Développement de la surveillance observationnelle

Delphine Rieutort

▶ **To cite this version:**

Delphine Rieutort. Développement de la surveillance observationnelle. Santé publique et épidémiologie. Université Grenoble Alpes, 2015. Français. ⟨NNT : 2015GREAS015⟩. ⟨tel-01288921⟩

HAL Id: tel-01288921

https://theses.hal.science/tel-01288921v1

Submitted on 15 Mar 2016

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.

HAL Authorization

UNIVERSITÉ GRENOBLE ALPES

THÈSE

Pour obtenir le grade de

DOCTEUR DE L'UNIVERSITÉ GRENOBLE ALPES

Spécialité : **Modèles, Méthodes et Algorithmes pour la Biologie, la Santé et l'Environnement**

Arrêté ministériel : 7 août 2006

Présentée par

Delphine RIEUTORT

Thèse dirigée par **Dominique J. BICOUT** et
Codirigée par **Régis DE GAUDEMARIS**

Préparée au sein du **Laboratoire TIMC IMAG UMR CNRS 5525**
Equipe Environnement et Prédiction de la Santé des Populations
dans **l'École Doctorale Ingénierie pour la Santé, la Cognition et l'Environnement**

Développement de la Surveillance Observationnelle

Thèse soutenue publiquement le **12 novembre 2015**

devant le jury composé de :

Pr Olivier FRANCOIS
Professeur des universités, BCM/TIMC-IMAG, Président de jury
Pr Lode GODDERIS
Professeur associé, faculté de médecine Leuven, Rapporteur
Pr Christophe PARIS
Professeur de médecine, CHU et Inserm Nancy, Rapporteur
Dr Nathalie VILLA-VIALANEIX
Chargée de recherche, INRA de Toulouse, Examinatrice
Pr Gérard LASFARGUES
Directeur général adjoint scientifique, ANSES Maison-Alfort, Examinateur
Pr Régis de GAUDEMARIS
Professeur des universités, Professeur de médecine, EPSP/TIMC-IMAG et CHU Grenoble, Directeur
Dr Dominique J. BICOUT
Docteur en Physique, Chercheur, EPSP/TIMC-IMAG, VetAgroSup Lyon, Directeur

Remerciements

Je remercie en premier lieu les membres du jury, ***Lode Godderis*** et ***Christophe Paris*** en qualité de rapporteurs, pour leurs remarques et corrections apportées à ce manuscrit ; ainsi que ***Olivier François*** en qualité de président, et ***Nathalie Villa-Vialaneix*** et ***Gérard Lasfargues***, en qualité d'examinateurs, pour avoir accepté de m'accorder du temps pour l'évaluation de ce travail.

Je tiens bien évidemment à remercier mes deux directeurs de thèse : ***Dominique Bicout*** et ***Régis de Gaudemaris***.

- Tout d'abord, ***Régis***. Je ne vous remercierai jamais assez pour m'avoir un jour dit en bas de l'amphi du cours d'épidémiologie « Nous on en fait dans notre équipe de l'épidémiologie, envoie moi un CV et on en discute ». Si j'avais su où cette discussion me mènerait... Merci pour ces cours de Master 1 qui m'ont éveillé à la recherche scientifique ! Merci également pour votre soutien lors de mes stages et de ma thèse, soutien scientifique ou moral d'ailleurs. Je vous souhaite une très bonne retraite et j'espère ne pas avoir fait de l'encadrement des doctorants un trop mauvais souvenir à emporter avec vous !
- ***Dominique***, je ne sais par où commencer ces remerciements. Merci pour ta confiance en premier lieu. J'ai eu la chance avec toi de pouvoir apprendre l'autonomie, à chercher par mes propres moyens avant d'appeler au secours, et pourtant d'avoir quand même une oreille attentive lorsque j'en avais besoin. Nous n'avions pas les mêmes façons de procéder, toi d'un naturel optimiste, et moi d'un naturel stressé, et pourtant, il me semble que nous avons réussi à faire marcher le binôme, malgré les quelques frayeurs que j'ai pu avoir ! Merci également de m'avoir ouvert les portes de la modélisation et de la programmation. Merci encore pour ta disponibilité, ton accessibilité et ton soutien tout au long de ces années.

Je tiens également à remercier l'***ANSES*** et les différents ***acteurs du RNV3P*** qui m'ont permis aujourd'hui de travailler sur leur réseau et leur base de données. Merci notamment à ***Lynda Larabi*** pour ses coups de mains dans la compréhension de la base de données à mes débuts, ainsi que les nombreux coups de téléphones échangés pour les extractions de la base.

Merci aussi aux acteurs du ***groupe IDEWE*** qui m'ont également permis de travailler avec leur base, notamment ***Martijn Schoutenden*** avec qui nous avons passé de très bons moments lors de sa mission chez nous.

Je souhaite également remercier tous les membres du ***COST Modernet*** qui m'ont donné l'opportunité de présenter mes travaux à l'international et qui m'ont toujours accueilli avec une grande sympathie.

Ensuite, je tiens à remercier ***l'équipe EPSP*** qui m'accueille depuis 2011, pour mon premier stage de Master 1…

- Tout d'abord, merci au ***Pr Anne Maître***, pour m'avoir accueilli au sein de l'équipe toutes ces années. Merci également pour toutes les discussions (non) scientifiques que nous avons pu avoir lors des repas. Merci également pour les cakes et les viennoiseries… !

- Merci également à ***Sylvette Liaudy***, documentaliste et « Tata Sylvette » dû aux nombreux conseils que tu m'as donné depuis toutes ces années... Merci aussi pour mes relectures de biblio et mes recherches d'articles…

- Un remerciement tout particulier pour ***Franck Balducci***. Que dire, que dire ? Tellement de choses ! Je te connais depuis deux fois plus longtemps que toi tu me connais !! Et ces années n'auraient pas été les mêmes si je n'avais pas pu taper à la porte de ton bureau pour te poser des questions extrêmement-non-scientifiques… Merci pour ton humour (ou celui des blagues que tu m'envoies, au choix !), ta bonne humeur au quotidien (sauf quand il fait trop chaud parce que tu es grincheux), ton soutien quand j'en avais besoin, etc, etc, etc. Il y en a trop à dire, je vais donc résumer en un immense merci, et j'espère bien que tu n'oublieras pas la doctorante blonde et maladroite que je suis ! Merci vieux-dinosaure-gyver-sappeyard, à très bientôt… !

- Merci à toute ***l'équipe du CHU*** pour votre accueil dans vos locaux. Notamment à ***Vincent Bonneterre*** pour nos discussions lors des congrès (ou autre) ainsi que pour tes nombreux conseils avisés ! Un remerciement particulier également à ***Christine Buisson*** et ***Chantal Heyer*** qui ont toujours été d'une gentillesse à toutes épreuves. Et enfin merci à ***Julie, Vanessa et Anissa*** pour votre accueil durant notre expulsion provisoire !

- Merci à ***Damien Barbeau*** également, pour tes longues discussions dans le bureau 305 à l'époque où tu n'étais pas encore président du monde ; ainsi qu'à ***Renaud Persoons***, pour ta gentillesse et ton humour.

Enfin je voudrais remercier les différents ***stagiaires et doctorants*** qui sont passés, et parfois restés, au sein de l'équipe et du bureau 305 !

- Merci à mes stagiaires que j'ai eu la chance d'encadrer durant mes années de thèse : ***Alice*** et ***Héloïse***. Ainsi qu'un remerciement plus particulier à ***Oriane*** pour tous les bons moments passés au travail ou au bureau…

- Un remerciement et clin d'œil également pour les doctorants « mâles » de l'équipe, ***Pascal, Etienne*** et ***Simon***. Les gars vous prenez le relais alors courage ! Simon je t'attends pour une partie de Molkky !

- Un dernier remerciement pour la dernière arrivée *géographiquement* dans les locaux ! Quelle belle manière de t'introduire, ***Marie*** la géographe, *humour*… ! Merci beaucoup pour cette dernière année passée avec nous, pour ta bonne humeur, ton humour*, indéfinissable*, et ton soutien que ce soit pour le pro, ou le perso... Merci pour tout, et surtout, Check Da Pool and Hits your Allehuha ! Caus' uptown funk… Aaaaahh !…

Enfin, que seraient des remerciements sans les remerciements aux ***proches*** qui nous soutiennent au jour le jour ? Sans aucun ordre de préférence parce que c'est trop dur…

- Je tiens donc à remercier ma ***maman***. Merci pour ton soutien, pour l'intérêt que tu portes à mon travail, et la fierté dont tu fais preuve à mon égard, même si je ne la comprends pas !
- Merci également à ***Mwimwi*** (oui j'ai osé !). Je ne suis pas sûre d'arriver à te remercier avec le peu de lignes qu'il me reste, alors pour faire court, merci d'exister dans ma vie. Tu es la preuve vivante que l'adage « loin des yeux loin du cœur » n'est qu'une excuse pour les personnes qui ne veulent pas affronter la distance ! Love, Kiss, Flux, Flex.
- Merci à toi ***Tof.*** On s'est rencontré il y a maintenant 8 ans, j'étais alors encore jeune et insouciante (je suis toujours insouciante, jeune un peu moins…). Merci pour ton amitié qui traverse les années (tel un éclair qui traverse le ciel), merci pour tous les supers moments que l'on a pu passer ensemble, merci à ton humour décalé (et parfois énervant !), merci pour tous les services rendus (je ne parle pas de ceux de l'enseignement), merci merci ! Bon j'ai essayé de faire mes remerciements dans ton style, mais je ne suis vraiment pas douée pour l'humour débile, désolée… Signé : Sista.
- Merci à ***Sylvie*** également, tout d'abord secrétaire de l'équipe, puis fidèle amie au fur et à mesure de ces années… Merci de ta présence et de ton soutien. Merci pour tes coups de gueule aussi… Haha ! Merci également à ***Dan***, même si je ne sais pas pourquoi je te remercie en fait… Haha ! Allez, je vous aime les coupains ! Une p'tite coinche ?
- Merci aussi à ***Vivi***, depuis toutes ces années tu es toujours fidèle au poste ! Merci pour tes allers-retours pour venir nous voir, pour ta bonne humeur, pour ton soutien au fil des ans… Ce n'est pas grand-chose comme remerciement, mais merci pour tout, vraiment.

Enfin, mes derniers remerciements vont bien évidemment à mon futur… Merci à toi ***Anthony*** pour ton soutien indestructible durant toutes ces années, ta compréhension pour les moments difficiles et ta bonne humeur du quotidien. Il est très difficile de te remercier pour tout ce que tu m'apportes en quelques lignes, alors simplement, merci. Merci d'être entré dans ma vie, et d'avoir choisi d'y rester… Pomme de reinette et pomme d'api…

« Ils ont pas de bol, quand même !
Mettre au point un truc pareil et tomber sur des cerveaux comme nous ! »

Franck Pitiot, Kaamelott, Livre IV, Les Pisteurs, écrit par Alexandre Astier.

Sommaire

ABREVIATIONS 13

FIGURES 15

TABLEAUX 19

GLOSSAIRE 21

AVANT-PROPOS 23

PARTIE I : INTRODUCTION A LA SURVEILLANCE GENERALE 27

1. LA SURVEILLANCE 29

1.1. Définitions 29

1.2. Les types de surveillance 31

1.2.1. La surveillance passive 31

1.2.2. La surveillance active 31

1.2.3. La surveillance Syndromique 32

1.3. Méthodes utilisées pour la Surveillance Epidémiologique 33

1.4. Exemple d'un réseau de surveillance active en santé-travail : RNV3P 34

1.4.1. Missions 34

1.4.2. Fonctionnement du réseau 35

2. PROBLEMATIQUE ET OBJECTIFS 39

2.1. Objectifs 41

PARTIE 2 : DEVELOPPEMENT DE LA METHODOLOGIE DE LA SURVEILLANCE OBSERVATIONNELLE 43

3. SURVEILLANCE OBSERVATIONNELLE 47

3.1. Description d'une base de données observationnelles 47

3.2. Déploiement de la Surveillance Observationnelle 49

3.3. Préparation et construction des données 52

3.3.1. Sélection des modalités et observations significatives 52

3.3.2. Construction des nœuds et des motifs 57

3.4. Analyse descriptive 62

3.4.1. Echantillon 62

3.4.2. Description des motifs 63

3.5. Analyse spectrale 63

3.5.1. Exposome des nœuds 64

3.5.2. « Reporting Index » et statut dynamique ... *65*
3.5.3. Spectrosome des motifs ... *69*
3.5.4. Arbre de la multi-association ... *70*
3.5.5. Spécificité de l'information ... *71*
3.5.6. Spectre des motifs ... *74*
3.5.7. Matrice primaire-secondaire ... *75*
3.6. Récapitulatif des indices de la Surveillance Observationnelle ... 76
4. CREATION DE L'OUTIL « OSA » ... 79
4.1. Choix de la plateforme d'implémentation de la méthodologie ... 79
4.2. Organisation du programme ... 79
4.2.1. Description générale ... *80*
4.2.2. Description détaillée de la partie d'introduction ... *82*
4.2.3. Description détaillée de la partie de paramétrage ... *85*
4.2.4. Description détaillée de la partie de traitement ... *91*
4.2.5. Description détaillée de la partie des exportations des résultats ... *93*
5. MODE D'EMPLOI DE L'OUTIL « OSA » ... 95
5.1. Conditions d'utilisations ... 95
5.2. Présentation et ouverture de l'interface ... 96
5.3. Choix de l'analyse et paramétrage des indicateurs ... 98
5.3.1. Choix de l'analyse ... *98*
5.3.2. Paramétrage des indicateurs ... *100*
5.3.3. Déroulement de l'analyse et exportation des résultats ... *102*
5.4. Fichiers des résultats ... 103

PARTIE 3 : APPLICATIONS AU RNV3P ... 107

6. PRESENTATION DE LA BASE DU RNV3P ... 111
6.1. La base de données RNV3P ... 111
6.1.1. Structure d'une observation et d'un PST ... *111*
6.1.2. Nomenclature de standardisation ... *112*
6.1.3. Démarche qualité ... *113*
6.1.4. Evolution du système d'information en 2013 ... *113*
6.2. Mise en place de la Surveillance Observationnelle ... 114
6.3. Détermination des paramètres pour le RNV3P ... 115
6.3.1. Nombre de recrutement de référence et paramètre $\boldsymbol{\nu}$ ... *115*
6.3.2. Détermination de l'imputabilité numérique ... *116*
6.3.3. Le paramètre $\boldsymbol{\lambda min}$ ... *118*
6.4. Synthèse des paramètres d'application de la SO au RNV3P ... 119
7. CANCER DE LA VESSIE ... 121
7.1. Analyse ... 122

7.2. Conclusion 135
8. ASTHME 137
8.1. Analyse 137
8.2. Conclusion 150
9. LYMPHOME NON HODGKINIEN 151
9.1. Analyse 152
9.2. Conclusion 164
9.3. Comparaison des résultats 164
9.3.1. Article soumis dans la revue American Journal of Industrial Medicine 165
9.3.2. Comparaison des résultats 189

PARTIE 4 : APPLICATIONS A LA BASE BELGE IDEWE 193

10. PRESENTATION DU GROUPE ET DE LA BASE IDEWE 197
10.1. Restructuration pour l'application de la Surveillance Observationnelle 197
10.2. Déploiement de la Surveillance Observationnelle pour IDEWE 199
10.2.1. Prise en compte de la variable d'imputabilité 199
10.2.2. Détermination des paramètres pour IDEWE 201
11. DOULEUR A LA GORGE 203
11.1. Analyse 203
11.2. Conclusion 215
12. AIDES-SOIGNANTS 217
12.1. Analyse 217
12.2. Conclusion 225
13. AGRICULTEURS 227
13.1. Analyse 227
13.2. Conclusion 234

PARTIE 5 : DISCUSSION ET PERSPECTIVES 237
SYNTHESE ET DISCUSSION 241
PERSPECTIVES 249
CONCLUSION 255

RÉFÉRENCES 257

VALORISATION DES TRAVAUX 263

ANNEXES 285

Abréviations

ACP	Analyse en Composante Principale
ANSES	Agence Nationale de Sécurité Sanitaire de l'Alimentation, de l'Environnement et du Travail
ARS	Agence Régionale de Santé
CCPP	Centre de Consultation de Pathologie Professionnelle
CIM-10	Classification Internationale des Maladies, 10è édition
CIRC	Centre International de Recherche sur le Cancer (version française de l'IARC)
CITP	Classification Internationale Type de Professions
CMR	Cancérigène Mutagène Reprotoxique
HAP	Hydrocarbures Aromatiques Polycycliques
IARC	International Agency of Research on Cancer (version anglaise du CIRC)
ID	Numéro d'Identification
INCa	Institut National du Cancer
INRS	Institut National de la Recherche et de Sécurité
INSEE	Institut National de la Statistique et des Etudes Economiques
INSERM	Institut National de la Santé Et de la Recherche Médicale
INVS	Institut de Veille Sanitaire
ITAVI	Institut Technique de l'Aviculture
LNH	Lymphome non hodgkinien
MSA	Mutuelle Sociale Agricole
NAF	Nomenclature des Activités Françaises
OMS	Organisation Mondiale de la Santé
ONCFS	Organisation Nationale de la Chasse et de la Faune Sauvage
OSA	Observational Surveillance Analysis
OSCOUR	Organisation de la Surveillance Coordonnée des Urgences
PST	Problème de Santé au Travail
RAISIN	Réseau d'Alerte d'Investigation et de Surveillance des Infections Nosocomiales
RI	Reporting Index
RNV3P	Réseau National de Vigilance et de Prévention des Pathologies Professionnelles
RR	Risque Relatif
SAGIR	Surveiller les maladies de la faune sauvage pour Agir
SIG	Système d'Information Géographique
SO	Surveillance Observationnelle
SurSaUD	Surveillance Sanitaire des Urgences et des Décès
TMS	Troubles Musculo-Squelettiques
UE	Unité Ecologique

Figures

Figure 1 : Organisation du manuscrit 24
Figure 2 : Localisation des CCPP et nombre de cas enregistrés dans la base du RNV3P entre 2001 et 2014. 36
Figure 3 : Exemple de structure générale d'une base de données observationnelle. 47
Figure 4 : Sélection des informations nécessaires à la Surveillance Observationnelle à partir d'une base de données observationnelle. 49
Figure 5 : Influence de $\boldsymbol{\nu}$ sur la relation entre le paramètre λ et l'imputabilité *I*. 55
Figure 6 : Schéma d'illustration de la construction des motifs des modalités explicatives. 60
Figure 7 : Exemple du suivi dans le temps de recrutement et d'effectif cumulé de cas. 62
Figure 8 : Schéma de construction d'un exposome. 64
Figure 9 : Schéma d'illustration de la définition de la valeur seuil $\boldsymbol{\xi RI}$ 66
Figure 10 : Exemple de calcul de $\boldsymbol{RI2}$ et $\boldsymbol{RI\infty}$ pour un motif en 2005. 67
Figure 11 : Schéma d'illustration du statut pour chaque motif en fonction de $\boldsymbol{RI\infty}$ et $\boldsymbol{RI2}$. 68
Figure 12 : Schéma de l'arbre de la multi-association de la Surveillance Observationnelle. 70
Figure 13 : Illustration des configurations correspondant aux paramètres $\boldsymbol{Hmin}$ et $\boldsymbol{Hmax}$ et spécificité associée. 73
Figure 14 : Illustration de la spécificité des motifs. 73
Figure 15 : Schéma de la construction du spectre des motifs. 74
Figure 16 : Schéma de la construction de la matrice primaire-secondaire. 75
Figure 17 : Schéma général de l'application OSA. 81
Figure 18 : Schéma détaillé des scripts d'introduction de l'outil OSA. 83
Figure 19 : Schéma détaillé des scripts de paramétrage de l'analyse OSA. 87
Figure 20 : Schéma détaillé des scripts de traitement de l'analyse OSA. 92
Figure 21 : Schéma détaillé des scripts de traitement des sorties graphiques et des rapports de résultats de l'outil OSA. 94
Figure 22 : Tableur prêt à l'utilisation pour la Surveillance Observationnelle via l'outil OSA. 96
Figure 23 : Présentation de l'interface de l'outil OSA. 97
Figure 24 : Présentation des fenêtres de paramétrage de l'extraction et de l'analyse. 99
Figure 25 : Présentation des fenêtres de paramétrage des longueurs des nœuds et motifs et calcul de significativité. 101
Figure 26 : Présentation des fenêtres de progression et de graphiques supplémentaires. 102
Figure 27 : Exemple d'un rapport PDF standardisé via le compilateur LaTeX. 105
Figure 28 : Structure d'un Problème de Santé au Travail au sein de la base RNV3P depuis 2001... 112
Figure 29 : Variation de la p-valeur (moyenne des modalités) selon $\boldsymbol{\nu}$ et le nombre d'occurrences de la modalité. 115
Figure 30 : Comparaison des distributions des valeurs d'imputabilités selon la distribution initiale et les formules d'imputabilité numérique. 117

Figure 31 : Variation du paramètre $\boldsymbol{\lambda min}$ et seuil de significativité du taux de recrutement sous l'imputabilité nulle. 118
Figure 32 : Recrutement et effectif cumulé des cas, des nuisances distinctes et des codes pathologies distincts pour le cancer de la vessie de 2001 à 2014. 122
Figure 33 : Exposome du cancer de la vessie en 2014. 123
Figure 34 : Caractérisation des motifs issus de l'exposome du cancer de la vessie en 2014. 127
Figure 35 : Spectrosome des motifs du cancer de la vessie en 2014. 129
Figure 36 : Spectre du cancer de la vessie en 2014. 130
Figure 37 : Exemple de trajectoire atypique de motifs identifiés pour le cancer de la vessie de 2001 à 2014 pour chaque catégorie de statut dynamique : persistant, en déclin, émergent et latent. 132
Figure 38 : Exemple de trajectoire dynamique de trois motifs d'ordre I, HAP, amiante et trichloroéthylène, et leur association d'ordre III, associés au cancer de la vessie, de 2001 à 2014. 134
Figure 39 : Recrutement et effectif cumulé des cas, des nuisances distinctes et des codes pathologies distincts pour l'asthme de 2001 à 2014. 138
Figure 40 : Exposome de l'asthme en 2014. 139
Figure 41 : Statut dynamique des 438 motifs de l'asthme en 2014. 142
Figure 42 : Arbre de la multi-association des 87 motifs actifs liés à l'asthme en 2014, dont 68 motifs d'ordre I actifs et 19 motifs d'ordre supérieurs actifs. 144
Figure 43 : Spectrosome des motifs de l'asthme en 2014. 145
Figure 44 : Spectre de l'asthme en 2014. 147
Figure 45 : Exemple de trajectoire atypique de motifs identifiés pour l'asthme de 2001 à 2014 pour chaque catégorie de statut dynamique : persistant, en déclin, émergent et latent. 148
Figure 46 : Exemple de trajectoire dynamique de deux motifs d'ordre I, poussière de farine, et acarien/mite/gale, et leur association d'ordre II, associés à l'asthme, de 2001 à 2014. 149
Figure 47 : Recrutement et effectif cumulé des cas, des nuisances distinctes et des codes pathologies distincts pour le LNH de 2001 à 2014. 152
Figure 48 : Exposome du LNH en 2014. 153
Figure 49 : Caractérisation des motifs issus de l'exposome du LNH en 2014. 156
Figure 50 : Spectrosome du LNH en 2014. 157
Figure 51 : Spectre du LNH en 2014. 159
Figure 52 : Matrice primaire-secondaire des motifs et des activités professionnelles associées au LNH en 2014. 160
Figure 53 : Exemple de trajectoire atypique de motifs identifiés pour le LNH de 2001 à 2014 pour chaque catégorie de statut dynamique : persistant, en déclin, émergent et latent. 162
Figure 54 : Exemple de trajectoire dynamique de deux motifs d'ordre I, solvants, diluants et amiante, et leur association d'ordre II, associés au LNH, de 2001 à 2014. 163
Figure 55 : Illustration de la structure initiale et restructurée de la base belge IDEWE. 198

Figure 56 : Recrutement et effectif cumulé des cas, des nuisances distinctes et des codes pathologies distinctes correspondant aux douleurs à la gorge et aux autres pathologies enregistrées au sein des mêmes observations de 2010 à 2013. 203
Figure 57 : Exposomes des douleurs de la gorge en 2013. 205
Figure 58 : Caractérisation des motifs issus de l'exposome des douleurs à la gorge en 2013. 210
Figure 59 : Spectrosome des motifs des douleurs à la gorge en 2013. 212
Figure 60 : Spectre des douleurs de la gorge en 2013. 213
Figure 61 : Exemple de trajectoire atypique de motifs identifiés pour les douleurs à la gorge de 2010 à 2013 pour chaque catégorie de statut dynamique : persistant, en déclin, émergent et latent. .. 214
Figure 62 : Recrutement et effectif cumulé des cas, des nuisances distinctes et du code d'activité professionnelle correspondant aux aides-soignants de 2010 à 2013. 217
Figure 63 : Exposome des aides-soignants en 2013. 218
Figure 64 : Caractérisation du statut dynamique des motifs issus de l'exposome des aides-soignants en 2013. 220
Figure 65 : Spectrosome des motifs des aides-soignants en 2013. 222
Figure 66 : Spectre des aides-soignants en 2013. 223
Figure 67 : Exemple de trajectoire atypique de motifs identifiés pour les aides-soignants de 2001 à 2013 pour chaque catégorie de statut dynamique : persistant, en déclin, émergent et latent. .. 224
Figure 68 : Recrutement et effectif cumulé des cas, des nuisances distinctes et du code d'activité professionnel correspondant aux agriculteurs de 2010 à 2013. 227
Figure 69 : Exposome des agriculteurs en 2013. 228
Figure 70 : Caractérisation du statut dynamique des motifs issus de l'exposome des agriculteurs en 2013. 230
Figure 71 : Spectrosome des motifs des agriculteurs en 2013. 231
Figure 72 : Spectre des agriculteurs en 2013. 232
Figure 73 : Exemple de trajectoire atypique de motifs identifiés pour les agriculteurs de 2010 à 2013 pour chaque catégorie de statut dynamique : persistant, en déclin, émergent et latent. 233
Figure 74 : Répartition de la distribution des résultats obtenus pour chaque $\Delta \boldsymbol{t}$ pour le cancer de la vessie selon les variations des paramètres $\boldsymbol{\lambda min}$ et $\boldsymbol{\nu}$. 243
Figure 75 : Illustration de la Surveillance Observationnelle appliquée au RNV3P à une exposition professionnelle : les Hydrocarbures Aromatiques Polycycliques. 246
Figure 76 : Illustration de la Surveillance Observationnelle dans le contexte de l'influenza aviaire dans la région de la Dombes. 252

Tableaux

Tableau 1 : Exemple des différents systèmes de surveillance à l'échelle nationale en France. 32
Tableau 2 : Valeurs de $\boldsymbol{\lambda min}$ en fonction du nombre de recrutement de référence critique $\boldsymbol{n\Delta t, cr}$ pour une modalité enregistrée avec l'imputabilité la plus faible. 56
Tableau 3 : Tableau de contingence du calcul du coefficient de φ pour deux modalités $\boldsymbol{H1}$ et $\boldsymbol{H2}$. 71
Tableau 4 : Variables et paramètres considérés pour l'application de la Surveillance Observationnelle à la base RNV3P. 119
Tableau 5 : Liste des 51 motifs actifs du cancer de la vessie identifiés par la Surveillance Observationnelle en 2014. 124
Tableau 6 : Liste des 87 motifs actifs de l'asthme identifiés par la Surveillance Observationnelle en 2014. 140
Tableau 7 : Liste des 40 motifs du LNH identifiés par la Surveillance Observationnelle en 2014. 154
Tableau 8 : Exemple d'une distribution d'imputabilité externe, allant de 0 à 3, renseignée par *n* spécialiste. 200
Tableau 9 : Variables et paramètres considérés pour l'application de la Surveillance Observationnelle à la base IDEWE..... 202
Tableau 10 : Synthèse des résultats de la Surveillance Observationnelle suivant la méthode pessimiste ou optimiste de l'attribution de la valeur d'imputabilité *a posteriori.* 204
Tableau 11 : Liste des 105 motifs identifiés par la Surveillance Observationnelle avec calcul « Pessimiste » (P) de l'imputabilité appliquée aux douleurs de la gorge, incluant les 78 motifs obtenus avec la version « Optimiste » (O) du calcul de l'imputabilité. 206
Tableau 12 : Liste des 44 motifs actifs des aides-soignants identifiés par la Surveillance Observationnelle en 2013. 219
Tableau 13 : Liste des 28 motifs actifs des agriculteurs identifiés par la Surveillance Observationnelle en 2013. 229
Tableau 14 : Résumé des différents résultats de la Surveillance Observationnelle appliquée au cancer de la vessie du RNV3P selon les variations des paramètres $\Delta\boldsymbol{t}$, $\boldsymbol{\lambda min}$ et $\boldsymbol{\nu}$. 242
Tableau 15 : Libellés des 12 unités écologiques étudiées dans l'étude en Dombes en 2008. 251

Glossaire

Cluster	Agrégats d'observations partageant toutes en commun la (ou les) même(s) modalité(s) explicative(s) primaire(s).
Cortège	Ensemble des modalités des variables explicatives primaires enregistrées pour une observation.
Exposome	Initialement définit comme le réseau de la totalité des expositions d'un individu, le terme définit ici un réseau relationnel des observations connectées entre elles si elles partagent au moins une modalité des variables explicatives primaires en commun.
Imputabilité	Information permettant d'attribuer une force d'association entre l'évènement de santé et les variables explicatives primaires.
Modalité	Correspond aux différentes valeurs distinctes que peut prendre une variable au sein d'une base de données.
Motif	Modalité ou combinaison de modalités représentative des variables explicatives primaires permettant de caractériser l'évènement de santé étudié.
Nœud	Ensemble d'observations partageant toutes en commun le même cortège de modalités des variables explicatives primaires.
Observations	Correspond à une ligne enregistrée au sein d'une base de données.
Ordre	Nombre de modalités présente au sein des cortèges de motifs : Ordre I, une modalité, Ordre II, deux modalités, etc.
Script	Ensemble de commandes et de fonctions utilisées au sein du logiciel R pour la création de l'interface et le déroulement de l'analyse OSA.
Spectre	Signature de l'évènement de santé ciblé caractérisé par les motifs représentatifs identifiés par la Surveillance Observationnelle.
Spectrosome	Réseau relationnel des motifs connectés entre eux s'ils partagent au moins une modalité de leur cortège en commun.
Variable	Colonne contenant une information spécifique pour chaque observation enregistrée au sein d'une base de données. De type quantitatif (exprime un nombre, une valeur, une quantité), ou qualitatif (représente une valeur nominale).
Variable cible	Variable à expliquer, correspondant à l'évènement de santé ciblé, et qui sera utilisée pour l'extraction d'un sous-échantillon pour la SO.
Variable explicative primaire	Variable explicative servant à la caractérisation de la variable cible via la création des nœuds et des motifs.
Variable explicative secondaire	Variable explicative permettant de générer la matrice primaire-secondaire.

Avant-propos

Les impacts de la population sur l'environnement et ceux de l'environnement sur la santé humaine sont à présent bien avérés, et représentent depuis les années 2000 de réels enjeux de Santé Publique. Il a été constaté ces dernières années des augmentations indéniables des prévalences de cancer, de maladies respiratoires, de troubles de la reproduction, mais également une constante dégradation de la qualité de l'air et de l'eau dans son milieu naturel. La thématique de Santé-Environnement a alors émergée au cœur de la société, à l'échelle nationale et mondiale, et notamment en France, avec le développement de départements spécifiques au sein des agences sanitaires ; ainsi que des plans nationaux tel que le Plan National de Santé Environnement lancé en 2004, dont la 3ème édition a été engagée pour les années 2015-2019. Cette prise de conscience collective a également permis l'émergence de nombreux systèmes de surveillance dont le but est d'améliorer la connaissance vis-à-vis des évènements sanitaires et environnementaux.

Cette surveillance, épidémiologique ou non, est devenue aujourd'hui un outil indispensable et déterminant dans les choix politiques de Santé Publique. Elle permet en effet d'évaluer l'état de santé des populations, d'identifier des groupes vulnérables ou à risque, de déceler l'apparition de processus pathologiques afin d'en étudier le développement dans le temps et l'espace, en vue de l'adoption de mesures appropriées de prévention ou de lutte. De nombreux réseaux de surveillance se sont alors développés, donnant naissance à de nombreuses bases de données, dont les utilisations restent encore pour le moment assez limitées.

Le terme de surveillance est un terme générique très large, incluant des actions de surveillance à un niveau local jusqu'à des systèmes de surveillance nationaux ou mondiaux. Ainsi, il existe de nombreuses définitions, des processus et des méthodologies d'analyse diverses, allant d'une analyse descriptive simple à des méthodologies plus complexes telles que la modélisation ou les systèmes d'information géographiques.

L'objectif de ce travail dans ce contexte est de pouvoir proposer une nouvelle méthodologie adaptée à la prise en compte optimale des bases de données dans un but de surveillance des évènements de santé.

Ce travail de thèse est organisé en **3 grandes parties** (Figure 1) :

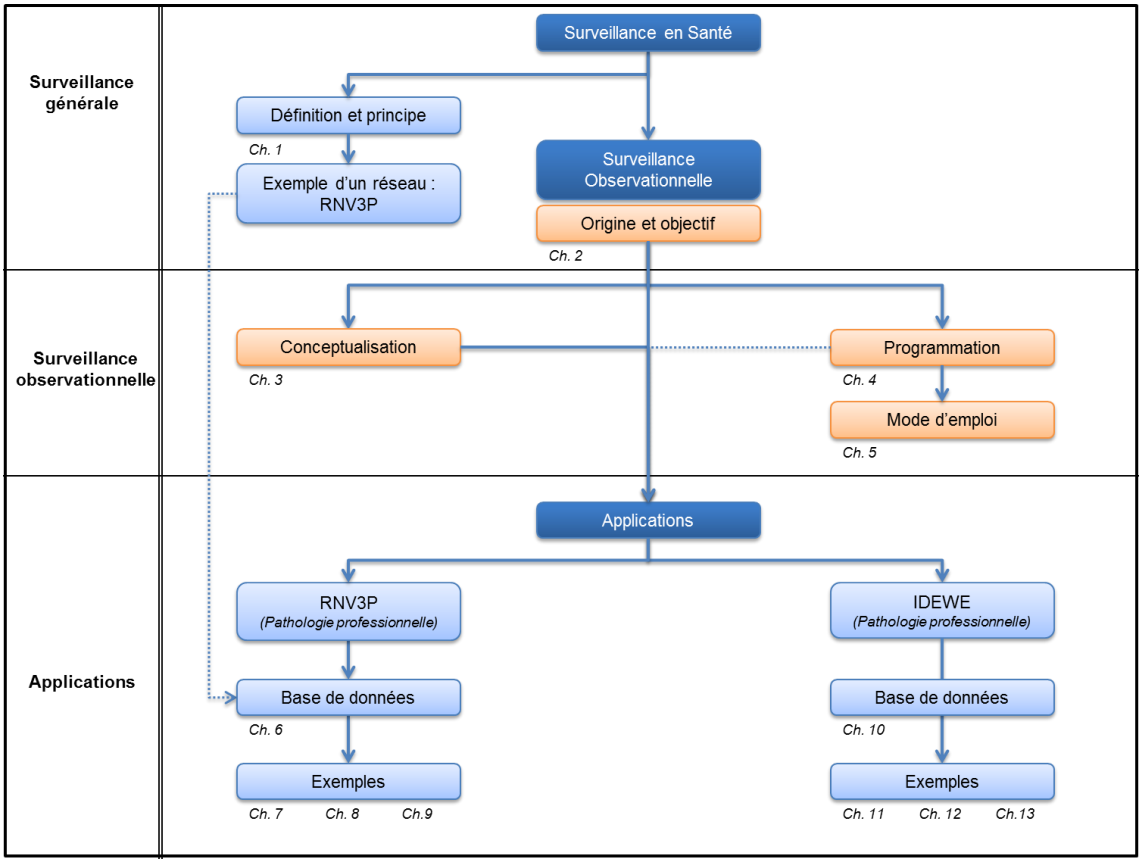


Figure 1 : Organisation du manuscrit

- Une partie introductive présentant la Surveillance Générale en santé **(Partie I)**.

Cette partie consiste en la définition de la Surveillance en santé ainsi que la présentation d'un réseau de surveillance active des maladies professionnelles *(Chapitre 1)* et la présentation des problématiques et objectifs sous-jacents au développement de la Surveillance Observationnelle *(Chapitre 2)*.

- Une partie consacrée au déploiement de la Surveillance Observationnelle **(Partie II)**.

Un chapitre est consacré à la conceptualisation théorique de la Surveillance ainsi qu'à la présentation des indicateurs associés développés *(Chapitre 3)*. Suivi d'un chapitre consacré à la présentation de l'intégration de la méthodologie sous la plateforme R, décrivant ainsi les différents processus développés afin de permettre une utilisation en routine de la méthodologie *(Chapitre 4)*. Enfin, une présentation de l'interface développée sous R et un mode d'emploi de cet outil sont présentés dans un dernier chapitre *(Chapitre 5)*.

- Une partie consacrée à la démonstration de la Surveillance Observationnelle au travers de différentes applications, permettant la démonstration de la portabilité de la méthodologie. Pour des raisons de structure, cette partie a été scindée en deux **(Parties III et IV)**.

La **partie III** est consacrée aux exemples d'applications à la base du RNV3P avec une présentation de la base de données *(Chapitre 6)* ainsi que la présentation de 3 exemples de surveillance de maladies professionnelles, différentes en termes d'expositions professionnelles :

- le ***cancer de la vessie***, pathologie chronique de forme stable à latence longue, dont les principales étiologies sont connues et reconnues au titre des maladies professionnelles *(Chapitre 7)* ;
- l'***asthme***, pathologie aigue à latence courte dont les principales étiologies sont connues et reconnues au titre des maladies professionnelles *(Chapitre 8)* ;
- le ***lymphome non hodgkinien***, pathologie chronique à distribution de latence (moyenne à longue), dont les étiologies sont mal définies et pour laquelle de nombreuses hypothèses existent *(Chapitre 9)*. Les résultats seront alors comparés avec une étude de la littérature.

La **partie IV** est consacrée à des exemples d'applications appliquées à la base de données professionnelles Belge IDEWE avec une présentation du groupe et de leur base de données *(Chapitre 10)* ainsi que trois exemples d'applications de la surveillance :

- la surveillance des expositions professionnelles en lien avec les ***douleurs à la gorge*** permettant la comparaison avec les exemples d'applications du RNV3P *(Chapitre 11)* ;
- la surveillance des pathologies associées à l'activité professionnelle des ***aides-soignants*** *(Chapitre 12)* afin de présenter l'adaptabilité de la méthodologie à différentes problématiques ;
- la surveillance des pathologies associées à l'activité des ***agriculteurs*** *(Chapitre 13).*

- Enfin une dernière partie sera consacrée à la discussion et aux perspectives de la méthodologie développée **(Partie V).**

Partie I : Introduction à la Surveillance Générale

« Un statisticien est une personne qui peut avoir la tête dans un four et les pieds pris dans la glace et dire qu'en moyenne il se sent bien »

Anonyme

1. LA SURVEILLANCE

La surveillance, aujourd'hui outil indispensable à la prise de décision par les pouvoirs publiques, s'inscrit dans des thématiques de Santé Publique et Environnementale pour lesquelles la santé des populations et son interaction avec l'environnement sont au cœur des préoccupations.

