-elles être ajoutées dans le questionnaire ?
 - est-il nécessaire de prévoir la modification d'un déplacement à partir du tableau récapitulatif final ?
- valider la présentation du questionnaire :
 - attrait du graphisme et de la mise en page des questions
 - lisibilité des instructions et facilité de saisie des données
 - stabilité de la présentation selon les navigateurs utilisés

IV Verbatims

1) Evaluer la durée du questionnaire : temps effectif et temps perçu pour saisir l'ensemble des réponses, impact de la barre de progression...

- Un peu plus de 10 minutes pour remplir le questionnaire. Honnêtement, cela ne semble ni long, ni difficile à remplir.
- Durée du questionnaire adaptée : entre 10 et 15 minutes, avec une durée perçue de l'ordre de 10 minutes...
- Le questionnaire n'est pas long, si peu de trajets effectués dans la journée. Sinon, c'est un peu lassant : au bout d'un moment, le « cobaye » risque d'abandonner.
- Questionnaire globalement un peu long. Idéalement, il aurait été préférable de pouvoir saisir l'ensemble des éléments relatifs à un déplacement sur une même page et non pas sur plusieurs.

- Bien le coup de la barre de progression, car elle avance vite.
- La barre de progression c'est pratique mais c'est un peu déprimant quand au bout de 10 minutes ça n'avance pas vite.
- La présence de la barre est attrayante au début, puis nous ne l'utilisons plus.

2) S'interroger sur la « lourdeur » du questionnaire : longueur, effet de lassitude, pertinence des questions...

- Pas de lassitude, mais il faut juste réfléchir un peu à ce qu'on a fait la veille
- Le questionnaire n'est pas long, à condition bien sûr de ne pas avoir une vie trépidante avec plein de déplacements en journée.
- Au début, ça va, car les questions changent souvent, mais quand faut faire le détail de tous les déplacements, c'est un peu redondant.

3) Evaluation du contenu du questionnaire : compréhension des questions et pertinence des réponses, choix des termes et tournures de phrases, interprétation du concept de déplacement...

- Pas de problème majeur de compréhension.
- Bonne compréhension du questionnaire.
- Doute à la question sur la figuration dans liste rouge/orange/annuaire France Télécom, car je me suis dit "les annuaires c'est toujours France Télécom ou c'est les pages jaunes comme marque aujourd'hui ?"
- Doutes sur le type de déplacements à inclure (petits déplacements ?).
- Les questions sont claires, mais demandent des réponses super détaillées des fois, ce qui est un peu fastidieux.
- Questionnaire détaillé, mais rébarbatif lorsque nous avons plusieurs déplacements à saisir.
- Il est difficile de se souvenir des adresses exactes des lieux, et nous avons tendance à ne pas renseigner tous les champs.

- Donner des indications sur le type de précision souhaitée (ex : pour les heures de début et de fin des déplacements).
- Il faut environ deux minutes pour trouver la façon de répondre au questionnaire.
- Il est difficile de se rendre compte tout de suite qu'il faut saisir l'adresse dans le cadre.
- Pour les boutons, mieux vaudrait des carrés à cocher et non des ronds à colorier, car ils sont parfois peu visibles et il est alors difficile de savoir si nous avons coché ou non la réponse souhaitée.

4) Vérifier l'impact des instructions et des aides visuelles sur le bon déroulement du questionnaire

- Aides visuelles et instructions claires dans l'ensemble.
- Je ne sais pas si c'est marqué au début du questionnaire, qu'il faut utiliser la flèche de droite pour avancer ?
- Il serait peut-être nécessaire que les instructions de navigation apparaissent sur la page 1 et non en page 2... pour que chaque personne sache que pour passer à l'étape suivante il faut cliquer sur la flèche jaune !
- Les consignes écrites en bleu sont davantage lisibles, alors que ce ne sont pas les consignes principales.
- Il faudrait renseigner l'internaute sur les enregistrements intermédiaires des données, afin de ne pas le décourager en cas d'interruption de la saisie.

5) Identifier les causes possibles d'abandon : points « difficiles » rencontrés, difficultés dans l'enregistrement des réponses...

- Quand on remplit les adresses : comme elles sont très détaillées (n°, type de rue, nom de la rue...), c'est plus pénible qu'un champ unique à remplir.
- Si on fait beaucoup de trajets, ça fait beaucoup d'adresses à saisir.
- La partie où il faut décrire l'ensemble des déplacements effectués est assez rébarbative et il est certainement difficile de se souvenir avec exactitude des horaires exacts...
- La saisie des adresses est globalement lourde.
- Les questions ont un caractère très personnel (tableau de composition du ménage, adresses exactes...).
- Le nombre de validations demandé est trop important. Par ailleurs, question négative qui pose problème, puisque la réponse a tendance à être négative (« Êtes-vous sûr(e) ne pas avoir effectué d'autres déplacements... ? »).
- Lorsque nous connaissons le temps nécessaire pour saisir un déplacement en totalité, nous avons tendance à omettre certains déplacements.

6) Evaluer la structure du questionnaire : pertinence des tests textes d'accroche initiale, de conclusion et de liaison

Rien de particulier.

7) Contrôler la cohérence du cheminement du questionnaire : succession logique des questions, filtres abusifs, choix impossibles...

- Questionnaire cohérent et logique.
- Que viennent faire les questions sur les téléphones et l'Internet ? L'enquête porte-t-elle bien sur les transports ou recherche-t-on à nous vendre quelque chose derrière ?

Annexes

- Lorsque nous saisissons « Validez les données », l'animation avec la puce jaune ne marche pas, et il est nécessaire de cliquer ensuite sur la flèche de navigation pour passer à l'écran suivant...
- Idem si nous cochons « Modifiez les données » : il faudrait aller directement aux données à modifier (le fait de cliquer sur la flèche suivante est illogique, puisque nous allons revenir en arrière pour modifier la saisie).

8) Valider la présentation du questionnaire : attrait du graphisme et de la mise en page, lisibilité des instructions, facilité de saisie des données, interactivité et navigation dans le questionnaire...

- Ce n'est pas super fun comme présentation, mais d'un côté, le sujet n'est pas non plus quelque chose de funky.
- C'est "fonctionnel" et ça se remplit facilement au moins.
- La mise en page est un peu bizarre, il y a des graphismes marrants (flèches) mais la police ne va pas trop avec et elle fait un peu vieillotte.
- Présentation peu accueillante, avec les couleurs blanc et jaune. Mais le graphisme des flèches est sympathique.
- La police des récapitulatifs est trop petite
- Sur la première page, les flèches et la croix sont trop petites par rapport au titre "enquête ménage déplacements".
- Dans l'ensemble, le résumé des informations à valider écrit en bleu est peu lisible (caractères trop petits) sauf sur le résumé concernant le téléphone et Internet où la police est plus lisible.
- La présentation d'une page n'est pas super attrayante pour les yeux (pas assez de couleurs qui attirent le regard).
- La navigation n'est pas vraiment interactive car on passe juste d'une page à une autre.
- Le design des petits ronds à cocher et les flèches pour avancer sont marrants.
- A la fin on est déçu, car au début du questionnaire figure un super schéma de déplacements avec des belles images, mais ça ne donne pas la même chose pour le trajet du « cobaye » à la fin, qui a juste droit une liste.
- Pourquoi les exemples avec les symboles (maison, travail...) ne sont-ils pas repris ensuite, ce serait plus attractif.
- Proposer un récapitulatif sous forme de graphique, et non sous forme de tableau.
- Mauvaise adaptation des pages du questionnaire à la taille de l'écran, et nécessité d'utiliser les ascenseurs horizontaux et verticaux.
- La flèche de navigation pour changer la page n'est pas toujours visible sur l'écran, et il est impératif de jouer avec l'ascenseur pour passer à la page suivante...
- Lorsque nous saisissons un déplacement, il faudrait proposer la saisie du retour de façon automatique. Cela éviterait de tout ressaisir (ex : domicile-travail, autre déplacement ? oui, est-ce le retour ? oui...).

V Lettre avis envoyée aux non-répondants à l'enquête standard



Lyon, le 27 mars 2006

Madame, Mademoiselle, Monsieur,

Le SYTRAL (Syndicat mixte des Transports pour le Rhône et l'Agglomération Lyonnaise) et ses partenaires réalisent, de novembre 2005 à avril 2006, une importante enquête sur les déplacements effectués quotidiennement par les résidents de l'aire métropolitaine lyonnaise. Cette enquête, qui s'inscrit dans le cadre des « Enquêtes Ménages Déplacements » réalisées périodiquement en France selon une méthodologie standardisée, sera après celles de 1976, 1985 et 1995, la quatrième dans l'agglomération lyonnaise.

Comme nous vous l'avions indiqué dans un précédent courrier, votre ménage a été retenu, par tirage au sort, pour faire partie de l'échantillon représentatif des ménages répondant à cette enquête. Malheureusement, les enquêteurs du bureau d'étude SOFRECO, missionnés par le SYTRAL, n'ont pas réussi à convenir avec vous d'un rendez-vous.

Afin de satisfaire à vos contraintes de disponibilité, nous vous proposons de répondre à cette enquête par Internet, en vous connectant sur le site www.enquete.let.fr et en saisissant votre identifiant et votre mot de passe que vous trouverez ci-dessous. Le questionnaire, d'une durée moyenne de 15 minutes, peut être rempli à tout moment par une personne de votre ménage ayant 11 ans et plus.

Les renseignements* que vous nous fournirez nous seront très utiles pour connaître vos besoins et vos préoccupations en matière de déplacements et étudier les améliorations à apporter notamment aux transports en commun, aux trajets en deux-roues ou à pied, à la circulation et au stationnement des automobiles. Nous vous rappelons que ces données resteront strictement confidentielles.

Pour toute information complémentaire, vous pouvez contacter le bureau d'études en charge de l'enquête au 01.55.48.91.24 du lundi au vendredi entre 14H00 et 19H00. Nous vous remercions de votre participation active et espérons vous retrouver prochainement sur le www.enquete.let.fr

Veuillez agréer, Madame, Mademoiselle, Monsieur, l'expression de notre considération distinguée.

Votre identifiant :

Votre mot de passe :

Bernard RIVALTA

Président du SYTRAL



Partenaires principaux

L'État, la Région Rhône-Alpes, les Départements du Rhône, de l'Ain et de l'Isère, le Grand Lyon, le Syndicat mixte du Schéma de Cohérence Territoriale du Beaujolais, le Syndicat des Transports Urbains du Nord-Isère, la Communauté d'Agglomération du Pays Viennois.

Partenaires associés

La Chambre de commerce et d'industrie de Lyon, le SCoT du Val-de-Saône-Dombes, le SCoT de la Dombes, le SCoT de l'agglomération Lyonnaise, le SCoT de l'Ouest Lyonnais, le SCoT du Bugey Côtière Plaine de l'Ain, le SCoT des Rives du Rhône, le SCoT Nord-Isère, le SCoT du Haut-Rhône Dauphinois.