1.1. Définitions

En 1920, Charles Edward Winslow définissait déjà la Santé publique telle que :

"Public health is the science and the art of preventing disease, prolonging life, and promoting physical health and efficiency through organized community efforts for the sanitation of the environment, the control of community infections, the education of the individual in principles of personal hygiene, the organization of medical and nursing service for the early diagnosis and preventive treatment of disease, and the development of the social machinery which will ensure to every individual in the community a standard of living adequate for the maintenance of health." [Winslow, 1920]

La Santé Publique était alors un concept dont le but principal était la gestion de la santé avec une prévention des maladies, un assainissement de l'environnement, une éducation sur l'hygiène personnelle, ainsi qu'une meilleure organisation des professionnels de santé et des dispositifs sociaux. La *nouvelle* Santé Publique prend à présent en charge toutes les dimensions administratives, sociales, politiques et économiques, de la santé. Elle repose en effet sur une meilleure compréhension de la façon dont les styles de vie et les conditions de vie déterminent la santé, notamment les facteurs sociologiques et environnementaux. Cette Santé Publique tend vers une prévention à la fois d'un point de vue des sciences du vivant et de celui des sciences humaines et sociales [Coquin, 2011].

La Santé-Environnement est un cas particulier de la Santé Publique. En 1989, l'OMS la définissait telle que :

« La santé environnementale comprend les aspects de la santé humaine qui sont déterminés par les facteurs de notre environnement. Elle concerne également la théorie et les pratiques d'évaluation et de contrôle des facteurs environnementaux susceptibles d'affecter la santé » [Salines et Eilstein, 2011]

La santé environnementale inclue alors aussi bien les agents chimiques, biologiques que l'environnement physique. Ainsi la santé-environnement peut-être applicable dans l'étude des effets des modifications de l'environnement général liées aux activités humaines, par exemple l'impact de la pollution, de la radioactivité industrielle, du rejet de déchets ou de pesticides qui impacteraient la qualité de l'eau, de l'air et des aliments. Elle peut également être appliquée dans l'étude de l'impact de l'environnement naturel, tel que les variations climatiques, le rayonnement solaire, les catastrophes naturelles ou l'émission de radon. Des organismes spécialisés appliquent quant à eux la Santé-Environnement dans un environnement spécifique, comme le milieu du travail par exemple. Et enfin, la Santé-Environnement peut inclure de manière large les effets sur la santé entraînés par des comportements individuels de la population, comme le tabagisme ou l'alcoolisme [Salines et Eilstein, 2011].

Dans ce contexte de préoccupations de la santé des populations et face aux différentes menaces liées à l'environnement pouvant impacter sur l'état de santé, la surveillance épidémiologique est définit comme :

« Un processus systématique de collecte, d'analyse et d'interprétation des données sur des évènements de santé spécifiques importants pour la planification, la mise en œuvre et l'évaluation des pratiques en santé publique, étroitement associées à leur juste diffusion à ceux qui ont besoin d'être informés » [Astagneau, 2011].

La surveillance a donc pour objectif de suivre l'état de santé ou les facteurs de risque d'une population définie, et en particulier de déceler l'apparition de processus pathologiques et d'en étudier le développement dans le temps et dans l'espace en vue de l'adoption de mesures appropriées de lutte. Les systèmes de surveillance sont alors des outils d'aide à la décision dans le domaine de la prévention et du contrôle des maladies et des évènements sanitaires émergents.

1.2. Les types de surveillance

Il existe un grand nombre de systèmes de surveillances adapté à l'évènement sanitaire spécifique étudié telle que la surveillance de la qualité de l'air, de la radioactivité, le plan canicule, la surveillance des décès hospitaliers, la pharmacovigilance, la toxicovigilance et autres. Malgré la diversité des systèmes, on distingue en général trois grands types de surveillance : la surveillance **passive**, la surveillance **active**, et la surveillance **syndromique**.

1.2.1. La surveillance passive

La surveillance passive est une surveillance qui repose sur la déclaration spontanée de cas ou suspicions de cas de l'évènement de santé étudié par les acteurs du système de surveillance. Cette surveillance prend source dans des données déjà existantes, le plus souvent recueillies au cours d'activité médicale ou dans des bases de données administratives. Elle est le plus souvent encadrée par un système législatif et ne provient pas d'une déclaration volontaire. Les systèmes les plus connus de ce type de surveillance sont les Centres Nationaux de Références [InVS, 2012a], pour l'observation des maladies transmissibles, et le système des maladies à Déclaration Obligatoire [InVS, 2012b] qui impose aux médecins de déclarer aux autorités les cas de maladies faisant partie de la liste définie par la réglementation (Tableau 1). Ce système de surveillance est adapté aux situations pour lesquelles il est nécessaire d'assurer une alerte précoce, en cas d'apparition ou de réapparition d'une maladie par exemple.

1.2.2. La surveillance active

La surveillance dite active ou programmée repose quant à elle sur une collecte active des données organisée spécifiquement par une structure désignée, encadrée ou non par un système législatif ou une agence sanitaire. Généralement, le système repose sur le volontariat des différents acteurs. Ces systèmes de surveillance active sont généralement utilisés par les réseaux de surveillance afin de couvrir une vaste zone géographique. Il en existe dans différents domaines tels que les infections nosocomiales avec le réseau RAISIN (Réseau d'alerte, d'investigation et de surveillance des infections nosocomiales) [InVS, 2013], la grippe, la varicelle, le zona, la borréliose de Lyme, etc., avec le réseau Sentinelle [INSERM et InVS, 2015], les zoonoses avec le réseau SAGIR (Surveiller les maladies de la faune sauvage pour Agir) [ONCFS, 2014] ou les maladies professionnelles avec le réseau RNV3P (Réseau National de Vigilance et de Prévention des Pathologies Professionnelles) [Bonneterre, 2009]

(Tableau 1). Cette surveillance est en général plus couteuse qu'une surveillance passive et dépend beaucoup de l'implication des membres du réseau.

1.2.3. La surveillance Syndromique

La surveillance syndromique est un type de surveillance plus récent, mis en place dans une volonté de prévention précoce d'évènements sanitaires émergents. Ce système est caractérisé par une absence d'objectifs spécifiques de recueil des données. Cette stratégie est en réalité construite autour de toutes les données disponibles recueillies ou non dans un but de surveillance. Les intervenants de ces systèmes de surveillance syndromique sont assistés par des procédures d'enregistrement automatique des données qui permettent la mise à disposition des données pour le suivi et l'analyse épidémiologique en temps réel ou proche réel. Ce système permet de générer des informations au jour le jour sur la santé des populations et de détecter des évènements inhabituels plus précocement qu'avec des méthodes de surveillance classique. Un exemple caractéristique de cette surveillance est le réseau SurSaUD (Surveillance Sanitaire des Urgences et des Décès) (Tableau 1) [InVS, 2015] qui centralise de manière quotidienne les informations issues des services d'urgence participant au réseau des urgences OSCOUR (Organisation de la Surveillance Coordonnée des Urgences), mais également les informations issues des consultations de SOS Médecins et enfin des données de mortalité par l'intermédiaire de l'INSEE (Institut National de la Statistique et des Etudes Economiques), afin de permettre une analyse fine des causes de décès et la génération d'alerte en cas d'évènement sanitaire inhabituel détecté.

Tableau 1 : Exemple des différents systèmes de surveillance à l'échelle nationale en France.

Système	Type	Domaine cible	Rattachement*
Centre nationaux de référence	Passif	Maladies transmissibles	InVS
Déclaration obligatoire	Passif	Liste de 31 maladies à signaler (botulisme, brucellose, chikungunya, dengue…)	ARS et InVS
Phyt'attitude	Actif	Signalement volontaire de symptomes par les agriculteurs	MSA
RAISIN	Actif	Maladies nosocomiales	InVS
RNV3P	Actif	Pathologies professionnelles et environnementales	ANSES
SAGIR	Actif	Santé de la faune, zoonose	Office National de la Chasse et de la Faune Sauvage
Sentinelle	Actif	Acte suicidaire, diarrhée aiguë, maladie de Lyme, oreillons, syndromes grippaux, urétrite masculine, varicelle et zona	Institut Pierre Louis d'Epidémiologie et de Santé Publique, de l'Inserm et de l'Université Pierre et Marie Curie, en collaboration avec InVS
SurSauD	Syndromique	Urgences et décès	InVS

**ANSES : Agence Nationale de la Sécurité Sanitaire de l'alimentation, de l'environnement et du travail*
ARS : Agence régionale de Santé ; InVS : Institut de Veille Sanitaire ; MSA : Mutuelle Sociale Agricole

1.3. Méthodes utilisées pour la Surveillance Epidémiologique

La surveillance repose sur 4 principes fondamentaux qui sont : la **collecte des données**, qui doit respecter des règles de confidentialité, d'éthique et de qualité selon le type de données recueillies ; **l'analyse de ces données** ; l'**interprétation** et enfin **la diffusion des résultats**. En terme d'analyse, les stratégies sont différentes et adaptées aux objectifs du système de surveillance. Dans certains cas de figures, l'utilisation de tests statistiques ou d'indicateurs épidémiologiques classiques tels que les valeurs d'incidence et de prévalence seront suffisantes, tandis qu'il sera parfois nécessaire de recourir à des méthodes plus complexes.

De nombreuses études se basent par exemple sur des analyses de séries temporelles des données dans un but de description, d'explication et de prévision. Une série temporelle est une série d'observations numériques (mesures) indicées par le temps. La collecte des données repose sur de multiples sources permettant le croisement des données. Cependant, elle est tributaire de problèmes de données manquantes, d'erreur de codification et de standardisation ou de différences de définitions entre les différentes sources des données et l'évolution des systèmes. Cette analyse essentiellement graphique, passe donc par plusieurs étapes de traitement. Une fois ces traitement effectués, les analyses de tendance et de régression peuvent être appliquées afin de répondre aux différentes hypothèses posées, comme par exemple une modification dans le temps de survenue de cas d'une maladie.

Les méthodes d'analyses spatiales sont également aujourd'hui couramment utilisées dans la surveillance épidémiologique. En effet, il a rapidement été mis en évidence l'importance de la localisation des populations dans la survenue d'évènement de santé. Les années 1990 ont permis l'essor de ces techniques grâce aux outils de Systèmes d'Information Géographiques (SIG) permettant alors de cartographier et de croiser des données de santé avec différentes informations telles que les données territoriales, administratives et démographiques. L'atout de ces méthodes est de pouvoir proposer une représentation visuelle de l'état de santé des populations et des facteurs potentiellement à risque pour cette santé, mais également de pouvoir appliquer des méthodes de géostatistiques. Les outils comme les études distance, de centralité, de diffusion ou de polarisation, permettent de mettre en évidence des caractéristiques spatiales non visibles par un simple croisement de bases de données. Ces méthodes peuvent également être couplées à de la modélisation afin de prédire par exemple la diffusion des polluants à partir d'une source, connaissant la typologie du terrain, la météorologie et la localisation des populations.

Enfin, la modélisation est également une discipline largement répandue dans la surveillance de la santé des populations. Elle a pour but de traduire un phénomène en « langage mathématique » afin de faciliter d'une part la description de ce phénomène, mais d'autre part la détection et la prédiction de phénomène émergent. Il est ainsi difficile de définir un champ d'action particulier puisqu'elle peut être appliquée à la prédiction de la diffusion des polluants, mais également à la surveillance des maladies transmissibles ou des zoonoses, à l'étude de l'implantation de vecteurs zoonotiques, à l'évaluation des risques toxicologiques pour l'humain, aux études budget espace-temps, ou encore au domaine de l'écologie et de l'écotoxicologie.

1.4. Exemple d'un réseau de surveillance active en santé-travail : RNV3P

Dans son rapport d'activité de 2011, l'assurance maladie avait recensé près de 50 000 maladies professionnelles par an depuis 2001. Une maladie « professionnelle » étant définie comme la conséquence directe à l'exposition à un risque physique, chimique, biologique d'un individu, ou résultant des conditions dans lesquelles il exerce son activité professionnelle [INRS, 2012]. Près d'un tiers des salariés auraient été exposés à au moins un produit chimique cancérogène au cours de la semaine précédant l'enquête SUMER en 2010 [Arnaudo, 2012]. De plus, on estime la part de cancer attribuable à des expositions professionnelles entre 4 et 8.5%, ce qui représente en France entre 14 000 et 30 000 nouveaux cas par an (sur les 355 000 nouveaux cas de cancers estimés en 2012), dont la moitié correspondrait à des cancers à létalité élevée [INCa, 2014]. La surveillance dans l'environnement de travail représente des enjeux importants des Plans Santé-Environnement et Plan Cancer, et c'est dans cette volonté de surveillance des conditions de travail des travailleurs que le réseau RNV3P a été mis en place à l'échelle nationale.

1.4.1. Missions

Le Réseau National de Vigilance et de Prévention des Pathologies Professionnelles (RNV3P) est un réseau expert de **Médecins du travail et de Spécialistes Hospitalo-Universitaires volontaires**, soutenu par l'Agence Nationale de Sécurité Sanitaire de l'Alimentation, de l'environnement et du travail (ANSES), enregistrant de manière systématique et standardisée des cas de Problèmes de Santé au Travail (PST). Ce réseau d'experts, créé en 2001, est né d'un besoin de mise en commun de leurs compétences et de leurs expériences dans le cadre de

l'amélioration des conditions de travail et a permis la création d'une base de données nationale pérenne regroupant tous les PST rencontrés à l'échelle nationale [Bonneterre, 2009].

Il possède 3 missions majeures :

- Activité de **veille épidémiologique active** qui consiste à suivre au cours du temps l'évolution de couples pathologies-expositions prédéfinis
- Activité de **veille clinique** qui permet aux partenaires du RNV3P, face à une nouvelle association pathologie-exposition(s), de rechercher la présence d'autres cas à l'échelle nationale et d'émettre un signalement pour une vigilance renforcée auprès des autres médecins du réseau
- Une **activité de recherche** pour élaborer des outils de mise en évidence d'associations pathologies-expositions potentiellement émergentes, notamment par la :
    - Surveillance des PST
    - Détection des PST potentiellement émergents
    - Prévention des menaces liées à l'activité professionnelle sur la santé

1.4.2. Fonctionnement du réseau

Les différents acteurs du réseau sont répartis à l'échelle nationale dans les Centres de Consultation de Pathologies Professionnelles (CCPP) des 31 Centres Hospitalier Universitaires de France (Figure 2). Les patients reçus en consultation sont adressés par une majorité de médecins du travail afin d'établir un avis d'expert sur le possible lien de causalité entre la pathologie diagnostiquée et une situation professionnelle particulière, comme par exemple une exposition à un produit Cancérigène-Mutagène-Reprotoxique (CMR) pour un cancer, ou un mouvement répété pour des Troubles Musculo-Squelettiques (TMS) ; mais peuvent également être adressés par des médecins généralistes ou spécialistes.

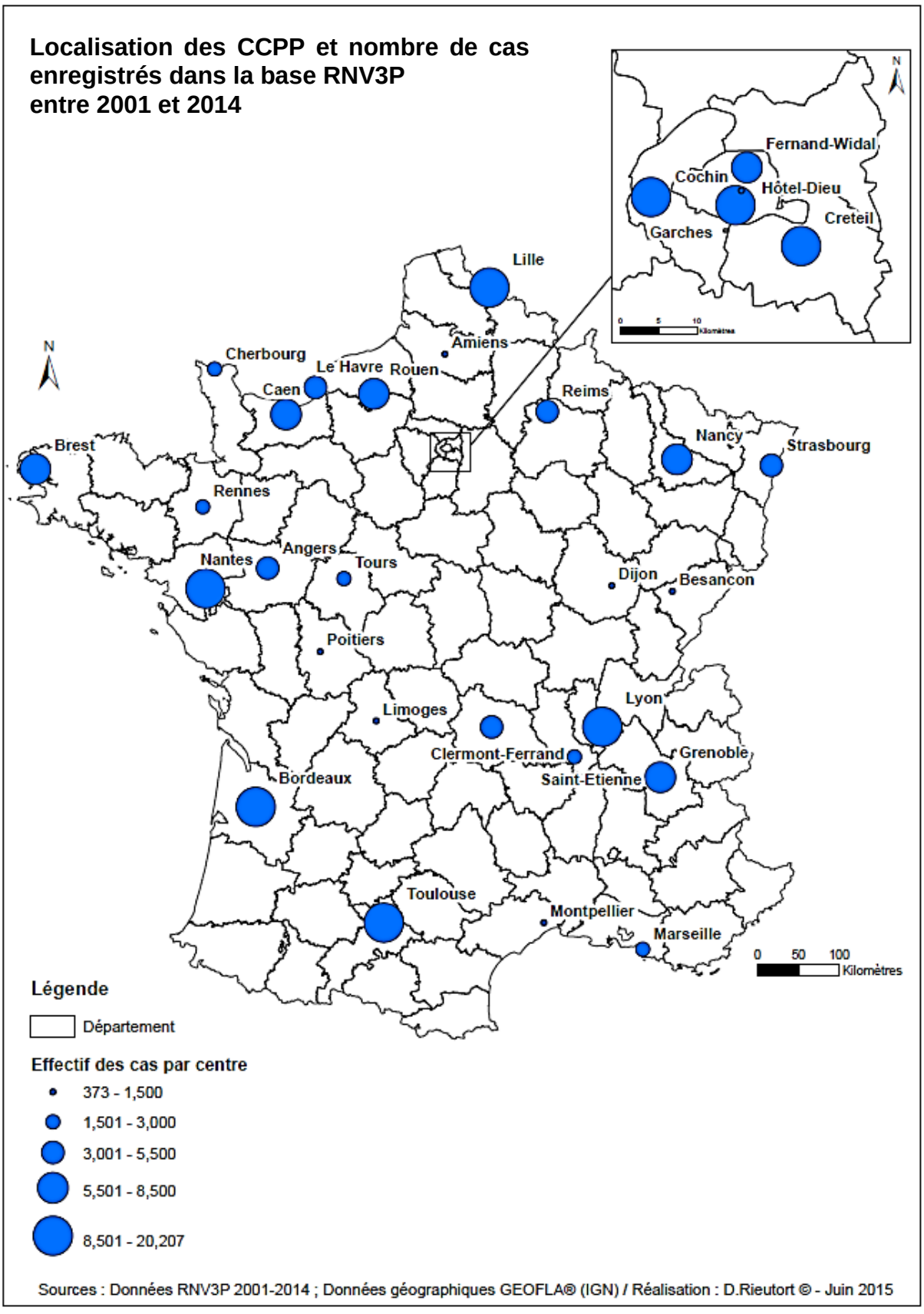


Figure 2 : Localisation des CCPP et nombre de cas enregistrés dans la base du RNV3P entre 2001 et 2014.

La majorité des patients s'adressant au RNV3P sont atteints de pathologies concernant des Tumeurs (code de la Classification internationale des Maladies-10 : C00-D48, 12.32%), des troubles mentaux et du comportement (F00-F99, 15.33%), des maladies de l'appareil respiratoire (J00-J99, 16.11%), des maladies de la peau et du tissu cellulaire sous-cutané (L00-L99, 8.79%) et des maladies du système ostéo-articulaire, des muscles et du tissu conjonctif (M00-M99, 15.18%).

Lors d'une consultation professionnelle, le médecin consultant effectue un « interrogatoire professionnel » du patient afin de retracer un *Cursus Laboris*, c'est-à-dire historique complet de son environnement professionnel.

Il commence par reconstituer le parcours professionnel en listant toutes les entreprises dans lesquelles le patient a exercé une activité professionnelle, ainsi que les différents postes qu'il a effectués au sein de ces entreprises. Le médecin expert va ensuite, à partir du *Cursus Laboris* détaillé du patient, dresser l'inventaire des expositions professionnelles, appelée nuisances, afin d'identifier les différents processus, agents chimiques, biologiques ou physiques, auxquelles le travailleur aurait pu être exposés directement ou indirectement, et qui seraient potentiellement en relation avec la pathologie diagnostiquée. Le médecin va ensuite « pondérer » ces liens par la recherche d'équipement de protection individuel ou collectif, par la durée d'exposition ainsi que la fréquence, afin d'estimer au mieux le réel risque encouru par le travailleur. Le médecin consultant doit également prendre connaissance des facteurs environnementaux hors profession afin de vérifier des potentiels facteurs de confusion ou d'interaction. Il sera ensuite nécessaire au médecin de comparer la chronologie des expositions avec la chronologie de l'apparition des symptômes et de la maladie afin de valider ou réfuter les potentielles causes mises en évidence. Au cours de l'établissement du *Cursus Laboris*, le médecin expert peut demander si nécessaire différents tests biologiques ou prélèvements et mesures sur le lieu de travail afin de confirmer ou infirmer des hypothèses potentiellement causales. En fonction des résultats, il pourra statuer sur le dossier concernant la causalité professionnelle et pourra, ou non, entamer des demandes d'indemnisation du patient.

Ce travail d'investigation est d'autant plus difficile qu'il est relativement influencé par différents paramètres intrinsèques, tel que l'état de santé parfois « dégradé » du patient ; la mémoire altérée, notamment dans le cas de latence longue entre l'exposition et la déclaration

de la pathologie ; la méconnaissance des produits utilisés au cours de la carrière ; ou la cessation des symptômes après un arrêt de travail dans le cas de pathologie à latence courte. Il est également très lié à la subjectivité et la connaissance du médecin. Un médecin spécialiste de la pathologie n'aura probablement pas le même *a priori* vis-à-vis de la maladie par rapport à un médecin non spécialiste. De même, un médecin ayant déjà rencontré le couple pathologie-nuisance en question orientera probablement plus l'interrogatoire dans ce sens. Cette subjectivité est un paramètre important à prendre en compte car peut engendrer une variabilité importante au sein du réseau pour une même association pathologie-nuisance.

Une fois l'interrogatoire terminé, un dossier est établi et traité informatiquement afin de codifier et standardiser les différentes informations avant leur intégration dans la base nationale (*Voir Partie 3* : *Applications au RNV3P, Chapitre 6, p.111*).

En 2014, près de 200 000 patients ont été enregistrés dans la base, dont environ 50% ont été considérés comme étant probablement ou potentiellement associés aux conditions de travail. La base de données obtenue peut ensuite être utilisée à des fins de surveillance (dite active) et de prévention des risques professionnels.

2. PROBLEMATIQUE ET OBJECTIFS

Les réseaux de surveillance s'étant multipliés ces dernières années pour répondre aux besoins des pouvoirs publiques, les données se sont également accumulées et il n'est plus rare de rencontrer des bases de données observationnelles contenant des milliers d'informations diverses et variées, et pour la plupart, jamais ou peu utilisées à des fins de surveillance. On parle parfois de « big data » ou « **données massives** », désignant un ensemble de données très volumineux, dont l'analyse de manière optimale via les outils classiques de gestions de données, statistiques ou épidémiologie devient difficile à mettre en place. L'enjeu que représentent ces bases de données en Santé Publique pour la surveillance et la prévention des menaces pour les populations est pourtant considérable. En effet, l'utilisation de ces bases de données de manière optimale, considérant toutes les informations recueillies, pourrait donner un nouvel essor à la surveillance ainsi que de nouvelles pistes de travail et de recherche. Des évènements sanitaires jusque-là silencieux ou spécifiques pourraient être mis en évidence par le simple fait de **l'utilisation d'une méthodologie adaptée**.

Cependant, étant donné le nombre de réseaux de surveillance existants, on trouve à présent des bases de données de tous types, de toutes tailles, ayant des objectifs, des méthodologies, des standardisations et des temporalités différentes. Certaines bases de données, dites « observationnelles », comprennent uniquement des données de déclarations, et il est donc difficile de trouver une méthodologie adaptée pour une utilisation optimale, et qui pourrait s'adapter à ces différents types de bases de données, sans considération de la « cible » de la surveillance. De plus, il est devenu évident que les évènements sanitaires liés à l'environnement et touchant les populations sont pour le plus souvent **d'origines multifactorielles**, pouvant agir en synergie ou alors comme facteur aggravant. Ces interactions entre les différents facteurs environnementaux est difficile à prendre en compte avec les méthodologies classiques qui ont pour habitude de cibler un comportement ou un évènement d'intérêt précis. **La question se pose alors de la prise en compte de la multi-information en tant qu'outil de veille sanitaire, et dans quelle mesure cette prise en compte serait possible.**

C'est dans cette dynamique d'amélioration de la prise en compte **de l'environnement global et de la multi-information** que la notion d'exposome, proposé par Christopher Wild en 2005 [Wild, 2005] s'est rapidement diffusée au sein de la communauté scientifique. Presque simultanément, en 2007 Bicout et al. développaient le concept d'Exposome comme un réseau

connecté d'expositions [Faisandier, 2007]. Concept issu de la science des réseaux et de la théorie des graphes, l'approche par exposome a pour objectif de prendre en compte la totalité des expositions à des facteurs environnementaux que rencontre un individu dès sa naissance, dans le but d'étudier les relations entre les expositions enregistrées et les potentielles menaces pour sa santé. Cette notion a par la suite été reprise sous différentes appellations adaptées aux sujets de recherches, comme le « diseasome » ou « disease human network » de Goh [Goh, 2007] (mettant en relation des pathologies humaines grâce aux mutations génétiques communes) ou encore le projet récent « HELIX » destiné à fournir le « early-life exposome » [Vrijheid, 2014] afin de mesurer les risques encourus durant la grossesse et les premières années de l'enfant et leur impact sur le développement futur. Il a été mis en évidence que cette approche propose un réel bénéfice pour l'investigation des causes environnementales des maladies chroniques [Rappaport, 2011], car elle a l'avantage de considérer les données comme un réseau, permettant de mettre rapidement en évidence les relations existantes et les caractéristiques particulières de l'évènement étudié grâce à la typologie des réseaux : clusters, densité, coefficient d'agglomération (« clustering »), et autres. La méthodologie de l'exposome semble donc être une méthodologie prometteuse, notamment pour la prise en compte d'évènements multiples, mais difficile à mettre en place car nécessitant une connaissance parfaite de la vie d'un individu.

2.1. Objectifs

L'objectif principal de ce travail est d'utiliser la notion de l'exposome précédemment développé dans les travaux exploratoires de Laurie Faisandier sur le RNV3P [Faisandier, 2007, Faisandier, 2011a, Faisandier, 2011b] afin de développer une nouvelle méthodologie de surveillance : la **Surveillance Observationnelle**. Cette surveillance, intégrant la notion d'exposome comme point de départ, a comme objectif de pouvoir être applicable et **généralisable à différentes bases de données, différents évènements d'intérêts** (de santé) mais surtout de pouvoir prendre en compte **les origines multifactorielles** des évènements de santé surveillés afin de générer des **hypothèses et signaux d'alertes statistiques** destinés aux acteurs des réseaux.

Les **objectifs** sont alors définis comme :

- Le **développement des bases théoriques** d'une surveillance et des indicateurs associés, adapté à une exploitation optimale des bases de données, notamment des bases de données de type observationnelle, généralisable, et tenant compte des origines multifactorielles des évènements d'intérêts ;

- L'implémentation de cette méthodologie dans **un outil permettant une utilisation en routine et systématique de la surveillance**, adaptable à différentes bases de données sanitaires pour répondre à différentes problématiques de santé publique.

Le travail présenté dans ce manuscrit correspond à l'utilisation de la méthodologie de l'exposome comme point de départ du développement de la Surveillance Observationnelle applicable à différentes bases de données et son application dans différentes problématiques de Santé-Environnement.

Partie 2 : Développement de la méthodologie de la Surveillance Observationnelle

« La connaissance s'acquiert par l'expérience, tout le reste n'est que de l'information. »

Albert Einstein

L'objectif de ce chapitre est de développer et de poser les bases d'une nouvelle méthodologie de surveillance : **la Surveillance Observationnelle**. Partant d'un travail exploratoire sur l'exposome développé dans de précédents travaux [Faisandier, 2007, Faisandier, 2011a, Faisandier, 2011b], une approche a été développée afin de permettre une exploitation optimale des bases de données de type « observationnelles » (ou de déclaration) permettant la prise en compte de la temporalité ainsi que de la multiplicité des informations disponibles. Cette surveillance permet le suivi dans le temps des différentes informations et caractéristiques associées à un évènement d'intérêt au sein d'une base de données.

Le principal objectif de ce développement est de proposer des indicateurs génériques adaptables à différents contextes, et à différentes bases de données, permettant l'identification et le suivi temporel de facteurs de « risques » associés à un évènement d'intérêt cible. La Surveillance Observationnelle consiste donc en 3 étapes majeures :

- L'**identification** des associations de facteurs de « risques » associés à l'évènement d'intérêt ou la variable étudié ;
- L'**analyse descriptive** de l'échantillon et des associations identifiées ;
- L'**analyse spectrale** comprenant la dynamique des associations identifiées et l'analyse des changements de leur structure.

Cette partie a pour but de présenter la conceptualisation théorique de la Surveillance Observationnelle par la présentation des différents processus de construction et calcul des indicateurs ; puis de présenter l'implémentation complète de la méthodologie dans la plateforme R jusqu'à la création d'un outil et d'une interface d'utilisation permettant l'application de la Surveillance Observationnelle.

3. SURVEILLANCE OBSERVATIONNELLE

3.1. Description d'une base de données observationnelles

Les bases de données appelées « observationnelles » sont des bases de données pérennes, issues d'une surveillance active, et recueillant un nombre important d'informations suite à des observations effectuées sur le terrain. C'est le cas, par exemple, des bases de données RNV3P et Phyt'attitude qui sont considérées comme bases de données observationnelles. Une base de données observationnelle est donc caractérisée par une diversité d'informations variables selon les évènements d'intérêt (santé) enregistrés. Par la suite, il sera utilisé le terme évènement de santé pour les évènements d'intérêt.

La Figure 3 illustre le contenu possible d'une base de données observationnelle enregistrée dans un format tableur (type Excel) (Figure 3). En lignes sont indiquées les ***observations*** correspondant aux différents enregistrements des évènements de santé au sein de la base et en colonnes sont indiquées toutes les diverses informations relatives à l'enregistrement sous forme de ***variables*** quantitatives ou qualitatives.

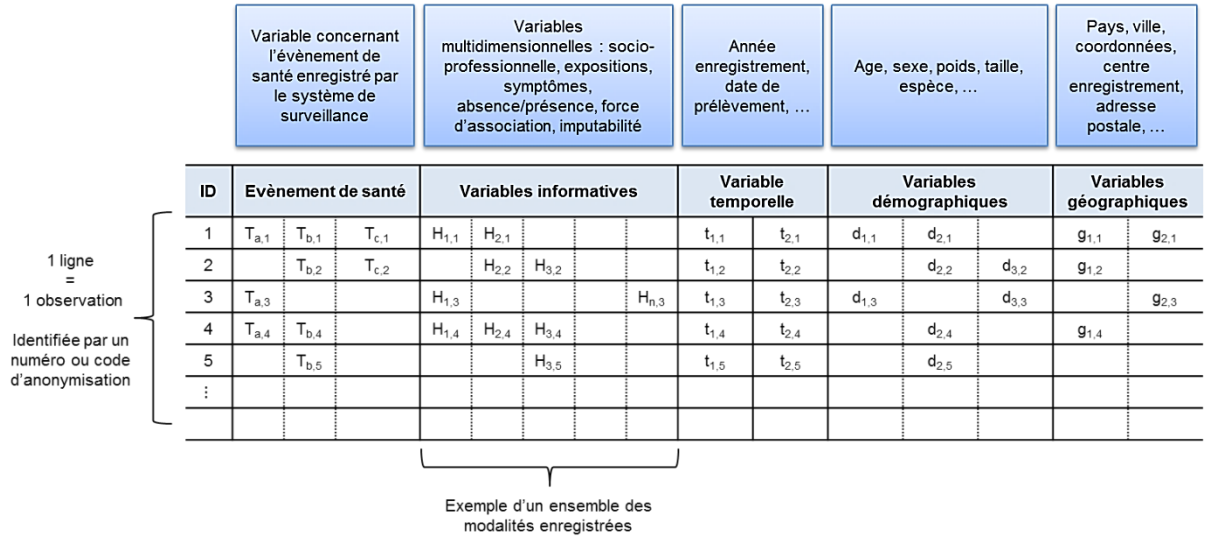


Figure 3 : Exemple de structure générale d'une base de données observationnelle.

On retrouve généralement dans un tel tableur :

- une variable d'**identification** (« ID »), ou d'anonymisation, des observations enregistrées ;

- des **variables relatives à l'évènement de santé** qui peuvent s'étendre sur plusieurs colonnes. Par exemple, il est possible d'enregistrer plusieurs pathologies ou plusieurs symptômes d'un même patient (au sein de la même observation) ou plusieurs types de prélèvements effectués ;

- des **variables apportant des informations complémentaires** à l'évènement de santé, telles que la catégorie socio-économique d'un ou groupe d'individus, les expositions d'un individu à des sources de polluants, la présence ou l'absence d'un individu dans un contexte particulier, des symptômes associées à une pathologie. Ces variables sont généralement standardisées et codées par des nomenclatures spécifiques.

  Il est également possible de rencontrer des variables permettant d'attribuer une force d'association entre la variable de l'évènement de santé et les variables associées, ces variables seront alors appelées ***imputabilité*** ;

- des **variables relatives à la temporalité**, telles que l'année d'enregistrement au sein de la base ou la date de prélèvement, etc ;

- des **variables démographiques** permettant d'identifier l'individu par son âge, son genre ou son espèce ;

- des **variables géographiques**, telles que le pays, la ville, des codes postaux, des coordonnées GPS, les centres d'enregistrement ou d'analyse.

Pour chaque groupe de ces variables qualitatives, il est enregistré différentes « mentions », par exemple différents symptômes, espèces ou pathologies, qui sont alors appelés les « ***modalités*** ». Les modalités correspondent aux différentes informations ou valeurs distinctes que peut prendre une variable.

> *Par exemple :*
> Pour la variable du « genre » dans une base de données de santé humaine, les modalités seront « homme » et « femme » ; pour une variable climatique, les modalités rencontrées pourraient être « pluie », « nuageux », « ensoleillé », etc.

Les bases de données sont donc caractérisées par des **observations en lignes**, chacune formée par un **ensemble d'informations regroupées en colonnes dans différentes variables**, et dans chacune de ces variables sont identifiées **différentes modalités**.

3.2. Déploiement de la Surveillance Observationnelle

La mise en œuvre de la Surveillance Observationnelle (SO) nécessite de sélectionner (choisir) et extraire deux éléments au moins à partir de la base de données ; une (ou plusieurs) variable(s) cible(s) et un ensemble de variables associées déclinées en modalités ; le but de la SO étant d'étudier et d'analyser la structure de l'association entre la variable cible et ses variables associées.

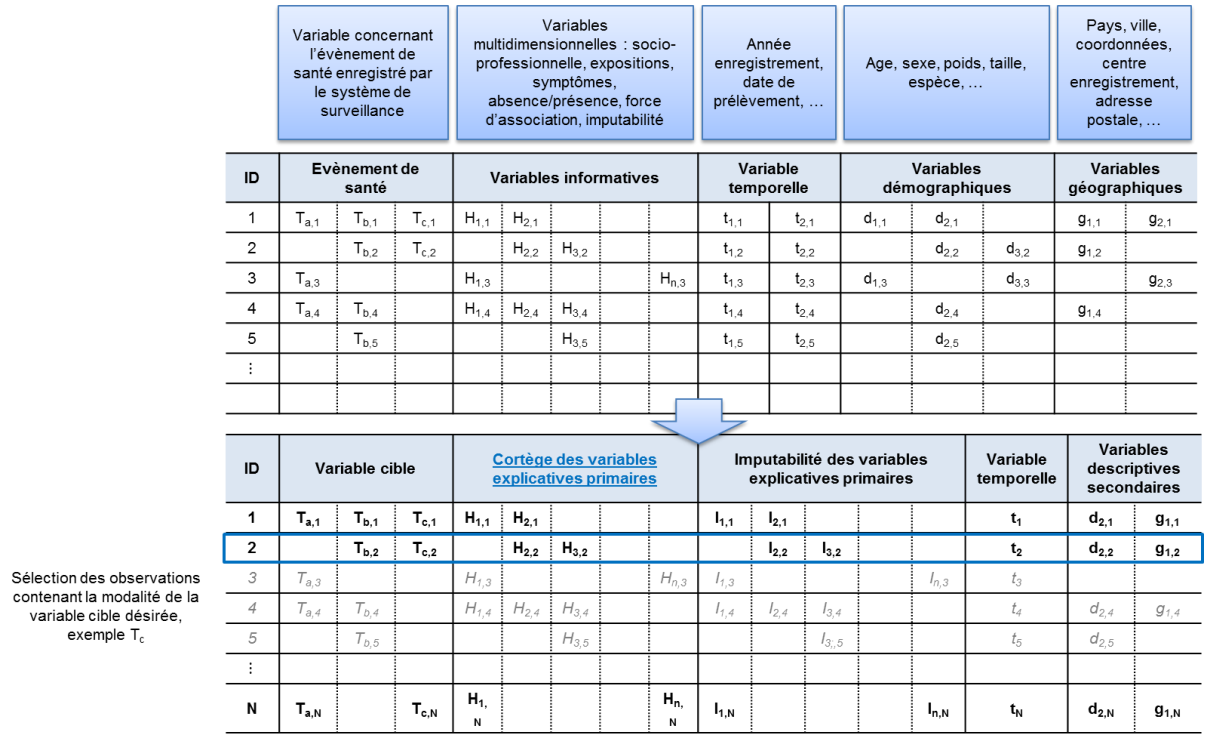


Figure 4 : Sélection des informations nécessaires à la Surveillance Observationnelle à partir d'une base de données observationnelle. L'encadré bleu illustre la définition d'un cortège de modalités associées.

La ***variable cible*** (qui peut être un ensemble de variables avec des modalités) est la variable qui va être étudiée par la Surveillance Observationnelle, à l'instar de la ***variable à expliquer*** en surveillance classique. Quant aux ***variables associées***, celles-ci représentent les ***variables explicatives*** de la Surveillance Observationnelle ; chaque variable étant déclinée en modalités. Ces variables peuvent être regroupées sous forme de ***cortège*** pour caractériser l'évènement d'intérêt surveillé, c'est-à-dire la variable cible (Figure 4). Un cortège (ou séquence) correspond donc à l'ensemble des informations ayant été enregistrées en même temps que l'évènement de santé ciblé (encadré bleu dans la Figure 4). Ces variables explicatives peuvent être subdivisées en ***variables explicatives primaires*** (qui sont utilisées pour expliquer la variable cible) et ***secondaires*** (qui peuvent être des variables de contextes associées aux variables primaires) ; cette subdivision pouvant être interchangeable.

Par exemple :

Dans un contexte de surveillance des maladies professionnelles, si la variable cible correspond à la variable pathologie, l'évènement de santé ciblé pourrait être une pathologie précise comme une dermatite de contact. Les modalités des variables explicatives primaires seraient les expositions professionnelles ayant conduit à cette dermatite de contact enregistrées sous forme de cortège.

Dans le processus de la SO, c'est le choix de la modalité de la variable cible qui permet de délimiter l'extraction du sous-échantillon de la base de données sur lequel sera appliquée l'analyse (Figure 4). Ce sous-échantillon comprend ***N observations*** (*N* lignes dans l'illustration) concernant la (les) modalité(s) de l'évènement de santé ciblé, accompagnées par un numéro d'identification, une ***variable temporelle*** (une ou plusieurs colonnes) et un ***cortège de taille H des modalités des variables explicatives (ou associées) primaires*** (*H* colonnes dans l'illustration) qui seront utilisées pour la caractérisation de l'évènement de santé ; l'inclusion et la taille des variables explicatives dépend de la base de données et du type de l'analyse à conduire. Selon la base de données, à chaque modalité d'une variable explicative peut-être associée une imputabilité et donc au cortège des modalités des variables explicatives primaires un ***cortège de taille I d'imputabilité*** traduisant la force d'association entre l'évènement de santé étudié et chacune des variables explicatives.

En résumé, pour faire la SO d'un évènement d'intérêt sur une période de temps *T*, il faut considérer un tableau de *N* lignes (pour les observations), 2 x *H* colonnes (*H* colonnes pour les cortèges des modalités des variables explicatives primaires et *H* colonnes pour le cortège des imputabilités associées), une colonne de variable temporelle (délimitant la période temporelle d'étude *T*) et, si nécessaire, d'autres colonnes pour les variables explicatives secondaires.

L'application de la Surveillance Observationnelle sur ces données extraites a pour but :

- D'**identifier au sein des cortèges des variables explicatives les modalités caractéristiques** de l'évènement de santé surveillé, par exemple une exposition professionnelle dans le cadre de l'étude d'une base de données de maladies professionnelles ;

- D'**identifier les combinaisons recensées entre ces modalités**, c'est-à-dire d'étudier les cortèges de modalités afin de mettre en évidence les modalités fréquemment **enregistrées ensemble au sein des mêmes cortèges**, représentant ainsi des ***motifs caractéristiques*** de modalités, afin de mettre en évidence par exemple de la multi-exposition environnementale ou professionnelle ;

- D'**étudier les différentes modalités identifiées ainsi que les combinaisons** de ces modalités afin de déterminer si oui ou non ces modalités et/ou combinaisons sont représentatives de l'évènement de santé étudié ;

- De **suivre dans le temps l'évolution des recrutements de ces modalités et/ou combinaisons** identifiées, ainsi que les éventuels changements au sein de leur structure, comme par exemple leur complexification suite au recrutement d'une nouvelle modalité ou la diminution de leur recrutement ;

- L'**étude de ces modalités et/ou combinaisons identifiées en fonction des autres variables explicatives secondaires** disponibles au sein de la base de données (Figure 4) par la caractérisation de leur présence/absence et leur dynamique en fonction de ces variables.

Cette Surveillance a donc pour objectif de fournir le ***spectre complet des modalités et/ou combinaisons*** identifiées et représentatives permettant de caractériser l'évènement de santé étudié, tout en étudiant les structures entre ces associations ainsi que leur évolution au cours du temps. Le processus de la SO s'effectue en 3 étapes distinctes :

- Une étape de **préparation des données** qui consiste à la formation des nœuds de l'exposome (mise en réseau de l'information) et la construction de motifs ;

- Une **analyse descriptive** des observations et caractéristiques ;

- Une **analyse spectrale** de la variable cible via les indicateurs de la Surveillance Observationnelle.

3.3. Préparation et construction des données

3.3.1. Sélection des modalités et observations significatives

A partir de l'échantillon extrait, la première étape de la SO consiste à préparer et sélectionner les observations (lignes) pertinentes et significatives qui seront utilisées dans la construction des nœuds (*voir 3.3.2.2*) et des motifs (*voir 3.3.2.3*) pour la restructuration des données en termes exploitables pour la SO.

Une observation est considérée **significative** si son cortège de variables explicatives primaires associées contient au moins une **modalité significative**. Une modalité d'une variable explicative primaire est significative si le nombre d'occurrences de cette modalité dans tout l'échantillon (*i.e.*, dans toutes les observations) satisfait un critère de significativité.

Soit $\boldsymbol{n_i}$ **le nombre total d'occurrences (recrutement)** de la modalité « i » (avec $i = 1,2,\dots,H$) d'une variable explicative associée à un évènement d'intérêt pendant toute la période d'étude T. Pour rappel, n_i représente le nombre total d'observations dans lesquelles la modalité considérée a été rapportée. Le taux de recrutement d'une modalité n'étant pas uniforme sur la période T, il est préférable, pour le calcul de la significativité, de travailler avec un ***nombre de recrutement de référence*** $\boldsymbol{n_{i,\Delta t}}$ dans un ***intervalle de temps de référence* Δt,** nombre plus stable, défini par :

$$n_{i,\Delta t} = \frac{n_i}{T} \times \Delta t \qquad (3.1)$$

Par exemple :
une modalité recrutée 20 fois en 10 ans aura un nombre de recrutement de référence de 2 pour un Δt=1 et 8 pour un Δt=4, et une modalité recrutée 2 fois en 10 ans aura un recrutement de référence de 0.2 pour un Δt=1 et 0.8 pour un Δt=4.

Pour être considérée dans l'analyse, la modalité devrait avoir un recrutement de référence au moins égal à 1, le choix de la plage temporelle de référence est alors décisif dans la conservation des modalités pour l'analyse.

Deux critères pour évaluer la significativité d'une modalité ont été utilisés selon l'existence ou non d'une ***variable d'imputabilité*** dans la base de données.

3.3.1.1. Critère de significativité basé sur la fréquence

En absence d'imputabilité, le calcul de la significativité des modalités du cortège des variables explicatives dans l'intervalle de temps de référence Δt est basé sur le calcul du risque relatif :

$$RR_i = \frac{\text{nombre d'occurences de la modalite } i}{\text{nombre d'occurences attendues de la modalité } i \text{ si recrutement homogène}} = \frac{n_{i,\Delta t}}{\frac{1}{H}\sum_{j=1}^{H} n_{j,\Delta t}} = \frac{n_i \times H}{\sum_{j=1}^{H} n_j} \quad (3.2)$$

où le nombre d'occurrences de la modalité i attendues sous l'hypothèse homogène de recrutement est obtenu comme le rapport de la somme des occurrences de toutes les modalités distinctes observées par le nombre H de modalités distinctes. Le critère de significativité s'annonce alors comme suit :

Une modalité est **significative** lorsque le risque relatif associé est tel que $\boldsymbol{RR_i \geq 1.96}$.

3.3.1.2. Critère de significativité basé sur la distribution d'imputabilité

Lorsque la variable d'imputabilité est renseignée, n_i est associé à une distribution d'imputabilité puisqu'à chaque modalité rapportée dans une observation est associée une imputabilité. Il est alors possible de l'utiliser dans le calcul de la significativité comme détaillé ci-après (*équation 3.3*).

Pour ce qui est du critère de significativité, Q_k dénote la probabilité d'avoir k recrutements d'une modalité donnée au sein de l'échantillon tel que $\sum_{k=0}^{\infty} Q_k = 1$. Pour chaque modalité i avec un nombre d'occurrences $n_{i,\Delta t}$ dans l'intervalle de référence Δt, la probabilité est définie telle que $p_i = P\left(\geq n_{i,\Delta t}\right)$ d'avoir un nombre d'occurrences supérieur ou égal à $n_{i,\Delta t}$:

$$p_i = P(\geq n_{i,\Delta t}) = \sum_{k=n_{i,\Delta t}}^{\infty} Q_k = 1 - \sum_{k=0}^{n_{i,\Delta t}-1} Q_k \qquad (3.3)$$

Le critère de significativité s'énonce alors comme suit :

Pour un risque **α** de 1ère espèce, couramment fixé à 5%, est défini :

$$si \begin{cases} p_i > \alpha \rightarrow recrutement\ aléatoire = modalité\ non-significative \\ p_i \leq \alpha \rightarrow recrutement\ non-aléatoire = modalité\ significative \end{cases}$$

Ainsi $p_i = \alpha$ permet de calculer **le nombre de recrutements critiques $n_{i,\Delta t,cr}$ acceptés comme aléatoires pour la modalité « *i* »**. Le critère de significativité peut s'écrire alternativement en termes de nombre d'occurrences :

$$\begin{cases} n_{i,\Delta t} > n_{i,\Delta t,cr} \rightarrow recrutement\ aléatoire = modalité\ non-significative \\ n_{i,\Delta t} \leq n_{i,\Delta t,cr} \rightarrow recrutement\ non-aléatoire = modalité\ significative \end{cases}$$

Dans tous les développements qui suivent dans ce mémoire, il a été supposé que la distribution des recrutements des modalités peut être décrite par une loi de Poisson de paramètre λ :

$$Q_k = e^{-\lambda}\frac{\lambda^k}{k!} \qquad (3.4)$$

où λ est donné par la relation empirique suivante :

$$\frac{\lambda - \lambda_{min}}{\lambda_{max} - \lambda_{min}} = \left(\frac{I - I_{min}}{I_{max} - I_{min}}\right)^{\nu} \qquad (3.5)$$

Dans cette relation, I représente **l'imputabilité moyenne** de la distribution d'imputabilité observée, les paramètres I_{min} et I_{max} sont déterminés par les différentes valeurs d'imputabilité directement rencontrées au sein de l'échantillon, tandis que les valeurs λ_{min}, λ_{max}, et ν peuvent être fixées *a priori*.

3.3.1.2.1. Choix du paramètre ν

La Figure 5 montre en variable réduite la variation de λ en fonction de I pour différentes valeurs de ν. Le paramètre ν permet ainsi d'**influer sur la relation entre le paramètre λ de la loi de Poisson et l'imputabilité de la modalité**. La relation est linéaire entre λ et I pour $\nu = 1$ et non linéaire autrement (Figure 5). Pour $\nu < 1$, λ atteint un palier plus rapidement en fonction de I : il y a moins de discrimination entre des valeurs faibles d'imputabilité et des valeurs fortes, préservant ainsi les petites valeurs d'imputabilité. A l'inverse, lorsque $\nu > 1$, λ croît plus lentement en fonction de I : plus la discrimination est forte, plus il faudra alors un nombre important d'enregistrements avec une imputabilité forte pour permettre de passer le seuil de significativité. Selon le problème étudié, le choix peut alors être fait de façon à avoir une dépendance appropriée de λ en fonction de I.

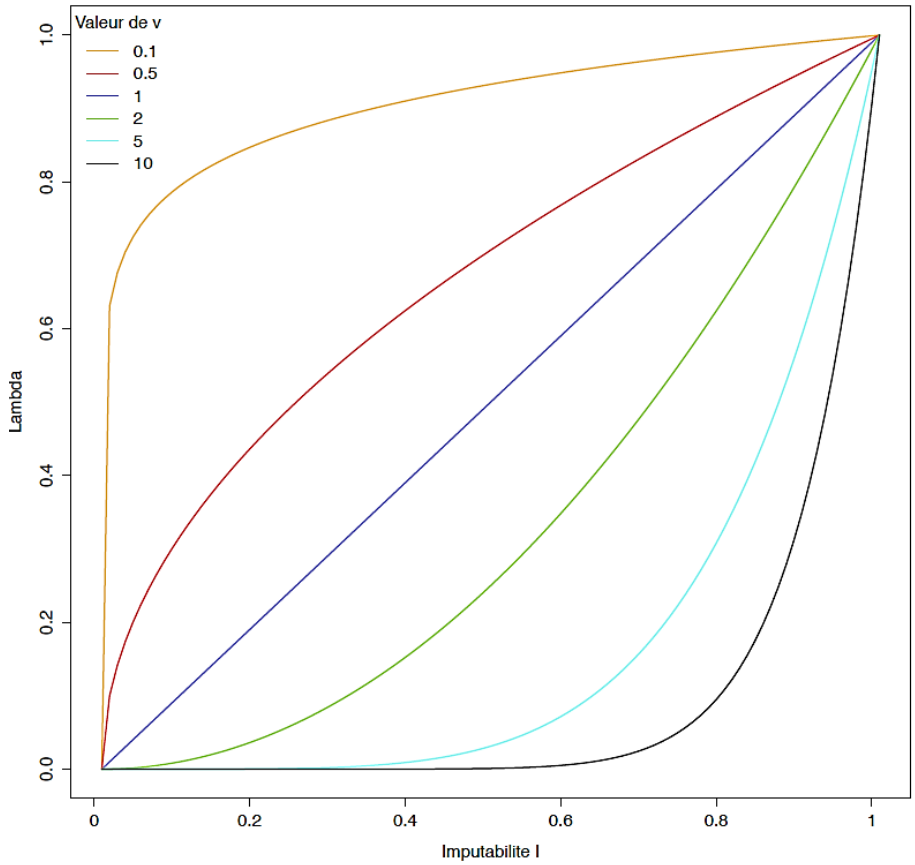


Figure 5 : Influence de ν sur la relation entre le paramètre λ et l'imputabilité I.

3.3.1.2.2. Choix des paramètres λ_{min} et λ_{max}

Les paramètres λ_{min} et λ_{max} sont définis de manière à fixer **les nombres d'occurrences critiques** correspondant respectivement à I_{min} et I_{max}.

Pour le paramètre λ_{max}, il a été utilisé le nombre d'occurrences critique $n_{\Delta t,cr} = 1$ et donc considéré qu'une modalité enregistrée **une seule fois avec une imputabilité maximale** était significative. Cela se traduit par $P(\geq n_{\Delta t,cr} = 1) = 0.05$, c'est-à-dire, $1 - exp\{-\lambda_{max}\} = 0.05$, ce qui donne : $\lambda_{max} = 0.05$.

Quant à λ_{min}, le Tableau 2 donne les λ_{min} correspondant au nombre d'occurrences minimum critiques $n_{\Delta t,cr}$ pour qu'une modalité **enregistrée avec une imputabilité minimale** soit significative. Les valeurs de λ_{min} en fonction de $n_{\Delta t,cr}$ sont obtenues à partir de la solution numérique de :

$$1 - \sum_{k=0}^{n_{\Delta t,cr}-1} \frac{(\lambda_{min})^k}{k!} exp\{-\lambda_{min}\} = 0.05 \qquad (3.6)$$

Tableau 2 : Valeurs de λ_{min} en fonction du nombre de recrutement de référence critique $n_{\Delta t,cr}$ pour une modalité enregistrée avec l'imputabilité la plus faible.

$n_{\Delta t,cr}$	λ_{min}
4	1.37
5	1.98
6	2.62
7	3.29
8	3.99
9	4.70
10	5.43

Par exemple :

Pour $\lambda_{min} = 1.37$, si la modalité a été enregistrée avec l'imputabilité la plus faible, il faudra 4 occurrences minimum pour pouvoir la considérer comme significative, tandis que pour $\lambda_{min} = 5.43$ il en faudra 10 minimum. Autrement dit, si pour une modalité enregistrée avec l'imputabilité la plus faible I_{min} il faudra utiliser $\lambda_{min} = 1.37$ si on choisit pour la significativité un nombre minimum d'occurrences de 4 et utiliser $\lambda_{min} = 5.43$ pour un nombre minimum d'occurrences de 10.

Plus la valeur du λ_{min} augmente, plus la sélection est drastique, permettant d'éliminer un maximum de « bruit », au risque de ne pas détecter des modalités émergentes. La valeur est donc à déterminer en fonction de la problématique et des caractéristiques de l'échantillon.

A partir de l'échantillon, **l'ensemble des modalités des variables explicatives primaires** sont identifiées et comptées afin d'établir un **nombre d'occurrence de référence sur un intervalle de temps de référence**. Si l'échantillon contient l'information sur l'imputabilité, celle-ci est utilisée afin de calculer la significativité de chaque modalité, sinon, la significativité est calculée à partir du risque relatif utilisant la fréquence relative de la modalité par rapport à toutes les autres modalités. Une **observation est conservée pour le reste de l'analyse si elle est significative**, *i.e.* si cette observation contient **au moins une modalité significative dans son cortège**.

3.3.2. Construction des nœuds et des motifs

Initialement définit comme un **réseau connecté d'expositions**, le terme « exposome » correspond ici à un **réseau relationnel des observations connectées entre elles si elles partagent au moins une modalité** des variables explicatives primaires en commun.

La restructuration et l'organisation des données par la méthode de l'exposome permettent d'étudier et de caractériser les associations entre un évènement donné et les modalités explicatives rapportées. La mise en réseau des données permet en effet la conservation des relations structurelles entre la variable cible et les modalités explicatives primaires rapportées, mais également entre les différentes modalités elles-mêmes.

3.3.2.1. Longueur des cortèges

Le nombre de modalités présentes au sein du cortège des variables explicatives d'une observation peut différer selon les bases de données et entre les observations d'un même échantillon. Le nombre de ces modalités possède une influence directe sur la construction des nœuds et des motifs servant à l'analyse spectrale de la Surveillance Observationnelle.

Par exemple :
Une base de données A pourrait limiter le nombre de variables explicatives à enregistrer à 5, tandis que la base B limiterait ce nombre à 10. De même, toutes les observations enregistrées au sein des bases de données A et B ne possèdent pas obligatoirement 5 ou 10 modalités explicatives pour l'évènement d'intérêt enregistré mais varient entre 1 et 5 pour la base A et entre 1 et 10 pour B.

Afin de permettre une cohérence des résultats, il est préférable de limiter le nombre de modalités présentes dans les cortèges. Bien qu'il existe des origines multifactorielles aux évènements d'intérêts, la multiplication de causes potentielles ne permet pas forcément de

répondre correctement à une problématique. Selon la question scientifique posée, il est alors judicieux de réduire la longueur de ces cortèges en fixant une longueur maximale « autorisée », c'est-à-dire un nombre maximal de modalités présentes au sein du cortège, appelée ***longueur d***. La réduction des cortèges passe alors par la « suppression » de certaines modalités (ou bien encore par la « conservation » des autres). Afin de faire un choix pertinent, la réduction de ces cortèges est basée soit sur la variable d'imputabilité si elle existe, soit sur le nombre d'occurrences de la modalité. Seront exclues une à une les modalités ayant soit l'imputabilité *I* la plus faible du cortège, soit le nombre d'occurrences n_i le plus faible, et ce, jusqu'à obtenir un cortège de modalités de la longueur désirée.

3.3.2.2. Construction des nœuds de l'exposome

Le terme « nœud » correspond à un **ensemble d'observations**, issues de l'échantillon, ayant toutes **en commun le même cortège de modalités** des variables explicatives primaires. Le nombre d'observations correspond alors au poids du nœud.

L'échantillon retenu d'après la significativité des modalités est considéré selon un réseau relationnel pour lequel chaque observation significative (contenant au moins une modalité significative) est représentée par un ***nœud*** $\boldsymbol{v = (p, h)^T}$, avec une ***variable cible p et une combinaison unique de modalités h***. Par convention, le vecteur *H* des modalités explicatives et associées à la variable cible peut comprendre de une à *d* modalités distinctes telle que $(h_1, h_2, h_3, h_4, \cdots, h_d)$. Chaque nœud *v* ainsi formé est alors pondéré par le ***nombre total*** $\boldsymbol{a_v}$ ***de copies identiques*** de chaque observation partageant ce même cortège de variables explicatives primaires.

Par exemple :

ID	Variable cible	Cortège des variables explicatives primaires			Imputabilité des variables explicatives primaires			Variable temporelle	Variables descriptives secondaires	
1	T_c	A	B		$I_{1,1}$	$I_{2,1}$		t_1	$d_{2,1}$	$g_{1,1}$
2	T_c	B	C		$I_{1,2}$	$I_{2,2}$		t_2	$d_{2,2}$	$g_{1,2}$
3	T_c	A	B	C	$I_{1,3}$	$I_{2,3}$	$I_{3,3}$	t_3		
4	T_c	B	C		$I_{1,4}$	$I_{2,4}$		t_4	$d_{2,4}$	$g_{1,4}$
5	T_c	A	B		$I_{1,5}$	$I_{2,5}$		t_5	$d_{2,5}$	
6	T_c	A	B		$I_{1,6}$	$I_{2,6}$		t_6	$d_{2,6}$	$g_{1,6}$
⋮										

Nœud 1
Nœud 2
Nœud 3

Pour cet exemple, 3 nœuds différents seront formés, le nœud « AB » comprenant les observations 1, 5 et 6, d'un poids égal à 3, le nœud « ABC » comprenant uniquement l'observation 3, d'un poids égal à 1, et le nœud « BC » comprenant les observations 2 et 4, d'un poids égal à 2.

3.3.2.3. Construction des motifs

Le « motif » est une **modalité ou une combinaison particulière et représentative** de modalités (variables explicatives) identifiée dans les cortèges de l'ensemble des nœuds et permettant de caractériser l'association évènement de santé ciblé ~ variables explicatives.

Afin de considérer toutes les associations de modalités possibles, un ensemble de combinaisons est généré à partir de l'ensemble des nœuds *V* précédemment créés (Figure 6). Le nombre de combinaison possible pour un nœud dépend de sa longueur *d*, c'est-à-dire du nombre de modalités qui le forment, tel que :

$$C_v = \sum_{i=1}^{d} \binom{d}{i} \tag{3.7}$$

Par exemple :

Un nœud composé de trois modalités « ABC » génère 7 combinaisons possibles : 3 modalités considérées séparément « A », « B » et « C », 3 associations de deux modalités « AB », « AC », « BC », et enfin une association triple « ABC ».

A partir de toutes les combinaisons possibles générées contenues dans le vecteur *C* et l'ensemble des nœuds *V*, la matrice $M = C \times V$ est générée, tel que l'élément $M_{ij} = a_{c,v}$ si la combinaison c_i est présente dans le nœud v_j, et 0 dans le cas contraire.

Ainsi pour chaque combinaison *c* créée est attribué le poids du nœud *v* à partir duquel elle a été générée. L'association ayant pu être retrouvée dans d'autres nœuds, les combinaisons distinctes sont identifiées et le poids total de chaque combinaison est calculé par :

$$w_{c,i} = \sum_{v=1}^{V} a_{c,v} \tag{3.8}$$

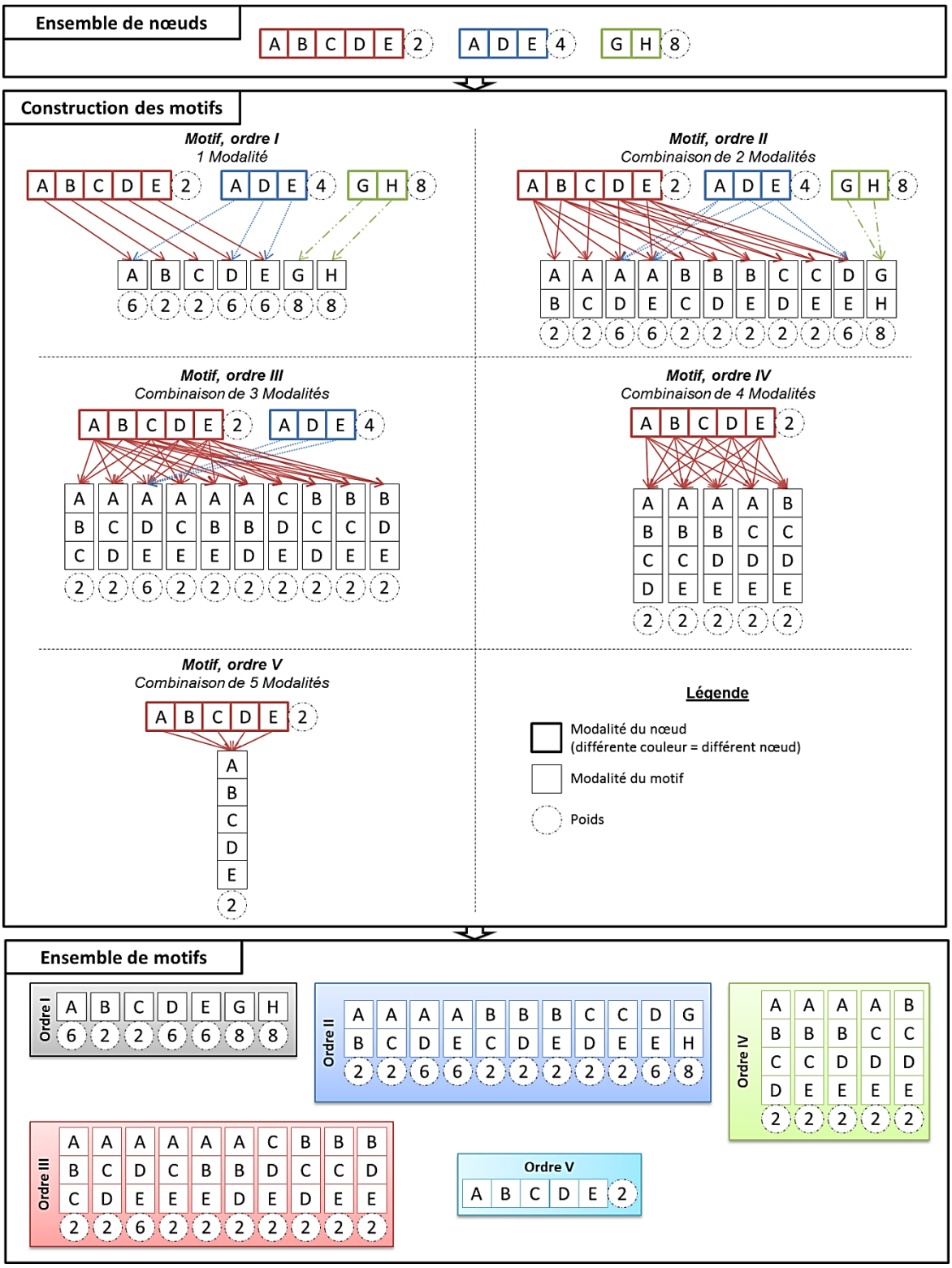


Figure 6 : Schéma d'illustration de la construction des motifs des modalités explicatives. Exemple avec $\boldsymbol{d} = \mathbf{5}$. Sont présentés ici 3 nœuds de longueur 5, 3 et 2 (comprenant respectivement 5, 3 et 2 modalités distinctes) dont les poids *a* sont de 2, 4 et 8 respectivement. A partir de chacun de ces nœuds, des combinaisons vont être générées successivement. D'abord des combinaisons comprenant 1 modalité (ordre I), puis de 2 modalités (ordre II), puis de 3 modalités (ordre III) uniquement pour les nœuds ayant $\boldsymbol{h} \geq \mathbf{3}$, puis 4 (ordre IV) pour les nœuds ayant $\boldsymbol{h} \geq \mathbf{4}$ et enfin 5 (ordre V) pour les nœuds ayant $\boldsymbol{h} \geq \mathbf{5}$, correspondant au maximum fixé *d*. Chaque combinaison distincte générée est ensuite identifiée afin d'attribuer un poids final $\boldsymbol{w_c}$ correspondant à la somme des poids de tous les nœuds à partir desquels elle a été générée.

Cependant, toutes les combinaisons ainsi générées ne sont pas incluses dans la suite de l'analyse. Seules les combinaisons représentatives sont utilisées pour former les motifs.

Une combinaison est **représentative** si :

$$q_i = \sum_{v=1}^{V} \left(a_{c,v}\right)^2 \tag{3.9}$$

Une combinaison satisfaisant cette condition est conservée en tant que ***motif m d'association représentatif*** formant l'ensemble M des motifs, chaque motif étant caractérisé par sa combinaison de ***modalités***, ***son poids*** $\boldsymbol{w_m}$ (nombre total de recrutements) et le ***nombre de nœuds u*** dans lesquels le motif apparaît :

$$u = \sum_{v=1}^{V} H\left(a_{c,v}\right) \text{ où } H(x) = \begin{cases} 1; x > 0 \\ 0; x \leq 0 \end{cases} \tag{3.10}$$

La condition appliquée sur les combinaisons a pour objet de prendre en compte les deux cas limites suivants. Soit une combinaison n'apparaît que dans un seul nœud, son poids dans le nœud devra alors être supérieur ou égal à 2 pour être conservée. A l'inverse, si la combinaison apparaît dans plusieurs nœuds mais avec un poids de 1 dans chacun des nœuds, il faudrait qu'elle apparaisse dans 4 nœuds au moins pour être conservée.

A partir d'un **ensemble de modalités** dans un enregistrement, les associations de modalités distinctes sont identifiées afin de former les **nœuds**, puis elles sont utilisées afin de construire des combinaisons de tailles différentes, dont les combinaisons représentatives sont conservées pour construire **les motifs d'associations** caractérisés par leur poids, le nombre de modalités les formant (ordre) et le nombre de nœuds dans lesquels ils ont été identifiés.

3.4. Analyse descriptive

3.4.1. Echantillon

A l'instar de tout système de surveillance, des analyses descriptives sont effectuées sur les différentes variables présentes au sein de la base, par exemple le suivi du recrutement au cours du temps et sa stratification selon les espèces ou le genre des individus. La Figure 7 illustre l'exemple du suivi d'un évènement d'intérêt au cours du temps avec en abscisse la variable temporelle, et en ordonnées à la fois le nombre de recrutement distinct par an (représenté sous forme de diagramme en bâtons) et l'effectif cumulé du recrutement (représenté par une courbe).

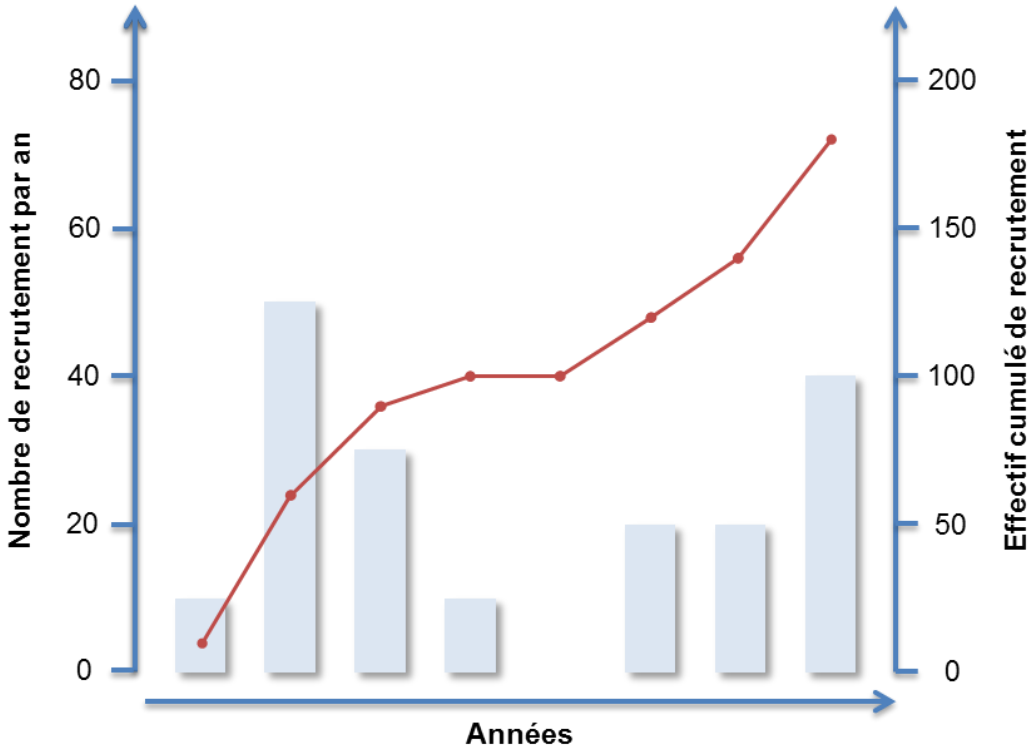


Figure 7 : Exemple du suivi dans le temps de recrutement et d'effectif cumulé de cas.

Il peut être également intéressant de faire l'analyse descriptive de l'évolution du nombre d'observations au cours du temps ainsi que le nombre de nœuds et de motifs qui en découlent. De par le caractère pérenne des bases de données observationnelles, le nombre d'observations au cours du temps ne pourra jamais régresser, cependant il est tout à fait possible d'obtenir des variations importantes du nombre de motifs générées selon les taux de recrutements observés.

3.4.2. Description des motifs

Concernant les motifs d'associations mis en évidence, il est également possible de les décrire selon trois principaux axes :

- le **nombre de nœuds u** dans lesquels ils ont été identifiés ;
- leur **poids total w_m** correspondant aux nombres d'observations dans lesquelles ils ont été identifiés ;
- leur **ordre** (ou longueur d), correspondant au nombre de modalité(s) formant le motif, ordre I pour une modalité, ordre II pour deux modalités, etc (Figure 6), identifiant ainsi les motifs « simples » des motifs « complexes » (ou « d'ordre supérieur »).

3.5. Analyse spectrale

L'analyse spectrale est le cœur de la Surveillance Observationnelle puisqu'elle permet le déploiement des différents indicateurs développés dans le but de décrire et caractériser un évènement de santé à une date *t* grâce aux modalités identifiées précédemment et de suivre l'évolution de leur structure au cours du temps. Elle se décline en 5 grandes parties :

- La construction graphique de **l'exposome à partir des nœuds** ;
- La **caractérisation dynamique des motifs** au cours du temps ;
- La construction graphique du **spectrosome à partir des motifs** pour permettre la visualisation de « cluster » ;
- La création d'un **arbre de la multi-information** permettant la visualisation des différents motifs identifiés ;
- L'obtention d'une signature de l'évènement de santé étudié par **la génération du spectre des motifs caractéristiques** de cet évènement de santé, synthétisant les différents indicateurs construits précédemment.

3.5.1. Exposome des nœuds

La méthodologie de l'exposome a été présentée précédemment dans les travaux de Laurie Faisandier [Faisandier, 2011a, Faisandier, 2011b]. A partir de l'ensemble des nœuds *V*, pour chaque nœud $v = (h_1, h_2, h_3, h_4, \cdots, h_d)$, on considère $B_{ij,\eta}$ (avec $0 \leq B_{ij,\eta} < n$), correspondant au nombre de modalités partagées par deux nœuds v_i et v_j apparaissant au moins η fois dans l'échantillon. On a alors $B_{ij,\eta} = \theta(a_{v,i} - \eta) \times \theta(a_{v,j} - \eta) \times \sum_{k,l=1}^{d} \delta_{h_{ik},h_{jl}}$ où $\delta_{h_{ik},h_{jl}}$ représente le symbole de Kronecker, qui compare les éléments h_i and h_j du cortège, avec $\delta_{h_{ik},h_{jl}} = 1$ si $h_{ik} = h_{jl}$ et $\delta_{h_{ik},h_{jl}} = 0$ si $h_{ik} \neq h_{jl}$; et le produit $\theta(a_{v,i} - \eta) \times \theta(a_{v,j} - \eta)$ qui assure que le nombre de copies de chaque nœud satisfasse $\eta \geq 1$, tel que $\theta(z) = 1$ si $z \geq 0$ et $\theta(z) = 0$ si $z < 0$. Les règles de connexion sont définies telles que deux nœuds v_i et v_j sont connectés si leur association $B_{ij,\eta} \geq D$ avec $1 \leq D \leq d - 1$, ou non connectés dans le cas contraire. Le D_η – exposome est défini comme un réseau non dirigé, caractérisé par une matrice adjacente $A = V \times V = (A_{ij})$ entre les nœuds v_i et v_j avec l'intersection $A_{ij} = 1$ si $B_{ij,\eta} \geq D$ et $A_{ij} = 0$ sinon.

Le réseau de l'exposome est décrit par le graphique $G = \{N, V, L, D, \eta\}$, avec *N* observations dans l'échantillon, décrit par *V* nœuds, connectés entre eux par *L* liens hétérogènes, partageant *D* modalités minimum et satisfaisant la limite η du nombre de copies identiques. Sur le graphique obtenu (Figure 8), la taille d'un nœud est proportionnelle à son poids, c'est-à-dire au nombre de copies identiques de ce cortège de modalités. Si un nœud n'est relié à aucun autre on parle de nœud isolé.

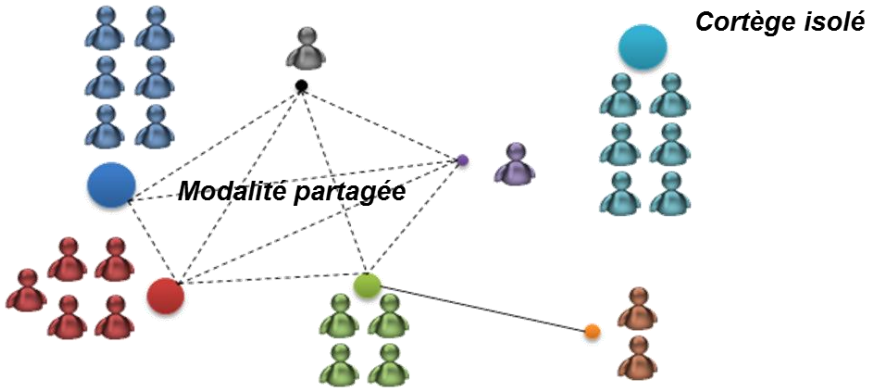


Figure 8 : Schéma de construction d'un exposome.

Cet exposome permet de **mettre en évidence des modalités communes aux différentes observations**. Il est ainsi possible d'identifier ce que l'on appelle des « **clusters** » sur l'exposome, qui correspondent à des agrégats d'observations partageant tous la même modalité.

L'exposome construit présente également toutes les caractéristiques classiques des réseaux et permet d'appliquer le calcul d'indicateur classique tel que :

- la **densité** du réseau : ratio entre le nombre de liens identifiés sur l'exposome et le nombre de liens maximum possible. Une densité égale à 1 correspond à un réseau pour lequel tous les nœuds seraient connectés entre eux ;
- le **degré de connectivité** des nœuds : correspond au nombre de nœuds auxquels est connecté un nœud et permet d'évaluer la popularité de celui-ci ;
- le **coefficient d'agglomération**, ou coefficient de clustering : permet d'évaluer la vraisemblance que les voisins d'un nœud soient également connectés entre eux.

Cependant, l'exposome est en pratique difficile à visualiser et à analyser car souvent très dense et les différents indicateurs permettent de fournir une « grille de lecture » plus claire de ce réseau.

3.5.2. « Reporting Index » et statut dynamique

L'indice de recrutement ou « Reporting Index » (RI) caractérise l'évolution au cours du temps des motifs : il permet d'indiquer **la proportion et l'importance du recrutement d'un motif *m* tenant compte du recrutement des autres motifs** à une année donnée.

Soit M le nombre total des motifs distincts dans la période d'observation allant de t_0 à t où chaque motif m est présent à la date i ($t_0 \leq i \leq t$) avec un poids $w_m(i)$. Le RI_m du motif m est calculé de la relation *:*

$$RI_m(t|t_0) = \sum_{i=t_0}^{\mathrm{t}} w_m(i) \Big/ \sum_{n=1}^{M} \left(\sum_{i=t_0}^{t} w_n(i) \right) \tag{3.11}$$

Noté que $\sum_{m=1}^{M} RI_m(t|t_0) = 1$ par construction. A ce RI est associé un seuil ξ_R inspiré de l'Analyse en Composante Principale (ACP) dans laquelle on détermine les axes principaux permettant d'expliquer au mieux l'information. Ici le seuil est calculé à partir de la relation :

$$\xi_{RI} = 1 \Big/ \sum_{m=1}^{M} [RI_m(t|t_0)]^2 \qquad (3.12)$$

La valeur ξ_{RI} correspond dans ce cas précis au nombre de motifs suffisants pour expliquer l'essentiel du $RI_m(t|t_0)$ sur la plage temporelle considérée. A l'instar du graphique des valeurs propre de l'ACP, les motifs sont classées par ordre décroissant de valeur de $RI_m(t|t_0)$ et la valeur ξ_{RI} (la « cassure ») est utilisée comme valeur limite pour $RI_m(t|t_0)$ séparant deux statuts dynamiques (Figure 9). Cette méthode a l'avantage de fournir un **seuil mobile**, adapté au différent profil de recrutement suivant les bases de données et les années.