* Conformément aux lois 51-711 du 7 juin 1951 et 78-17 du 6 janvier 1978, modifiées par la loi n° 2004-801 du 6 août 2004, la confidentialité des informations communiquées est strictement garantie. Un droit d'accès et de rectification peut être exercé pendant le délai où ces informations sont gardées sous forme nominative auprès du SYTRAL 21, boulevard Vivier Merle - BP 3044 - 69 399 LYON Cedex 03

VI Questionnaire web

**ENQUETE WEB SUR LES DEPLACEMENTS DANS
L'AGGLOMERATION LYONNAISE**

Bonjour,

Nous réalisons une enquête sur les pratiques de déplacements dans l'agglomération lyonnaise, pour le compte du SYTRAL, autorité organisatrice des transports urbains.

Les informations sur la mobilité quotidienne en milieu urbain sont rares. C'est pourquoi vos réponses sont particulièrement utiles pour connaître vos besoins en termes de déplacements et étudier les améliorations à apporter.

Vos réponses resteront strictement confidentielles, et leur exploitation sera réalisée de façon anonyme. Les renseignements recueillis sont réservés uniquement aux fins de l'enquête et ne seront communiqués à aucun organisme extérieur.

Cette enquête prend moins de 20 minutes.

Nous vous remercions par avance du temps que vous voudrez bien nous consacrer.

Veuillez cliquer sur la flèche suivante pour avancer dans le questionnaire

En vertu des dispositions de la loi n° 78-17 du 06/01/1978, relative à l'informatique, aux fichiers et aux libertés, nous vous informons que vous disposez d'un droit d'accès, de modification, de rectification et de suppression des données qui vous concernent.

Vous pouvez à tout moment interrompre le questionnaire en cliquant sur la croix. Il vous sera possible de le reprendre à l'endroit où vous vous êtes arrêté(e), en retournant sur le site www.enquete.let.fr et en saisissant de nouveau votre identifiant et votre mot de passe.

Vous pouvez librement naviguer dans le questionnaire au moyen des flèches de navigation situées en bas de l'écran. Utilisez-les au lieu des touches de votre navigateur.

Pour sélectionner vos réponses, il suffit de cliquer sur la puce. Vous devez répondre à chaque question, avant de passer à la page suivante. En cas d'oubli d'une réponse, un message d'alerte s'affiche à l'écran.

Si vous possédez une souris avec molette, faites attention car l'usage de la molette peut modifier l'option choisie dans les listes déroulantes.

Pour toute information complémentaire, nous sommes à votre écoute au 01 55 48 91 24, du lundi au vendredi de 14h à 19h.

1. Avant de vous interroger sur vos déplacements, nous allons nous intéresser à votre ménage. Le ménage se compose de toutes les personnes qui habitent dans votre logement.

i Les personnes qui résident habituellement au moins trois nuits par semaine (du lundi au vendredi inclus) dans un autre logement fixe ne doivent pas être prises en compte.

Pouvez-vous inscrire ci-dessous le nombre de personnes qui habitent dans votre logement - y compris vous-même ?

2. Pouvez-vous inscrire dans le tableau ci-dessous le prénom, l'âge, le sexe et l'occupation de toutes les personnes qui habitent dans votre logement, en commençant par vous même ?

i Si un des enfants a moins d'un an, indiquez « 1 ».

Prénom	Âge	Sexe	Occupation
..... ans	.. M .. F	Travail à temps plein
			Travail à temps partiel
			Apprentissage, formation, stage
			Étudiant (e)
			Scolaire jusqu'au bac
			Chômeur(se), recherche d'emploi
			Retraité(e)
			Reste au foyer
			Enfant non scolarisé
			Autre Précisez

3. Combien y a-t-il de voitures dans votre ménage ?

i Sont concernés les véhicules de tourisme et les petits véhicules utilitaires possédés ou à libre disposition.

Si votre ménage ne possède pas de voiture, veuillez saisir « 0 ».

..... voitures

4. Pouvez-vous nous indiquer l'adresse de votre domicile ?

i Veuillez noter le plus précisément possible.

Les champs suivis du symbole « * » doivent être renseignés, pour valider la saisie de votre adresse et accéder à l'écran suivant.

N° :.....

Voie * :.....

Lieu dit :.....

Code postal ou département * :.....

Commune * :.....

5. Merci de vérifier l'exactitude des informations saisies ci-dessous, puis de les valider ou des les modifier.

i Après avoir saisi votre réponse, il est nécessaire de cliquer sur la flèche de navigation pour passer à l'écran suivant.

Validez les données	
Modifiez les données	

Si « Validez les données » -> Question 6

Si « Modifiez les données » -> Retour à la question 1

6. Les renseignements demandés dans cette partie ne concernent que vous. Merci de ne plus tenir compte des autres personnes qui composent votre ménage.

i Sauf indication contraire, une seule réponse doit être cochée pour chaque question.

Si vous n'êtes pas prénom généré par l'ordinateur conformément à la première ligne du tableau de composition du ménage, veuillez sélectionner votre prénom dans le menu ci-joint.

<i>Prénom généré par l'ordinateur</i>	Prénom 1
	Prénom 2
	Prénom 3
	Prénom

Filtre : si âge au moins égal à 16 ans

7. Possédez-vous le permis de conduire voiture ou pratiquez-vous la conduite accompagnée ?

i Seul le permis de catégorie B (tourisme) est concerné.

... oui
... non
... NSP

Filtre : si occupation actuelle différente de scolaire jusqu'au bac ou étudiant

8. Quel est votre niveau d'études ?

Primaire	
Secondaire (de la 6 ^{ème} à la 3 ^{ème} , CAP)	
Secondaire (de la seconde à la terminale, BEP)	
Supérieur jusqu'à BAC + 2	
Supérieur BAC + 3 et plus	
Pas d'études	
Autre	
Précisez	

Filtre : si occupation actuelle égale à scolaire jusqu'au bac ou étudiant

8bis. Quelle formation suivez-vous ?

Primaire	
Primaire	
Secondaire (de la 6 ^{ème} à la 3 ^{ème} , CAP)	
Secondaire (de la seconde à la terminale, BEP)	
Supérieur jusqu'à BAC + 2	
Supérieur BAC + 3 et plus	
Pas d'études	
Autre	
Précisez	

Filtre : si occupation actuelle différente d'apprentissage, formation, stage, scolaire jusqu'au bac ou étudiant

9. Quelle est ou quelle a été votre profession principale ?

i En cas de doutes sur votre profession, veuillez consulter l'aide ci-dessous.

Agriculteurs	
Artisans, commerçants, chefs d'entreprise	
Cadres et professions intellectuelles supérieures	
Professions intermédiaires	
Employés	
Ouvriers	
Autre	
Précisez	

Catégorie	Détail
Agriculteurs exploitants	Agriculteurs sur petite, moyenne ou grande exploitation
Artisans, commerçants et chefs d'entreprise	Artisans, Commerçants et assimilés, Chefs d'entreprise de 10 salariés ou plus
Cadres et professions intellectuelles supérieures	Professions libérales, Cadres de la fonction publique, Professeurs, Professions scientifiques, Professions de l'information, des arts et des spectacles, Cadres administratifs et commerciaux d'entreprise, Ingénieurs et cadres techniques d'entreprise
Professions intermédiaires	Professeurs des écoles, Instituteurs et assimilés, Professions intermédiaires de la santé et du travail social, Clergé, Religieux, Professions intermédiaires administratives de la fonction publique, Professions intermédiaires administratives et commerciales des entreprises, Techniciens, Contremaîtres, Agents de maîtrise
Employés	Employés civils et agents de service de la fonction publique, Policiers et militaires, Employés administratifs d'entreprise, Employés de commerce, Personnels des services directs aux particuliers
Ouvriers	Ouvriers qualifiés de type industriel ou artisanal, Chauffeurs, Ouvriers qualifiés de la manutention, du magasinage et du transport, Ouvriers non qualifiés de type industriel ou artisanal, Ouvriers agricoles

Filtre : si travailleur à temps plein, à temps partiel, apprentissage, formation, stage, scolaire jusqu'au bac ou étudiant

10. Quelle est l'adresse de votre lieu habituel de travail ou d'études ?

i Veuillez noter le plus précisément possible.

Les champs suivis du symbole « * » doivent être renseignés, pour valider la saisie de votre adresse et accéder à l'écran suivant.

Nom d'établissement :.....
 N° :.....
 Voie * :.....
 Lieu dit :.....
 Code postal ou département * :.....
 Commune * :.....

A défaut d'une adresse précise, donnez le plus de renseignements possibles sur le lieu (nom de l'établissement, lieu public le plus proche, croisement de rues le plus proche...), en indiquant toujours la commune et le département dans les champs à gauche.

--

Filtre : si travailleur à temps plein, à temps partiel, apprentissage, formation, stage scolaire jusqu'au bac ou étudiant et si voiture accessible par le ménage et possession du permis de conduire ou pratique de la conduite accompagnée.

11. En général pour vous rendre sur votre lieu de travail ou d'études, utilisez-vous la voiture en tant que conducteur ?

... oui
 ... non

Filtre : si utilisation de la voiture en tant que conducteur pour se rendre sur le lieu de travail

12. En général, sur votre lieu de travail ou d'études, rencontrez-vous des problèmes de stationnement ?

i Vous pouvez choisir plusieurs réponses

Oui	
Non, j'ai une place réservée	
Non, il y a une offre de stationnement importante à proximité	
Non, compte tenu de mes horaires	
Autre	
Précisez	

Filtre : si voiture accessible par le ménage et possession du permis de conduire ou pratique de la conduite accompagnée

13. En semaine (du lundi au vendredi), avec quelle fréquence utilisez-vous la voiture en tant que conducteur ?

Tous les jours ou presque	
Au moins 2 déplacements par semaine	
Au moins 2 déplacements par mois	
Exceptionnellement	
Jamais	

14. En semaine (du lundi au vendredi), avec quelle fréquence utilisez-vous la voiture en tant que passager ?

Tous les jours ou presque	
Au moins 2 déplacements par semaine	
Au moins 2 déplacements par mois	
Exceptionnellement	
Jamais	

15. En semaine (du lundi au vendredi), avec quelle fréquence utilisez-vous les transports en commun ?

Tous les jours ou presque	
Au moins 2 déplacements par semaine	
Au moins 2 déplacements par mois	
Exceptionnellement	
Jamais	

16. En semaine (du lundi au vendredi), avec quelle fréquence utilisez-vous un deux roues ?

i Veuillez ne pas tenir compte de l'utilisation faite pour les loisirs.

Tous les jours ou presque	
Au moins 2 déplacements par semaine	
Au moins 2 déplacements par mois	
Exceptionnellement	
Jamais	

Filtre : si fréquence d'utilisation d'un deux-roues = « Tous les jours ou presque » ou « Au moins deux déplacements par semaine ».

17. Quel type de deux-roues utilisez-vous ?

Vélo	
Deux-roues à moteur < à 50cm ³	
Deux-roues à moteur > à 50cm ³	
Autre, précisez	

18. Merci de vérifier l'exactitude des informations saisies ci-dessous, puis de les valider ou des les modifier.

i Après avoir saisi votre réponse, il est nécessaire de cliquer sur la flèche de navigation pour passer à l'écran suivant.

Validez les données	
Modifiez les données	

Si « Validez les données » -> OK

Si « Modifiez les données » -> Retour à la question 6

Nous allons maintenant vous demander de décrire dans l'ordre chronologique les déplacements que vous avez effectués entre 04h00 du matin Mardi 23/05/2006 et 04h00 du matin Mercredi 24/05/2006.

Merci de nous indiquer pour chaque déplacement :

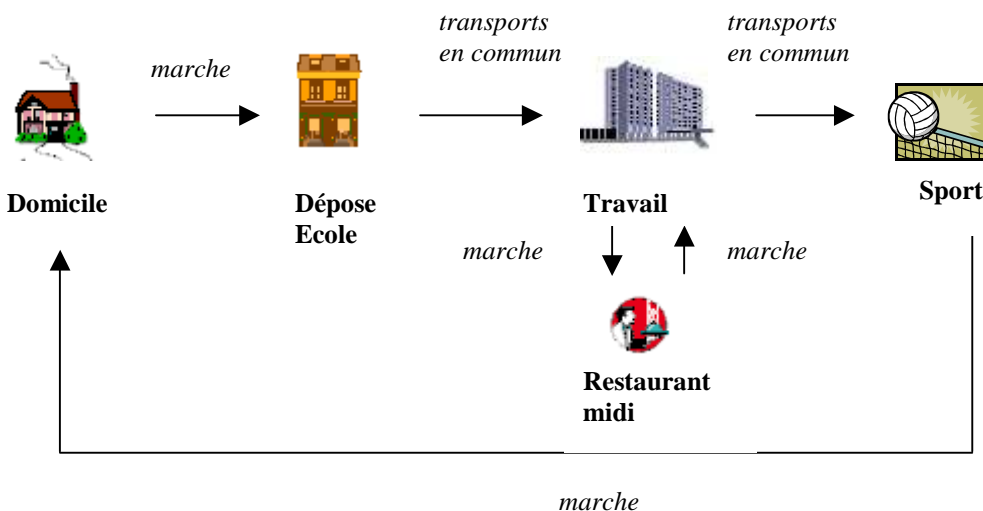
- l'adresse des lieux de départ et d'arrivée, en donnant le plus d'informations possibles (adresse exacte, lieu public ou station de transport en commun à proximité, croisement de rues le plus proche....) ;
- votre heure de départ et d'arrivée de chaque endroit ;
- l'ensemble des moyens de transport utilisés ;
- l'activité principale que vous avez effectuée à destination.

i Vos réponses sont utiles, même si vous n'avez pas quitté votre domicile durant cette période ; en cas de doutes sur un déplacement, merci de l'inclure dans votre descriptif.

L'exemple suivant est purement illustratif ; il ne doit en aucun cas limiter vos réponses.

Exemple : l'individu a effectué 6 déplacements au cours de la journée, dont 2 en transports collectifs et 4 à pied (chaque flèche correspond à un déplacement).

:



A la fin de chaque déplacement saisi, il vous sera possible de modifier, ajouter ou supprimer des informations.

A la fin du questionnaire, vous pourrez visualiser un tableau récapitulant l'ensemble des déplacements saisis.

Les explications ci-dessus sont consultables à tout instant par l'intermédiaire du lien «AIDE».

RAPPEL :

- Tous les déplacements sont à considérer, y compris pour les petits arrêts (achats, dépose d'un passager, promenade...) et les retours au domicile.
- Si vous gardez votre voiture dans un parking et finissez votre trajet à pied, sélectionnez la voiture particulière et la marche comme moyens de déplacement.
- Dans le cas d'une simple promenade sans arrêt à un endroit précis, seul le lieu le plus éloigné de votre point de départ doit être retenu.

19. Pouvez-vous nous indiquer le lieu où vous étiez Jour J à 04h00 du matin?

Domicile	
Travail ou études	
Autre	

Si Domicile -> Question 20.

Si Travail ou études -> Pouvez-vous confirmer / modifier l' adresse ci-dessous, ou à défaut saisir l' adresse de votre lieu de travail ou d' études?

Si autre -> Pouvez-vous indiquer précisément le lieu où vous étiez ?

Nom d'établissement :.....
N° :.....
Voie * :.....
Lieu dit :.....
Code postal ou département * :.....
Commune * :.....

A défaut d'une adresse précise, donnez le plus de renseignements possibles sur le lieu (nom de l'établissement, lieu public le plus proche, croisement de rues le plus proche...), en indiquant toujours la commune et le département dans les champs à gauche.

--

Déplacement n°1 :

20. Quand êtes-vous parti(e) de cet endroit ?

Jour j	
Jour J+1	
Je n'ai pas quitté cet endroit entre 4h00 du matin Jour J et 4h00 du matin Jour J+1	

Si « Je n'ai pas quitté cet endroit entre 4h00 du matin Jour J et 4h00 du matin Jour J+1 » :

Etes-vous sûr(e) d'avoir déclaré l'ensemble de vos déplacements entre 04h00 Jour J et 04h00 Jour J+1, y compris pour acheter du pain, le journal... pour accompagner quelqu'un, pour retourner à votre domicile ou pour toutes autres raisons...?

Oui	
Non	

Si « oui » -> Question 34

Si « non » -> Retour à la question 20

21. A quelle heure avez-vous quitté cet endroit ?

i Merci de respecter le format hh / mm

	H	
--	---	--

22. Quel(s) moyen(s) de transport avez-vous utilisé pour ce déplacement?

i Veuillez cocher l'ensemble des moyens de transport utilisés.

<input type="checkbox"/> Marche à pied	
<input type="checkbox"/> Vélo	
<input type="checkbox"/> Conducteur de véhicule particulier	
<input type="checkbox"/> Passager de véhicule particulier	
<input type="checkbox"/> Fourgon, camionnette, camion	
<input type="checkbox"/> Deux roues	
<input type="checkbox"/> Métro	
<input type="checkbox"/> Tramway	
<input type="checkbox"/> Bus ou autocars	
<input type="checkbox"/> Train	
<input type="checkbox"/> Taxi	
<input type="checkbox"/> Autre	
<i>Précisez</i>	

Filtre : si conducteur de voiture particulière

23. Avez-vous payé le stationnement lorsque vous êtes arrivé(e)s à destination?	... oui ... non ... NSP ... autre, précisez
--	--

24. Pouvez-vous nous indiquer le lieu où vous êtes allés ?

Domicile	
Travail ou études	
Autre	

Si Domicile -> Question 26.

Si Travail ou études -> Pouvez-vous confirmer / modifier l' adresse ci-dessous, ou à défaut saisir l' adresse de votre lieu de travail ou d' études?

Si autre -> Pouvez-vous indiquer précisément le lieu où vous étiez ?

Nom d'établissement :
N° :
Voie * :
Lieu dit :
Code postal ou département * :
Commune *:

A défaut d'une adresse précise, donnez le plus de renseignements possibles sur le lieu (nom de l'établissement, lieu public le plus proche, croisement de rues le plus proche...), en indiquant toujours la commune et le département dans les champs à gauche.

--

25. Quelle était votre activité principale à cet endroit ?

i Si votre travail nécessite des tournées professionnelles (ex : chauffeur livreur, conducteur de taxi, facteur, agent EDF...), merci de nous indiquer précisément l'heure de début et de fin de votre activité professionnelle, ainsi que tous les trajets réalisés pour motifs personnels. En revanche, il n'est pas nécessaire d'indiquer chacun de vos déplacements pour raison professionnelle.

Travail	
Etudes	
Achats	
Démarches personnelles (administratif, santé, recherche d'emploi...)	
Loisirs ou visites	
Restauration hors du domicile	
Aller chercher / déposer quelqu'un	
Tournée professionnelle	
Autre	
Précisez	

Annexes

Sous-menus :

=> <i>Etudes</i>	=> <i>Achats</i>	=> <i>Loisirs</i>
Collège	Grands magasins et leurs galeries marchandes	Activités sportives, culturelles ou associatives
Lycée	Hypermarchés et supermarchés	Promenade, « lèche-vitrines »
Universités et grandes écoles	Petits et moyens commerces, marchés	Visite à des parents ou à des amis

26. Quand êtes-vous arrivé(e) à cet endroit ?

<i>Jour J</i>	
<i>Jour J + 1</i>	
<i>Autre jour</i>	

27. A quelle heure êtes-vous arrivés à cet endroit ?

(L'heure de départ saisie était ...)

i Merci de respecter le format hh / mm

	H	
--	---	--

28. Merci de vérifier l'exactitude des informations saisies ci-dessous, puis de les valider ou des les modifier avant de passer au déplacement suivant.

i Après avoir saisi votre réponse, il est nécessaire de cliquer sur la flèche de navigation pour passer à l'écran suivant.

Validez ce déplacement	
Modifiez ce déplacement	
Effacer ce déplacement	

Si « Validez les données » -> Question 30

Si « Modifiez les données » -> Retour à la question 19

Si « Effacez les données » -> OK

29. Voulez-vous vraiment supprimer ce déplacement ?

Oui	
Non	

Si « oui » -> Retour à la question 28

Si « non » -> Retour à la question 19

30. Avez-vous effectué un autre déplacement entre 04h00 Jour J et 04h00 Jour J+1?

Oui	
Non	

Si « oui » -> Question 31

Si « non » -> Tableau récapitulatif

Déplacement n°2 :

31. Quand avez-vous quitté cet endroit ?

Jour j	
Jour J+1	
Je n'ai pas quitté cet endroit entre 4h00 du matin Jour J et 4h00 du matin Jour J+1	

Reprendre les questions 21 à 30

Tableau récapitulatif

Validez la chaîne de déplacements	
Ajoutez un nouveau déplacement	

Si « Ajoutez un nouveau déplacement » :

32. Pouvez-vous nous indiquer le lieu où vous étiez?

Domicile	
Travail ou études	
Autre	

Reprendre les questions 21 à 30

Si « Validez la chaîne de déplacements » :

33. Etes-vous sûr(e) d'avoir déclaré l'ensemble de vos déplacements entre 04h00 Jour J et 04h00 Jour J+1, y compris pour acheter du pain, le journal... pour accompagner quelqu'un, pour retourner à votre domicile ou pour toutes autres raisons...?