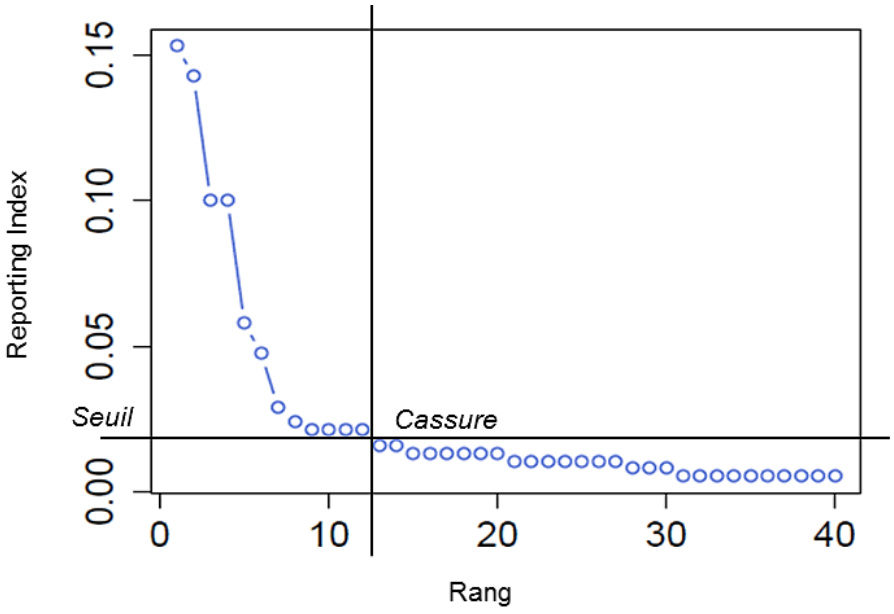


Figure 9 : Schéma d'illustration de la définition de la valeur seuil ξ_{RI}.

Deux limites temporelles ont été utilisées pour caractériser les motifs :

- $\boldsymbol{RI_{\infty}}$ **: un aspect pérenne** en tenant compte du nombre total de recrutement du motif sur la totalité de la période d'étude couverte dans l'échantillon :

$$RI_{\infty,m} = \lim_{t_0 \to -\infty} RI(t|t_0) \tag{3.13}$$

- $\boldsymbol{RI_2}$ **: un aspect plus récent** permettant d'identifier des émergences ou des disparitions en tenant compte uniquement des recrutements un intervalle de temps donné (ici $t-1$ et t) :

$$RI_{2,m} = RI(t|t-1) \tag{3.14}$$

Ci-dessous l'exemple de calcul selon la plage temporelle considérée pour le calcul du *RI* (Figure 10).

	2002	2003	2004	2005	2006	2007	2008	2009
Incidence annuelle du motif	2	4	5	**2**	0	1	0	2
Totalité des enregistrements	10	15	20	**12**	10	40	13	5

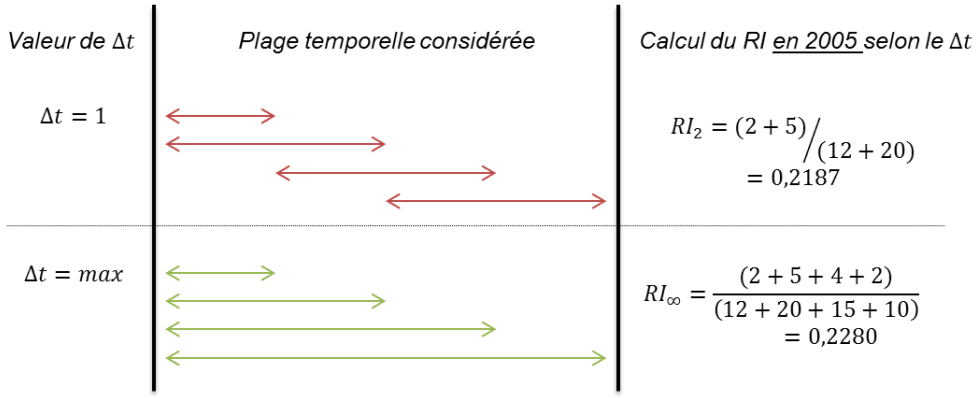


Figure 10 : Exemple de calcul de $\boldsymbol{RI_2}$ et $\boldsymbol{RI_{\infty}}$ pour un motif en 2005. En 2005, tenant compte des enregistrements des autres motifs, le motif est caractérisé par un $\boldsymbol{RI_{\infty} = 0.2280}$ et un $\boldsymbol{RI_2 = 0.2187}$.

Ces deux aspects ont permis d'identifier **4 statuts dynamiques particuliers** suivant la valeur de leur de RI_∞ et RI_2 respectifs par rapport au seuil ξ_{RI} tel que :

- **A** : un motif dont RI_∞ et RI_2 sont tous deux au-dessus des valeurs seuils. Ceci correspond à un **motif persistant** car largement recruté sur la totalité de la période étudiée, ainsi que sur les deux dernières années.
- **B** : un motif dont RI_∞ est supérieur au seuil et RI_2 inférieur au seuil. Correspond à un **motif en déclin** car historiquement recruté sur la totalité de la période, mais de manière moins importante les deux dernières années.
- **C** : un motif dont RI_∞ est inférieur au seuil mais dont RI_2 est supérieur au seuil. Correspond à un **motif en émergence** car peu recruté sur la totalité de la période, mais de manière plus importante sur les deux dernières années.
- **D** : un motif dont RI_∞ et RI_2 sont tous deux inférieurs aux seuils. Correspond à un **motif latent**, inactif, dont le recrutement a été faible sur la totalité de la période ainsi que sur les deux dernières années. Il représente pourtant une hypothèse de travail car potentiellement en émergence dans les dates à venir.

La combinaison des seuils de RI_∞ et RI_2 sur un graphique permet d'identifier les 4 statuts dynamiques (Figure 11).

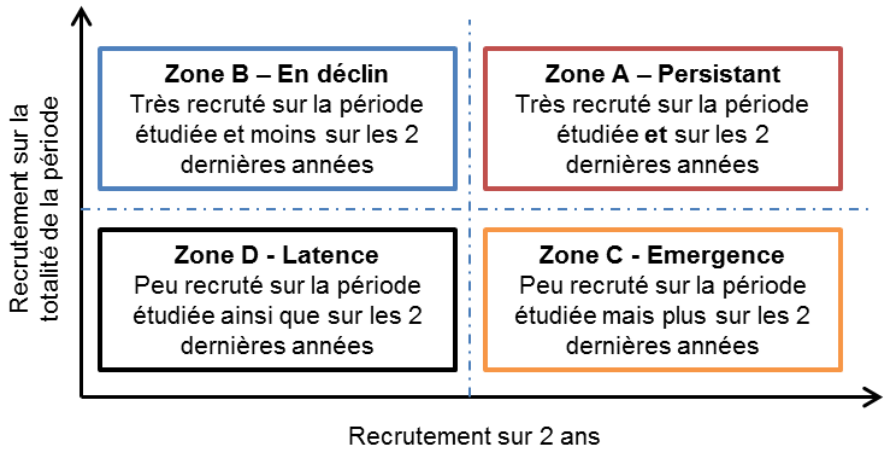


Figure 11 : Schéma d'illustration du statut pour chaque motif en fonction de RI_∞ et RI_2.

Pour chaque motif, la trajectoire dynamique peut être modélisée afin de suivre l'évolution de son statut au cours du temps. Il est possible de suivre l'émergence des motifs et comprendre l'émergence d'une association en observant la trajectoire des motifs d'ordre I la formant.

Pour l'analyse à une année *t*, il est généralement plus aisé de se focaliser sur les motifs dit **« actifs »** c'est à dire les motifs ayant été caractérisés comme « Persistant », « En déclin » ou « Emergent ». Cependant, les motifs « Latents » ayant été actifs l'année précédente sont également conservés au sein des motifs actifs afin de conserver un « historique des trajectoires ».

3.5.3. Spectrosome des motifs

Le spectrosome des motifs permet de visualiser sous forme de réseau la multi-information, grâce à la mise en évidence de **motifs dit « isolés » ou en « agrégats »**, c'est-à-dire partageant des modalités communes avec d'autres motifs d'ordre supérieur.

La procédure de construction du spectrosome des motifs est la même que pour la construction de l'exposome des nœuds *(voir 3.3.2.2, p.58*), en utilisant cette fois-ci les motifs en tant que nœuds du réseau, afin d'établir les liens entre eux.

Le spectrosome est donc conceptuellement différent puisqu'il est essentiellement formé par les motifs sous forme de clusters imbriqués ou de nœuds isolés. Le nœud **isolé** permet de mettre en évidence un contexte spécifique associé à l'évènement de santé puisque les modalités composant le motif **ne sont partagées par aucun autre motif**. Un **cluster** est quant à lui définit ici comme un **agrégat de deux motifs au minimum partageant au moins un élément de leur cortège**. La particularité de ces clusters est qu'ils sont majoritairement formés par des motifs d'ordre supérieur. Dans le cas d'un motif d'ordre I, il ne peut y en avoir qu'un au sein d'un cluster et celui-ci permet « d'identifier ce cluster ». Les clusters permettent d'identifier les différentes conditions dans lesquelles est rencontré le motif d'ordre I en question et d'identifier les complexifications de l'information.

Par exemple :
Si un motif d'ordre I « A » permet la formation de 5 autres motifs d'ordre II, un cluster sera alors formé entre ces 6 motifs, et le cluster sera appelé par le nom du motif d'ordre I : « A ».

3.5.4. Arbre de la multi-association

L'arbre de la multi-association est un indicateur graphique inspiré d'un arbre phylogénétique et développé dans le but **d'évaluer la part de la multi-information** pour la variable cible. Il permet de **visualiser les différents motifs et associations de motifs identifiés sur le spectrosome des motifs**.

Tous les motifs actifs d'ordre I sont représentés sur la gauche du graphique, puis toutes les associations actives identifiées entre ces motifs d'ordre I sont également annotés afin d'illustrer les différents motifs d'ordre supérieur à I (Figure 12).

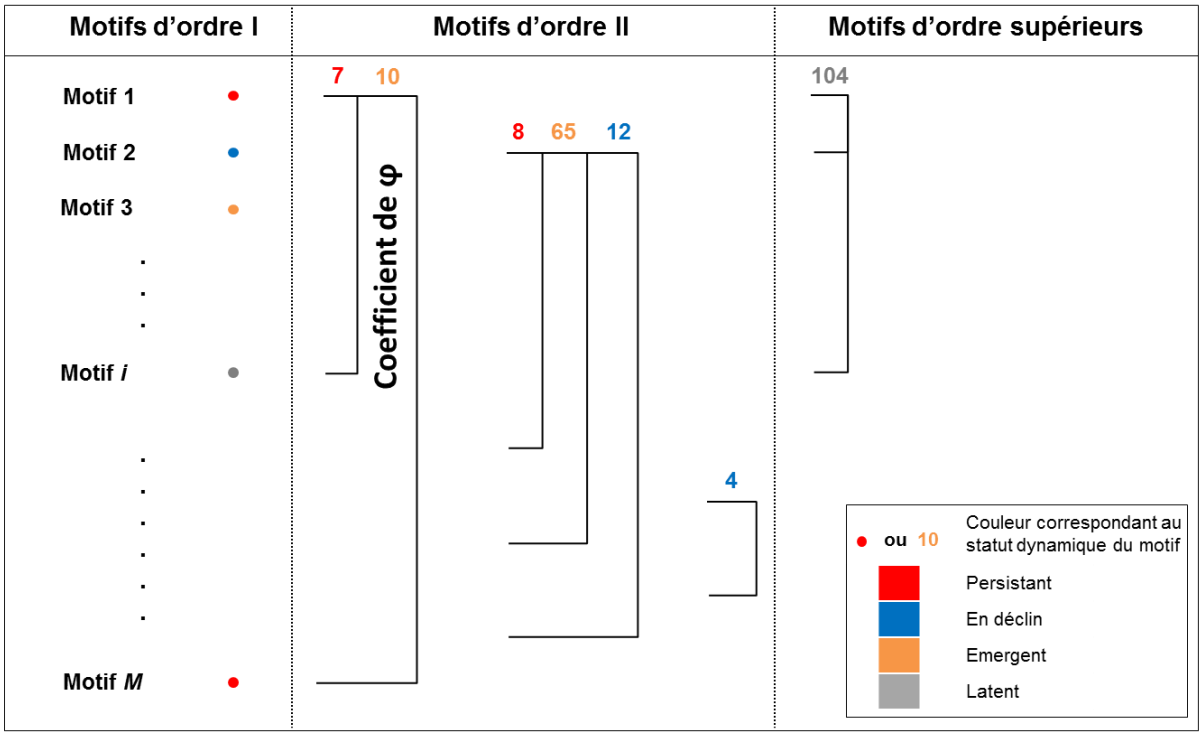


Figure 12 : Schéma de l'arbre de la multi-association de la Surveillance Observationnelle. A gauche sont indiqués les motifs d'ordre I, à droite, les branches représentent les associations identifiées entre les différents motifs d'ordre I pour former les motifs d'ordres supérieurs. Les numéros correspondent au numéro d'identification des motifs, identiques sur chaque graphique de la Surveillance Observationnelle. La couleur correspond au statut dynamique des motifs : rouge pour un motif persistant, bleu pour un motif en déclin, orange pour un motif émergent, gris pour un motif latent. Pour les motifs d'ordre II le coefficient de φ est également indiqué.

Un coefficient de phi φ [Cramer, 1946] est calculé pour chaque association d'ordre II afin de déterminer la force et la significativité de la relation tel que :

$$\varphi = \frac{(a \times d) - (b \times c)}{\sqrt{(a+c) \times (b+d) \times (a+b) \times (c+d)}} \; et \; \varphi^2 > 3.84 \qquad (3.15)$$

Avec *a* correspondant au nombre d'observations dans lesquels sont retrouvés les modalités h_1 et h_2 ; *b* correspondant au nombre d'observations avec la modalité h_1 uniquement ; *c* correspondant au nombre d'observations contenant la modalité h_2 uniquement, et *d* le nombre d'observations ne contenant aucune de ces deux modalités (Tableau 3). La significativité de la force d'association est établie si $\varphi^2 > 3.84$.

Tableau 3 : Tableau de contingence du calcul du coefficient de φ pour deux modalités H_1 et H_2.

	h_2	$\overline{h_2}$	
h_1	a	b	a+b
$\overline{h_1}$	c	d	c+d
	a+c	b+d	Total observations

3.5.5. Spécificité de l'information

Le spectrosome des motifs permet donc de pouvoir identifier des motifs isolés, ne partageant aucune modalité avec d'autres, ou des agrégats de motifs partageant au moins une modalité. Il est ainsi mis en évidence une spécificité particulière pour chaque motif, allant d'un motif spécifiquement retrouvé dans une condition avec le motif isolé, et un motif « ubiquitaire » qui est identifié dans plusieurs associations et qui permet la formation d'un cluster. Pour chaque motif un indicateur de spécificité de l'information a alors été développé. Cet indicateur est adapté de l'indice de biodiversité de Shannon [Shannon, 1948], indice permettant de **quantifier l'hétérogénéité** de la biodiversité du milieu d'étude et d'en observer l'évolution dans le temps.

L'indicateur de la spécificité permet de caractériser la façon dont le motif est rapporté au sein de la base de données avec l'évènement de santé ciblé : spécifique d'une combinaison particulière ou ubiquitaire et permettant la formation de nombreux agrégats.

La spécificité est calculée pour chaque motif *m* par la relation :

$$S_m = \frac{H_{max} - H_m}{H_{max} - H_{min}} \tag{3.16}$$

Avec :

$$H_m = -\sum_{v=1}^{u_m} \left(\frac{a_{m,v}}{w_m} \times \log \frac{a_{m,v}}{w_m} \right) \tag{3.17}$$

La valeur de $H_{max,m}$ est donné par :

$$H_{max,m} = \log(u_m) \tag{3.18}$$

$H_{max,m}$ correspond à une configuration dans laquelle le poids du motif serait réparti de manière **uniforme** (Figure 13).

Par exemple :
un motif de poids égal à $w_m = 10$ et identifié dans $u_m = 10$ nœuds dont le poids de chaque nœuds serait égal à 1.

La valeur de $H_{min,m}$ a été définie telle que :

$$H_{min,m} = \log\left(\frac{w_m}{w_m - u_m + 1}\right) + \left(\frac{u_m - 1}{w_m} \times \log(w_m - u_m + 1)\right) \tag{3.19}$$

$H_{min,m}$ correspond à une configuration dans laquelle un motif aurait été vu **essentiellement dans un seul nœud** avec un poids important, et le reste du poids réparti sur les autres nœuds (Figure 13).

Par exemple :
un motif de poids $w_m = 10$, identifié dans $u_m = 5$ nœuds, dont 4 avec un poids de 1 et le dernier avec un poids égal à 6.

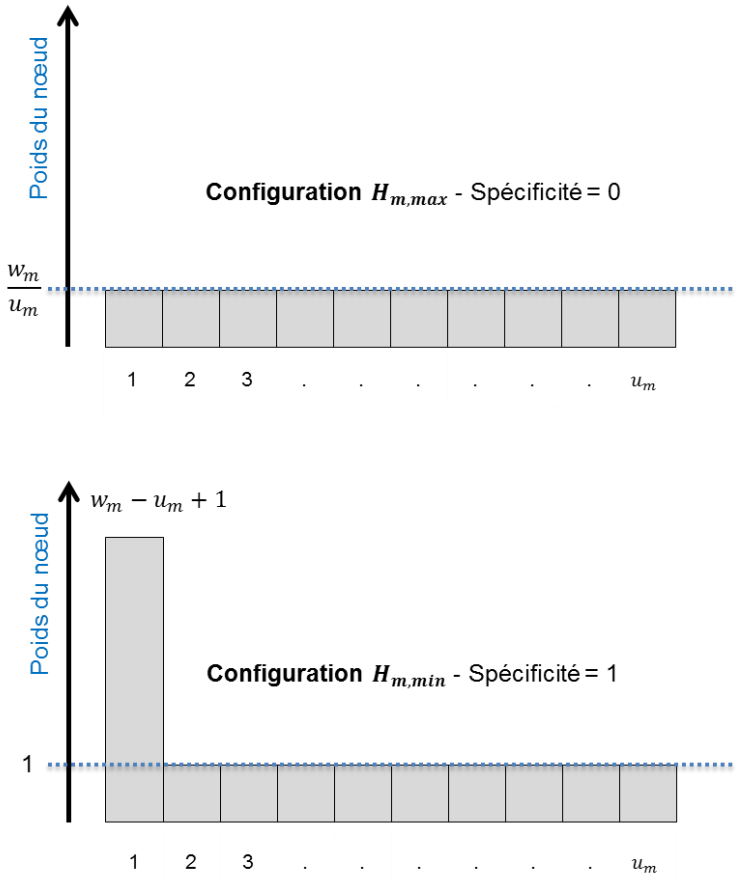


Figure 13 : Illustration des configurations correspondant aux paramètres H_{min} et H_{max} et spécificité associée.

Si $S = 1$, la spécificité met en évidence une configuration dans laquelle l'information est portée par une **seule association caractéristique** (un seul nœud du spectrosome), c'est à dire que le motif est enregistré essentiellement dans un seul nœud (cortège) comme dans le cas de $H_{min,m}$. A l'opposé, si $S = 0$, la spécificité met en évidence une configuration dans laquelle l'information **est répartie au sein de la base dans plusieurs nœuds**, ou associations. Ainsi, au-dessus du seuil de 0.5, on considère le motif comme tendant vers une association spécifique et donc un contexte spécifique, tandis qu'en dessous, c'est un motif plutôt ubiquitaire retrouvé dans de nombreuses autres associations (Figure 14).

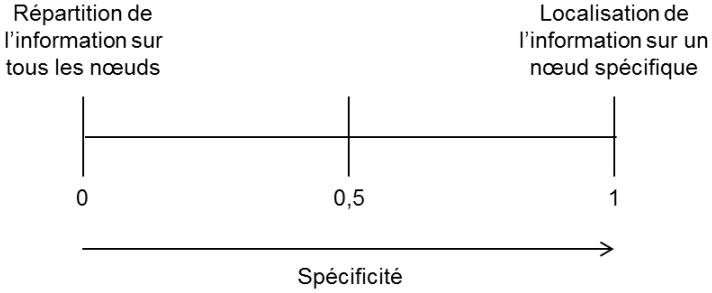


Figure 14 : Illustration de la spécificité des motifs.

3.5.6. Spectre des motifs

Le spectre des motifs représente la **signature de l'association entre la variable cible et les modalités explicatives** par la synthèse des motifs et des différents indicateurs modélisés. Chaque spectre sera différent suivant la variable cible et les variables explicatives primaires étudiées.

Tous les motifs sont représentés et classés par leur spécificité, leur statut dynamique, leur poids et leur ordre (Figure 15). Il est ainsi possible d'identifier rapidement des groupes de motifs spécifiques, ou très représentés ou très actifs.

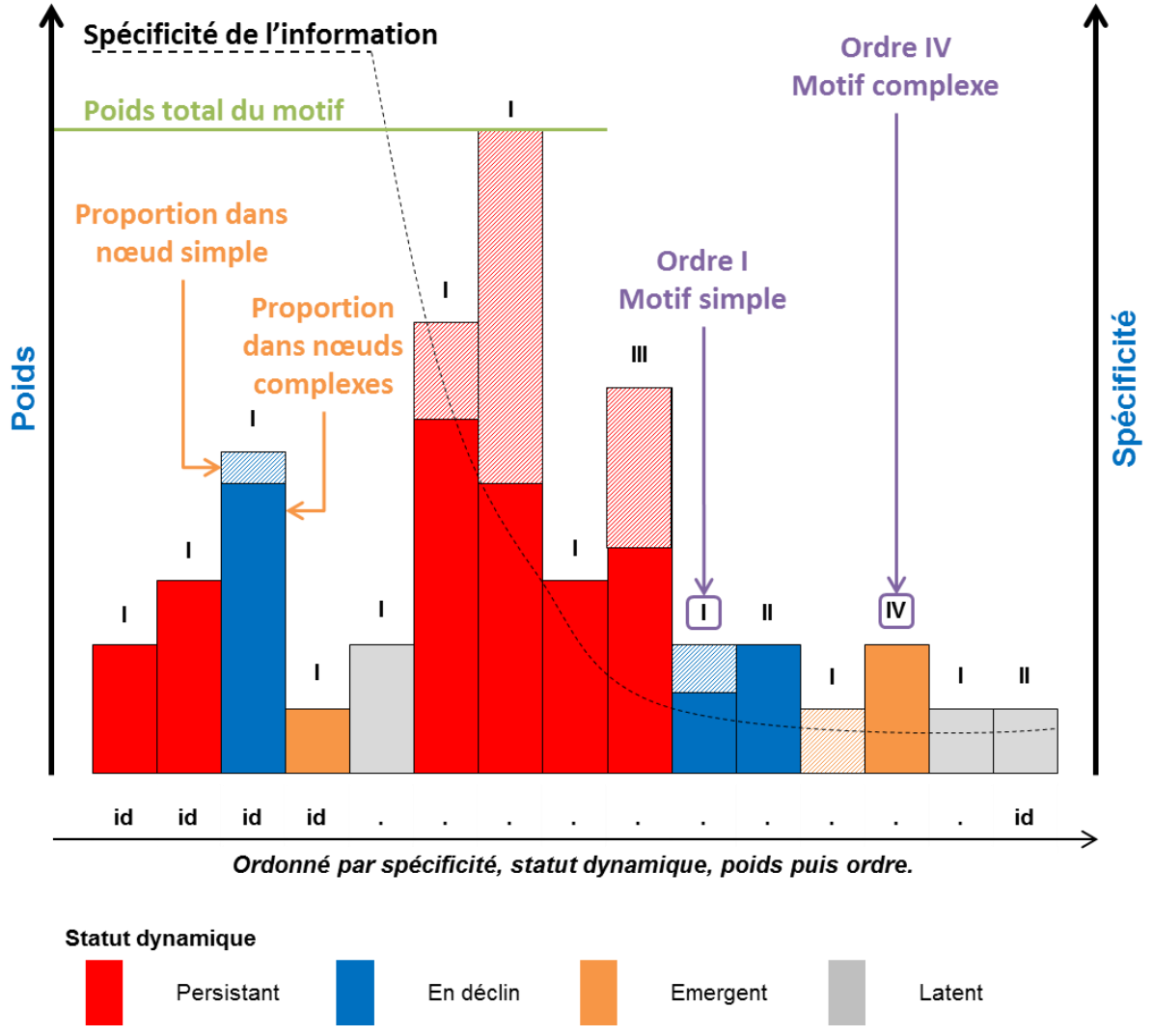


Figure 15 : Schéma de la construction du spectre des motifs.

La Surveillance Observationnelle permet l'établissement de différents indicateurs servant de grille de lecture au **spectre final**, signature de l'évènement de santé étudié. L'**exposome** permet d'identifier les relations et les motifs spécifiques, le **spectrosome et l'arbre de la multi-association** permettent l'identification spécifique des motifs simples et d'ordre supérieur et les différentes **trajectoires dynamiques** des motifs permettent de comprendre le statut dynamique actuel de chacun des motifs.

3.5.7. Matrice primaire-secondaire

Une fois les motifs construits et caractérisés, il est alors possible de les étudier en fonction des autres variables explicatives secondaires disponibles au sein de l'échantillon initial (Figure 4) grâce à la construction de la ***matrice primaire-secondaire***. Soit la matrice en 3 dimensions $M_{PS} = M \times D_{sec} \times T$, permettant d'effectuer le suivi de l'ensemble des motifs M et d'une variable explicative secondaire D_{sec} en fonction du temps T.

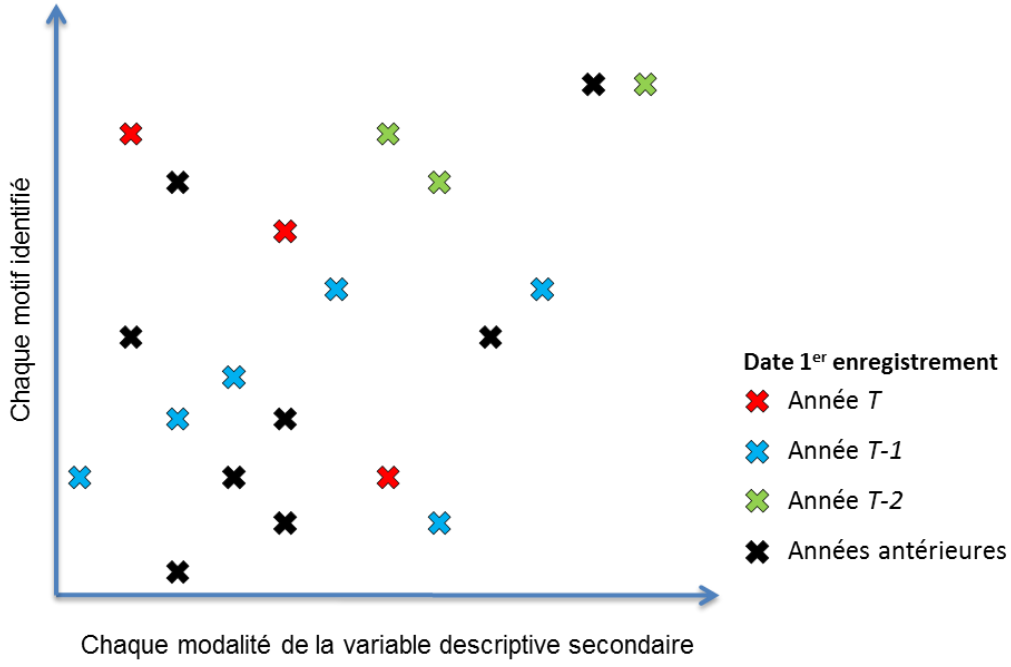


Figure 16 : Schéma de la construction de la matrice primaire-secondaire.

Par exemple :

Dans le cas d'une surveillance appliquée à une pathologie professionnelle, les motifs renvoyant aux expositions professionnelles identifiées, il sera alors possible d'étudier la présence ou l'absence des motifs au sein des activités professionnelles considérées comme variable explicative secondaire.

3.6. Récapitulatif des indices de la Surveillance Observationnelle

Indices	Intitulés	Partie consacrée
N	Nombre d'observations (lignes) au sein de l'échantillon correspondant à l'évènement de santé ciblé	3.2
H, h	Ensemble des modalités identifiées *h* en association avec l'évènement de santé étudié correspondant aux variables explicatives	3.2
I	Variable d'imputabilité pouvant accompagner les modalités explicatives	3.2
$\boldsymbol{n_i}$	Nombre d'occurrences de la modalité *i*	3.3.1
T	Plage temporelle totale couverte par l'étude	3.2
$\boldsymbol{\Delta t}$	Plage temporelle de référence	3.3.1
$\boldsymbol{n_{i,\Delta t}}$	Nombre d'occurrences de référence de la modalité *i* sur la plage temporelle de référence Δt	3.3.1
$\boldsymbol{RR_i}$	Risque relatif de la modalité *i* pour le calcul de la significativité par occurrence	3.3.1.1
$\boldsymbol{p_i}$	Probabilité de la modalité *i* dans le calcul de la significativité par imputabilité	3.3.1.2
$\boldsymbol{n_{i,\Delta t,cr}}$	Nombre d'occurrences de référence critique de la modalité *i* sur la plage temporelle de référence Δt	3.3.1.2
$\boldsymbol{\lambda, \lambda_{min}, \lambda_{max}, \nu}$	Paramètre λ de la loi de Poisson utilisée pour le calcul de la significativité par imputabilité et paramètres ν, λ_{min} et λ_{max} associés	3.3.1.2.1 ; 3.3.1.2.2
d	Nombre de modalités maximal autorisé pour la formation d'un cortège : longueur du cortège	3.3.2.1
V, v	Ensemble des nœuds *v* correspondant aux différents cortèges distincts identifiés	3.3.2.2
$\boldsymbol{a_v}$	Poids du nœud *v*	3.3.2.2
C,c	Ensemble des combinaisons *c* formés à partir des nœuds *v*	3.3.2.3
$\boldsymbol{W_{c,i}}$	Poids de la combinaison *i* correspondant à la somme totale des poids des nœuds dans lesquelles elle a été identifiée	3.3.2.3
$\boldsymbol{q_i}$	Condition à valider par la combinaison *i* afin d'être considérée comme motif représentatif	3.3.2.3
M,m	Ensemble de motifs *m* correspondant aux combinaisons représentatives construites à partir des nœuds	3.3.2.3
$\boldsymbol{W_{m,i}}$	Poids du motif *i*	3.3.2.3
$\boldsymbol{u_m}$	Nombre de nœuds dans lesquels le motif a été identifié	3.3.2.3
$\boldsymbol{RI_{\Delta t,m}}$	« Reporting Index » calculé de façon pérenne ou sur un intervalle Δt donné	3.5.2
A, B, C, D	Statut dynamique du motif calculé en fonction des R_∞ et R_2	3.5.2
$\boldsymbol{\varphi}$	Coefficient d'association calculé pour chaque motif d'ordre II	3.5.4
S	Spécificité calculé pour chaque motif selon la distribution de son information au sein de l'échantillon	3.5.5
$\boldsymbol{H, H_{min}, H_{max}}$	Indices de Shannon pour le calcul de la spécificité de chaque motif	3.5.5

RESUME

La Surveillance Observationnelle appliquée aux bases de données observationnelles permet d'obtenir un ensemble d'indicateurs permettant de caractériser un évènement spécifique de santé en termes de temporalité et de multi-informations. Ainsi à partir d'une base de données et d'un échantillon extrait, la Surveillance Observationnelle fournit :

- Un **exposome des nœuds** permettant la restructuration des données sous forme d'un réseau de nœuds pour l'application de la Surveillance Observationnelle ;
- L'identification des modalités ou combinaisons de modalités caractéristiques de l'évènement de santé appelés **motifs ;**
- La mise en évidence de la **mono ou multi-information** de l'évènement ciblé par l'étude de la longueur des motifs ;
- La structure des associations existantes entre les motifs via le **spectrosome ;**
- La caractérisation **dynamique** de ces motifs : persistance, émergence, déclin, latence ;
- La caractérisation de leur mode de recrutement via le calcul de la **spécificité.**

Ces informations sont alors résumées sur la signature caractéristique de l'évènement de santé : **le spectre des motifs.**

4. CREATION DE L'OUTIL « OSA »

Dans le cadre d'une utilisation en routine de la Surveillance Observationnelle, la création d'un outil polyvalent pouvant être généralisable et applicable aux différentes bases de données s'est rapidement révélé nécessaire. Cet outil a été nommé **OSA pour « Observational Surveillance Analysis »**.

4.1. Choix de la plateforme d'implémentation de la méthodologie

Il a été choisi d'implémenter la méthodologie dans le logiciel R [R Core Team, 2013], logiciel libre de traitement de données, d'analyses statistiques, de modélisation et de programmation mettant en œuvre le langage de programmation S. Ce logiciel est dit « open source » : le code source est à la disposition du grand public et une collaboration importante entre les programmeurs permet l'amélioration du code source et le partage des changements avec les différents membres de la communauté d'utilisateurs. Cette communauté propose régulièrement de nouvelles mises à jour ainsi que des améliorations via des « *packages* », collections de fonctions, de données et de codes, qui permettent d'apporter des méthodologies supplémentaires au logiciel de base. Le logiciel R est également compatible avec le langage « *LaTeX* », langage et composeur de texte permettant le formatage de PDF et l'obtention des résultats mis en forme de manière automatique et standardisée.

Le langage R est « orienté objet », ou « interprété », et les traitements s'effectuent via des lignes de commande. L'utilisation du logiciel nécessite un minimum de connaissance dans la syntaxe de ce langage, c'est pourquoi la création d'une interface a été nécessaire pour une meilleure utilisation en routine par des utilisateurs non familiarisés avec le logiciel.

4.2. Organisation du programme

L'ensemble de la méthodologie a été implémenté dans des « scripts », représentant un ensemble de commandes et de fonctions servant d'une part à la création de l'interface et d'autre part au déroulement de l'analyse d'un point de vue calculatoire. Pour le bon fonctionnement de l'outil, chaque étape a été développée dans des scripts indépendants permettant un chargement plus fragmenté et une rapidité d'exécution plus fluide.

4.2.1. Description générale

Le programme est divisé en 4 grandes parties exécutrices (Figure 17) : la partie **d'introduction ou d'accueil,** la partie de **paramétrage,** celle de **traitement** puis celle **d'exportation des résultats**, réparties sur un total de 23 scripts :

- La partie ***introduction***, répartie sur 2 scripts, permet le chargement global des *packages* et de l'interface, la définition des paramètres par défaut, la gestion de l'environnement de travail et des importations de tableur par l'utilisateur (*Voir 4.2.2, p.82*).

- La partie ***paramétrage***, répartie sur 4 scripts, permet de gérer les paramètres d'extraction d'un sous-échantillon (si besoin), de paramétrer les éléments de l'analyse souhaités (indicateurs) ainsi que de paramétrer les seuils de significativité des calculs (*Voir 4.2.3, p.85*).

- La partie ***traitement***, répartie sur 6 scripts, permet le fonctionnement de l'analyse en elle-même en effectuant l'extraction, le traitement des données, la création des nœuds et des motifs et les calculs des indicateurs nécessaires pour la suite des résultats (*Voir 4.2.4, p.91*).

- La partie ***résultat***, répartie sur 11 scripts, permet de générer chacun des indicateurs de la Surveillance Observationnelle selon les choix effectués lors du paramétrage. Elle permet également de générer des graphiques supplémentaires ainsi qu'un rapport d'analyse automatique et standardisé (*Voir 4.2.5, p.93*).

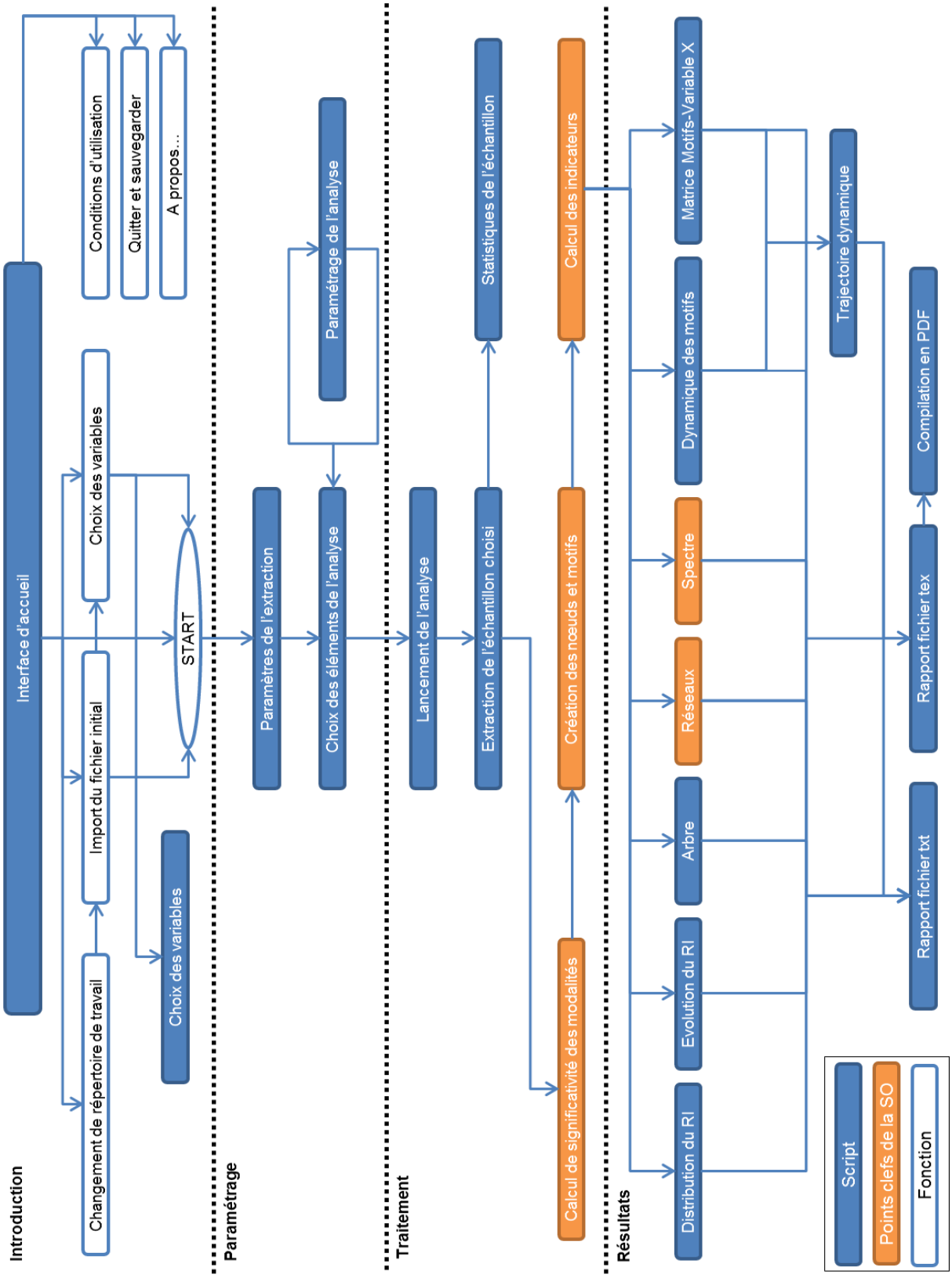


Figure 17 : Schéma général de l'application OSA. Le programme est scindé en 4 parties distinctes : l'introduction et la mise en place de l'interface, la partie paramétrage de l'analyse et des phases calculatoires, la partie calculatoire et la partie génération des résultats.

4.2.2. Description détaillée de la partie d'introduction

La première partie du programme représente l'accueil et consiste en la mise en place des outils nécessaires à la création de l'interface, mais également à la définition des paramètres initiaux de l'analyse, qui seront modulables par la suite (Figure 18). Le programme propose une interaction avec l'utilisateur pour la définition d'un environnement de travail (dossier spécifique dans lequel tous les fichiers de résultats seront exportés), l'importation d'un fichier (base de données ou simple échantillon), le choix des variables (identification, temporelle, variable cible et explicatives, imputabilité). Enfin, sont proposés des informations génériques sur l'analyse, les conditions d'utilisation et des informations concernant l'équipe de développement.

4.2.2.1. Les packages et paramètres initiaux

L'application est programmée afin de télécharger et charger automatiquement les *packages* nécessaire au bon déroulement de l'analyse si ceux-ci sont absents de la machine utilisateur : « statnet » [Handcock, 2015] pour les réseaux (exposome et spectrosome) et « xtable » pour la mise en forme de tableau. Le *package* nommé « tcltk », présent dans la version de base du logiciel est également chargé afin de permettre la création de l'interface et la gestion des différentes interactions avec l'utilisateur.

Les paramètres initiaux de l'analyse, pouvant être modifiés, ont été fixés empiriquement tels que :

- λ_{min}, fixé à 1.37 pour un $n_{i,\Delta t,cr} = 1$;
- la longueur maximum *d* des nœuds et motifs fixée à 5 ;
- le choix de l'extension des résultats en « PDF » pour une meilleure définition ;
- le nombre de digits (caractères) à utiliser lors de l'identification des liens de l'exposome et du spectrosome afin d'alléger la légende, fixé à 5.

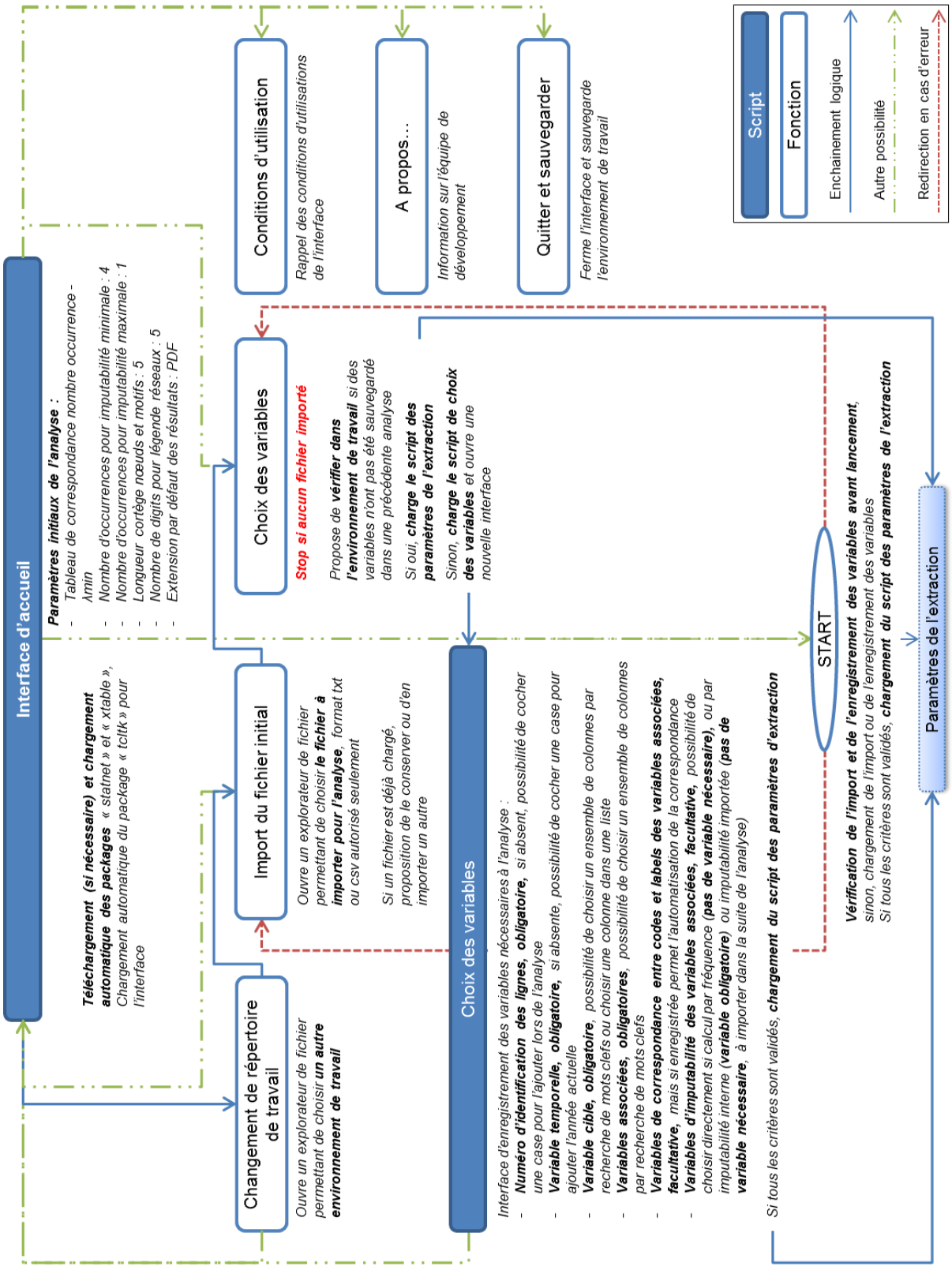


Figure 18 : Schéma détaillé des scripts d'introduction de l'outil OSA. Script de lancement de l'interface, importation de fichier, choix des variables et gestion du répertoire de travail.

4.2.2.1. Les options utilisateur

Afin de ne pas surcharger l'application avec un trop grand nombre de *packages*, l'importation de fichiers est filtrée afin d'afficher uniquement les fichiers dont l'extension est reconnue par défaut par le logiciel R : « csv » et « txt ». Le programme est également défini afin de stopper l'analyse dans le cas où les conditions d'importation et de définition des variables ne seraient pas remplies. Afin d'éviter des manipulations à répétition lors d'analyses multiples, le programme propose également de vérifier l'existence dans l'environnement de travail d'un fichier déjà importé ainsi que des variables stockées permettant de procéder directement au paramétrage de l'analyse (*voir 4.2.3, p.85*).

Dans le cas où aucune variable ne serait enregistrée au sein de l'environnement de travail, le programme permet d'en sélectionner de nouvelles en proposant la liste des variables identifiées dans le fichier importé. Il est nécessaire de remplir plusieurs variables à caractères obligatoires et il est possible d'ajouter des variables complémentaires pour l'analyse.

- **Les variables obligatoires :**
    - ***Identification :*** l'identification des observations (lignes) est primordiale pour l'analyse. Cette variable existe dans la plupart des bases de données, cependant, dans le cas contraire, il est possible d'ajouter cette variable.
    - ***Temporelle :*** la variable temporelle est nécessaire pour le calcul des indicateurs dynamique. Dans le cas où cette variable ne serait pas disponible, il est possible d'ajouter une variable contenant l'année en cours, cependant, les résultats concernant le dynamisme des motifs ne pourront être pris en compte.
    - ***Variable cible :*** Cette variable est indispensable afin de permettre le choix de la modalité cible sur laquelle sera appliquée la Surveillance Observationnelle. Il est possible de choisir une seule variable dans une liste ou bien un ensemble de variables correspondant à cette variable cible par recherche de mots clefs.
    - ***Variable explicatives primaires :*** Les variables explicatives primaires sont également obligatoires. Il est possible de choisir l'ensemble des variables par recherche de mots clefs également.

- **<u>Les variables facultatives :</u>**
    - ***Identification des codes :*** Les variables permettant une correspondance entre les codes des variables explicatives et leur label peuvent également être indiquées à cette étape du traitement. Dans le cas où elles n'existeraient pas, il est possible d'importer un fichier de correspondance avant le lancement de l'analyse.
    - ***Variable d'imputabilité :*** Les variables permettant d'indiquer une force d'association entre la variable cible et les variables explicatives primaires, telle que l'imputabilité, peuvent être indiquées à ce moment du traitement. Si les variables sont renseignées, le calcul de la significativité sera automatiquement paramétré sur un calcul par imputabilité (*voir 3.3.1.2, p.53*). Si les variables ne sont pas renseignées, l'utilisateur a le choix de demander un calcul par fréquence (*voir 3.3.1.1, p.53*), ou par imputabilité tout de même, sous réserve qu'il importe un fichier contenant des informations sur l'imputabilité avant le lancement de l'analyse.

Enfin, il est possible à l'utilisateur à partir de cette fenêtre de :

- quitter et de sauvegarder l'environnement de travail correspondant à son analyse, environnement de travail qui permettrait ainsi lors d'une prochaine analyse de ne pas avoir à réimporter un fichier ou redéfinir les variables ;
- lancer la suite du programme de la Surveillance Observationnelle.

4.2.3. Description détaillée de la partie de paramétrage

La seconde partie du programme concerne le paramétrage de l'analyse, c'est-à-dire la définition des paramètres d'extraction (si nécessaire) et des paramètres pour le calcul de la significativité, la longueur des nœuds et des motifs désirés, et enfin le choix des indicateurs souhaités (Figure 19).

4.2.3.1. Paramètres d'extraction

Il est possible de travailler avec un fichier déjà prêt, dans ce cas, seule la définition d'un nom d'identification de l'analyse est souhaitée, ou d'extraire un sous-échantillon concernant

l'évènement d'intérêt à partir de la base de données importée. Dans ce cas précis, en plus du nom de l'analyse, des critères d'extraction sont requis :

- ***Les codes de la modalité de la variable cible*** correspondant à l'évènement d'intérêt désiré[1]. Il est possible d'indiquer plusieurs codes à condition qu'ils possèdent le même nombre de caractères (digits). De même, il est possible de choisir entre une modalité spécifique correspondant à ce code, ou bien un ensemble de modalité (famille) commençant par ce code.

> *Par exemple :*
> Dans le cas d'une analyse effectuée sur une base de données de pathologies, il serait possible d'extraire tous les cas de tumeurs du sang (famille), ou bien uniquement les leucémies (modalité spécifique).

  Il est également possible d'ajouter d'autres codes contenant un nombre de digits différents, ou bien de supprimer une modalité spécifique d'une famille.

> *Par exemple :*
> Dans le cas de l'analyse sur les tumeurs du sang, l'utilisateur pourrait choisir d'extraire tous les cas, sauf ceux concernant les leucémies.

- ***La plage temporelle*** désirée pour l'analyse, par défaut, le programme propose la plage temporelle allant de la plus ancienne année identifiée à la plus récente.

- ***Des variables explicatives complémentaires.*** Le programme propose également de pouvoir extraire selon deux autres variables si cela est nécessaire afin de permettre des analyses spécifiques.

> *Par exemple :*
> Il est possible d'extraire des observations en fonction du genre ou de l'espèce des individus, d'un centre de recueil particulier, d'un mode de recueil, etc.

[1] Il est préférable de travailler directement sur des codes plutôt que du « texte » pour éviter les problèmes de compatibilité à des niveaux nationaux par exemple, pour lesquels il existe plusieurs opérateurs.

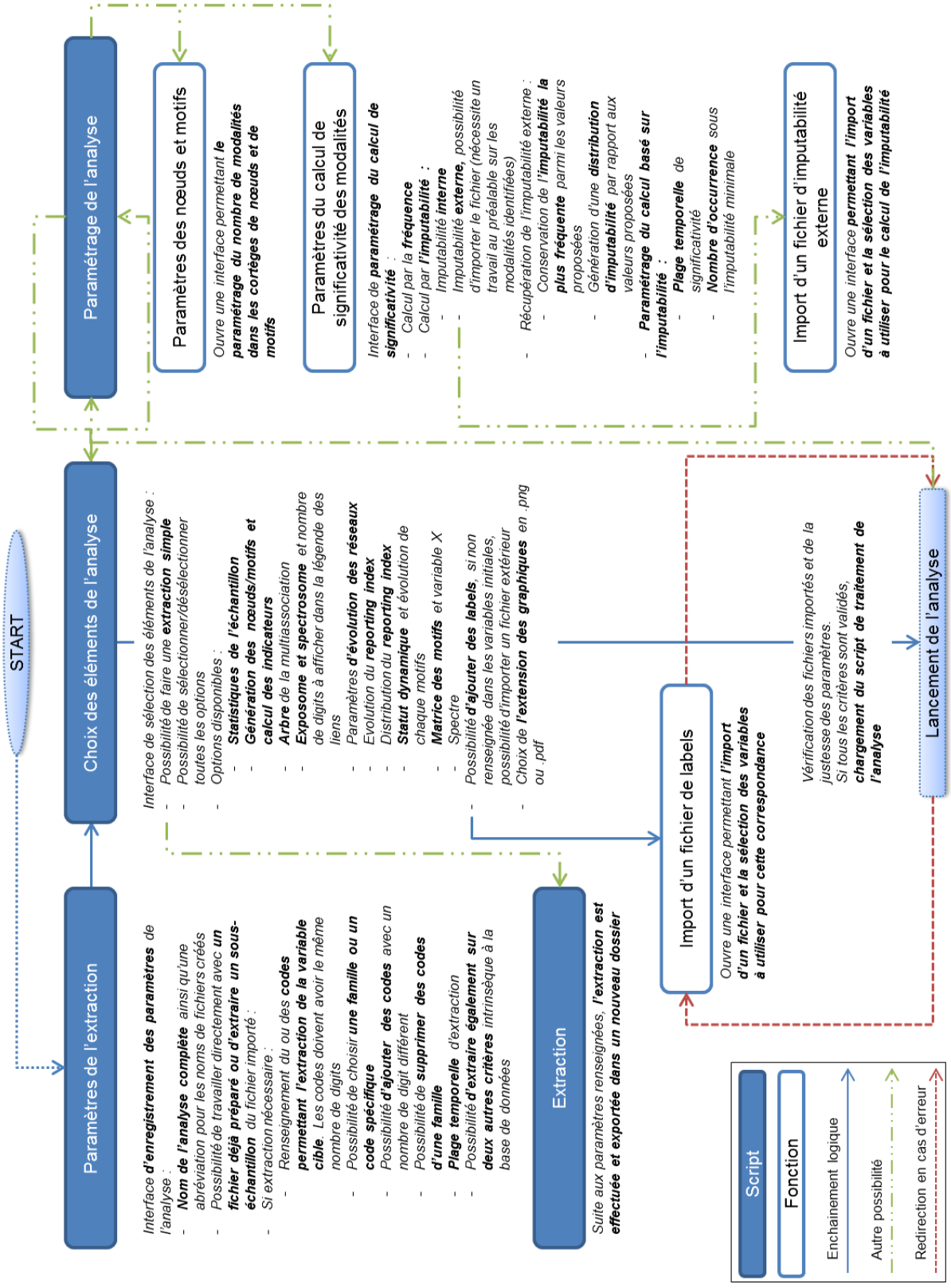


Figure 19 : Schéma détaillé des scripts de paramétrage de l'analyse OSA. Paramètres d'extraction et paramètres des calculs de l'analyse.

4.2.3.2. Paramètre de l'analyse

Le paramétrage de l'analyse va permettre à l'utilisateur d'une part de choisir la longueur *d* maximale des cortèges des nœuds et des motifs qui vont être construits puis analysés dans le processus de la Surveillance Observationnelle, et d'autre part de paramétrer le calcul de la significativité des nuisances.

Il y a 4 grandes possibilités de calcul de significativité des nuisances :

- ***Calcul basé sur la fréquence*** des modalités : ce calcul de risque relatif s'effectue uniquement sur la fréquence de chacune des modalités les unes par rapport aux autres et ne nécessitent aucun apport particulier de fichier (*Voir 3.3.1.1, p.52*).

- ***Calcul basé sur l'imputabilité interne*** du fichier : ce calcul est effectué automatiquement si l'utilisateur a renseigné l'existence de variables d'imputabilité entre la variable cible et les variables explicatives primaires (*Voir 3.3.1.2, p.53*).

- ***Calcul basé sur l'imputabilité externe***: ce calcul nécessite l'apport d'un fichier externe contenant des variables renseignant une force d'association entre les différentes modalités des variables explicatives et la variable cible à expliquer. Cela nécessite une extraction et un prétraitement au préalable afin de prendre connaissance de toutes les modalités rapportées comme associées à l'évènement d'intérêt et de leur attribuer une force d'association *a posteriori*. Cette étape peut être effectuée par plusieurs personnes donnant ainsi une distribution de force d'association.
  A partir de cette distribution d'imputabilité, il existe deux possibilités d'exploitation[2] :

  - ***Majoritaire :*** l'imputabilité d'une modalité correspond à la valeur d'imputabilité la plus fréquente parmi la distribution. En cas d'égalité, il est alors possible de choisir de privilégier un risque fort en privilégiant une imputabilité forte (pessimiste), ou au contraire de privilégier une imputabilité faible (optimiste).

  - ***Distribution :*** à partir de la distribution d'imputabilité, une nouvelle distribution d'imputabilité est générée, dont la taille est égale à la fréquence de la modalité n_i.

[2] Cette notion sera abordée de manière plus détaillée dans la partie IV avec l'exemple de la base IDEWE

Pour le calcul de l'imputabilité, il est ensuite possible de paramétrer les seuils critiques de significativité, c'est-à-dire l'intervalle temporelle de référence Δt pour le calcul du nombre de recrutement de référence $n_{i,\Delta t}$ (*équation (3.1)*) ainsi que les occurrences minimales nécessaires pour être considéré significatif sous l'imputabilité minimale (*Voir 3.3.1.2, p.53*), le programme faisant automatiquement la correspondance en valeur de λ_{min} pour le calcul (*équation (3.5)*).

4.2.3.3. Choix des indicateurs

Le programme a été construit dans le but de pouvoir proposer la liste complète des indicateurs développés dans le cadre de la SO mais également de pouvoir proposer des analyses « à la carte ». Il est ainsi possible d'effectuer une extraction simple d'un sous-échantillon, en vue d'analyse future, ou bien de lancer une analyse fragmentée ou complète. L'analyse est programmée pour gérer les dépendances entre les différents indicateurs, et certaines étapes sont paramétrées pour s'effectuer sans intervention de l'utilisateur : le calcul de la significativité des modalités, la création des motifs et le calcul des indicateurs lorsqu'une sortie graphique est demandée. Les différents choix de l'utilisateur sont alors :

- ***Extraction simple :*** le sous-échantillon correspondant aux critères enregistrés d'extraction est extrait et exporté dans un dossier et tableur nommé d'après le nom de l'analyse choisi par l'utilisateur.

- ***Statistiques de l'échantillon :*** une exportation graphique du recrutement et de l'effectif cumulé du nombre d'observations dans l'échantillon, du nombre distinct de codes de la variable cible, et enfin le nombre distinct de modalités identifiées au sein des variables explicatives primaires associées à la variable cible.

- ***Génération des nœuds et motifs :*** étape obligatoire pour toute sélection d'indicateurs spécifiques à la Surveillance Observationnelle. Lorsque cette étape est lancée, le calcul des indicateurs est alors automatique : coefficient d'association φ, « Reporting Index » pour le statut dynamique (détermination des motifs actifs et inactifs) et spécificité en fonction de l'indice de biodiversité de Shannon.

- ***Graphique des indicateurs*** de la Surveillance Observationnelle (*voir 3.5, p.63*) :

  - ***Evolution du « Reporting Index » :*** exportation graphique de l'évolution du RI_2 glissant pour les motifs actifs ;

  - ***Distribution du « Reporting Index » :*** exportation graphique du classement des motifs par leur RI_2 et leur RI_∞ ;

  - ***Statut dynamique des motifs*** : exportation graphique du statut dynamique de tous les motifs pour la dernière année considérée en fonction des RI_2 et RI_∞ et leurs seuils associés ξ_{RI} ;

  - ***L'arbre de la multi-association*** : exportation graphique avec indication du statut dynamique des motifs actifs et affichage de la significativité des motifs d'ordre I à l'état de modalité ;

  - ***Exposome des nœuds et/ou spectrosome des motifs :*** possibilité d'exporter graphiquement l'exposome des nœuds et/ou le spectrosome des motifs. Le réseau étant généré par un algorithme sa spatialisation peut parfois différer d'une analyse à une autre, tout en conservant les mêmes paramètres internes. L'utilisateur a alors la possibilité de choisir la spatialisation qui lui convient le mieux (placement des labels notamment). Un tableur est exporté contenant les informations sur les différents liens identifiés du réseau ;

  - ***Paramètres de l'exposome et spectrosome :*** il est également possible d'obtenir les paramètres de l'évolution de l'exposome et spectrosome, même si ceux-ci n'ont pas été exportés graphiquement ;

  - ***Spectre des motifs :*** exportation graphique du spectre des motifs de l'évènement d'intérêt étudié permettant une synthèse globale des indicateurs des motifs identifiés : statut dynamique, spécificité, poids, ordre, simple ou en association. Possibilité d'obtenir le spectre de la totalité des motifs et/ou le spectre des motifs actifs uniquement ;

  - ***Matrice primaire-secondaire*** : l'utilisateur a la possibilité d'exporter la matrice primaire-secondaire permettant le croisement entre les motifs et la variable explicative secondaire au choix présente dans l'échantillon.

- ***Correspondance des codes et labels :*** Si la base de données importée ne proposait pas de variables permettant la correspondance entre les codes de standardisation utilisés et un libellé, l'utilisateur a la possibilité d'importer un fichier externe permettant d'effectuer ces correspondances.

- ***Choix de l'extension graphique :*** l'utilisateur a la possibilité de choisir une extension des résultats graphiques sous format « PDF » (par défaut) ou sous format « PNG ».

4.2.4. Description détaillée de la partie de traitement

La partie traitement du programme correspond à l'analyse spectrale de la Surveillance Observationnelle, mettant en application toutes les requêtes de l'utilisateur en fonction des paramètres précédemment enregistrés (Figure 20). Cette partie a la particularité de permettre à l'utilisateur de suivre en temps réel l'évolution de l'analyse en lui donnant les chiffres clefs ainsi qu'en sollicitant de sa part des informations nécessaire au déroulement.

4.2.4.1. Calcul de la significativité des modalités

Une fois l'extraction effectuée, le programme effectue l'exportation du graphique de statistique (si requis) ainsi que le traitement sous-jacent du calcul concernant la significativité des différentes modalités distinctes identifiées. Ce calcul dépend du choix de l'utilisateur : un calcul par le risque relatif, ou bien un calcul basé sur une variable d'imputabilité (*Voir théorie : 3.3.1, p.52 et script : 4.2.3.2, p.88*). Une fois les modalités significatives identifiées, une étape de « nettoyage » est appliquée afin de réduire la taille des cortèges des nœuds selon la taille maximale désirée, les modalités ayant la plus grande imputabilité (ou la plus grande fréquence selon la méthode de calcul) sont conservées et les doublons possibles sont corrigés.

4.2.4.2. Création des nœuds et des motifs

Les nœuds, associations distinctes de modalités des variables explicatives, sont construits, et permettent la génération de toutes les combinaisons (*Voir 3.3.2 p.57*). A partir de toutes les combinaisons possibles, les motifs représentatifs sont identifiés (*Voir 3.3.2.3, p.59*), puis conservés dans un tableur permettant leur analyse dans la suite du traitement. De même, toutes les modalités distinctes ayant servies à la construction des motifs significatifs sont conservées et représentent l'ensemble des motifs d'ordre I.

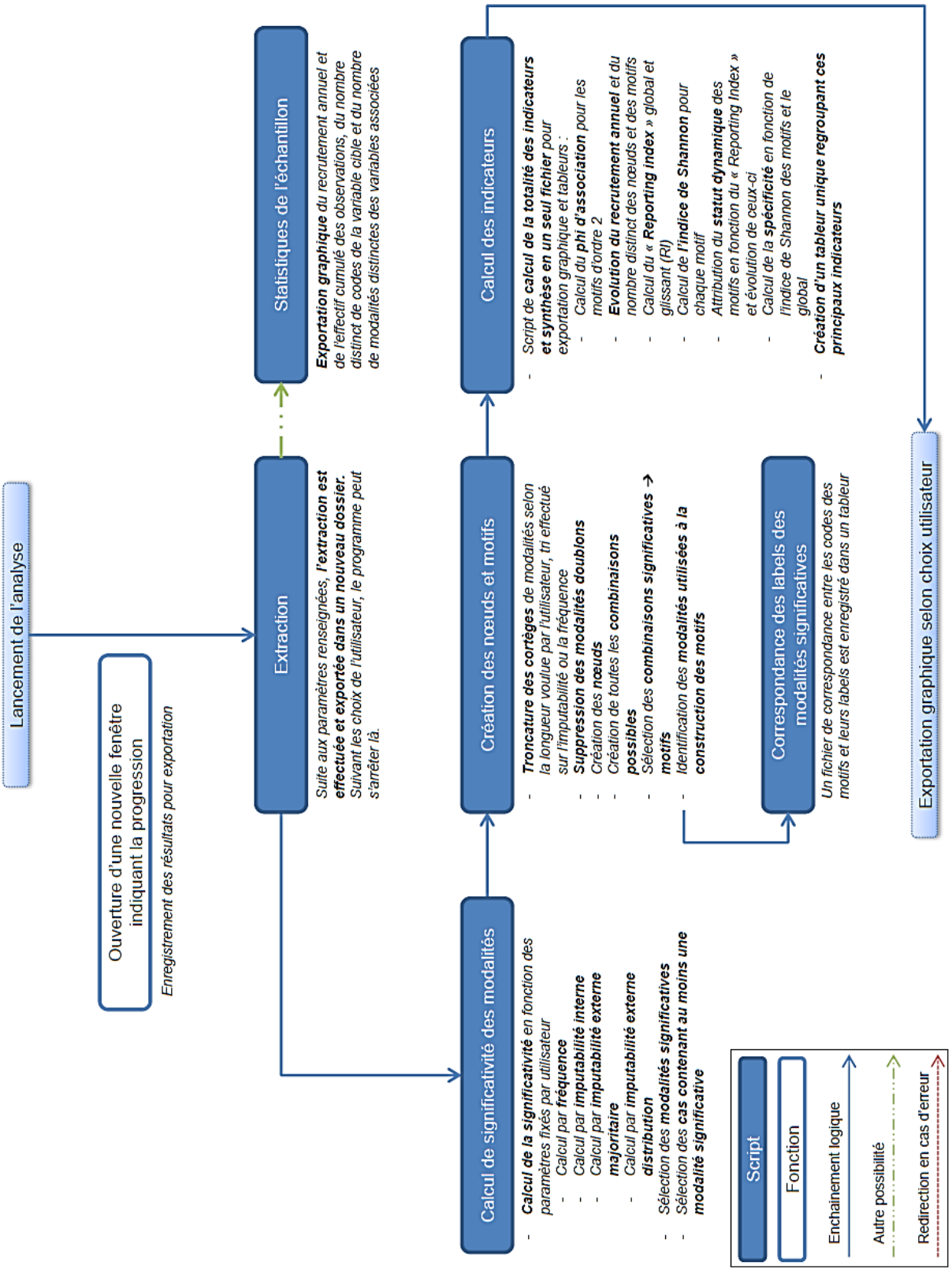


Figure 20 : Schéma détaillé des scripts de traitement de l'analyse OSA. Extraction, statistiques, calcul de significativité, création des nœuds et motifs, indicateurs.

4.2.4.3. Calcul des indicateurs

Les indicateurs des motifs identifiés sont calculés dans un unique et même script et serviront à la conception de tous les graphiques des indicateurs sélectionnés par l'utilisateur :

- Calcul du coefficient ***φ d'association*** pour les motifs d'ordre 2 : Indicateur utilisé pour l'arbre de la multi-association (*Voir 3.5.4, p.70*) ;

- Calcul du ***« Reporting Index »*** : en fonction du recrutement annuel (incidence) et de l'effectif cumulé des motifs précédemment calculé, les deux valeurs de RI_∞ et RI_2 sont modélisés pour chaque motif et chaque année, permettant d'obtenir la valeur seuil ξ_{RI} et les statuts dynamique, puis de générer les graphiques suivant : évolution du RI_2 de chaque motif et distribution de RI_∞ et RI_2 (*Voir 3.5.2, p.65*) ;

- Calcul de la ***Spécificité S en fonction des indices de biodiversité de Shannon*** de chaque motif *m* : H_m, $H_{min,m}$ et $H_{max,m}$ (*Voir 3.5.5, p.71*).

Un tableur regroupant les principaux indicateurs est généré et exporté dans le dossier créé pour l'analyse.

4.2.5. Description détaillée de la partie des exportations des résultats

Cette partie du programme correspond uniquement à l'exportation des graphiques standardisés ayant été choisi par l'utilisateur lors du paramétrage en utilisant le fichier unique de synthèse des indicateurs de la SO générée lors de l'étape de calcul précédente. L'utilisateur aura la possibilité de suivre l'évolution de l'analyse, de préciser dans le cas d'une imputabilité externe majoritaire de privilégier les valeurs fortes ou faibles de l'imputabilité (*Voir 4.2.3.2, p.88*), de générer différents exposomes/spectrosomes afin d'obtenir la meilleure spatialisation des nœuds et labels et de générer ou non les trajectoires de tous les motifs en fonction du temps.

Une fois les graphiques exportées, l'utilisateur en est informé et peut alors décider d'ajouter des graphiques supplémentaires descriptifs ou d'exporter des spectrosomes ciblés sur des motifs en particulier. Le programme offre également la possibilité d'exporter un rapport automatique rappelant les principales informations de l'analyse, sous forme de bloc-notes, ou sous forme de PDF si l'utilisateur dispose d'une librairie *LaTeX*.

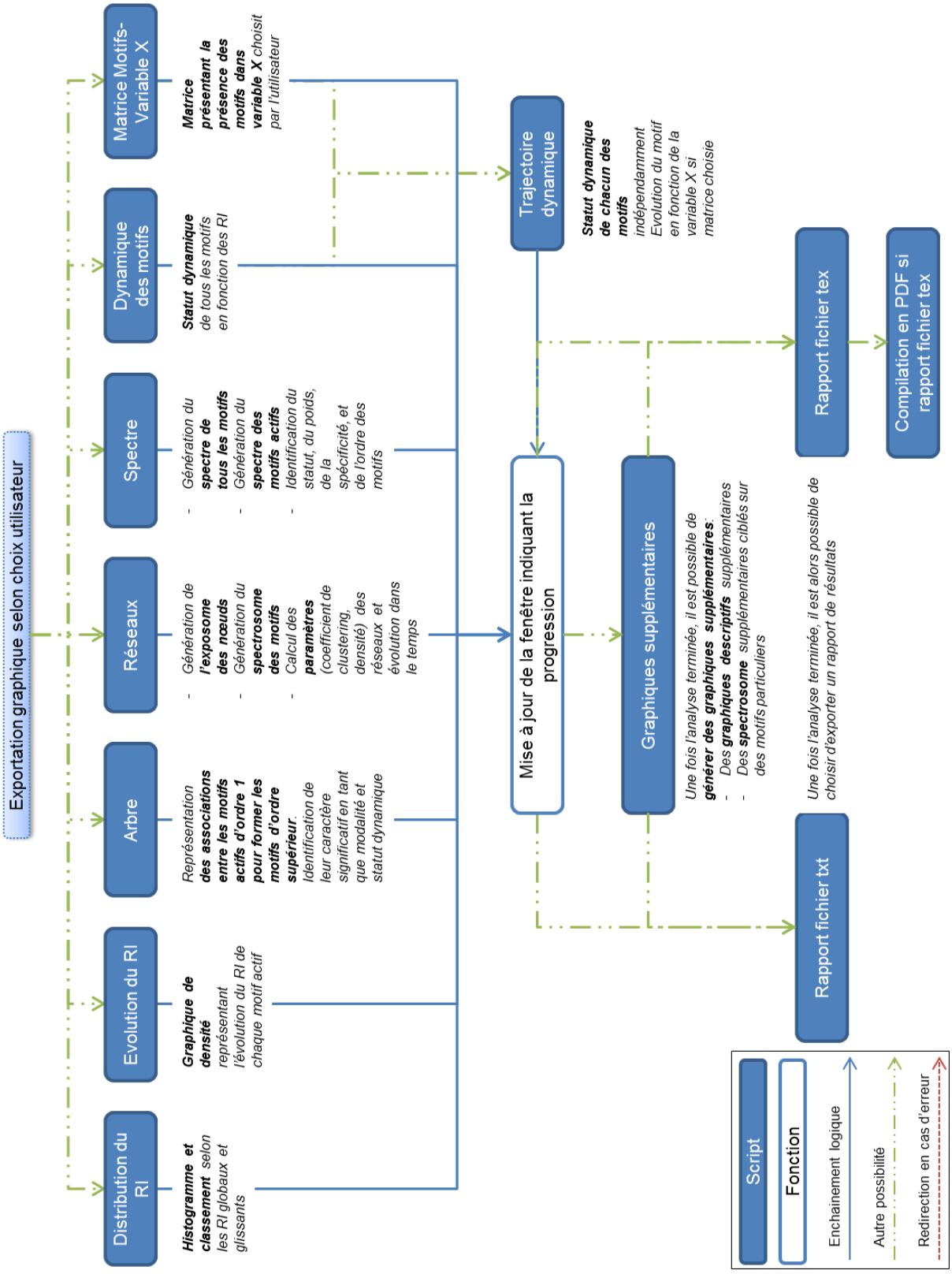


Figure 21 : Schéma détaillé des scripts de traitement des sorties graphiques et des rapports de résultats de l'outil OSA. Traitement des scripts choisis par utilisateur, possibilité de générer un rapport style bloc note ou un pdf si bibliothèque *LaTeX* à disposition.

5. MODE D'EMPLOI DE L'OUTIL « OSA »

L'interface a été développé grâce au *package* « tcltk » qui permet de créer une interface utilisateur personnalisée et permettant l'utilisation des différents scripts composant le programme. **Le langage choisi pour le développement de cet outil « OSA » est l'anglais**, permettant une utilisation lors de collaboration internationale par exemple. Ce chapitre a pour objectif de présenter de manière pratique le déroulement d'une analyse effectuée avec OSA.

5.1. Conditions d'utilisations

Afin de pouvoir utiliser cette interface en routine, il est nécessaire de posséder, en plus du logiciel R, **la totalité des 23 scripts** servant au bon fonctionnement du programme ainsi que les différents *packages* utilisés : « tcltk » pour l'interface, « statnet » pour la génération des exposome/spectrosome et « xtable » pour la mise en forme des tableaux sous *LaTeX*. L'application est programmée pour télécharger ces deux derniers *packages* si ceux-ci ne sont pas présents sur l'ordinateur de l'utilisateur.

Concernant le type de fichier compatible avec l'application, afin d'éviter une surcharge de *packages* nécessaires pour l'application, les restrictions basiques de R ont été conservées telles que seuls les fichiers bloc-notes dont l'extension est « .txt » et les tableurs issus des feuilles de calculs dont l'extension est « .csv » peuvent être utilisés. Ceci implique la nécessité de vérifier les bonnes extensions des fichiers avant importation. Il est également préférable de disposer d'un **fichier standardisé** avec des nomenclatures standard des variables afin d'éviter des erreurs liées à des syntaxes différentes des libellés entre acteurs du réseau.

Il est nécessaire d'avoir un fichier prêt à l'emploi pour lancer l'analyse (Figure 22). En effet, cette interface n'est pas prévu pour créer des nouvelles colonnes pour l'analyse, un utilisateur connaissant le langage R pourra le faire avant de lancer le processus de surveillance directement dans l'environnement de travail, cependant pour un non-utilisateur, une manipulation dans un logiciel de tableur sera nécessaire.

Variable	ID	Temporelle	Cible Cible	…	Descriptive primaire Descriptive primaire	…	Descriptive secondaire Descriptive secondaire	…
Fonction	Permettant l'identification	Permettant l'analyse dynamique	Variable à expliquer sur laquelle sera appliquée l'analyse de la SO		Variables explicatives utilisées pour la caractérisation de la variable cible à expliquer		Permettant d'effectuer des analyses complémentaires des motifs	
Structure de l'information			Chaque information devra appartenir à des colonnes séparées		Chaque information devra appartenir à des colonnes séparées		Chaque information devra appartenir à des colonnes séparées	
Type de variable	Numérique ou en caractère	Numérique	De préférence standardisée grâce à une nomenclature et libellé associé à disposition		De préférence standardisée grâce à une nomenclature et libellé associé à disposition		De préférence standardisée grâce à une nomenclature et libellé associé à disposition	

Figure 22 : Tableur prêt à l'utilisation pour la Surveillance Observationnelle via l'outil OSA.

De plus, pour permettre la construction des cortèges des nœuds et motifs, il est nécessaire de répartir chacune des informations de manière indépendante dans des **colonnes distinctes** pour le traitement. Enfin, si l'utilisateur choisit d'effectuer un calcul via l'imputabilité, celle-ci doit impérativement être au **format numérique ou ordinal** car l'étape de significativité exige un calcul qui ne peut être effectué sur une variable nominale et doit être répartie sur le même nombre de colonnes que le nombre de variables explicatives primaires. Si le calcul choisit fait appel à une imputabilité externe, il est nécessaire de posséder un autre fichier contenant la liste des modalités explicatives, leur code (et label de préférence), ainsi qu'une distribution de l'imputabilité associée dans des colonnes distinctes pour chaque évaluateur a posteriori.

5.2. Présentation et ouverture de l'interface

L'ouverture du programme se fait directement dans le logiciel R, soit en utilisant les fonctionnalités fournies par le logiciel pour importer des scripts, soit en tapant l'unique ligne de commande permettant l'importation du script de démarrage et l'ouverture de l'outil OSA :

```
source("Chemin_utilisateur\Welcome.R")
```

« Chemin_utilisateur » faisant référence au chemin du fichier dans lequel se trouve le script nommé « Welcome.R », du type « C:\\Utilisateurs… ».

Ainsi, à l'ouverture de l'outil, la fenêtre d'accueil apparaît et l'utilisation se fera par la suite uniquement via cette l'application (Figure 23A).