Oui	
Non	

Si « non » -> « Veuillez cliquer sur Ajouter un nouveau déplacement », puis retour à la question 32.

Si « oui » :

34. Pour terminer, nous allons vous demander des renseignements sur votre équipement téléphonique. Avez-vous le téléphone fixe au domicile ?	... oui ... non ... ne souhaite pas répondre
---	--

35. Etes-vous inscrit ?

Sur liste rouge	
Sur liste orange	
Sur l'annuaire des abonnés « France Télécom » (ou en cours d'inscription)	
Autre, précisez	

36. Avez-vous une connexion Internet au domicile ?	... oui ... non ... ne souhaite pas répondre
---	--

37. S'agit-il d'une connexion haut débit ?	... oui ... non ... NSP
---	-------------------------------

38. Possédez-vous un téléphone portable ? <i>i Il s'agit des téléphones portables que vous utilisez, à titre personnel ou professionnel.</i>	... oui ... non ... ne souhaite pas répondre
--	--

39. Pouvez-vous nous indiquer dans quelle tranche se situe le montant des revenus annuels nets de votre ménage ? <i>i Sont pris en considération les primes, 13ème mois, revenus annexes, prestations sociales</i>	
--	--

Moins de 10 000 €	
De 10 000 € à moins de 20 000 €	
De 20 000 € à moins de 30 000 €	
De 30 000 € à moins de 40 000 €	
De 40 000 € à moins de 60 000 €	
60 000 € et plus	
NSP	

40. Merci de vérifier l'exactitude des informations saisies ci-dessous, puis de les valider ou des les modifier.

i Après avoir saisi votre réponse, il est nécessaire de cliquer sur la flèche de navigation pour passer à l'écran suivant.

41. Vous avez choisi de répondre à cette enquête par Internet. Quelle(s) raison(s) ont motivé votre choix ?

i Vous pouvez choisir plusieurs réponses

Je n'ai jamais été contacté par un enquêteur pour fixer un rendez-vous pour une enquête à mon domicile
Je suis rarement disponible aux créneaux proposés par les enquêteurs
Je ne souhaite pas recevoir un enquêteur à mon domicile
Je préfère répondre sur le web
Je n'ai pas le temps
Autre, précisez

42. Connaître votre adresse e-mail nous permettrait de vous contacter rapidement pour obtenir d'éventuelles précisions concernant vos réponses.

J'accepte de saisir mon adresse e-mail.

Je n'accepte pas de saisir mon adresse e-mail.

Filtre : si accepte de saisir l'adresse e-mail

Email		@	
--------------	--	---	--

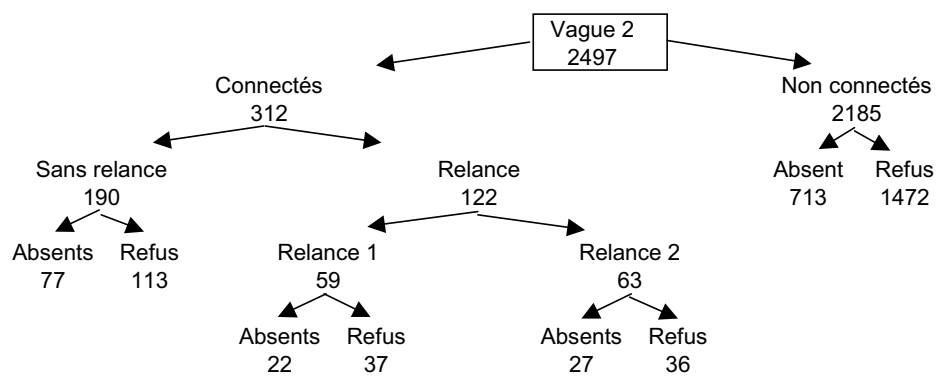
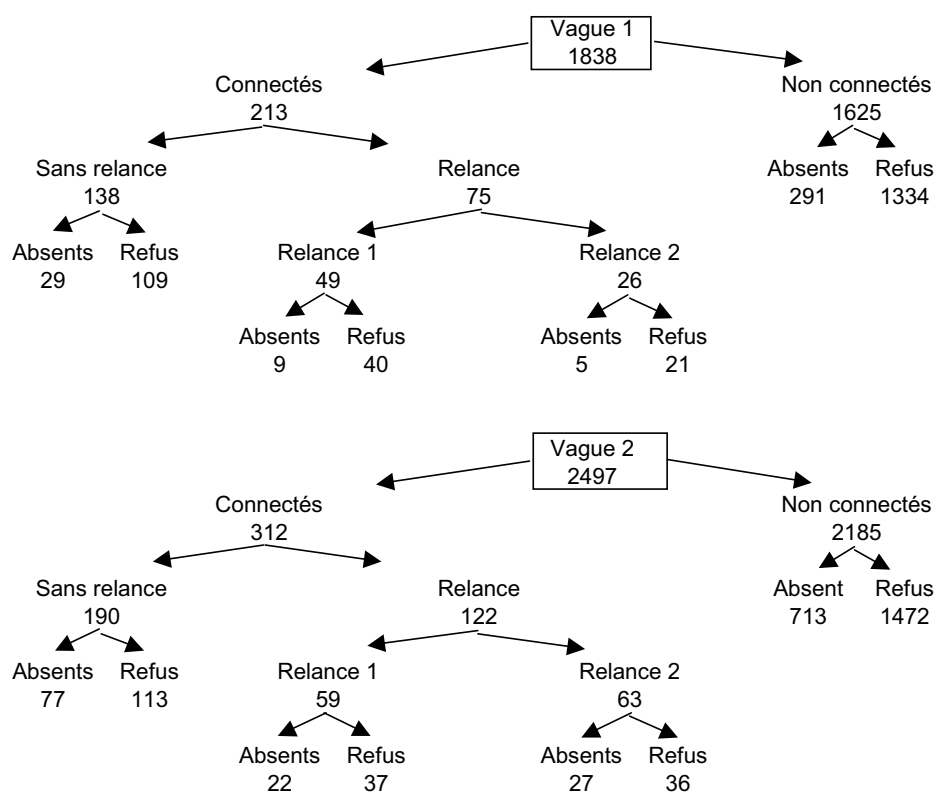
Cette enquête est maintenant terminée !

Merci beaucoup d'avoir consacré du temps pour remplir ce questionnaire. Vos réponses vont contribuer au succès de notre étude.

Pour tout commentaire sur ce questionnaire, adressez vos remarques à :
let@let.ish-lyon.cnrs.fr

Veillez cliquer sur la flèche de navigation pour quitter le questionnaire.

VII Vagues de courriers web



Source : Questionnaire web (2005)

VIII Répartition des répondants selon la réponse à la question : 'Souhaitez-vous communiquer votre adresse e-mail?'

<i>Sexe</i>	Oui		Non		Total	
Femme	76	53%	133	59%	209	57%
Homme	68	47%	91	41%	159	43%
Vide	1				1	
Total	145		224		369	

<i>Age</i>	Oui		Non		Total	
< 30 ans	27	19%	36	16%	63	17%
30-60	106	74%	164	73%	270	73%
> 60 ans	11	8%	24	11%	35	10%
Vide	1				1	
Total	145		224		369	

<i>Occupation</i>	Oui		Non		Total	
Actif	106	74%	154	69%	260	71%
Inactif	11	8%	23	10%	34	9%
Retraité	15	10%	32	14%	47	13%
Etudiants-Scolaires	12	8%	15	7%	27	7%
Vide	1				1	
Total	145		224		369	

<i>PCS</i>	Oui		Non		Total	
Artisans-Commerçants	5	3%	9	4%	14	4%
Cadres et PIS	62	43%	70	32%	132	37%
Profession intermédiaire	13	9%	26	12%	39	11%
Employés	45	31%	81	37%	126	35%
Ouvriers	6	4%	15	7%	21	6%
Etudiants-Scolaires-Apprentis	13	9%	16	7%	29	8%
Vide	1		7		8	
Total	145		224		369	

<i>Nombre de personnes</i>	Oui		Non		Total	
1 personne	34	23%	56	25%	90	24%
2 personnes	40	28%	72	32%	112	30%
3 personnes	30	21%	27	12%	57	15%
4 personnes	30	21%	51	23%	81	22%
5+ personnes	11	8%	18	8%	29	8%
Total	145		224		369	

<i>Niveau d'éducation</i>	Oui		Non		Total	
Diplôme supérieur	102	70%	146	65%	248	67%
Diplôme non supérieur	43	30%	78	35%	121	33%
Total	145		224		369	

<i>Revenus annuels</i>	Oui		Non		Total	
< 10 000 €	5	3%	12	5%	17	5%
[10 000 - 20 000 [€	23	16%	38	17%	61	17%
[20 000 - 30 000 [€	29	20%	33	15%	62	17%
[30 000 - 40 000 [€	33	23%	32	14%	65	18%
[40 000 - 60 000 [€	24	17%	27	12%	51	14%
> 60 000 €	16	11%	7	3%	23	6%
Non réponse	15	10%	75	33%	90	24%
Total	145		224		369	

IX Répartition des personnes de 10 ans et plus par enquête et par caractéristique socio-économique

		INSEE	EMD 06
Sexe	Hommes	47,2%	47,5%
	Femmes	52,8%	52,5%
Age	< 20 ans	14,7%	15,3%
	20 - 29 ans	18,9%	14,8%
	30 - 59 ans	45,1%	45,4%
	> 59 ans	21,3%	24,5%
Occupation	Actif	45,7%	47,3%
	Etudiant et Scolaire	20,3%	19,7%
	Chômeur	6,5%	4,5%
	Inactif	9,4%	4,4%
	Retraité	18,1%	24,1%
PCS	Agriculteur, exploitant	0,1%	0,1%
	Artisan, commerçant et chef d'entreprise	5,0%	3,7%
	Cadre et profession intellectuelle supérieure	13,9%	22,4%
	Profession intermédiaire	21,2%	24,9%
	Employés	24,6%	26,0%
	Ouvriers	18,7%	16,7%
	Autre, sans activité professionnelle	16,5%	6,3%
Diplôme	En cours	20.3%	19.9%
	Aucun diplôme	14.9%	1.7%
	CEP	10.8%	7.1%
	CAP, BEPC	23.7%	19.0%
	BAC, brevet professionnel	10.4 %	18.8%
	BAC+2	8.9%	13.7%
	Diplôme supérieur	11.0%	19.8%