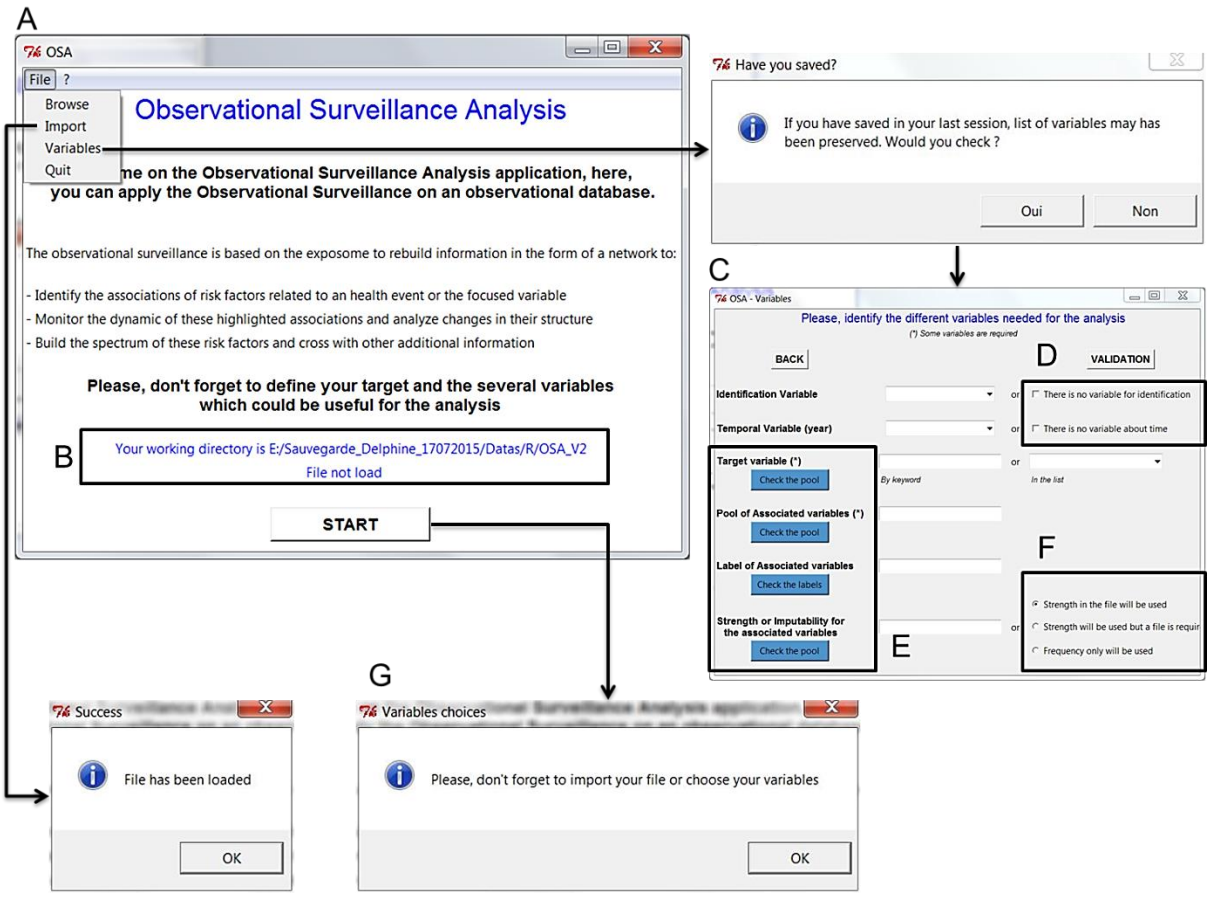


Figure 23 : Présentation de l'interface de l'outil OSA.

Cette fenêtre correspond à la première partie du programme, c'est-à-dire l'introduction (*Voir 4.2.2, p.82*) et elle permet de préparer le déploiement de la Surveillance Observationnelle sur le fichier utilisateur. Elle indique à l'utilisateur le chemin de l'environnement de travail ainsi que la présence d'un fichier importé ou non au sein de cet environnement (Figure 23B). Les menus déroulants en haut à gauche permettent de changer le répertoire de travail, d'importer un fichier et de choisir les variables à utiliser pour l'analyse.

Le choix des variables s'effectuent via une autre fenêtre d'interface (Figure 23C). Celle-ci propose des listes déroulantes contenant les différents noms des variables du fichier importé pour la variable d'identification, temporelle et la variable cible. Cependant, il est également proposé à l'utilisateur d'ajouter une variable d'identification et/ou de temporalité si celle(s)-ci n'existent pas en cochant les cases correspondantes (Figure 23D). De même, l'utilisateur a la possibilité de faire une recherche par mots-clefs s'il désire non pas une seule variable cible, mais un ensemble de variable cible en indiquant un mot-clef et en cliquant sur « Check the

pool » (Figure 23E). L'interface proposera alors une liste de toutes les colonnes contenant ce mot-clef. Cette fonctionnalité est également proposée pour la sélection des variables cibles, de leur label et des forces d'associations.

> *Par exemple :*
> Si les variables à utiliser sont toutes sous le même format tel que : « Code_exposition1 », « Code-exposition2 », « Code_exposition3 », et ainsi de suite, l'utilisateur pourra faire la recherche « Code_exposition » pour obtenir l'ensemble de ces variables.

Les renseignements des variables apportant des informations concernant les libellés des codes et la force d'association ne sont pas des variables obligatoires pour cette analyse. Cependant, dans le cas où il n'existerait pas de variable concernant une imputabilité, l'utilisateur peut cocher soit le calcul par fréquence, soit le calcul par une imputabilité externe, dans quel cas il sera obligé d'importer un fichier externe par la suite (Figure 23F).

Une fois les variables choisies, le bouton « VALIDATION » permet l'enregistrement de la sélection pour leur future utilisation et ouvre automatiquement la fenêtre des paramètres d'extraction, démarrant ainsi la deuxième partie du programme. Si l'utilisateur clique sur le bouton « START » sans avoir importé ni renseigné les variables, il sera invité à le faire (Figure 23G).

5.3. Choix de l'analyse et paramétrage des indicateurs

5.3.1. Choix de l'analyse

Après avoir cliqué sur le bouton « VALIDATION » de la fenêtre du choix des variables, la fenêtre des paramètres d'extraction apparaît (Figure 24).

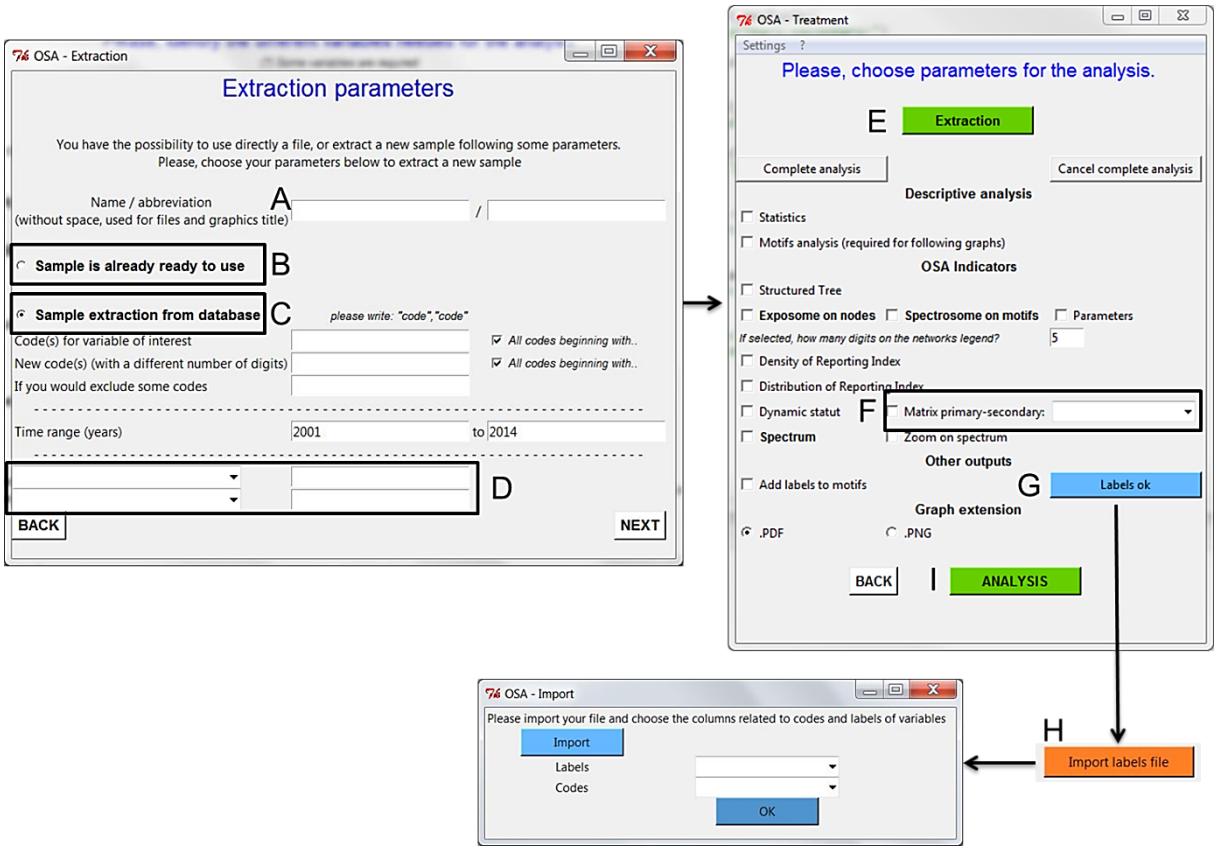


Figure 24 : Présentation des fenêtres de paramétrage de l'extraction et de l'analyse.

Elle permet d'indiquer tout d'abord le nom de l'analyse ainsi qu'une abréviation qui sera utilisé pour créer les dossiers et fichiers de résultats, identifiant chaque analyse distinctement et évitant ainsi de remplacer des fichiers existants (Figure 24A).

Deux possibilités sont proposées à l'utilisateur :

- Travailler sur un fichier déjà préparé (Figure 24B)
- Effectuer une extraction d'un sous-échantillon (Figure 24C)

Dans le cas d'une extraction, l'utilisateur est invité à renseigner les différentes informations nécessaires, c'est-à-dire les codes de la variable cible correspondant à l'évènement de santé qu'il souhaite analyser et la plage temporelle souhaitée pour cette étude. Il peut également choisir de faire une extraction plus spécifique en ajoutant une ou deux autres variables en tant que critères d'extraction supplémentaires (Figure 24D).

Par exemple :
Si la variable cible est une pathologie, l'utilisateur pourra utiliser un code correspondant à une pathologie particulière, ou bien plusieurs codes correspondant à un ensemble de pathologie, et potentiellement extraire uniquement les cas concernant les hommes ou les cas dont l'enregistrement a été effectué dans un centre spécifique du réseau de surveillance.

Une fois les paramètres renseignés, le bouton « NEXT » permet de les enregistrer et d'ouvrir la fenêtre qui permet cette fois-ci le choix des indicateurs de l'analyse spectrale.

L'utilisateur a alors la possibilité de faire une extraction simple en vue d'une préparation de fichier par exemple (Figure 24E), ou de lancer une analyse complète, ou encore de choisir uniquement certains indicateurs. Selon les indicateurs choisis, certaines conditions sont requises, par exemple si l'utilisateur choisit de tracer la matrice permettant un croisement entre les motifs et une variable explicative secondaire, la variable explicative à croiser doit être renseignée grâce à la liste déroulante (Figure 24F). De même, si l'utilisateur choisit d'ajouter une correspondance entre les codes utilisés par l'analyse et des libellés, il peut être nécessaire d'importer un fichier permettant cette correspondance si cela n'a pas été fait lors de l'enregistrement des variables à utiliser. Si la correspondance est déjà faite, le bouton affiche alors « Labels ok » sur un fond bleu (Figure 24G), sinon « Import labels file » sur un fond orange (Figure 24H).

5.3.2. Paramétrage des indicateurs

Le menu déroulant en haut à gauche permet le paramétrage de la modélisation des indicateurs. Le menu « Settings » (Figure 25A) permet d'accéder soit aux paramètres pour le calcul de la significativité des modalités (Figure 25C), soit au paramétrage de la longueur maximale des cortèges autorisée (Figure 25B).

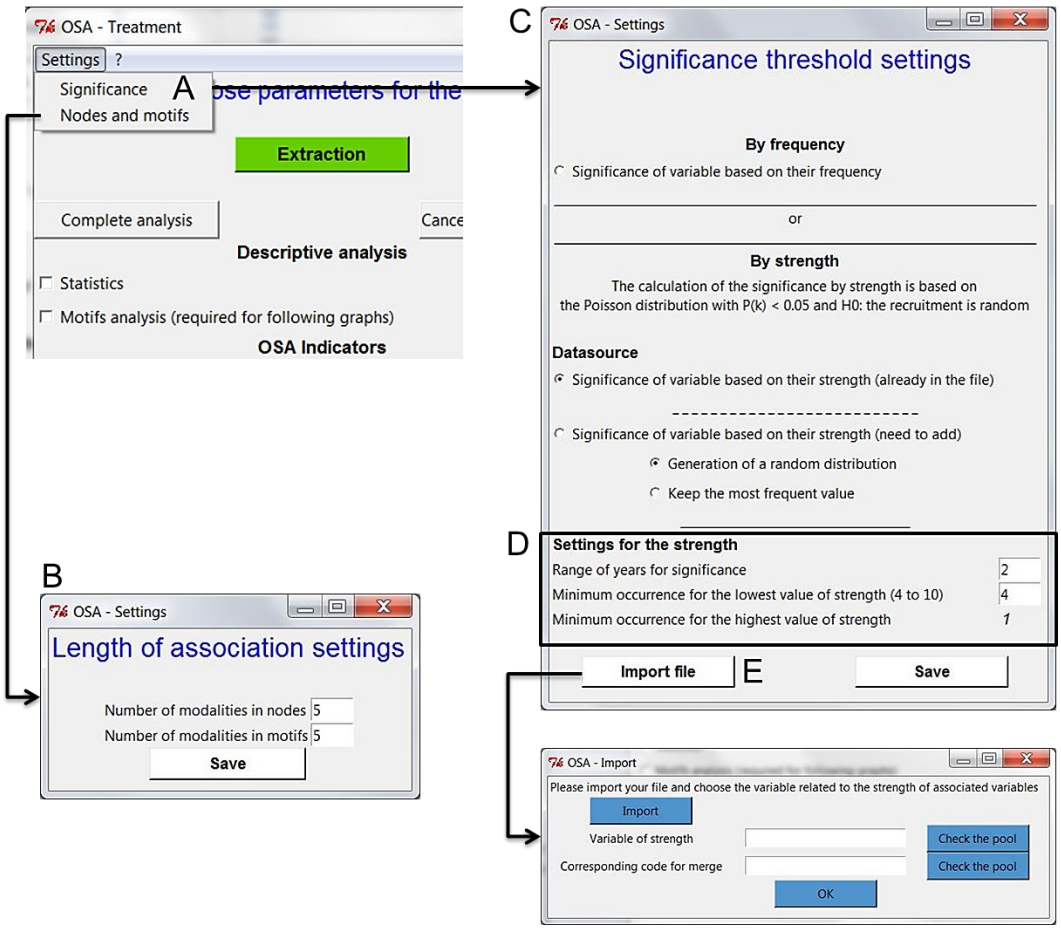


Figure 25 : Présentation des fenêtres de paramétrage des longueurs des nœuds et motifs et calcul de significativité.

Pour le calcul de la significativité par imputabilité, l'utilisateur a la possibilité de paramétrer le nombre d'occurrences critique sous imputabilité minimale pour considérer la significativité (correspondant au λ_{min}) et la plage temporelle de référence Δt pour le calcul du $n_{i,\Delta t}$ de chaque modalité (Figure 25D). Il peut également choisir d'importer un fichier s'il a fait le choix de travailler avec une imputabilité externe (Figure 25E) et dans ce cas, choisir une méthode de récupération de l'imputabilité (majoritaire ou distribution) (*Voir 4.2.3.2, p.88*).

Le bouton « ANALYSIS » de la fenêtre de traitement lance alors l'analyse si tous les critères sont validés (Figure 24I), dans le cas contraire, un message d'erreur apparaît en indiquant le renseignement manquant. De même, il est demandé pour chaque analyse si les paramètres ont été vérifiés afin d'encourager l'utilisateur à prendre connaissance de ces possibilités de modulation.

5.3.3. Déroulement de l'analyse et exportation des résultats

Lors du lancement de l'analyse, une nouvelle fenêtre apparaît (Figure 26), et informe l'utilisateur en temps réel de l'avancement de l'analyse, le déroulement total peut durer entre 2 minutes pour un échantillon de 300 observations à plusieurs dizaines de minutes pour un échantillon de plusieurs milliers d'observations. Ce temps est également dépendant de la longueur maximale autorisée pour la génération des nœuds et motifs. Plus la longueur est grande, plus les possibilités sont nombreuses et plus l'analyse sera longue. Cette fenêtre permet également de donner des informations clefs sur l'analyse et de rappeler les différents indicateurs sélectionnés et paramètres fixés (Figure 26B).

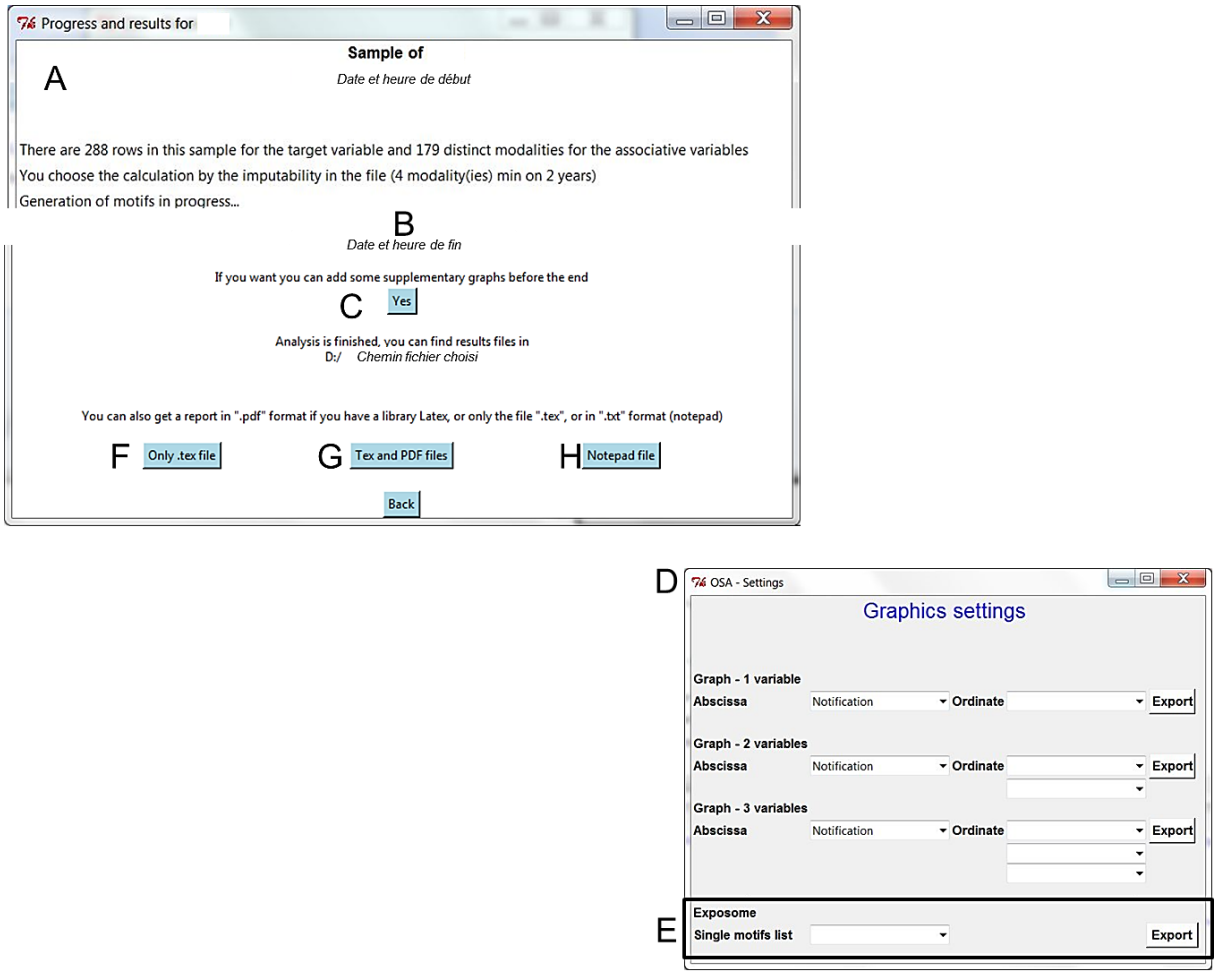


Figure 26 : Présentation des fenêtres de progression et de graphiques supplémentaires.

Cette fenêtre permet également un échange interactif avec l'utilisateur :

- Si une analyse a été effectuée précédemment, l'application demandera à l'utilisateur s'il désire utiliser les mêmes nœuds et motifs créés précédemment.

- La construction de l'exposome étant effectué grâce à un algorithme, la disposition des nœuds et des labels peut parfois être difficile à lire. L'application propose alors à l'utilisateur de générer un nombre non limité d'exposome afin de choisir la disposition satisfaisante pour la présentation des résultats.

Une fois le traitement effectué, la fenêtre indique le chemin d'accès permettant de retrouver les différents résultats exportés et ce dossier est ouvert automatiquement. L'application propose également d'obtenir de nouveaux graphiques avant de finaliser l'étude (Figure 26C). Si l'utilisateur choisit d'obtenir de nouveaux graphiques, une nouvelle fenêtre s'affiche alors (Figure 26D) lui permettant par exemple de tracer des graphiques permettant de suivre une ou plusieurs variables selon le temps ou bien des spectrosomes spécifiques de certains motifs identifiés (Figure 26E).

Enfin, une fois l'analyse et les graphiques supplémentaires finalisés, l'utilisateur peut choisir de générer un rapport automatique et standardisé, soit sous format bloc note (Figure 26H), soit sous format PDF s'il dispose d'une librairie *LaTeX* (Figure 26G). Il peut également décider de générer simplement le fichier « .tex » utilisable pour *LaTeX* pour pouvoir le compiler plus tard s'il le souhaite (Figure 26F).

5.4. Fichiers des résultats

Pour une analyse complète, sans graphiques supplémentaires, les résultats sont composés de :

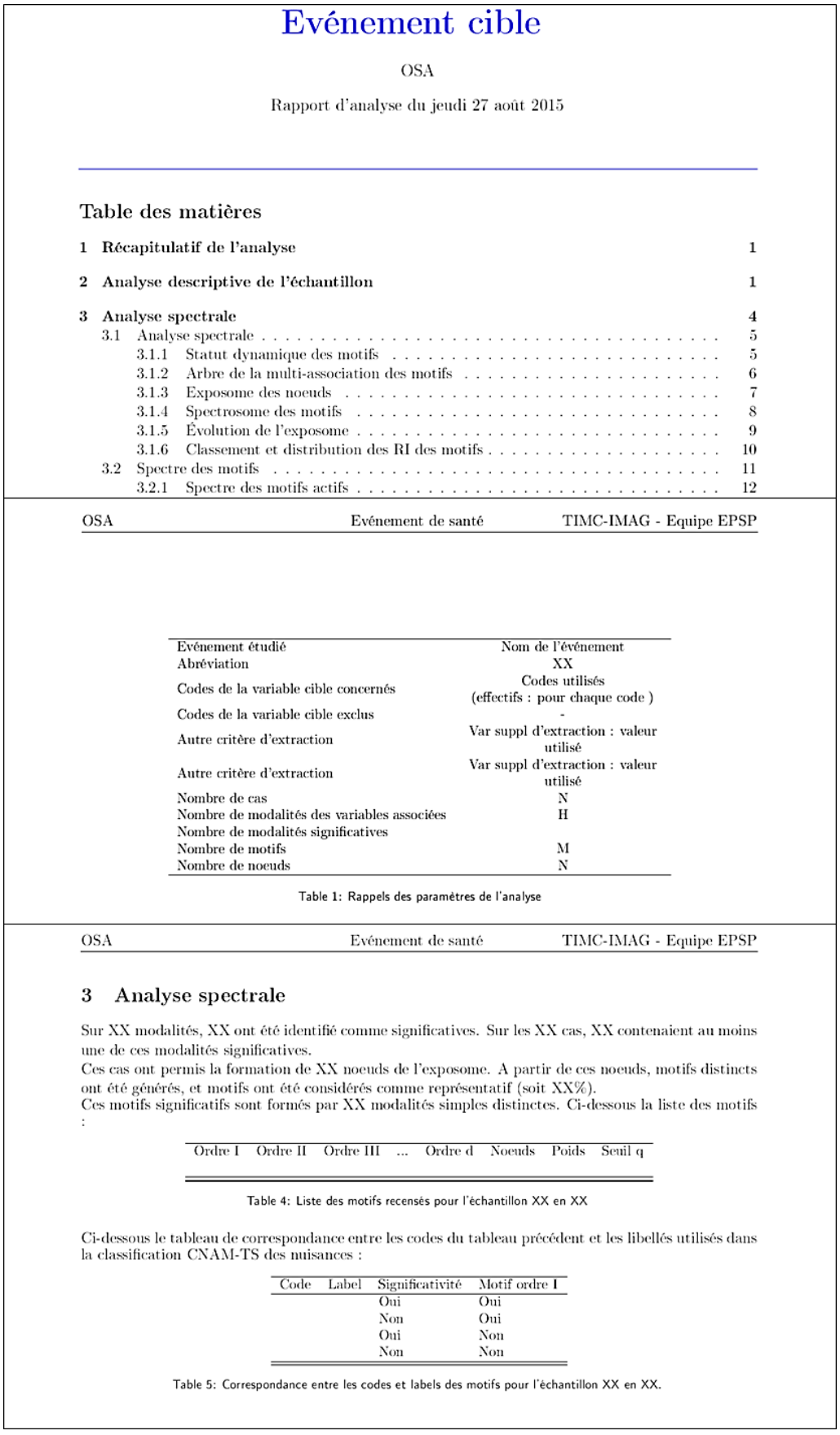
- 13 fichiers de type tableurs :
    - l'échantillon extrait à partir de la base de données, contenant uniquement les observations concernant l'évènement d'intérêt ;
    - l'évolution du nombre de recrutement de la variable cible, notamment si plusieurs codes ont été utilisés pour l'extraction du sous-échantillon ;
    - l'évolution du nombre de recrutement de chacune des modalités recensées pour les variables explicatives primaires ;

  - un tableur rappelant de manière générale l'évolution du nombre de cas et du nombre des différentes modalités de la variable cible et des variables explicatives primaires ;
  - le tableur des nœuds distincts formés à partir des observations ;
  - la synthèse des motifs représentatifs et les indicateurs modélisés pour la dernière année de la plage temporelle ;
  - un tableur contenant l'évolution du « Reporting index » sur la totalité de la plage temporelle étudiée ;
  - un fichier contenant les liens de l'exposome des nœuds ;
  - un fichier contenant les liens du spectrosome des motifs ;
  - les deux fichiers des paramètres d'évolution des deux réseaux ;
  - la matrice de croisement des motifs et d'une variable explicative secondaire ;
  - le fichier de correspondance entre les codes et les libellés.

- 14 fichiers graphiques ***au minimum*** :
  - graphique descriptif de l'évolution du nombre de cas et de recrutement des modalités de la variable cible et variables explicatives primaires ;
  - l'arbre de la multi-association ;
  - l'exposome des nœuds ;
  - le spectrosome des motifs ;
  - les deux graphiques de l'évolution des paramètres de l'exposome et du spectrosome ;
  - les deux graphiques du degré de connectivité des nœuds de l'exposome et du spectrosome ;
  - le classement des valeurs de « Reporting Index » et la distribution de cet indicateur ;
  - le graphique représentant le statut dynamique des motifs ;
  - l'évolution du « Reporting index » pour chaque motif actif, les trajectoires distinctes étant optionnelles ;
  - le spectre des motifs et le spectre des motifs actifs ;
  - la matrice.

- 1 fichier bloc note pour le rapport si demandé ;

- 1 fichier tex (pour compilateur LaTeX), et PDF, si demandé.



Evénement cible

OSA

Rapport d'analyse du jeudi 27 août 2015

Table des matières

1 Récapitulatif de l'analyse 1

2 Analyse descriptive de l'échantillon 1

3 Analyse spectrale 4

3.1 Analyse spectrale . . . 5

3.1.1 Statut dynamique des motifs . . . 5

3.1.2 Arbre de la multi-association des motifs . . . 6

3.1.3 Exposome des noeuds . . . 7

3.1.4 Spectrosome des motifs . . . 8

3.1.5 Évolution de l'exposome . . . 9

3.1.6 Classement et distribution des RI des motifs . . . 10

3.2 Spectre des motifs . . . 11

3.2.1 Spectre des motifs actifs . . . 12

Evénement étudié	Nom de l'événement
Abréviation	XX
Codes de la variable cible concernés	Codes utilisés (effectifs : pour chaque code)
Codes de la variable cible exclus	-
Autre critère d'extraction	Var suppl d'extraction : valeur utilisé
Autre critère d'extraction	Var suppl d'extraction : valeur utilisé
Nombre de cas	N
Nombre de modalités des variables associées	H
Nombre de modalités significatives	
Nombre de motifs	M
Nombre de noeuds	N

Table 1: Rappels des paramètres de l'analyse

OSA Evénement de santé TIMC-IMAG - Equipe EPSP

3 Analyse spectrale

Sur XX modalités, XX ont été identifié comme significatives. Sur les XX cas, XX contenaient au moins une de ces modalités significatives.
Ces cas ont permis la formation de XX noeuds de l'exposome. A partir de ces noeuds, motifs distincts ont été générés, et motifs ont été considérés comme représentatif (soit XX%).
Ces motifs significatifs sont formés par XX modalités simples distinctes. Ci-dessous la liste des motifs :

Ordre I	Ordre II	Ordre III	...	Ordre d	Noeuds	Poids	Seuil q

Table 4: Liste des motifs recensés pour l'échantillon XX en XX

Ci-dessous le tableau de correspondance entre les codes du tableau précédent et les libellés utilisés dans la classification CNAM-TS des nuisances :

Code	Label	Significativité	Motif ordre I
		Oui	Oui
		Non	Oui
		Oui	Non
		Non	Non

Table 5: Correspondance entre les codes et labels des motifs pour l'échantillon XX en XX.

Figure 27 : Exemple d'un rapport PDF standardisé via le compilateur LaTeX.

RESUME – OSA

Afin de permettre l'utilisation de la méthodologie de la Surveillance Observationnelle en routine ainsi que l'adaptabilité à diverses bases de données et problématiques, il a été décidé de développer un outil « Observational Surveillance Analysis » (OSA).

L'outil, développé sous la plateforme **R** permet la génération d'une interface. La création de cet outil a permis notamment la flexibilité des choix des variables cibles et explicatives, facilitant son usage en routine.

Cet outil OSA permet également la génération des résultats de manière automatique et standardisée :

- Un ensemble de tableurs permettant le récapitulatif des phases calculatoires
- Un ensemble de graphiques des indicateurs
- Un rapport de synthèse en fichier bloc note et/ou PDF

Partie 3 : Applications au RNV3P

La surveillance dans le milieu professionnel est une déclinaison de la Surveillance en Santé Publique, et est définie comme étant le suivi et l'analyse systématique et permanent d'un problème de santé et de ses déterminants professionnels à l'échelle d'une population. Les principaux objectifs de la surveillance épidémiologique des risques professionnels sont donc d'établir des indicateurs permettant de quantifier le poids de l'activité professionnelle sur l'état de santé de la population générale, de repérer des secteurs et des activités à risque élevé, d'alerter sur d'éventuels problèmes en relation avec le travail, connus ou émergents, et enfin d'évaluer les dispositifs de prévention et de réparation. Les objectifs sont donc essentiellement étiologiques car on chercher à identifier des facteurs professionnels susceptibles d'augmenter le risque de développer des maladies et à quantifier celui-ci.

L'objectif de cette partie est donc de présenter des exemples d'application de la Surveillance Observationnelle appliqué à un contexte de surveillance des Maladies Professionnelles. Une maladie professionnelle est définie comme la conséquence d'une exposition, plus ou moins longue, à un risque existant dans l'exercice normal de l'activité. Trois pathologies ont été choisies au sein de la base de données du réseau expert RNV3P :

- Le ***cancer de la vessie***, pathologie chronique de forme stable à latence longue, dont les principales étiologies sont connues et reconnues au titre des maladies professionnelles ;

- L'***asthme***, pathologie aigue à latence courte dont les principales étiologies sont connues et reconnues au titre des maladies professionnelles ;

- Le ***lymphome non hodgkinien (LNH)***, pathologie complexe et chronique à distribution de latence (moyenne à longue), regroupant de nombreux sous-types et dont les étiologies sont encore mal connues.

L'objectif est alors de pouvoir surveiller des pathologies et les caractériser par leur profil d'expositions professionnelles particulier tout en tenant compte de la potentielle multi-exposition présente sur le lieu de travail. L'intérêt de présenter trois exemples différents est de pouvoir illustrer l'adaptabilité de la méthodologie à un contexte, mais également à des pathologies différentes en termes de connaissances, de latence et de multi-expositions. L'exemple du LNH, dont les étiologies sont encore mal définies, sera approfondi grâce à un travail supplémentaire effectué afin de comparer les résultats obtenus avec la littérature.

6. PRESENTATION DE LA BASE DU RNV3P

Créé en 2001, le Réseau National de Vigilance et de Prévention des Pathologies Professionnelles (RNV3P) est un réseau expert de Médecins du travail et de Spécialistes Hospitalo-Universitaires, soutenu par l'agence sanitaire ANSES, enregistrant de manière systématique et standardisée des cas de Problèmes de Santé au Travail (PST) (*voir 1.4, p. 34*).

6.1. La base de données RNV3P

6.1.1. Structure d'une observation et d'un PST

Une observation au sein de la base de données du RNV3P correspond à ***l'ensemble des informations recueillies pour un patient reçu dans un CCPP***, on retrouve notamment des informations :

- ***Administratives :*** son genre ; son âge ; sa localisation géographique ; l'année de consultation et d'enregistrement ; prise en charge sociale ;
- ***Hospitalières :*** le CCPP de recueil ; la spécialisation du médecin expert ; l'origine du médecin adressant ;
- ***Cliniques :*** la pathologie diagnostiquée ; la conclusion sur l'aspect professionnel de la pathologie ; les situations, professionnelles ou non, identifiées comme potentiellement causale : les nuisances ;
- ***Professionnelle :*** son statut professionnel ; son secteur d'activité et sa profession actuelle ; son secteur d'activité et sa profession potentiellement en cause si conclusion en « pathologie professionnelle ».

On appelle un Problème de Santé au Travail (PST) une observation dont la conclusion clinique a été définie comme étant une « **pathologie professionnelle** » par les médecins experts du réseau, **conséquence d'une exposition, plus ou moins longue, à un risque existant dans l'exercice normal de l'activité**.

Le PST est défini comme une association entre une pathologie diagnostiquée (p) et un contexte professionnel composite (e) constitué d'un cortège nuisances (H) correspondant aux

expositions professionnelles potentiellement associées à la pathologie, et l'activité professionnelle (*a*) correspondant aux secteur (*s*) et à la profession (*o*) du patient, potentiellement en relation avec la pathologie diagnostiquée. Chaque nuisance est également pondérée par une force d'association, l'imputabilité (*f*), évaluée par le médecin en fonction de la littérature et de l'historique professionnel (et environnemental) du patient, comprise entre « pas de relation » et « relation certaine ».

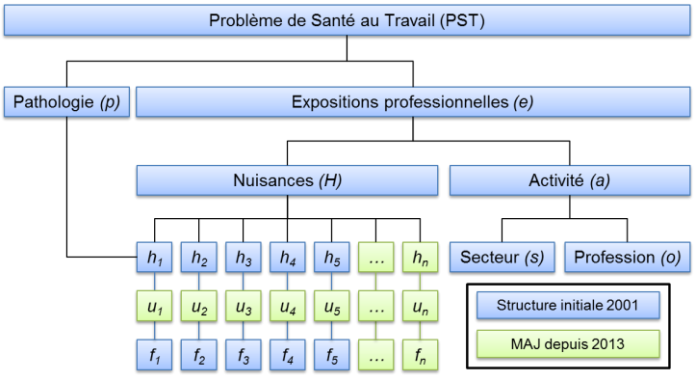


Figure 28 : Structure d'un Problème de Santé au Travail au sein de la base RNV3P depuis 2001.

6.1.2. Nomenclature de standardisation

Chacune des variables au sein de la base est standardisée et codée grâce à des thésaurus administratifs reconnus, tels que la Classification Internationale des Maladies (CIM-10, 3 digits[3]) pour les pathologies, la Nomenclature des Activités Françaises (NAF, 4 digits) pour les secteurs d'activités, le Code International des Types de Professions (CITP, 4 digits) pour les professions et enfin un code imbriqué inspiré par le thésaurus de la Caisse Nationale d'Assurance Maladie pour les expositions (thésaurus utilisé pour les indemnisations des tableaux de maladies professionnelles). Chacun de ces thésaurus est dit « imbriqué », chaque caractère correspondant à un niveau d'information de plus en plus précis.

Par exemple :

Les codes imbriqués des expositions du benzène et toluène :

2 – Agents chimiques organiques *(niveau 1)*

2***1*** – Hydrocarbures et dérivés *(niveau 2)*

21***3*** – Hydrocarbures aromatiques et dérivés *(niveau 3)*

213***1*** – Hydrocarbures aromatiques monocycliques *(niveau 4)*

2131***0100*** – Benzène *(niveau 5)*

2131***0200*** – Toluène *(niveau 5)*

[3] Un digit correspond à un caractère, un espace ou un chiffre numérique

Ces différents codes seront utilisés dans les analyses afin de permettre une meilleure standardisation des résultats, notamment dans les légendes des exposomes et spectrosomes pour lesquelles les codes seront utilisés de manière « tronquée » afin de présenter les grandes familles d'expositions et alléger la représentation graphique.

6.1.3. Démarche qualité

La base de données nationale RNV3P comprend la totalité des dossiers des patients vus dans les centres, cependant, la base de données sur laquelle les analyses statistiques sont effectuées au sein du RNV3P correspond à une sous-partie de la base de données initiale suivant des critères strictes de qualité. Chaque cas doit être validé par un médecin sénior du réseau. Pour une validation, il est nécessaire que chaque enregistrement de situations professionnelles (nuisances) possède une valeur d'imputabilité associée et tous les dossiers en attente de conclusion clinique sont exclus.

La base de données professionnelle utilisée comprend uniquement les cas dont la conclusion médicale a été définie comme « pathologie professionnelle » soit environ 50% des observations.

6.1.4. Evolution du système d'information en 2013

A sa création en 2001, la base recensait un maximum de 5 expositions professionnelles potentielles pour un PST. Cependant, depuis le nouveau système d'information et la migration en 2013 de la base de données sur un serveur plus performant (Oracle), la possibilité de renseigner un nombre illimité de nuisances a été ajoutée (Figure 28). Ainsi, pour chaque cas, il est désormais possible d'enregistrer toutes les hypothèses envisagées tout en préservant la plausibilité de la relation grâce à la valeur d'imputabilité. De plus, l'information des nuisances a été scindée en deux catégories au lieu d'une seule : la substance et son usage. Il est désormais possible de faire la distinction entre l'usage d'un solvant et la molécule du solvant, ce qui était auparavant difficile et conduisait à des associations « famille-molécule » au sein des cortèges de nuisances.

6.2. Mise en place de la Surveillance Observationnelle

Dans le cadre de l'application de la Surveillance Observationnelle au RNV3P la transposition des termes est la suivante :

- **Variable cible** : variable renseignant la **pathologie** diagnostiquée, codée par le thésaurus CIM-10 ;
- **Variables explicatives associées** : les variables renseignant les **expositions professionnelles** suspectées, le terme *modalité renvoyant aux expositions professionnelles distinctes* recensées ;
- **Imputabilité :** Variable renseignant pour chaque exposition l'imputabilité potentielle entre la pathologie et l'exposition, et attribuée par les médecins experts.

Les nœuds et motifs seront alors formés par des cortèges d'expositions professionnelles et serviront à la caractérisation des pathologies professionnelles étudiées. Les objectifs de la SO transposés au contexte du RNV3P sont :

- L'identification des associations pathologies-expositions professionnelles par **l'identification d'expositions professionnelles ou de multi-expositions professionnelles** associées à la pathologie
- **L'analyse descriptive** des cas enregistrés au sein de la base concernant la pathologie professionnelle ciblée ainsi que l'analyse descriptive des expositions et combinaisons d'expositions associées à la pathologie
- **L'analyse spectrale** de la pathologie et la mise en évidence des apparitions ou extinctions des associations d'expositions professionnelles en lien avec cette pathologie

La variable d'imputabilité étant disponible au sein du RNV3P, le calcul de significativité des nuisances (ici les modalités), s'effectuent en utilisant le calcul de l'imputabilité (*voir 3.3.1.2, p. 53, équation 3.3*). Cette variable d'imputabilité, éant une variable ordinale, comprise entre « 0 » et « 3 », il a été nécessaire de calculer une imputabilité numérique à utiliser dans le calcul ainsi que de déterminer les paramètres $\lambda_{min}, \Delta_t, \nu$ adaptés à cette base de données.

6.3. Détermination des paramètres pour le RNV3P

La longueur *d* maximale choisit pour la formation des nœuds et des motifs a été de 5 afin de conserver l'intégralité des informations anciennement enregistrées dans l'ancien système d'information du RNV3P.

6.3.1. Nombre de recrutement de référence et paramètre ν

De par le statut pérenne de la base, l'intervalle temporel de référence Δt choisi a été fixé de manière empirique à 2 ans afin de permettre un bon équilibre entre de nouvelles expositions enregistrées et se mettant en place et des vieilles expositions correspondant à des anciennes hypothèses de travail. Le paramètre ν a quant à lui était fixé à 0.5 permettant ainsi de conserver des nuisances avec des faibles imputabilités pour évaluer leur potentiel en tant qu'hypothèse de travail (Figure 29).

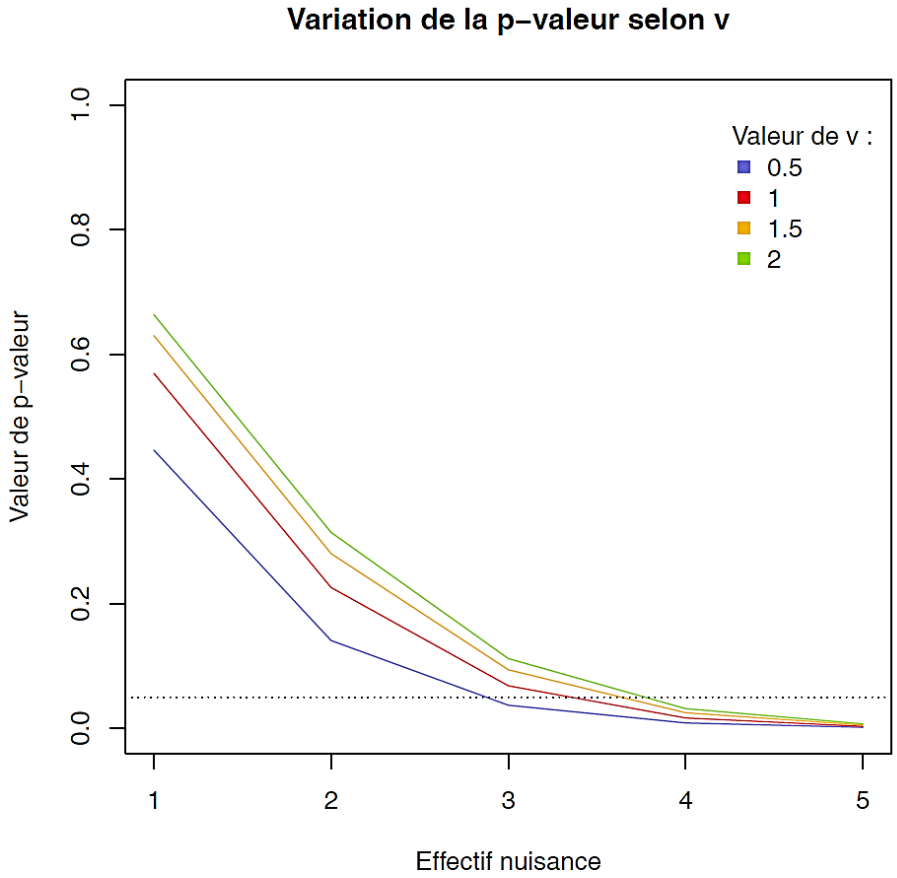


Figure 29 : Variation de la p-valeur (moyenne des modalités) selon ν et le nombre d'occurrences de la modalité.

6.3.2. Détermination de l'imputabilité numérique

Afin de calculer le paramètre λ (*voir 3.3.1.2, p. 53 ; et 3.3.1, équation (3.5)*), une imputabilité moyenne et « numérique » a été calculée à partir de l'imputabilité ordinale initiale (« 0 » à « 3 ») de la base de données pour chacune des modalités recrutées. Plusieurs possibilités de calcul ont été évaluées afin de respecter au mieux la distribution initiale des imputabilités : moyenne arithmétique, somme des carrés moyens et moyenne harmonique.

Soit x_i l'effectif de chaque valeur d'imputabilité I_i rencontrée pour une modalité :

- Moyenne arithmétique :

$$I = \frac{(I_0 \times x_0) + (I_1 \times x_1) + (I_2 \times x_2) + (I_3 \times x_3)}{x_0 + x_1 + x_2 + x_3} \quad (6.20)$$

- Somme des carrés moyens :

$$I = \sqrt{\frac{({I_1}^2 \times n_1) + ({I_2}^2 \times n_2) + ({I_3}^2 \times n_3) + ({I_4}^2 \times n_4)}{n_1 + n_2 + n_3 + n_4}} \quad (6.21)$$

- Moyenne harmonique :

$$I = 1 \Big/ \frac{\left({x_0}/{I_0}\right) + \left({x_1}/{I_1}\right) + \left({x_2}/{I_2}\right) + \left({x_3}/{I_3}\right)}{x_0 + x_1 + x_2 + x_3} \quad (6.22)$$

Des tirages aléatoires de 1 000 valeurs ont été effectués pour 4 profils de probabilité différents, chacun privilégiant une valeur d'imputabilité particulière : 0, 1, 2 et 3 (Figure 30). Les distributions de l'imputabilité numérique ont été générées pour chacun de ces tirages et chacune des formules permettant ainsi la comparaison visuelle de chacune des méthodes.

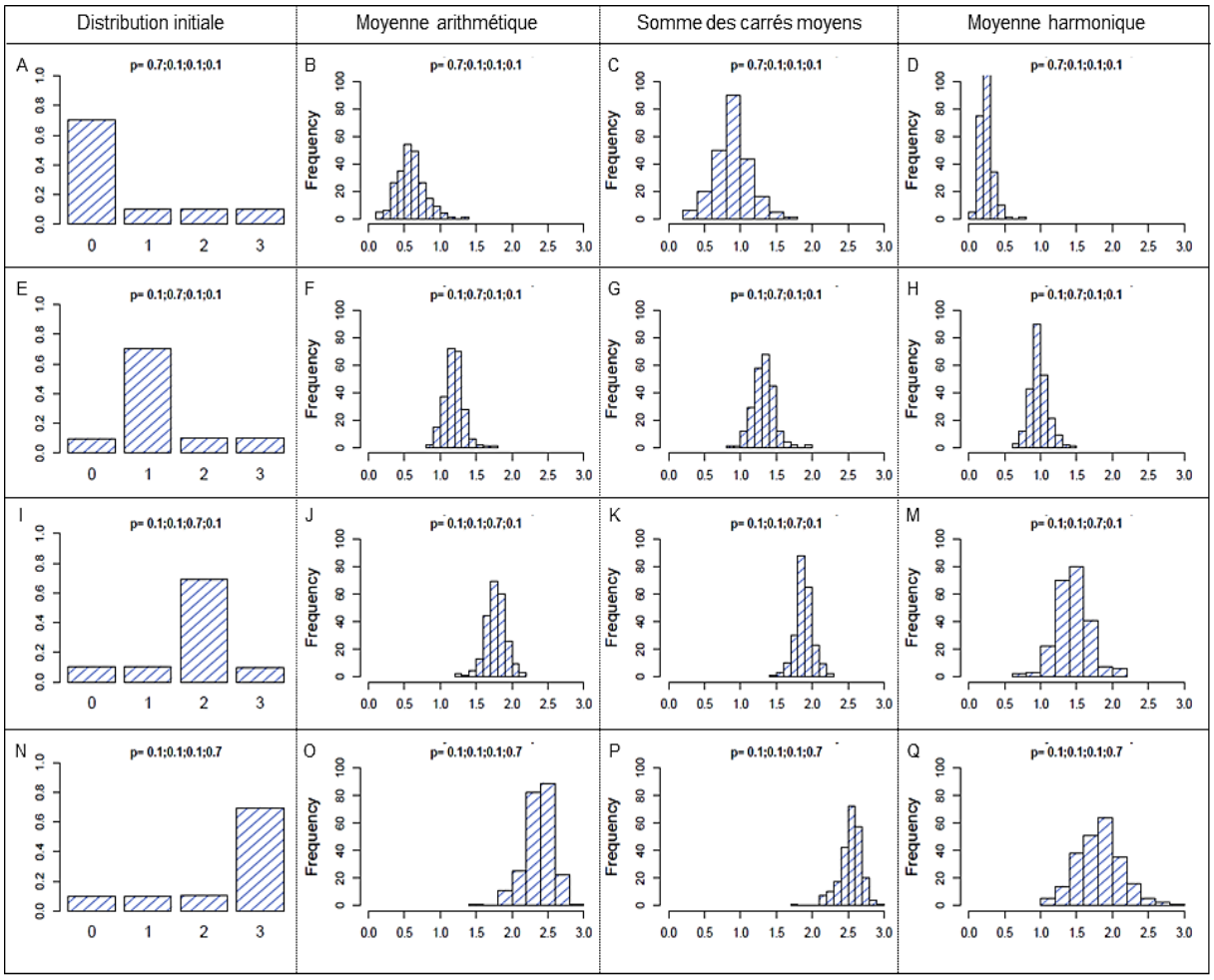


Figure 30 : Comparaison des distributions des valeurs d'imputabilités selon la distribution initiale et les formules d'imputabilité numérique. A, E, I, N : Distribution initiale pour les distributions de probabilités centrées sur les valeurs 0, 1, 2 et 3 respectivement. B, F, J, O : Distribution de l'imputabilité numérique calculée selon la moyenne arithmétique (équation 6.20) pour les distributions de probabilités centrées sur les valeurs 0, 1, 2 et 3 respectivement. C, G, K, P : Distribution de l'imputabilité numérique calculée selon la somme des carrés moyens (équation 6.21) pour les distributions de probabilités centrées sur les valeurs 0, 1, 2 et 3 respectivement. D, H, M, Q : Distribution de l'imputabilité numérique calculée selon la moyenne harmonique (équation 6.22) pour les distributions de probabilités centrées sur les valeurs 0, 1, 2 et 3 respectivement.

Le calcul de l'imputabilité par la moyenne harmonique (*équation 6.22*) n'a pas été retenu car lors d'une distribution d'imputabilité très prononcée en faveur de l'imputabilité maximale, la distribution avait tendance à s'étirer de 1 à 3 (Figure 30Q), sous-estimant ainsi la valeur de l'imputabilité tandis que les deux autres calculs restaient plus fidèles à la distribution initiale (Figure 30O et P). De même, le calcul par la somme des carrés moyens (*équation 6.21*) avait tendance à surestimer les valeurs d'imputabilité numérique dans le cas d'une distribution prononcée en faveur de l'imputabilité minimale (Figure 30C), ce qui pourrait conduire à une surestimation du risque en conservant des modalités dont l'imputabilité serait considérée comme nulle. De plus, les calculs de la somme des carrés moyens et la moyenne harmonique nécessitaient la transformation des valeurs d'imputabilité afin d'être applicable.

Le calcul retenu de l'imputabilité moyenne I d'une modalité avec n_i cas dans chaque modalité est donc celui de la moyenne arithmétique (Figure 30B, F, J, O) car impliquant une conservation brute de la valeur de l'imputabilité et une distribution plus fidèle à la distribution initiale.

6.3.3. Le paramètre λ_{min}

Le paramètre λ_{min} a été étudié afin de déterminer le seuil adéquat permettant de conserver raisonnablement des nuisances d'imputabilité faible pouvant servir d'hypothèse de travail, tout en écartant un trop grand bruit de fond (Figure 31).

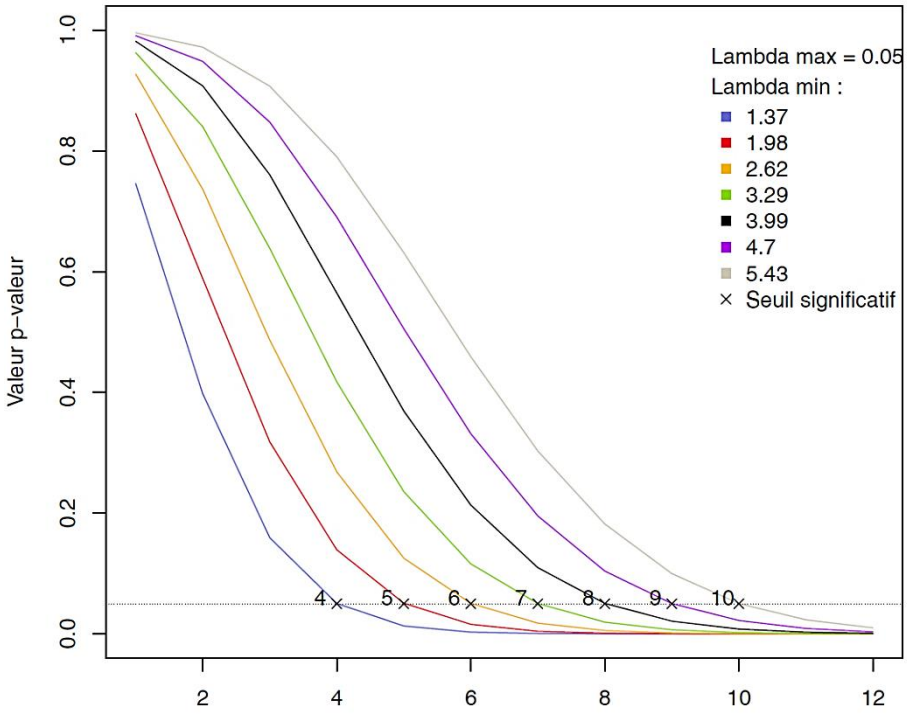


Figure 31 : Variation du paramètre λ_{min} et seuil de significativité du taux de recrutement sous l'imputabilité nulle.

Les paramètres retenus pour le RNV3P ont été de $\lambda_{min} = 1.37$. Pour rappel, $\lambda_{max} = 0.05$. Cela correspond à, sur l'intervalle de temps de référence Δt, un nombre de recrutement $n_{i,\Delta t,cr}$ de référence supérieur ou égal à 4 pour une imputabilité minimale 0 et égal à 1 pour une imputabilité maximale 3.

6.4. Synthèse des paramètres d'application de la SO au RNV3P

Tableau 4 : Variables et paramètres considérés pour l'application de la Surveillance Observationnelle à la base RNV3P.

SO	RNV3P
Variable cible à expliquer	Pathologie professionnelle
Variables explicatives primaires (modalités formant les nœuds et motifs)	Substances ou Usages
Significativité	
Intervalle temporel Δt	2 ans
Nombre de recrutement de référence nécessaire sous imputabilité nulle ($n_{i,\Delta t,cr}$)	4 ($\lambda_{min} = 1.37$)
Nombre de recrutement de référence nécessaire sous imputabilité maximale ($n_{i,\Delta t,cr}$)	1 ($\lambda_{max} = 0.05$)
Paramètre ν	0.5
Longueur de nœuds *d* maximale	5
Longueur de motifs *d* maximale	5

7. CANCER DE LA VESSIE

Le cancer de la vessie est un cancer plutôt commun, notamment dans les pays développés, affectant le plus souvent les hommes. Approximativement 90% des cancers sont des cancers urothéliaux, les 10% restants étant typiquement constitué de carcinome épidermoïde et adénocarcinome [Kiriluk, 2012]. Le cancer de la vessie est considéré comme une pathologie de forme stable à latence longue avec des expositions associées bien connues et reconnues au titre des maladies professionnelles.

Les taux d'incidence et de mortalité, ajustés sur l'âge, dans la plupart des pays développés sont de 9.5/100 000 et 2.5/100 000 [IARC, 2012]. En France, le cancer de la vessie représente le 4^{e} cancer incident pour les hommes avec un taux ajusté selon l'âge de 15.1/100 000, contre 2.2/100 000 pour les femmes (19^{e} cancer incident). En termes de mortalité, le cancer de la vessie est 7^{e}, avec un taux normalisé selon l'âge de 5/100 000 pour les hommes et 16^{e} chez les femmes avec un taux de 1/100 000 [IARC, 2012].

Le facteur de risque le plus connu du cancer de la vessie serait le tabac, avec plus d'un tiers des nouveaux cas liés au tabagisme. Cependant, la fraction attribuable à l'activité professionnelle dans le développement du cancer de la vessie est également lourde puisqu'elle a été estimée entre 2 et 14% par l'Institut National du Cancer (INCa) en 2012 [INCa, 2012] et entre 4 et 20% pour les hommes d'après la revue de la littérature de Kogevinas et al. en 2014 [Kogevinas et Garcia-Closas, 2014].

Dès les années 1950, les amines aromatiques (utilisées dans les industries de teinture) ont été identifiés comme un risque important pour le développement du cancer de la vessie [Case, 1954, Kogevinas et Garcia-Closas, 2014], notamment la benzidine, la napthylamine, le 4-amino-biphenyl, l'ortho-toluidine, l'ortho-tolidine et le 4-4'méthylenebis(2-chloroaniline) [S.C. Brown, 2011, IARC, 2008, Massardier-Pilonchery, 2013, Reulen, 2008], tous classés comme cancérigène certains ou probables par le CIRC [IARC, 2010]. D'autres expositions professionnelles ont également été mises en évidence telles que les hydrocarbures aromatiques polycycliques (HAP), les huiles, les produits agricoles et les solvants [T. Brown, 2012, Colt, 2011, Dryson, 2008, Kogevinas et Garcia-Closas, 2014].

Le cancer de la vessie étant un cancer de forme stable dont les expositions sont relativement bien connues, il était intéressant d'appliquer la Surveillance Observationnelle pour obtenir le spectre des expositions associées afin de comparer les résultats avec la littérature et d'apporter

des hypothèses de travail supplémentaires en s'appuyant sur les données de qualité du réseau RNV3P. L'analyse de cette pathologie sera plus développée afin de servir de grille de lecture pour les autres exemples.

Les codes de la Classification Internationale des Maladies (CIM-10) utilisés pour cette analyse étaient : C65*, C66*, C67* et D090.

7.1. Analyse

En 2014, au sein des 107 713 PST (observation avec conclusion **professionnelle**) de la base RNV3P et des 14 055 cancers professionnels, 1 006 concernaient le cancer de la vessie (Figure 32), dont 943 hommes et 63 femmes, avec une moyenne d'âge de 63.34 et 60.84 ans respectivement. Un nombre total de 331 nuisances distinctes ont été enregistrées depuis 2001, majoritairement des agents chimiques inorganiques (n=41), des agents chimiques organiques tels que les hydrocarbures et dérivés (n=38), les esters (n=9) et la famille des amines, imines et dérivés (n=23). De nombreux matériel et procédés industriels ont également été enregistrés (n=157) tels que les peintures et vernis, les additifs pour carburant et les produits pesticides. Enfin, 199 activités furent identifiées, notamment dans le secteur de la construction.

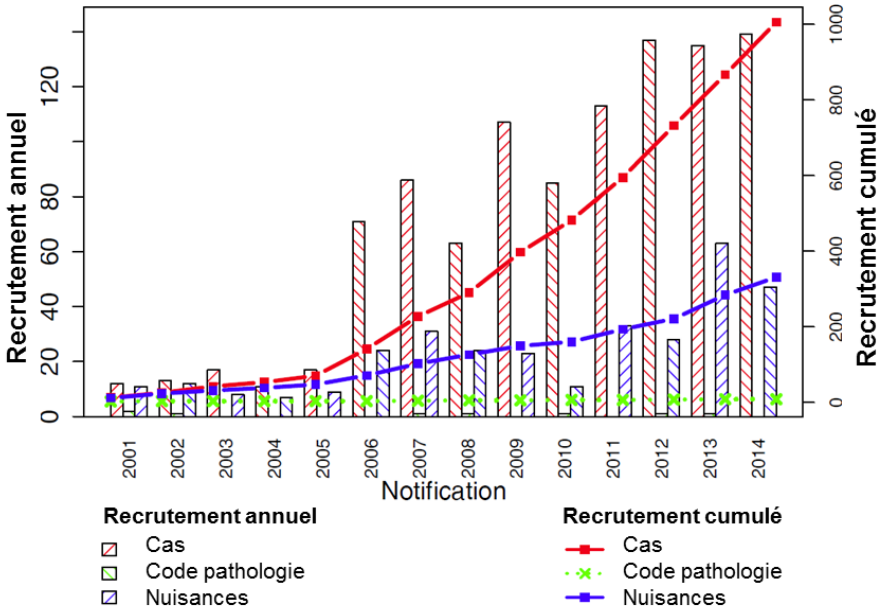


Figure 32 : Recrutement et effectif cumulé des cas, des nuisances distinctes et des codes pathologies distincts pour le cancer de la vessie de 2001 à 2014.

Parmi les 331 nuisances distinctes, 37 furent considérées comme significatives par le calcul basé sur l'imputabilité et 880 PST possédaient au moins une de ces nuisances dans leur

cortège. A partir de ces 880 PST, 477 cortèges distincts d'expositions professionnelles ont été identifiés, formant les nœuds de l'exposome pour un total de 31 059 liens entre ces nœuds (Figure 33). Sur la figure de l'exposome, la couleur des liens représente une famille d'expositions au 3ème niveau d'imbrication du thésaurus utilisé (*voir exemple 6.1.2, p.112*), par exemple, la couleur rouge correspond au code « 213 : Hydrocarbure aromatique et dérivé ». La couleur des liens permettent d'identifier visuellement les clusters majoritaires de l'exposome[4].

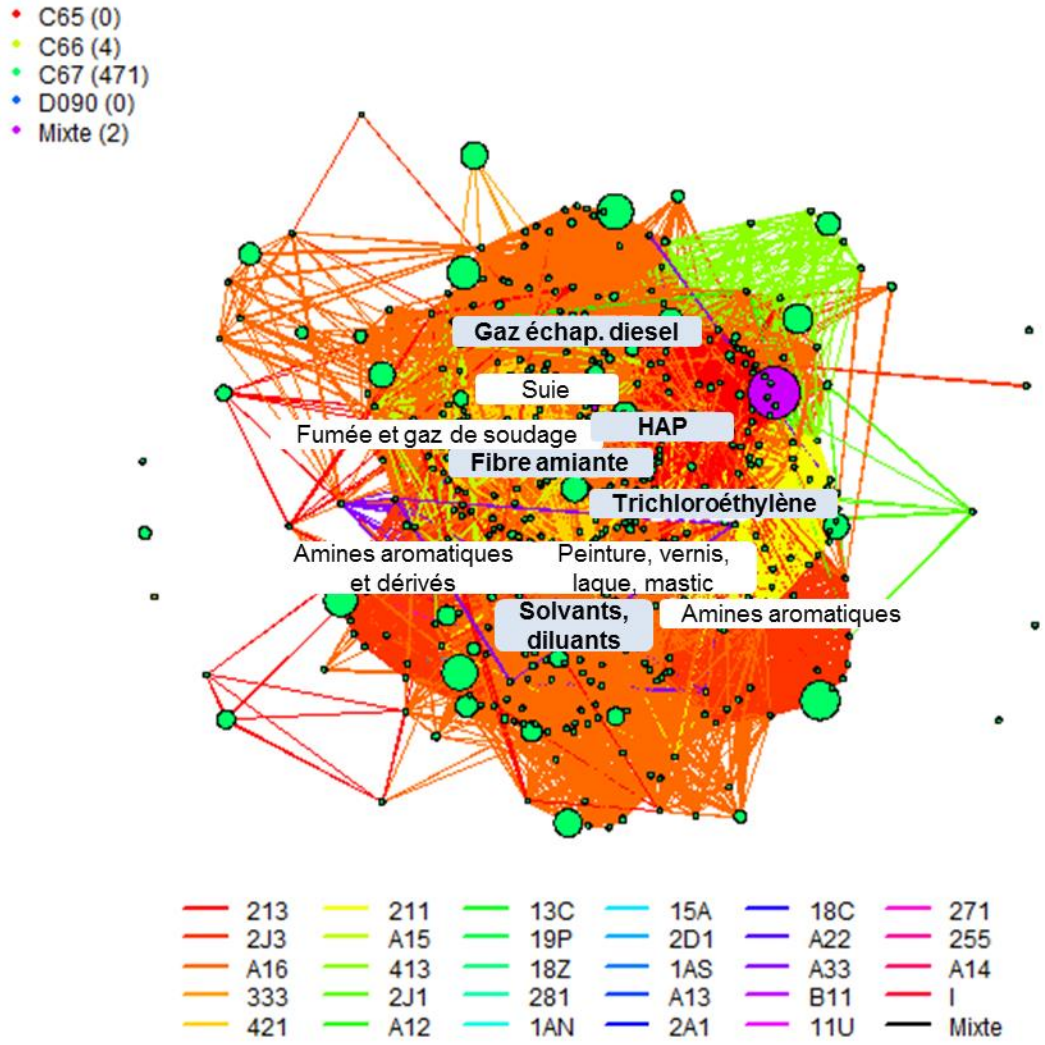


Figure 33 : Exposome du cancer de la vessie en 2014. La taille des nœuds correspond au poids des nœuds (nombre de PST identiques), la couleur correspond aux différents codes pathologies. La couleur des liens correspond aux différentes familles (3e niveau du code imbriqué) de nuisances identifiées en relation avec le cancer de la vessie et permettent d'identifier les clusters de l'exposome. Cinq clusters majoritaires ont été identifiés : « Gaz d'échappement diesel » (A16H0510), « HAP » (21320000), « Fibre d'amiante » (42110000), « Trichloroéthylène » (21170204) et « Solvants, diluants » (A1660000).

[4] Pour rappel, un cluster est un agrégat (ou groupe) de nœuds partageant exactement la même nuisance ou association de nuisances et met ainsi en évidence expositions (exposome des nœuds) ou motifs (spectrosome des motifs) particuliers.

La majorité correspondait à des nuisances de type « Matériel et produit divers » (Code A16), notamment les solvants, huiles et carburants ; la famille des « hydrocarbures aromatiques et dérivés » (Code 213) ; la famille des « hydrocarbures aliphatiques et dérivés » (Code 211) et « l'amiante » (Code 4211000). Six nœuds isolés, c'est-à-dire ne partageant aucune nuisance en commun avec les autres nœuds, ont été mis en évidence : le « 4,4'-méthylene-bis-orthochloraniline » (2JZ00000), l'association du « 2-naphtylamine » (2J310C02) et du « chlorure de diméthylaminobenzène diazonium » (2K230100), les « dérivés hydroxyle d'amine aromatique » (2J350000), les « autres dérivés d'amine et d'imine » (2JZ00000), les « produits dégagés dans les procédés de fonderie » (A1400000) et les « oxydant et décolorant (cosmétique) » (A16L0120).

Cent-soixante-six motifs d'expositions professionnelles ont été identifiés, c'est-à-dire 166 nuisances ou associations de nuisances ayant été rapportées comme potentiellement associées au cancer de la vessie. Un numéro d'identification a été attribué à chacun en fonction de leur poids, permettant de les identifier dans tous les résultats issus de la Surveillance Observationnelle. La liste complète des 166 motifs d'expositions professionnelles identifiés est disponible dans le tableau en Annexe 1, la liste partielle des 51 motifs actifs est disponible dans le Tableau 5. Parmi ces 166 motifs, 54 correspondaient à des motifs d'ordre I (une seule nuisance au sein du cortège), impliquant un total de 112 motifs d'ordre supérieurs à I (combinaison de nuisances au sein du cortège) : 93 motifs d'ordre II et 19 motifs d'ordre III.

Tableau 5 : Liste des 51 motifs actifs du cancer de la vessie identifiés par la Surveillance Observationnelle en 2014.

ID	Ordre I	Ordre II	Ordre III	Nœuds	Poids	Statut
1	hydrocarbure aromatique polycyclique (HAP)	-	-	128	285	A
2	amiante (fibre)	-	-	137	160	A
3	solvant, diluant	-	-	86	106	A
4	diesel gaz d'échappement	-	-	64	103	A
5	trichloroéthylène	-	-	83	96	A
6	amine aromatique et dérivé	-	-	41	85	A
7	peinture, vernis, laque, mastic	-	-	51	83	A
8	suie et autre produit dégagé lors de l'utilisation de produit noir	-	-	52	81	B
9	amine aromatique	-	-	41	78	A
10	HAP	amiante (fibre)	-	45	59	A
11	fumée et gaz de soudage	-	-	41	55	A
12	huile de coupe, fluide d'usinage	-	-	34	50	A
13	matière colorante	-	-	30	48	A

14	huile et graisse lubrifiante	-	-	33	37	A
15	encre	-	-	25	33	A
16	huile, graisse et fluide d'usinage industriel	-	-	28	31	A
17	houille, charbon	-	-	20	29	A
18	HAP	diesel gaz d'échappement	-	17	28	A
19	amiante (fibre)	solvant, diluant	-	23	26	A
20	HAP	suie et autre produit dégagé lors de l'utilisation de produit noir	-	18	26	B
21	trichloroéthylène	HAP	-	24	25	A
22	trichloroéthylène	amiante (fibre)	-	24	24	A
23	huile et graisse lubrifiante minérale	-	-	20	22	A
24	gaz d'échappement	-	-	20	22	A
25	peinture, vernis, laque, mastic	solvant, diluant	-	18	22	B
26	amiante (fibre)	fumée et gaz de soudage	-	19	21	A
27	amiante (fibre)	diesel gaz d'échappement	-	18	21	A
28	HAP	amine aromatique et dérivé	-	8	19	B
29	tabac	-	-	7	17	B
30	produit pesticide (biocide, insecticide, rodenticide, fongicide, engrais)	-	-	10	16	B
31	HAP	amine aromatique	-	10	15	A
32	amiante (fibre)	peinture, vernis, laque, mastic	-	15	15	A
33	amiante (fibre)	suie et autre produit dégagé lors de l'utilisation de produit noir	-	13	15	B
34	HAP	fumée et gaz de soudage	-	12	14	B
35	poussière de bois	-	-	13	13	C
36	solvant, diluant	encre	-	8	13	D
37	autre HAP	-	-	8	12	D
38	amiante (fibre)	huile et graisse lubrifiante	-	12	12	C
39	huile et graisse lubrifiante	diesel gaz d'échappement	-	11	12	D
41	HAP	peinture, vernis, laque, mastic	-	10	11	D
44	amiante (fibre)	huile de coupe, fluide d'usinage	-	11	11	D
47	produit noir	-	-	8	10	D
51	amiante (fibre)	huile et graisse lubrifiante minérale	-	10	10	C
52	hydrocarbure aromatique et dérivé	-	-	5	9	D
63	solvant, diluant	huile et graisse lubrifiante	-	8	9	C
65	trichloroéthylène	HAP	amiante (fibre)	9	9	C
71	amiante (fibre)	gaz d'échappement	-	8	8	D
79	goudron	-	-	6	6	C
82	trichloroéthylène	fumée et gaz de soudage	-	5	6	D
92	silice	-	-	5	5	D
109	amiante (fibre)	goudron	-	5	5	C

Le statut dynamique des motifs a été caractérisé pour l'année 2014 (Figure 34A), 26 motifs ont été identifiés comme **persistants** (id : 1-7,9-19, 21-24, 26-27, 31-32), 8 en **déclin** (id : 8, 20, 25, 28-30, 33-34), 7 en **émergence** (id : 35, 38, 51 63, 65, 79, 109) et 125 en état de **latence**, dont 10 ayant été considérés comme actifs en 2013 (id : 36-37, 39, 41, 44, 47, 52, 71, 82, 92). **Cinquante et un motifs ont donc été caractérisés comme actifs** et ont été représentés sur l'arbre de la multi-association (Figure 34B), dont 25 d'ordre supérieur à I. Les motifs les plus importants rapportés comme potentiellement associés au cancer de la vessie étaient les « Hydrocarbures Aromatiques Polycycliques » (HAP) (id: 1) d'un poids de 285 PST, enregistrés en association dans 30 autres motifs ; « l'amiante » (id : 2), enregistrée dans 160 PST, participant à la formation de 34 autres motifs ; les « solvants, diluants » (id : 66) avec un poids de 106 PST, participant à la formation de 21 autres motifs ; les « gaz d'échappement diesel » (id : 4) avec 103 enregistrements et participant à la formation de 16 autres motifs ; et enfin le « trichloroéthylène » avec 96 PST enregistrés et participant à la formation de 20 autres motifs.

Parmi les 7 motifs émergents, deux correspondaient à des motifs d'ordre I : « Poussière de bois » et « « Goudron » (id : 35 et 79) ; 4 à des motifs d'ordre II, impliquant majoritairement « l'amiante » et les « huiles et graisses lubrifiantes » (id : 38, 51, 63, 79) et un motif d'ordre III : l'association « Trichloroéthylène/HAP/Amiante » (id : 65).

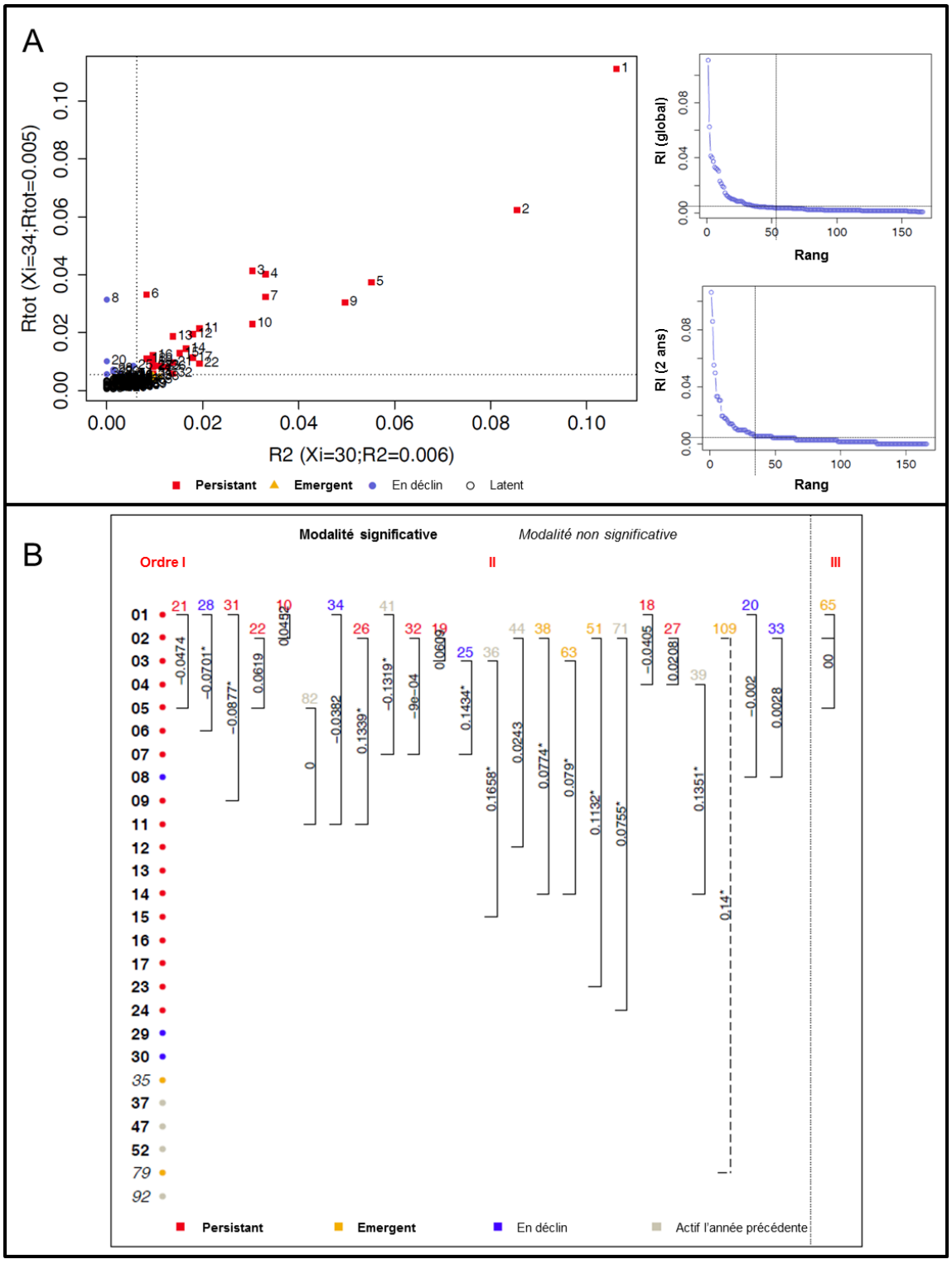


Figure 34 : Caractérisation des motifs issus de l'exposome du cancer de la vessie en 2014. A) Statut dynamique des 166 motifs en 2014. Les délimitations des RI ont été calculé à 0.006 pour les deux années glissantes et 0.005 pour le RI global. Il y avait 26 motifs persistants, 8 en déclin, 7 en émergence et 125 en latence. B) Arbre de la multi-association des 51 motifs actifs liés au cancer de la vessie en 2014, avec 26 motifs d'ordre I actifs, 24 motifs d'ordre II actifs et 1 motif d'ordre III actifs. ***Les numéros indiqués correspondent au numéro d'identification du Tableau 5.***

Le spectrosome des motifs d'expositions professionnelles associés au cancer de la vessie est disponible en Figure 35. Contrairement à l'exposome des nœuds de la Figure 33, seuls les 166 motifs identifiés par l'analyse de la SO ont été représentés, permettant la mise en évidence des grandes familles de nuisances associées au cancer de la vessie ainsi que la visualisation de la structure de ces différentes associations. Il est à noter que les motifs formant des clusters sur l'exposome étaient les mêmes motifs ayant été mis en évidence par leur poids : les « HAP », « l'amiante », les « solvants et diluants », les « échappements de gaz de diesel » et le « trichloroéthylène ». Ainsi, le cancer de la vessie est caractérisée par d'importants motifs d'ordre I participant également à la formation de nombreux autres motifs de multi-expositions, démontrant un caractère très ubiquitaire de ces expositions.

Le résultat final de la Surveillance Observationnelle pour le cancer de la vessie correspond au spectre des motifs d'expositions professionnelles (Figure 36), synthétisant pour chaque motif identifié les différents indicateurs de la surveillance : spécificité, statut dynamique par la couleur, poids total (hauteur de la barre), ordre du motif et implication dans la multi-exposition (hachure pour la mono-exposition et couleur pleine pour la multi-exposition). Seuls les motifs actifs ainsi que les 10 motifs latents en 2014, mais actifs en 2013, identifiés dans le Tableau 5 ont été présentés sur cette signature de l'exposition du cancer de la vessie afin d'en faciliter la lisibilité.