Source : EMD Lyon 2006 et INSEE - RGP 1999

X Comparaison des marges dans l'échantillon et dans la population

Variable	Modalité	Marge échantillon	Marge population	Pourcentage échantillon	Pourcentage population
Age	< 30 ans	4011	56	30.22	16.77
	[30,60] ans	6278	247	47.31	73.95
	> 60 ans	2982	31	22.47	9.28
Diplôme	Supérieur	5072	227	38.22	67.96
	Non supérieur	8199	107	61.78	32.04
Connexion internet	Oui	7804	267	58.80	79.94
	Non	5467	67	41.20	20.06
Nombre de personnes	1	2049	86	15.44	25.75
	2	4120	99	31.05	29.64
	3	2304	50	17.36	14.97
	4	2566	71	19.34	21.26
	5	2232	28	16.82	8.38
PCS	Cadres	1336	97	10.07	29.04
	Employés	1727	92	13.01	27.54
	Autre emploi	3015	49	22.72	14.67
	Sans emploi	7193	96	54.20	28.74
Possession du permis	Oui	9780	307	73.69	91.92
	Non	3491	27	26.31	8.08
Téléphone portable	Oui	9114	292	68.68	87.43
	Non	4157	42	31.32	12.57

Source : EMD web - Lyon 2006

XI Répartition des individus par enquête après redressement de l'échantillon face-à-face

		Enquête web	Enquête face-à-face	Khi-deux de pearson valeur	p-value
Sexe	Homme	56,8%	56,3%	0,037	0,847
	Femme	43,2%	43,7%		
Occupation	Etudiant	5,4%	5,5%	13,256	0,103
	Apprenti, stage	0,5%	0,1%		
	Temps partiel	9,5%	11,8%		
	Temps plein	60,6%	59,3%		
	Chômeur	3,8%	4,4%		
	Femme au foyer	4,9%	3,3%		
	Retraité	12,8%	11,3%		
	Scolaire	1,9%	2,9%		
	Autre	50,0%	1,4%		
Revenus	Moins de 10 000 euros	6,1%	10,9%	8,246	0,143
	De 10 000 à 20 000 euros	21,9%	22,4%		
	De 20 000 à 30 000 euros	22,2%	22,9%		
	De 30 000 à 40 000 euros	23,3%	20,5%		
	De 40 000 à 60 000 euros	18,3%	16,3%		
	Plus de 60 000 euros	8,2%	7,0%		
Lieu de travail	1 - Centre	26,6%	21,0%	16,479	0,058
	2 - Hypercentre	30,3%	28,9%		
	3 - 1ère cour. ouest	11,4%	11,5%		
	4 - 1ère cour. est	9,3%	9,7%		
	5 - 2ème cour. ouest	8,3%	6,7%		
	6 - 2ème cour. est	5,5%	6,0%		
	7 - 3ème cour. ouest	2,4%	5,7%		
	8 - 3ème cour. est	1,0%	2,5%		
	9 - 4ème cour. est	2,4%	4,5%		
	10 - externe	2,8%	3,5%		
Nombre de voitures du ménage	Aucune	9,2%	10%	0,669	0,716
	1	40,7%	42,1%		
	2 et plus	50,1%	47,9%		

Source : EMD web - Lyon 2006

XII Espérance d'une loi tronquée

L'espérance d'une loi normale tronquée en s est :

$$E(\epsilon_i | \epsilon_i > s) = \int_{\epsilon_i=s}^{\infty} \epsilon_i f(\epsilon_i | \epsilon_i > s) .d\epsilon_i \quad (97)$$

$$E(\epsilon_i | \epsilon_i > s) = \int_{\epsilon_i=s}^{\infty} \epsilon_i \frac{\phi(\epsilon_i)}{Prob(\epsilon_i > s)} .d\epsilon_i \quad (98)$$

$$E(\epsilon_i | \epsilon_i > s) = \int_{\epsilon_i=s}^{\infty} \epsilon_i \frac{\phi(\epsilon_i)}{[1 - \Phi(\epsilon_i)]} .d\epsilon_i \quad (99)$$

$$E(\epsilon_i | \epsilon_i > s) = \frac{1}{1 - \Phi(s)} * \frac{1}{\sqrt{2\Pi}} * \int_{\epsilon_i=s}^{\infty} \epsilon_i e^{(\frac{-\epsilon_i^2}{2})} .d\epsilon_i \quad (100)$$

$$E(\epsilon_i | \epsilon_i > s) = \frac{1}{1 - \Phi(s)} * \frac{1}{\sqrt{2\Pi}} * \left[-e^{(\frac{-\epsilon_i^2}{2})} \right]_s^{\infty} \quad (101)$$

$$E(\epsilon_i | \epsilon_i > s) = \frac{1}{1 - \Phi(s)} * \frac{1}{\sqrt{2\Pi}} * e^{\frac{-\epsilon_i^2}{2}} \quad (102)$$

$$E(\epsilon_i | \epsilon_i > s) = \frac{\phi(s)}{1 - \Phi(s)} = \frac{\phi(s)}{\Phi(-s)} \quad (103)$$

Par analogie :

$$E(\epsilon_i | \epsilon_i \leq s) = -\frac{\phi(s)}{\Phi(s)} \quad (104)$$

XIII Statistiques descriptives - personnes mobiles de 18 ans et plus

Les tableaux suivants donnent quelques statistiques descriptives par enquête sur les variables retenues : nombre d'observations, valeur minimum et maximum, moyenne et écart-type pour les variables continues (tableaux 102 et 103), modalités et effectifs, pour les variables nominales (tableaux 104 et 105).

Variabes	Obs.	Moyenne	Ecart-type	Minimum	Maximum
Age	9 811	43,35	17,62	18	98
Nb d'enfants du ménage	9 811	0,68	1,03	0	7
Nb de voitures du ménage	9 811	0,68	0,39	0	4
Nb de personnes du ménage	9 811	2,79	1,41	1	10
Distance du domicile au centre de l'agglomération (m)	9 811	7 628	5 168	377	21 937
Densité de la zone de résidence (habitants au km2)	9 811	5 908	5 875	134	21 059

TAB. 102 – Statistiques descriptives des variables continues (échantillon face-à-face)

Variabes	Obs.	Moyenne	Ecart-type	Minimum	Maximum
Age	270	43,35	13,07	18	83
Nb d'enfants du ménage	270	0,62	0,89	0	3
Nb de voitures du ménage	270	0,80	0,35	0	2
Nb de personnes du ménage	270	2,62	1,29	1	6
Distance du domicile au centre de l'agglomération (m)	270	6 846	4 931	377	21 815
Densité de la zone de résidence (habitants au km2)	270	5 566	5 909	161	21 059

TAB. 103 – Statistiques descriptives des variables continues (échantillon web)

Variabes	Observations	Modalités	Pourcentage
Sexe	9 811	Homme	52,7%
		Femme	47,3%
Possession du permis	9 811	Oui	85,8%
		Non	14,2%
Vendredi	9 811	Oui	15,2%
		Non	84,8%
Activité	9 811	Actif : centre	23,9%
		Actif : périphérie	29,0%
		Actif : non précis	2,8%
		Inactif	44,3%
Revenus déclarés	9 811	Oui	65,6%
		Non	34,3%
Connexion internet	9 811	Oui	58,1%
		Non	41,9%
Liste de téléphone	9 811	Annuaire	69,3%
		Liste	20,8%
		Pas de téléphone	9,9 %
Téléphone portable	9 811	Oui	73,1%
		Non	26,9%
Niveau d'étude	9 811	Supérieur	35,7%
		Non supérieur	54,6%
		En cours	9,7%
Diplôme	9 980	Supérieur	42,7%
		Non supérieur	57,3%

TAB. 104 – Statistiques descriptives des variables nominales (échantillon face-à-face)

Variabes	Observations	Modalités	Pourcentage
Sexe	270	Femme	56,3%
		Homme	43,7%
Possession du permis	270	Oui	94,4%
		Non	5,6%
Vendredi	270	Oui	34,8%
		Non	65,2%
Activité	270	Actif : centre	46,3%
		Actif : périphérie	29,3%
		Actif : non précis	0,4%
		Inactif	24,1%
Revenus déclarés	270	Oui	79,3%
		Non	20,7%
Connexion internet	256	Oui	85,2%
		Non	14,8%
Liste de téléphone	252	Annuaire	54,0%
		Liste	37,7%
		Pas de téléphone	8,3 %
Téléphone portable	253	Oui	88,9%
		Non	11,1%
Niveau d'étude	270	Supérieur	67,8%
		Non supérieur	27,0%
		En cours	5,2%
Diplôme	270	Supérieur	72,2%
		Non supérieur	27,8%

TAB. 105 – Statistiques descriptives des variables nominales (échantillon web)

Lexique

CAPI : Computer Assisted Personal Interviewing (Enquête face-à-face assistée par ordinateur)

CATI : Computer Assisted Telephone Interviewing (Enquête téléphonique assistée par ordinateur)

CAWI : Computer Assisted Web Interview (Enquête web assistée par ordinateur)

CERTU : Centre d'Etudes sur les Réseaux, les Transports, l'Urbanisme et les constructions publiques

CETE : Centre d'Etudes Techniques de l'Equipement

CNIL : Commission Nationale de l'Informatique et des Libertés

EMD : Enquête Ménages Déplacements

FILOCOM : Fichier des LOGements par COMMunes

GPS : Global Positionning System (Système de géolocalisation par satellite)

INSEE : Institut National de la Statistique et des Etudes Economiques

SCOT : Schéma de COhérence Territoriale

SYTRAL : SYndicat mixte des Transports pour le Rhône et l'Agglomération Lyonnaise

TCL : Transports en Commun Lyonnais

Table des figures

1	Les différentes sources de biais dans les enquêtes web	27
2	Modèle conceptuel des déterminants de l'intention de répondre en ligne	35
3	Deux situations de mixité du recueil de données	37
4	Evolution du taux de pénétration de l'ordinateur et d'internet dans les ménages français	44
5	Taux d'équipement à internet au domicile	45
6	Evolution des inégalités d'équipement des ménages français - Indice de Gini	48
7	Exemple d'un cycle de déplacements quotidiens	71
8	Cycle des questions relatives à un déplacement	71
9	Choix de l'activité exercée	72
10	Les 9 zones du périmètre du SCOT de Lyon	87
11	Dates de connexion des internautes	90
12	Evolution du nombre d'abandons selon la progression dans le questionnaire	95
13	Heures de connexion des internautes la semaine	99
14	Heures de connexion des internautes le week-end	100
15	Temps moyen de saisie des réponses en ligne (mn)	101
16	Répartition des ménages par enquête et par nombre de voitures	117
17	Répartition des ménages de l'enquête web par zone de résidence	121
18	Répartition des ménages de l'enquête face-à-face par zone de résidence	122
19	Répartition des ménages par nombre de personnes et par enquête	123
20	Répartition des ménages par nombre de voitures et par enquête	124
21	Répartition des ménages par nombre de voitures par personne de 18 ans et plus	124
22	Répartition des ménages par tranche de revenus annuels nets . .	126
23	Répartition des ménages selon leur inscription sur une liste té- léphonique	127
24	Répartition des ménages selon la possession d'une connexion internet	128
25	Répartition des ménages selon leur possession d'un téléphone portable	128
26	Répartition des personnes par sexe et par enquête	129
27	Répartition des personnes par tranche d'âge et par enquête . . .	130
28	Répartition des personnes par occupation et par enquête	131

29	Répartition des personnes par PCS et par enquête	132
30	Répartition des personnes par niveau d'étude et par enquête . .	132
31	Répartition des personnes selon la possession du permis de conduire et par enquête	133
32	Répartition des personnes par enquête et par fréquence d'utili- sation de la voiture conducteur en semaine	138
33	Répartition des personnes par enquête et par fréquence d'utili- sation de la voiture passager en semaine	139
34	Répartition des personnes par enquête et par fréquence d'utili- sation des transports collectifs en semaine	140
35	Répartition des personnes par enquête et par fréquence d'utili- sation des deux-roues en semaine	140
36	Répartition des personnes par nombre de déplacements et par enquête	142
37	Répartition des déplacements quotidiens par mode et par enquête	143
38	Répartition des déplacements quotidiens par motif à destination et par enquête	145
39	Répartition des déplacements quotidiens par enquête et par mo- tif à l'origine, lorsque le motif à la destination est le domicile . .	146
40	Nombre de déplacements quotidiens moyen par heure de début et par enquête	147
41	Distribution des sorties quotidiennes du domicile par enquête . .	153
42	Répartition des sorties quotidiennes du domicile par 1er mode et par enquête	154
43	Répartition des sorties quotidiennes du domicile par motif à la destination et par enquête	156
44	Répartition des personnes par enquête et par équipement en téléphone fixe après redressement	165
45	Répartition des personnes par enquête et par déclaration du revenu après redressement	166
46	Répartition des personnes par enquête et par jour de référence des déplacements après redressement	166
47	Répartition des personnes par enquête et par fréquence d'utili- sation de la voiture conducteur en semaine après redressement .	168
48	Répartition des personnes par enquête et par fréquence d'utili- sation de la voiture passager en semaine après redressement . .	168
49	Répartition des personnes par enquête et par fréquence d'utili- sation des transports collectifs en semaine après redressement .	169
50	Répartition des personnes par enquête et par fréquence d'utili- sation des deux-roues en semaine après redressement	170
51	Répartition des individus par nombre de déplacements et par enquête après redressement	170
52	Répartition des déplacements des individus par mode et par enquête après redressement	172
53	Répartition des déplacements des individus par motif à la des- tination et par enquête après redressement	173

Table des figures

54	Répartition des déplacements des individus par motif à l'origine et par enquête après redressement lorsque le motif à la destination est le domicile	173
55	Répartition des déplacements des individus par heure de début et par enquête après redressement	175
56	Répartition des individus par nombre de sorties du domicile et par enquête après redressement	177
57	Répartition des sorties du domicile par 1er mode et par enquête après redressement	179
58	Répartition des sorties du domicile par motif principal et par enquête après redressement	180
59	Hypothèse de relations entre l'usage des nouvelles technologies, les caractéristiques socio-économiques et le comportement de mobilité des répondants	207
60	Relation entre l'âge des répondants et le nombre de déplacements moyen par enquête	210
61	Relation entre le nombre de personnes du ménage et le nombre de déplacements moyen par enquête	211
62	Relation entre le nombre d'enfants du ménage et le nombre de déplacements moyen par enquête	212
63	Relation entre la distance du domicile au centre de l'agglomération (km) et le nombre de déplacements moyen par enquête	214
64	Relation entre le nombre de voitures du ménage par personne de 18 ans et plus et le nombre de déplacements moyen par enquête	216
65	Relation entre la distance du domicile au centre de l'agglomération et la densité de la zone de résidence pour l'échantillon face-à-face	218
66	Relation entre la distance du domicile au centre de l'agglomération et la densité de la zone de résidence pour l'échantillon web	219
67	Relation entre le nombre d'enfants et le nombre de personnes du ménage	219
68	Relation bivariée entre le nombre de personnes, l'éloignement du domicile par rapport au centre et le nombre de déplacements des individus	238
69	Relation entre le genre des répondants et le nombre de déplacements (enquête face-à-face)	255
70	Relation entre l'activité des répondants et le nombre de déplacements par enquête	255
71	Relation entre la déclaration des revenus annuels du ménage et le nombre de déplacements par enquête	256
72	Relation entre la possession du permis de conduire ou d'un téléphone portable et le nombre de déplacements pour les répondants en face-à-face	257
73	Relation entre le taux d'immobiles et l'âge des répondants par enquête	258

74	Relation entre le taux d'immobiles et le nombre de personnes du ménage par enquête	258
75	Relation entre le taux d'immobiles et le nombre d'enfants du ménage par enquête	259

Liste des tableaux

1	Les quatre principales sphères d'activité	2
2	Les différentes méthodes de sélection de l'échantillon pour réaliser une enquête web	22
3	Accès à internet en France et au sein de l'Union européenne . .	28
4	Avantages et limites du média web pour la réalisation d'enquêtes	35
5	Comparaison des enquêtes en face à face et par web	40
6	Lieux de connexion des internautes français	47
7	Qualité de géobornage des adresses saisies par les répondants . .	78
8	Précision du calcul de distance	78
9	Vitesses moyennes de référence par mode en km	81
10	Taux de réponse global enquête face-à-face sur le périmètre du SCOT de Lyon	85
11	Taux de réponse enquête face-à-face par zone sur le périmètre du SCOT de Lyon	86
12	Les différentes raisons du choix du web comme média de réponse	88
13	Etude comparée des taux de réponse dans des enquêtes mixtes .	89
14	Taux d'abandon cumulé du questionnaire web	91
15	Comportement de réponse des répondants partiels	95
16	Répartition des répondants selon le nombre de non-réponses partielles au questionnaire	97
17	Occurrence des non-réponses aux questions proposant la modalité de réponse 'Ne sait pas' ou 'Non-réponse'	97
18	Evolution du nombre de déplacements selon le nombre de relances effectué	98
19	Durée de connexion globale du questionnaire web	100
20	Durée de connexion du questionnaire web par bloc de réponses (en mn)	101
21	Comportement de réponse selon le nombre de connexions	102
22	Equipement internet selon le nombre de connexions	103
23	Résultats possibles d'un test d'hypothèses	112
24	Taille moyenne des ménages	117
25	Statistiques descriptives de la variable 'Nombre de personnes' . .	122
26	Statistiques descriptives de la variable 'Revenus moyens par UC'	126
27	Répartition des personnes par lieu de travail habituel ou d'études et par enquête	134
28	Nombre moyen de déplacements quotidiens pour l'enquête web .	143

29	Nombre moyen de déplacements quotidiens pour l'enquête face-à-face	143
30	Nombre moyen de déplacements quotidiens par mode et par enquête	144
31	Nombre moyen de déplacements quotidiens par motif à la destination et par enquête	146
32	Distance moyenne des déplacements quotidiens calculée à vol d'oiseau par enquête	148
33	Durée moyenne des déplacements quotidiens calculée à vol d'oiseau par enquête	149
34	Budgets temps (mn) et distances (km) quotidiens des individus par enquête	149
35	Répartition des déplacements quotidiens par Origine-Destination pour l'enquête web	150
36	Répartition des déplacements quotidiens par Origine-Destination pour l'enquête face-à-face	151
37	Répartition des personnes par lieu de résidence et lieu habituel de travail ou d'études pour l'enquête web	151
38	Répartition des personnes par lieu de résidence et lieu habituel de travail ou d'études pour l'enquête face-à-face	152
39	Nombre moyen de sorties quotidiennes du domicile pour l'enquête web	152
40	Nombre moyen de sorties quotidiennes du domicile pour l'enquête face-à-face	153
41	Durée moyenne des sorties quotidiennes du domicile par enquête	153
42	Distance moyenne des sorties du domicile par enquête	154
43	Nombre moyen de sorties quotidiennes du domicile par 1er mode et par enquête	155
44	Nombre moyen de sorties quotidiennes du domicile par motif à la destination et par enquête	156
45	Les fonctions de distance disponibles	162
46	Modèle logit du média de réponse à l'enquête (web)	163
47	Effectifs des modalités des variables de calage	164
48	Nombre moyen de déplacements par enquête après redressement	171
49	Nombre moyen de déplacements par mode et par enquête après redressement	172
50	Nombre moyen de déplacements par motif à la destination et par enquête après redressement	174
51	Distance moyenne des déplacements par enquête après redressement	175
52	Durée moyenne des déplacements par enquête après redressement	176
53	Budgets temps et distances par enquête et par personne après redressement	176
54	Nombre de sorties pour motif 'Travail' et budget temps de travail moyen par actif et par enquête après redressement	176
55	Nombre moyen de sorties par enquête après redressement	177

Liste des tableaux

56	Durée moyenne des sorties par enquête après redressement . . .	178
57	Distance moyenne des sorties par enquête après redressement . . .	178
58	Nombre moyen de sorties du domicile par 1er mode et par enquête après redressement	180
59	Nombre moyen de sorties par motif à la destination et par enquête après redressement	181
60	Nombre de déplacements moyen par genre et par enquête	210
61	Nombre de déplacements moyen par niveau d'éducation et par enquête	212
62	Nombre de déplacements moyen par niveau d'activité et par enquête	213
63	Nombre de déplacements moyen par possession d'un téléphone portable et par enquête	215
64	Nombre de déplacements moyen par possession du permis de conduire et par enquête	215
65	Nombre de déplacements moyen effectué le vendredi par enquête	216
66	Nombre de déplacements moyen effectué par enquête, selon que l'individu déclare ou non son niveau de revenus	217
67	Régression linéaire appliquée à l'ensemble de l'échantillon (web et face-à-face)	220
68	Significativité de la régression linéaire appliquée à l'ensemble de l'échantillon (web et face-à-face)	220
69	Régression linéaire appliquée à l'échantillon face-à-face	221
70	Significativité de la régression linéaire appliquée à l'échantillon face-à-face	221
71	Régression linéaire appliquée à l'échantillon web	222
72	Significativité de la régression linéaire appliquée à l'échantillon web	222
73	Equation de sélection - Modèle probit	226
74	Indicateurs de la qualité du modèle de sélection	228
75	Probabilités de choix du mode de réponse	229
76	Analyse de la mobilité (Echantillon face-à-face)	231
77	Significativité de la régression utilisée dans la seconde étape (échantillon face-à-face)	231
78	Analyse de la mobilité (Echantillon web)	232
79	Significativité de la régression utilisée dans la seconde étape (échantillon web)	233
80	Modèle non contraint	235
81	Significativité du modèle non contraint	236
82	Modèle contraint	236
83	Significativité du modèle contraint	236
84	Modèle stable	237
85	Significativité du modèle stable	237
86	Distribution du nombre de déplacements par personne et par enquête	244

87	Statistiques descriptives des variables continues (échantillon face-à-face)	253
88	Statistiques descriptives des variables continues (échantillon web)	253
89	Statistiques descriptives des variables nominales (échantillon face-à-face)	254
90	Statistiques descriptives des variables nominales (échantillon web)	254
91	Modèle de Poisson simple - Echantillon face-à-face	260
92	Modèle de Poisson tronqué - Echantillon face-à-face	260
93	Modèle de Hurdle (binomial) - Echantillon face-à-face	261
94	Maximum de vraisemblance - Echantillon face-à-face	261
95	Significativité du modèle Hurdle - Echantillon face-à-face	261
96	Modèle de Poisson simple - Echantillon web	262
97	Modèle de Poisson tronqué - Echantillon web	263
98	Modèle de Hurdle (binomial) - Echantillon web	263
99	Maximum de vraisemblance - Echantillon web	263
100	Significativité du modèle Hurdle - Echantillon face-à-face	264
101	Mise en perspective, par échantillon, des déterminants de la décision de mobilité et de son intensité	265
102	Statistiques descriptives des variables continues (échantillon face-à-face)	339
103	Statistiques descriptives des variables continues (échantillon web)	339
104	Statistiques descriptives des variables nominales (échantillon face-à-face)	340
105	Statistiques descriptives des variables nominales (échantillon web)	341

Table des matières

Avant-propos	ii
Introduction générale : un renouvellement des besoins sur les données de mobilité	1
I Analyse de la mobilité : enjeux et état des lieux	1
I.1 Les composantes de la mobilité	2
I.2 L'Enquête Ménages Déplacements : une méthodologie éprouvée	2
II De nouvelles exigences méthodologiques	3
III Des répondants de plus en plus réticents	4
III.1 La difficulté de constituer une base de sondage	5
III.2 Une certaine lassitude des enquêtes	6
III.3 Des répondants au profil particulier	6
III.4 Une précision des données remise en cause	8
IV Le potentiel des nouvelles technologies	9
IV.1 Une évolution des modes de recueil de données	9
IV.2 L'essor des enquêtes web	10
V Le cadre de recherche	11
I La mise en ligne d'un questionnaire : fondements théoriques et empiriques	15
Introduction de la partie I	17
Chapitre 1 : Le potentiel du web pour les enquêtes ménages déplacements	18
I Apports et limites du web pour les enquêtes de mobilité	20
I.1 Une grande diversité d'enquêtes	20
I.2 Pourquoi un tel engouement pour les enquêtes web ?	22
I.2.1 Un coût marginal quasiment nul	23
I.2.2 L'interactivité du questionnaire	23
I.2.3 Des données de qualité	24
I.2.4 Des données collectées rapidement	25
I.2.5 Un mode moins contraignant	26
I.2.6 Le suivi du comportement de réponse	26
I.3 Des limites importantes	27

I.3.1	Le biais de couverture	27
I.3.2	Le biais d'échantillonnage	29
I.3.3	Le biais de non-réponse	30
I.3.4	Le biais de mesure	31
I.3.5	Les diverses sources d'incertitudes	32
I.3.6	La sécurité des données	33
I.4	Un avenir prometteur	34
II	Vers une méthodologie d'enquête mixte	36
II.1	Une diversité de protocoles d'enquête	36
II.2	Pour diminuer le taux de non-réponse	37
II.2.1	En limitant le niveau d'effort requis	38
II.2.2	En conservant la pertinence des réponses obtenues	39
II.3	Le web : un mode complémentaire du face-à-face ?	39
II.3.1	Deux modes aux atouts différents	40
II.3.2	Intérêt de la mixité du web et du face-à-face	41
II.3.3	Des problèmes de comparabilité	42
III	Les caractéristiques des internautes français	43
III.1	L'ordinateur, un bien de consommation largement diffusé	43
III.2	Une progression rapide des connexions internet dans les ménages français	45
III.2.1	Des français au bord de la route du haut débit	46
III.2.2	Quels accès hors du domicile ?	46
III.3	Des usages très variés, selon les caractéristiques des internautes	46
III.4	Vers une réduction des inégalités	48
IV	Conclusion	49

Chapitre 2 : La mise en place d'un module web dans l'enquête ménages déplacements de Lyon **51**

I	Considérations méthodologiques	52
I.1	Quelques recommandations	52
I.1.1	Un questionnaire accessible au plus grand nombre	52
I.1.2	Une introduction de qualité	53
I.1.3	Une présentation adaptée au média web	54
I.1.4	Une progressivité des questions	56
I.1.5	Limiter les non-réponses	57
I.1.6	Les sauvegardes partielles des réponses	58
I.1.7	Les filtres et tests de cohérence : personnalisation des questions et qualité des données recueillies	59
I.1.8	La mise en oeuvre du pré-test	59
I.2	Les principales difficultés rencontrées	62
I.2.1	Des interviews en face-à-face traditionnellement longues	62

	I.2.2	Un questionnaire standard relativement complexe	63
	I.2.3	Un seul répondant sélectionné dans le ménage	64
	I.2.4	Un cadre législatif précis	65
	I.2.5	Contraintes de réalisation	65
II		La rédaction des questions	66
	II.1	Présentation du questionnaire web	66
	II.2	La collecte d'informations sur les caractéristiques du ménage et du répondant	67
	II.2.1	Les caractéristiques générales du ménage	67
	II.2.2	Les caractéristiques du répondant	68
	II.2.3	La fréquence d'usage des modes de transport	68
	II.3	La collecte d'informations sur les déplacements individuels	69
	II.3.1	Une approche mixte déplacements-activités	69
	II.3.2	Le déplacement, une notion difficile à appréhender	70
	II.3.3	Le choix du mode et de l'activité	72
	II.3.4	La chronologie des déplacements	73
III		Administration du questionnaire en ligne	74
	III.1	L'importance de la prise de contact	74
	III.2	La codification des adresses	75
	III.2.1	Le recueil de données précises	75
	III.2.2	Le géobornage : des résultats mitigés	77
	III.2.3	Un calcul des durées et distances parcourues proche de la réalité	78
	III.3	La validation du questionnaire	79
	III.4	L'imputation des données	80
	III.4.1	Les valeurs aberrantes	80
	III.4.2	Les valeurs manquantes	80
	III.4.3	Les techniques d'imputation	80
IV		Conclusion	81

Chapitre 3 : Premiers résultats sur le comportement de réponse des internautes **83**

I		Un taux d'échec global à l'enquête en face-à-face non négligeable	84
	I.1	Le calcul du taux de réponse	84
	I.2	Estimation de la cible théorique de l'enquête web	85
	I.3	Des disparités spatiales	86
	I.4	Les raisons de la non participation à l'enquête en face-à-face	86
II		Une expérience encourageante	89
	II.1	Les taux de réponse constatés dans les protocoles d'enquêtes mixtes	89
	II.2	Une réactivité importante	90
	II.3	Un taux de réponse satisfaisant	90
	II.4	La conversion des non-répondants	92

III	Analyse de la non-réponse	93
III.1	Non-réponse totale et non-réponse partielle	93
III.2	Les causes d'abandon	94
III.3	Un taux de non-réponse partielle modéré	96
III.4	Des relances efficaces	98
IV	Caractéristiques des connexions web	99
IV.1	Des moments de connexions privilégiés	99
IV.2	Une durée de connexion étroitement liée au nombre de déplacements saisis	100
IV.3	Peu de connexions multiples	102
V	Conclusion	103
 Conclusion de la partie I		 106
 II Analyse comparative des échantillons web et face-à-face : quels enseignements ?		 107
 Introduction de la partie II		 109
 Chapitre 4 : Analyse comparative des répondants web et face-à-face		 110
I	Rappels théoriques sur les tests statistiques	112
I.1	Test du Khi-deux	113
I.2	Test de comparaison de deux proportions	114
I.3	Test de comparaison de deux moyennes	115
II	Qui sont les non-répondants à l'enquête en face-à-face ?	116
II.1	Des ménages de taille modeste, mais davantage motorisés	116
II.2	Les hommes, les jeunes et les non diplômés : des profils peu concernés par l'enquête en face-à-face	118
II.3	Les enseignements	118
III	Les ménages internautes : quelles spécificités ?	119
III.1	Des différences pas toujours significatives selon le lieu de résidence du ménage	119
III.2	Des ménages internautes de plus grande taille	122
III.3	Des ménages internautes très motorisés	123
III.4	Des revenus en moyenne plus élevés pour la population web	125
III.5	Des ménages web très équipés en moyens de communication	127
III.5.1	Un équipement en téléphonie fixe au domicile comparable	127
III.5.2	Des ménages web fortement connectés à internet	127
III.5.3	Des ménages web fortement équipés en téléphonie mobile	128
IV	Les répondants web : quel profil ?	129
IV.1	Une population plus féminine ?	129

IV.2	Une population dominée par les personnes en âge de travailler	130
IV.3	Des internautes majoritairement actifs	131
IV.4	Une sur-représentation des professions dites supérieures et des employés	131
IV.5	Un niveau d'étude élevé pour les internautes	132
IV.6	Une population fortement détentrice du permis de conduire	133
IV.7	Un lieu de travail habituel ou d'études plus fréquemment central	133
V	Conclusion	134
Chapitre 5 : Analyse comparative de la mobilité		136
I	Habitudes d'utilisation des modes de transport en semaine . . .	138
I.1	Une utilisation très fréquente de la voiture en tant que conducteur	138
I.2	Mais une utilisation moins fréquente de la voiture en tant que passager	139
I.3	Un usage plus modéré des transports collectifs	139
I.4	Une utilisation des deux-roues quasi-marginale	140
I.5	Conclusion sur les habitudes de déplacement	141
II	Caractéristiques des déplacements quotidiens	141
II.1	Un nombre de déplacements plus faible	141
II.2	Modes de déplacement : une utilisation plus fréquente de la voiture au détriment de la marche	143
II.3	Motifs de déplacements : un poids accru du motif 'Travail', au détriment du motif 'Etudes' principalement . . .	145
II.4	Heure de départ des déplacements : une pointe horaire plus marquée le matin et plus tardive le soir	147
II.5	Des déplacements plus longs en distance et en temps . .	148
II.6	Des échanges entre zones accrues, surtout avec le centre...	150
II.7	...dû à des localisations plus éclatées du domicile et du travail	151
III	Spécificités du web en termes de sorties quotidiennes	152
III.1	Des sorties moins fréquentes	152
III.2	Des sorties nettement plus longues en temps et en distance	153
III.3	Premier mode des sorties du domicile	154
III.4	Davantage de sorties 'Travail', au détriment des sorties 'Etudes', 'Loisirs' et 'Accompagnement'	155
IV	Conclusion	157
Chapitre 6 : Le redressement de l'échantillon en face-à-face : quels impacts sur la mobilité ?		159
I	Les aspects théoriques du calage sur marges	160
I.1	Principes de la méthode	160
I.2	Résolution	161
II	Application du calage sur marges à l'échantillon en face-à-face .	162
II.1	Sélection des variables de calage	162

II.2	Détermination des poids de calage	163
II.3	Des différences socio-économiques disparaissent	164
II.4	...alors que d'autres subsistent	165
II.4.1	Une présence moindre dans l'annuaire des abonnés France Télécom	165
II.4.2	Une propension plus élevée à déclarer leurs revenus	166
II.4.3	Une plus forte proportion de déplacements le vendredi	166
II.4.4	Conclusion sur le redressement de l'échantillon en face-à-face	167
III	Des comportements de mobilité spécifiques	167
III.1	Utilisation des modes de transport	167
III.1.1	Une utilisation très fréquente de la voiture en tant que conducteur	167
III.1.2	Mais une utilisation moins fréquente de la voiture en tant que passager	168
III.1.3	Un usage plus modéré des transports collectifs	169
III.1.4	Une utilisation des deux-roues peu fréquente, mais supérieure à celle de l'enquête en face-à-face	169
III.2	Des internautes moins mobiles	170
III.3	Une mobilité plus faible à pied	171
III.4	Motifs de déplacements : moins d'accompagnements et de loisirs	173
III.5	Pointe horaire du matin plus marquée et du soir plus tardive	174
III.6	Des déplacements plus longs en distance et en temps	175
III.7	Des journées plus longues pour les actifs	176
III.8	Un nombre de sorties du domicile plus faible	177
III.9	Des sorties plus longues en durée et en distance	178
III.10	Premier mode des sorties du domicile : une sous-estimation des sorties à pied chez les internautes	179
III.11	Motif principal des sorties du domicile : une prépondérance du motif 'travail'	180
IV	Conclusion	181
 Conclusion de la partie II		 185
 III La comparabilité des données : enjeux et méthodes de correction		 187
 Introduction de la partie III		 189

Chapitre 7 : Impact du mode d'enquête sur la mesure des com-		
portements de mobilité		190
I	La formalisation du problème	192
II	Le biais de sélection de l'échantillon	194
II.1	L'endogénéité du mode d'enquête au niveau de mobilité .	194
II.1.1	Un effet direct du mode d'enquête non observable	195
II.1.2	Application du modèle de Rubin	196
II.2	Les méthodes d'estimation disponibles	196
II.2.1	La sélection sur variables observables	197
II.2.2	La prise en compte de variables inobservables .	197
II.2.3	La méthode des variables instrumentales	197
II.3	La méthode en deux étapes	199
II.3.1	Développements théoriques	199
II.4	Test de l'existence significative d'un biais de sélection . .	203
II.4.1	Sens et intensité du biais de sélection	204
II.4.2	Conclusion : Les limites de la méthode en deux étapes	205
III	Modèle explicatif ne tenant pas compte du biais de sélection . .	206
III.1	Les variables disponibles pour l'analyse de la mobilité . .	206
III.1.1	Variables sociodémographiques	207
III.1.2	Variables caractéristiques de l'équipement en télécommunication	208
III.1.3	Variables caractéristiques de la mobilité	208
III.1.4	Variables caractéristiques du choix du mode de réponse	209
III.2	Analyse de la mobilité par un modèle de régression li- néaire multiple	209
III.2.1	Hypothèses sur la forme des variables explicatives	209
III.2.2	Le problème de la colinéarité des variables ex- plicatives	217
III.2.3	Présentation des résultats du modèle de régres- sion linéaire multiple	219
III.2.4	Premières conclusions	222
IV	Modèle explicatif incluant le biais de sélection	223
IV.1	Première étape : équation de sélection	223
IV.1.1	Le modèle probit	223
IV.1.2	Estimation du modèle de sélection des répon- dants	226
IV.1.3	Analyse de sensibilité	228
IV.2	Deuxième étape : équation d'intérêt	229
IV.2.1	Analyse de la mobilité pour l'échantillon en face-à-face	230
IV.2.2	Analyse de la mobilité pour l'échantillon web .	231
IV.2.3	Interprétation des coefficients des inverses du ratio de Mills	233
IV.3	Test de stabilité du modèle	234

IV.3.1	Comparaison d'un modèle contraint et non contraint	234
IV.3.2	Test du rapport des vraisemblances	234
IV.3.3	Application aux données de l'enquête	235
IV.3.4	Formulation d'un modèle stable	237
V	Quantification de l'impact du mode d'enquête sur la mobilité	238
V.1	Les différentes composantes de l'influence du mode d'enquête sur la mobilité	238
V.1.1	Impact direct du mode d'enquête	238
V.1.2	Impact des interactions entre le mode et certaines variables explicatives	238
V.2	Les profils types de répondants	239
VI	Conclusion	240

Chapitre 8 : Intérêt du modèle "Hurdle" pour la compréhension des comportements de mobilité 242

I	Pour poser le problème	244
II	Formalisation économétrique	246
II.1	La loi de Poisson	246
II.2	Le modèle de Poisson censuré	247
II.3	Modèle Hurdle	249
II.4	Evaluation de la pertinence du modèle Hurdle	250
III	Application à l'équation de mobilité quotidienne	251
III.1	Les variables disponibles pour l'analyse	251
III.1.1	Définition de l'immobilité	251
III.1.2	Description des variables	252
III.1.3	Statistiques descriptives	253
III.1.4	Hypothèses sur l'influence des différents facteurs	254
III.2	La population des répondants face-à-face	259
III.3	La population des répondants web	262
III.4	Synthèse	264
IV	Conclusion	265

Conclusion de la partie III 270

Conclusion générale : le web, un média d'avenir pour les enquêtes ménages déplacements 271

I	Portée et limites des choix méthodologiques	272
I.1	Un protocole d'enquête mixte : web et face-à-face	272
I.2	Le fort potentiel du web pour les enquêtes de mobilité	273
I.3	Nouveau média, nouvelles contraintes	274
II	Les principaux enseignements de l'approche empirique	275
II.1	Le profil des répondants web	276
II.2	Des comportements de mobilité très différents	276
II.3	Les interprétations possibles de ces résultats	278
III	L'intérêt de l'approche économétrique	279
III.1	L'identification du biais de sélection	279
III.2	La prise en compte des déterminants de l'immobilité	280

Table des matières

III.3	Les limites de notre cadre conceptuel	281
IV	Les perspectives de recherche	282
Bibliographie		285
Annexes		309
I	Evolution du nombre d'internautes en France	311
II	Evolution des inégalités d'accès aux nouvelles technologies . . .	312
III	Guide d'entretien semi-directif	313
IV	Verbatims	315
V	Lettre avis envoyée aux non-répondants à l'enquête standard . .	318
VI	Questionnaire web	319
VII	Vagues de courriers web	333
VIII	Répartition des répondants selon la réponse à la question : 'Souhaitez-vous communiquer votre adresse e-mail ?'	334
IX	Répartition des personnes de 10 ans et plus par enquête et par caractéristique socio-économique	335
X	Comparaison des marges dans l'échantillon et dans la population	336
XI	Répartition des individus par enquête après redressement de l'échantillon face-à-face	337
XII	Espérance d'une loi tronquée	338
XIII	Statistiques descriptives - personnes mobiles de 18 ans et plus .	339
Lexique		342
Table des figures		344
Liste des tableaux		349
Table des matières		353

Résumé : Les taux de réponse des enquêtes ménages déplacements tendent à décroître dans le temps, et il semble peu probable que cette tendance s'inverse à l'avenir. Ces dernières années, les méthodes utilisées pour recueillir les données de mobilité ont évolué, afin de prendre en compte d'une part la nécessité d'obtenir des informations fiables et suffisamment précises pour nourrir des modèles de plus en plus complexes et d'autre part l'intégration des nouvelles technologies dans les protocoles d'enquête (web, GPS...). La combinaison de différents médias s'est imposée comme un moyen d'améliorer la qualité des données produites à moindre coût, en permettant une augmentation du taux de réponse global. Mais la question de la comparabilité des données dans le temps et entre les différents modes reste entière. Nous avons proposé de mener une enquête web en parallèle de l'enquête ménages déplacements réalisée à Lyon en face-à-face en 2006. Le but était de demander aux individus qui ont refusé de recevoir un enquêteur à leur domicile ou qu'il n'a pas été possible de joindre durant la première vague d'interviews de répondre en ligne. Les deux principaux objectifs de cette recherche sont de tester la faisabilité d'une enquête web auprès des non-répondants et de comparer les données de mobilité des deux échantillons web et face-à-face. L'analyse montre que les internautes sont moins mobiles que les individus qui répondent en face-à-face (3,00 vs. 3,63 déplacements quotidiens), avec des différences plus ou moins marquées selon les modes et les motifs. Si les différences socio-économiques des répondants peuvent expliquer en partie cet écart (les internautes ont un profil particulier), l'hypothèse d'une sous-déclaration imputable au média web n'est pas à exclure. Pour approfondir l'analyse, il est nécessaire de mobiliser des techniques économétriques. Le modèle du biais de sélection, dont les paramètres sont estimés à l'aide de la procédure en deux étapes développée par Heckman semble prometteur, puisqu'il nous permet par exemple d'isoler l'effet des différences socio-économiques de celui lié au mode d'enquête sur la mobilité quotidienne.

Mots clés : Enquêtes ménages déplacements, questionnaire web, protocoles d'enquêtes mixtes, mobilité quotidienne, comparabilité des données, biais de sélection.

Abstract : Response rates for household travel surveys are tending to fall, and it seems unlikely that this trend will be reversed in the future. In recent years, travel data collection methods have evolved in order to obtain reliable data that are sufficiently detailed to feed increasingly complex models, and in order to integrate new technologies into survey protocols (Internet, GPS...). Combining different media is an obvious low-cost way of improving data quality as it increases the overall response rate. But the question of the comparability of data over time and between different survey modes remains unresolved. We initiated a project of a web survey in parallel of the household travel survey conducted in face to face in Lyon in 2006. The idea is to propose to those households who refuse to respond or are not contactable after a certain number of attempts to respond by the web. The two main objectives of this research are to test the feasibility of a web survey for non-respondents and compare mobility results of both survey modes. Our analysis shows that the Internet respondents travelled less than the face-to-face respondents (3.00 vs 3.63 daily trips), and that the differences between the two groups varied according to the travel mode and trip purpose. While part of this difference can be explained by socioeconomic disparities (the Internet respondents had a specific profile) we cannot exclude the possibility of under-reporting due to the web medium. To take analysis further, it is necessary to apply econometric techniques. The sample selection model, whose parameters must be estimated using the two-stage procedure developed by Heckman, provide a promising avenue as, for example, it allows us to isolate the effect on daily travel of socio-economic differences from that of survey mode.

Keywords : Household travel surveys, web surveys design, mixed modes surveys, daily travel behaviour, data comparability, selection bias.