Quinze motifs ont été caractérisés comme spécifiques avec une information tendant à être localisée (id : 21, 23, 25-27, 29-31, 34, 39, 41, 47, 52, 63, 82), tandis que 11 ont été caractérisés comme non localisés comme par exemple la « poussière de bois » (id : 35) retrouvée dans 13 nœuds différents, chacun ayant un poids égal à 1. Les motifs d'expositions professionnelles les plus importants du cancer de la vessie ont une spécificité intermédiaire. Cependant, il est à noter que les motifs « amiante » et « solvants et diluants » (id : 2 et 3) tendent vers une non spécificité de l'information avec une valeur de spécificité inférieure à 0.5, et une moyenne de poids par nœud inférieure à 2 représentant la tendance de ces expositions à être enregistrées avec de nombreuses autres expositions.

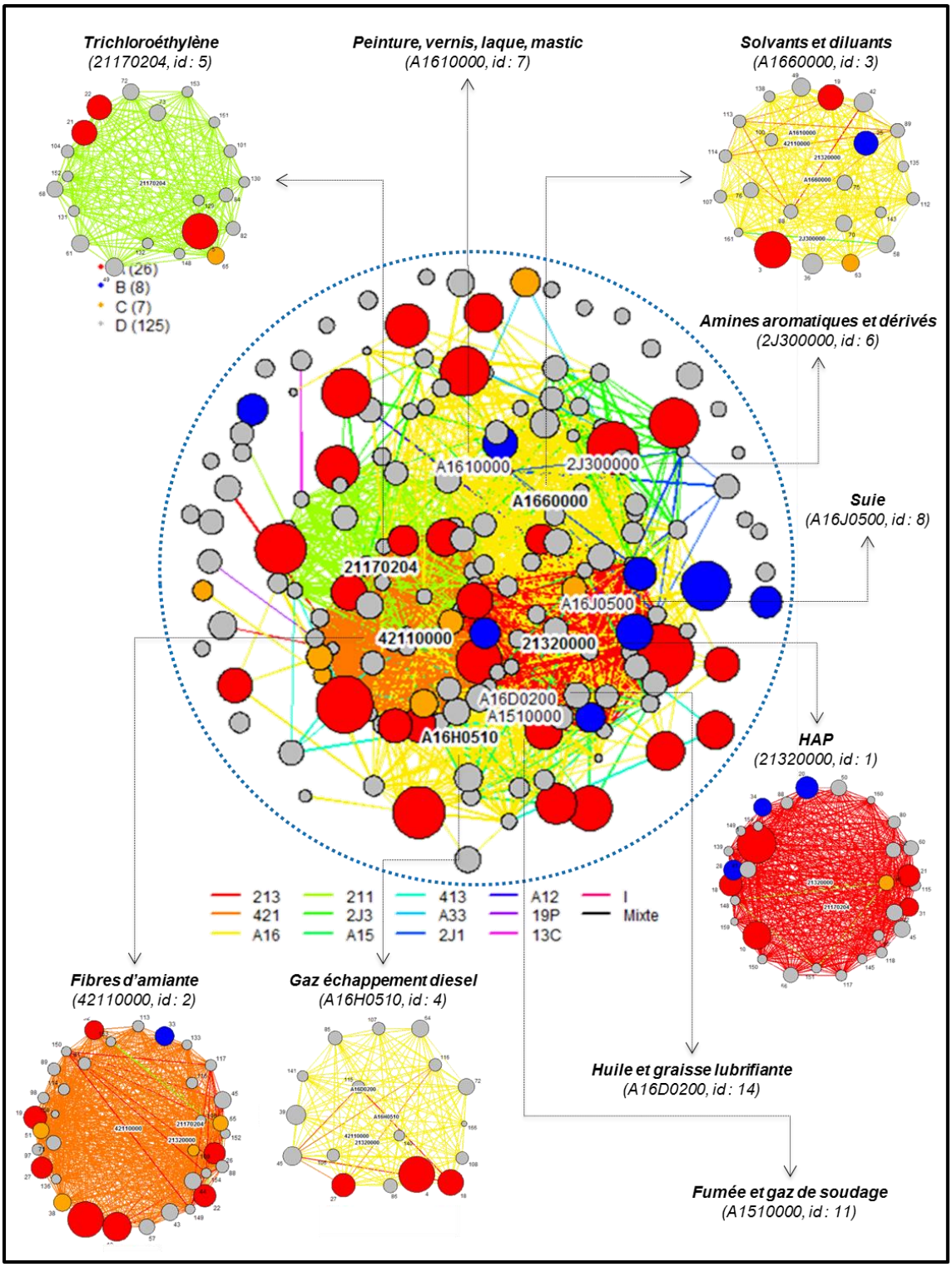


Figure 35 : Spectrosome des motifs du cancer de la vessie en 2014. Au centre le spectrosome des motifs d'expositions professionnelles associés au cancer de la vessie. La taille des nœuds correspond au poids du motif et la couleur correspond au statut dynamique du motif. La couleur des liens correspond aux différentes familles de nuisances identifiées en relation avec le cancer de la vessie. Cinq clusters ont été identifiés en raison du nombreux nombre de liens auxquels ils participent, de gauche à droite et de haut en bas : trichloroéthylène (5), solvants et diluants (3), HAP (1), fibres d'amiante (2) et gaz d'échappement diesel (4).

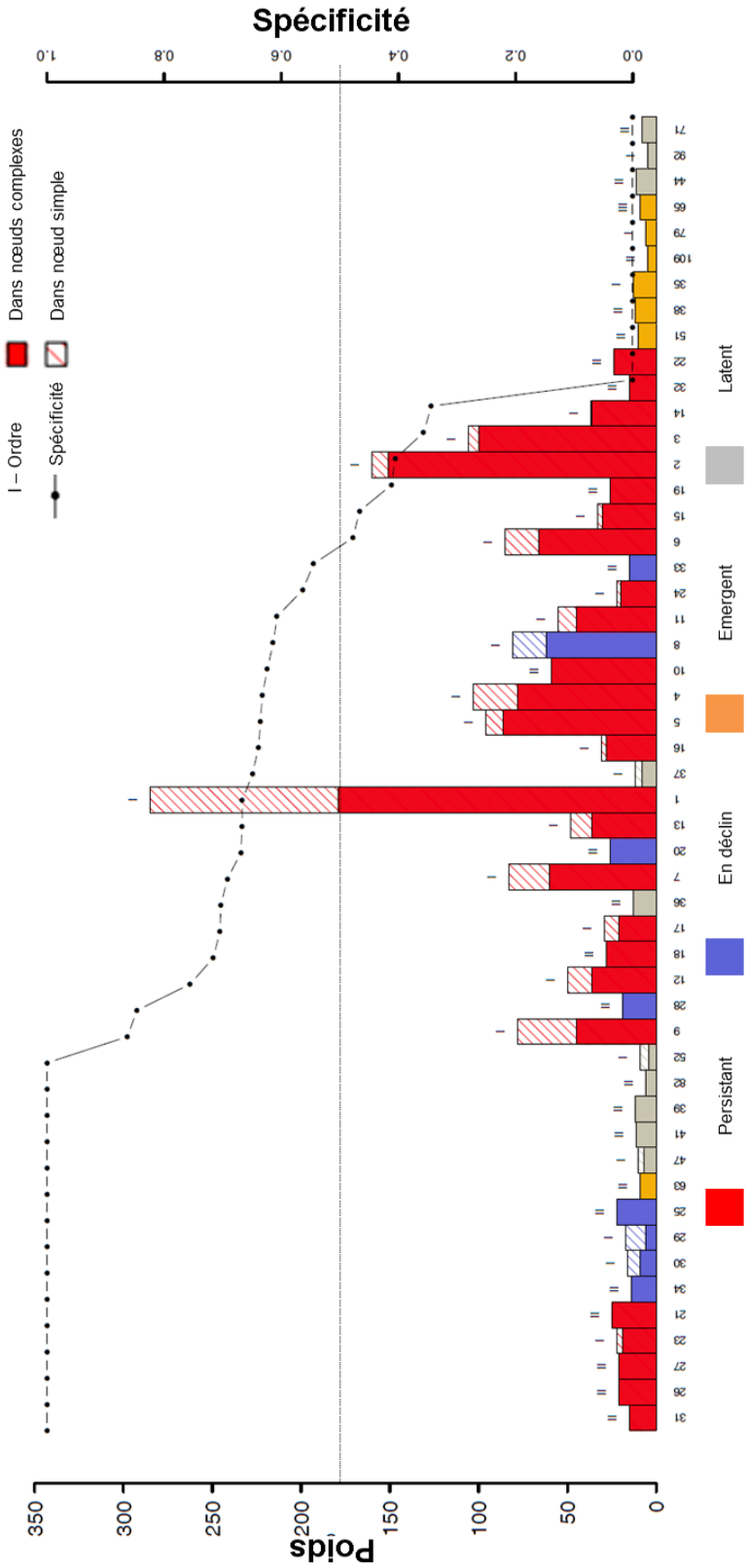


Figure 36 : Spectre du cancer de la vessie en 2014. Le spectre du cancer de la vessie a été caractérisé par 51 motifs actifs (dont 26 persistants, 8 en déclin, 7 en émergence et 10 latents mais actifs en 2013), parmi lesquels 26 étaient des motifs d'ordre I et 25 des motifs d'ordre supérieur. Quinze motifs avaient une localisation de leur information spécifique c'est-à-dire privilégiant une association en particulier, 11 étaient plutôt ubiquitaires avec une information apportée par différents nœuds de manière homogène et les autres étaient des motifs intermédiaires, dont les motifs 2 et 3 tendaient vers une spécificité avec une valeur supérieure à 0.5.

Dans le contexte professionnel du RNV3P, il était intéressant de croiser les informations concernant les motifs avec l'activité professionnelle (variable descriptive secondaire), permettant ainsi de caractériser un « contexte professionnel » à risque associant les expositions professionnelles et les activités à risque. Cependant, la matrice primaire-secondaire permettant ce croisement des motifs en fonction des activités professionnelles n'a pas été inclus du à sa taille trop importante (166 motifs pour 399 activités professionnelles)[5]. A titre d'exemple, le motif numéro 65 associant les « HAP », « l'amiante » et le « trichloroéthylène » est apparu en 2010 avec une première activité professionnelle « Conducteurs de machines à fabriquer du ciment » dans le secteur du « commerce de vaisselle et verrerie », puis a été ensuite identifié en 2012 dans l'activité « Mécaniciens et ajusteurs de machines agricoles et industrielles » en « production et distribution de combustible gazeux », en 2013 dans l'activité « Maçon » dans le secteur de la « construction », « Contrôleur de machine à vissage » dans le secteur de la « fabrication de vis et de boulon », puis en 2014 avec les activités professionnelles « Soudeurs et oxycoupeurs » dans le domaine de la « fabrication de produits chimiques inorganiques », « électromécanicien » dans le « transport ferroviaire », « constructeur de maison » dans le « bâtiment » et enfin « mineurs et conducteurs d'installations » dans le secteur de « l'extraction de houille ».

Lors du calcul des statuts dynamiques des motifs, il est également possible d'obtenir la trajectoire dynamique du motif, retraçant sur les différentes années d'étude l'évolution de son statut dynamique (Figure 34A) et de coupler cette information à l'activité professionnelle. Les trajectoires atypiques (présentant le plus de variations) du cancer de la vessie sont disponibles en Figure 37. La courbe bleue représente l'évolution du statut dynamique du motif, les courbes rouges et vertes de la partie haute représentent respectivement les valeurs de RI_{∞} et RI_2 pour chaque motif, et les courbes pointillées représentent les valeurs seuils ξ_{RI}. Les histogrammes de la partie basse représentent quant à eux le recrutement du motif au sein d'activité professionnelle particulière : l'orange clair correspondant à l'enregistrement du motif avec une activité professionnelle déjà rencontrée pour ce motif, l'orange foncé correspondant à l'enregistrement du motif en association avec une nouvelle activité professionnelle n'ayant pas encore été rencontrée pour ce motif particulier.

[5] Un exemple de matrice est présentée dans le Chapitre 9 : Lymphome non hodgkinien.

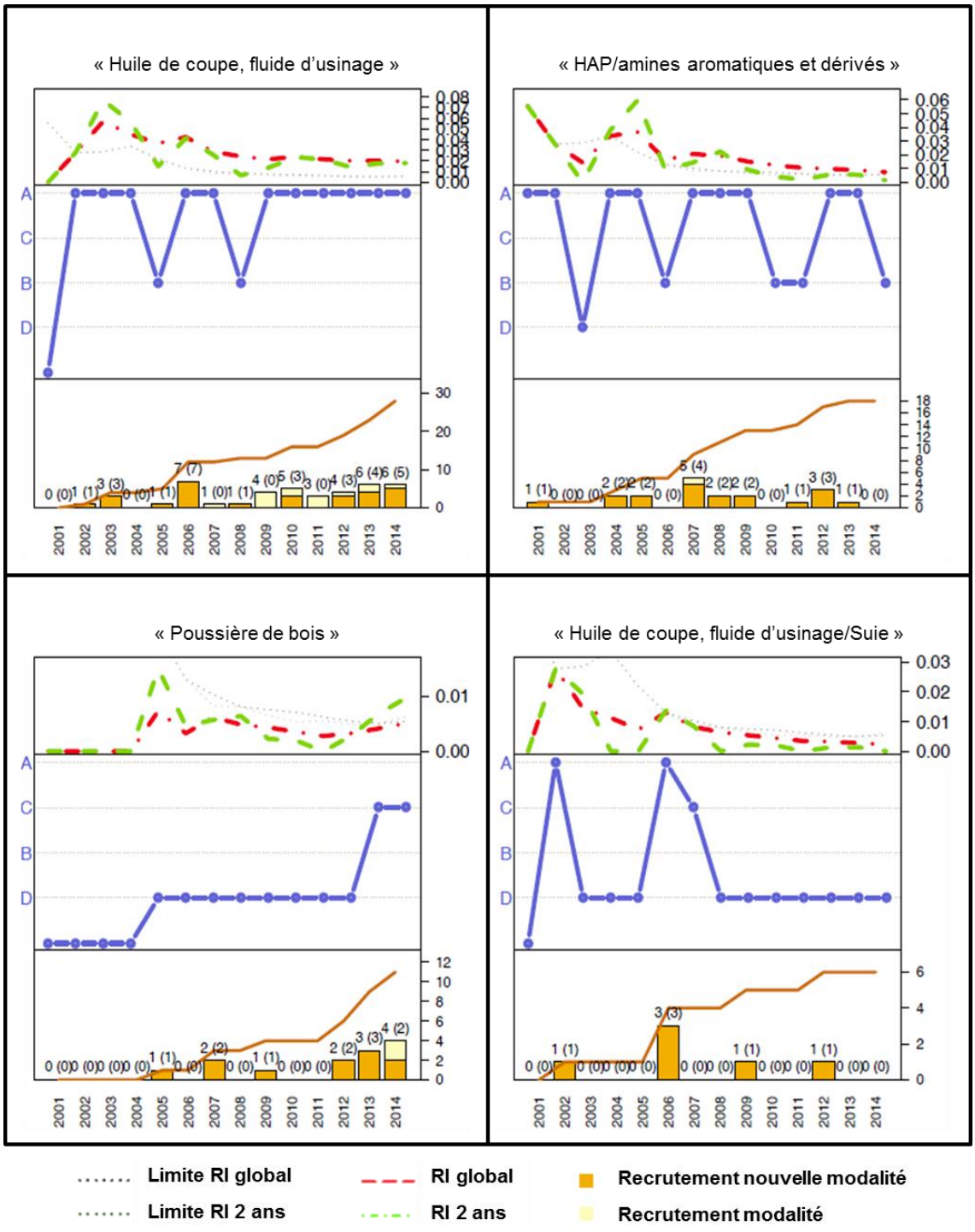


Figure 37 : Exemple de trajectoire atypique de motifs identifiés pour le cancer de la vessie de 2001 à 2014 pour chaque catégorie de statut dynamique : persistant, en déclin, émergent et latent. Exemple du motif « Huile de coupe, fluide d'usinage » (persistant), « HAP/amines aromatiques et dérivés » (en déclin), la « poussière de bois » (émergence) et le motif « Huile de coupe, fluide d'usinage/Suie » (latent). La courbe bleue correspond à l'évolution du statut dynamique des motifs : A – persistant, B – en déclin, C – émergent et D – latent. Les courbes rouge et verte correspondent respectivement aux valeurs de $\boldsymbol{RI}_{\infty}$ et $\boldsymbol{RI}_{\boldsymbol{2}}$ pour le motif à l'année donnée, les courbes pointillées correspondent aux limites $\boldsymbol{\xi}_{\boldsymbol{RI}}$ calculées en fonction des autres recrutements pour l'année donnée. La partie basse correspond à l'évolution du recrutement des motifs dans des activités professionnelles particulières, distinguant les nouvelles activités professionnelles et les activités déjà rencontrées.

Il est ainsi visible que les trajectoires dynamiques peuvent parfois être très instables, variant sans cesse d'une année à l'autre suivant les enregistrements effectués et suivant l'évolution de la connaissance pour chacune des étiologies différentes. Sont ainsi présentés les exemples des motifs « Huile de coupe, fluide d'usinage » persistant depuis 2009 ; le motif des « HAP/amines aromatiques et dérivés » dont la trajectoire est très fluctuante et régulièrement en déclin, comme c'est le cas en 2014 ; la « poussière de bois » apparue 2006 et en émergence depuis 2013 ; et le motif « Huile de coupe, fluide d'usinage/Suie » latent depuis 2008 malgré une forte apparition en 2001.

En Figure 38 est présenté un autre exemple, celui du motif émergent « HAP/amiante/trichloroéthylène » (id : 65), permettant l'illustration et l'intérêt de suivre des expositions professionnelles particulières, mais également de suivre l'évolution de leur association rapportée au cours du temps. Globalement les 3 motifs d'ordre I ont été caractérisés par un statut persistant, malgré la fluctuation du recrutement du motif du « trichloroéthylène » en 2003 et 2004, et leur association d'ordre III est apparue en 2010. Cependant, il est à noter que les différentes associations d'ordre II sont apparues respectivement en 2005, 2006 et 2008 pour les « HAP/amiante », les « HAP/trichloroéthylène » et l'«amiante/trichloroéthylène ». Ainsi cette association d'ordre III ne semble pas être « apparue » sous l'impulsion du recrutement des autres. Ces motifs d'ordre I sont déjà reconnus en tant que facteur de risques, mais leur association pourrait avoir d'autres conséquences cliniques en cas de synergie. Enfin, cette association n'a été retrouvée que dans 9 activités professionnelles différentes tandis que les différents motifs d'ordre I ont été retrouvés respectivement dans 125, 87 et 62 activités distinctes.

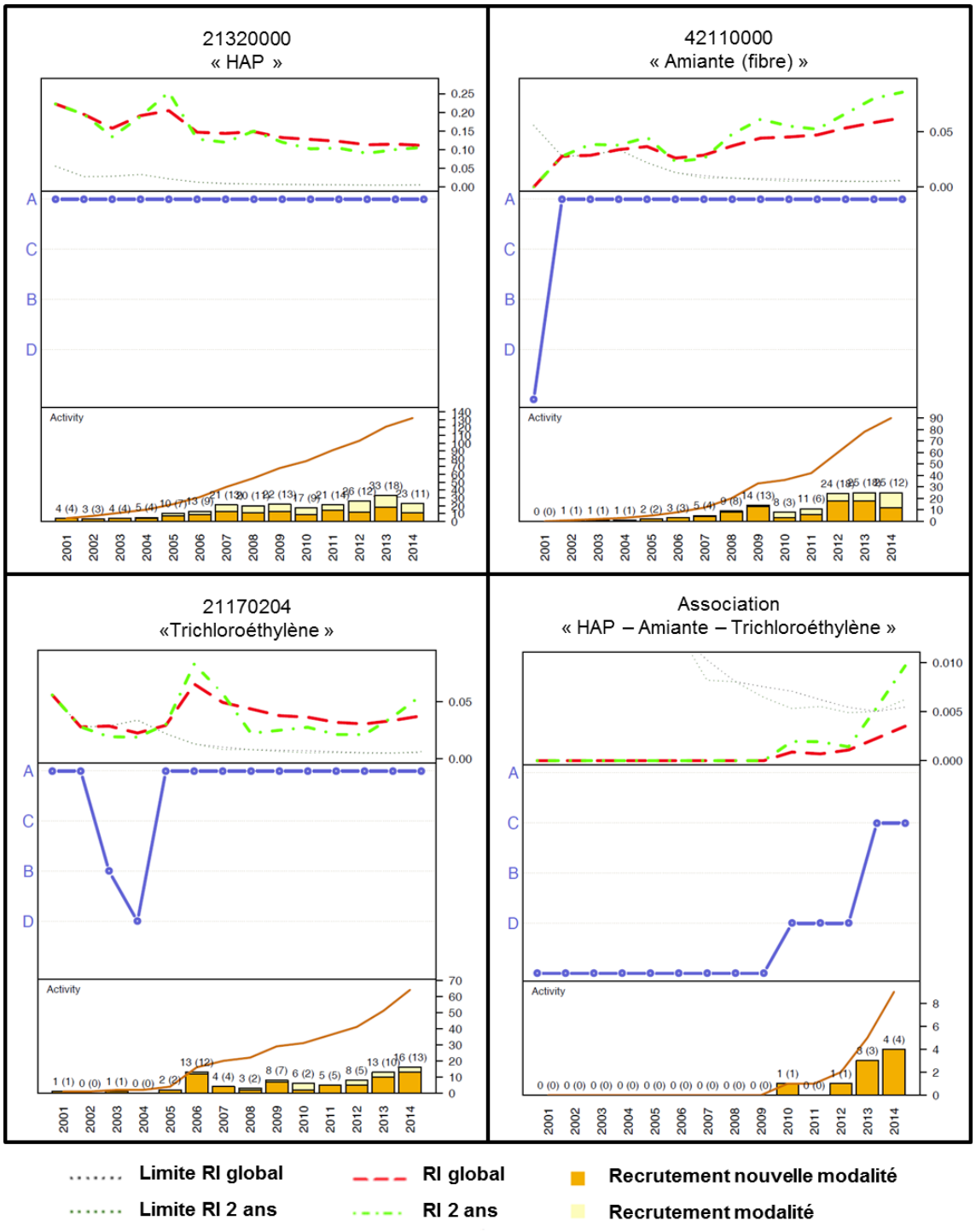


Figure 38 : Exemple de trajectoire dynamique de trois motifs d'ordre I, HAP, amiante et trichloroéthylène, et leur association d'ordre III, associés au cancer de la vessie, de 2001 à 2014. La courbe bleue correspond à l'évolution du statut dynamique des motifs : A – persistant, B – en déclin, C – émergent et D – latent. Les courbes rouge et verte correspondent respectivement aux valeurs de RI_{∞} et RI_2 pour le motif à l'année donnée, les courbes pointillées correspondent aux limites ξ_{RI} calculées en fonction des autres recrutements pour l'année donnée. La partie basse correspond à l'évolution du recrutement des motifs dans des activités professionnelles particulières, distinguant les nouvelles activités professionnelles et les activités déjà rencontrées.

7.2. Conclusion

Le cancer de la vessie, cancer professionnel avéré et dont de nombreuses expositions sont connues et reconnues au titre des maladies professionnelles et de l'assurance maladie, a été caractérisé par **un très grand nombre de motifs, 166, dont 51 ont été considérés comme actifs en 2014**. Le cancer de la vessie a été caractérisé par **une forte multi-exposition car 67.5% des motifs étaient d'ordre supérieur à I**. La Surveillance observationnelle a permis non seulement de mettre en évidence les motifs les plus importants rapportés comme associés au cancer de la vessie au sein de la base RNV3P : par ordre décroissant d'importance, **les HAP, l'amiante, les solvants et diluants et le trichloroéthylène, mais également des huiles, des gaz d'échappement et de la fumée de soudage**, cohérents avec la littérature, mais également de caractériser leur statut dynamique et ainsi l'évolution de la connaissance. Ainsi, il a été mis en évidence l'émergence d'un motif particulier associant « HAP/Amiante/Trichloroéthylène » nécessitant une investigation supplémentaire afin de comprendre la raison de cette « multi-exposition ».

L'application de la méthodologie de la Surveillance Observationnelle à une pathologie à latence longue et chronique, mais dont les expositions sont connues et reconnues, a permis d'éprouver la méthodologie et de la comparer avec la littérature, mais également de démontrer l'importance de la prise en compte des multi-expositions (dans ce cas présent) associées à un évènement de santé.

8. ASTHME

L'asthme, maladie chronique des voies aériennes représente une importante part des maladies respiratoires avec plus de 200 millions de personnes concernées dans le monde [WHO, 2013]. La prévalence en France pour l'adulte est estimée à 6% [Finger, 2012] et 15 à 20% de ces asthmes seraient attribuables à des expositions professionnelles [Toren et Blanc, 2009]. Environ 400 expositions professionnelles sont reconnues comme facteur de risques, leur nombre augmentant chaque année [Finger, 2012], impliquant majoritairement des farines, du bois, le latex et des allergènes d'origine végétale ou animale, pouvant se rencontrer dans de nombreux environnement professionnels différents.

L'asthme étant différent du cancer de la vessie précédemment étudié, c'est à dire une pathologie aigue, à latence courte mais dont les expositions sont également bien connues, il était intéressant d'appliquer la méthode de la Surveillance Observationnelle. L'objectif étant d'étudier les expositions professionnelles associées à l'asthme contracté dans un contexte professionnel et à partir des données de qualité du RNV3P.

Les codes CIM-10 utilisés pour cette analyse sont : J45*, en excluant les codes J451 et J458, correspondant aux critères de sélections d'un précédent travail présenté dans le rapport scientifique de l'ANSES en 2011 [ANSES, 2011].

8.1. Analyse

Parmi les 107 713 PST de la base RNV3P et les 7 203 cas de maladies chroniques des voies respiratoires inférieures, les cas d'asthme ont représenté un total de 5 336 PST (Figure 39), dont 2 969 hommes et 2 367 femmes, avec une moyenne d'âge de 40.43 et 39.86 ans respectivement. Un nombre total de 714 nuisances distinctes ont été enregistrées depuis 2001, majoritairement des agents chimiques organiques (ester, amine aliphatique ou aromatique) (n=170), des agents biologiques (champignon, moisissure, animaux et végétaux) (n=102) et des produits ou matériaux ou des processus industriels (produit d'entretien ou d'hygiène) (n=303). De même, 638 activités furent identifiées, notamment dans le secteur de l'industrie agricole et alimentaire (n=19) et le secteur de la santé et de l'action sociale (n=24).

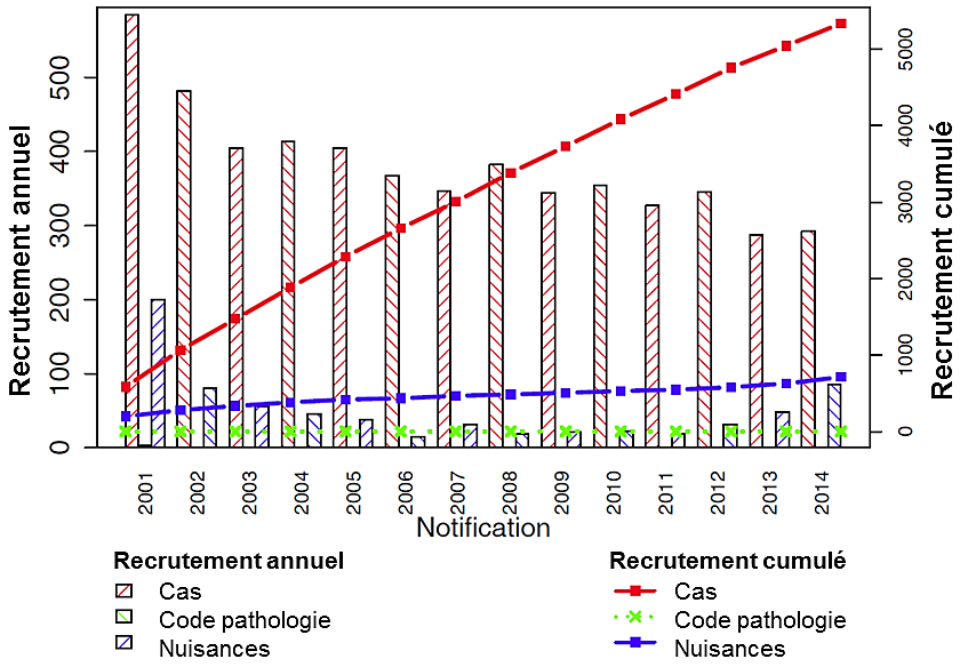


Figure 39 : Recrutement et effectif cumulé des cas, des nuisances distinctes et des codes pathologies distincts pour l'asthme de 2001 à 2014.

Parmi les 714 nuisances distinctes, 236 furent considérées comme significatives par le calcul basé sur l'imputabilité, et 4 959 PST possédaient au moins une de ces nuisances dans leur cortège. A partir de ces 4 959 PST, 1 544 cortèges distincts d'expositions professionnelles ont été identifiés pour former les nœuds de l'exposome pour un total de 63 628 liens entre ces nœuds (Figure 40). La couleur des liens de l'exposome correspond cette fois-ci à des familles d'expositions professionnelles au 2ème niveau d'imbrication, par conséquent plus générale que les familles de l'exposome du cancer de la vessie, à cause de la diversité très importante des expositions professionnelles rencontrées.

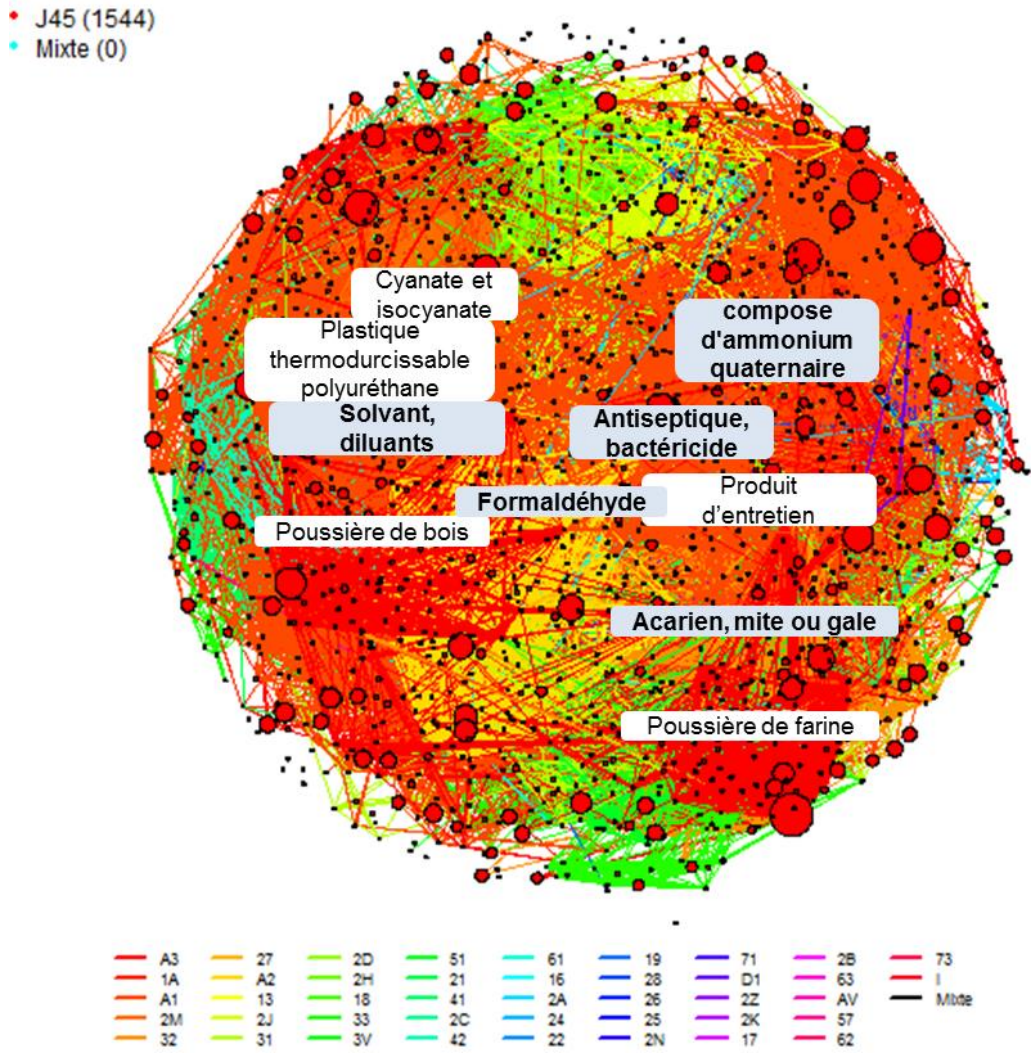


Figure 40 : Exposome de l'asthme en 2014. La taille des nœuds correspond au poids des nœuds (nombre de PST identiques), la couleur correspond aux différents codes pathologies. La couleur des liens correspond aux différentes familles (2e niveau du code imbriqué) de nuisances identifiées en relation avec l'asthme et permettent d'identifier les clusters de l'exposome, les 5 majoritaires étaient : « Solvants, diluants » (A1660000), « Antiseptique et bactéricide » (A16L0350), « composé d'ammonium quaternaire » (1AN06600), « formaldéhyde » (27110100) et « acarien, mite ou gale » (32110110).

La majorité des liens identifiés entre les nœuds correspondait à des nuisances de type « Matériel et produit industriel divers » (Code A16), dont les peintures, les solvants et les produits d'entretien ou d'hygiène ; puis la famille « Azote et ses composés inorganiques » (Code 1AN) dont les composés d'ammonium quaternaire ; les « acarien ou mite ou gale » (Code 32110110) et les « produits d'origine végétale » (Code A33) telle que la poussière végétale. Enfin 36 nœuds isolés ont été mis en évidence dont 5 ayant été enregistrés au moins deux fois : les « composés chimiques carcinogènes » (A16M0100), l'association « ail » et « oignon » (33350100 et 33350L00), les « adhésifs, colle polyuréthane » (A1630700), le « fromage » (A31P0200) et les « huiles essentielles » (A33D0000).

Un total de 436 motifs a été identifié comme motifs d'expositions professionnelles, la liste et leur numéro d'identification est disponible dans le tableau en Annexe 2. La liste partielle des 87 motifs **actifs** est disponible dans le Tableau 6. Parmi ces 436 motifs, 196 correspondaient à des motifs d'ordre I (une seule modalité), impliquant un total de 240 motifs d'ordre supérieurs à I : 217 motifs d'ordre II et 24 motifs d'ordre III.

Tableau 6 : Liste des 87 motifs actifs de l'asthme identifiés par la Surveillance Observationnelle en 2014.

ID	Ordre I	Ordre II	Ordre III	Nœuds	Poids	Statut
1	poussière de farine	-	-	68	769	A
2	composé d'ammonium quaternaire	-	-	117	440	A
3	cyanate et isocyanates	-	-	79	298	A
4	oxydant et décolorant (cosmétique)	-	-	35	274	A
5	acarien ou mite ou gale	-	-	99	221	A
6	produit d'entretien	-	-	74	218	A
7	produit capillaire	-	-	37	212	A
8	poussière de bois	-	-	76	203	A
9	antiseptique, bactéricide pour la peau (sauf SHA solutions hydro-alcooliques)	-	-	92	198	A
10	protéine de latex	-	-	67	179	B
11	solvant, diluant	-	-	100	173	A
12	formaldéhyde	-	-	88	163	A
13	plastique thermodurcissable polyuréthane	-	-	73	159	A
14	peinture, vernis, laque, mastic	-	-	66	153	A
15	amine aliphatique	-	-	64	136	A
16	lessive détergent	-	-	51	134	A
17	poussière inorganique	-	-	66	128	A
18	fumée et gaz de soudage	-	-	44	108	A
19	glutaraldéhyde	-	-	32	87	A
20	poussière organique	-	-	46	85	A
21	chlore	-	-	29	72	A
22	persulfate	-	-	23	68	A
23	enzyme	-	-	15	65	B
24	résine époxydique	-	-	35	62	A
25	adhésif, colle	-	-	37	61	A
26	chrome	-	-	43	59	B
27	champignon, moisissure	-	-	33	58	A
28	colophane	-	-	28	57	B
29	graminée céréale	-	-	33	53	A
30	huile de coupe, fluide d'usinage	-	-	25	48	B
31	acarien ou mite ou gale	poussière de farine	-	14	47	A
32	composé d'ammonium quaternaire	amine aliphatique	-	16	46	B

33	méthacrylate de méthyle	-	-	26	45	A
34	pollens	-	-	31	44	A
35	amiante (fibre)	-	-	38	43	A
36	produit capillaire	oxydant et décolorant (cosmétique)	-	9	43	B
37	poussière de chantier	-	-	22	42	A
38	enzyme	poussière de farine	-	7	41	B
39	poussière de ciment, aluminosilicate de calcium	-	-	26	40	A
40	détergent désinfectant pour locaux (sols, murs, surfaces) domestique et professionnel	-	-	18	38	A
41	composé d'ammonium quaternaire	glutaraldéhyde	-	17	38	B
42	nickel	-	-	29	37	B
43	caoutchouc, latex naturel	-	-	26	37	A
44	teinture (cosmétique)	-	-	9	37	B
45	rat	-	-	16	36	A
46	encre	-	-	23	35	B
47	adhésif, colle cyanoacrylate	-	-	20	34	B
48	composé d'ammonium quaternaire	antiseptique, bactéricide pour la peau (sauf SHA solutions hydro-alcooliques)	-	14	34	B
49	trichlorure d’azote	-	-	8	33	A
50	poussière de papier carton	-	-	16	33	A
51	composé d'ammonium quaternaire	produit d’entretien	-	12	33	B
52	peinture, vernis, laque, mastic sans précision en phase solvant	-	-	23	31	D
55	ambiance thermique froide	-	-	18	29	D
59	poussière d'origine végétale	-	-	15	28	C
61	silice	-	-	21	27	C
62	peinture, vernis, laque, mastic	solvant, diluant	-	14	26	D
64	acide chlorhydrique	-	-	23	24	C
65	méthacrylate	-	-	12	24	C
67	ambiance humide	-	-	21	24	C
69	décapant	-	-	19	23	D
71	blé	-	-	13	22	C
75	ef3 physique	-	-	15	20	D
81	sel de sodium de la N-chloro-p-toluènesulfonamide	-	-	5	19	D
83	fourrure, feutre	-	-	9	19	D
106	tabac	-	-	6	15	C
108	produit de décomposition thermique de matière plastique	-	-	6	15	D
109	décolorant, éclaircissant capillaire	-	-	12	15	C
111	composé d'ammonium quaternaire	détergent désinfectant pour locaux (sols, murs, surfaces) domestique et professionnel	-	7	15	C
112	oxydant et décolorant (cosmétique)	teinture (cosmétique)	-	4	15	D
118	seigle	-	-	9	14	C
125	arbre et plante	-	-	8	13	C

131	acarien ou mite ou gale	graminée céréale	-	8	12	C
139	persulfate	décolorant, éclaircissant capillaire	-	8	11	C
140	acarien ou mite ou gale	pollens	-	10	11	C
149	chat	-	-	8	10	D
162	blé	seigle	-	4	9	C
168	persulfate	oxydant et décolorant (cosmétique)	-	6	9	C
174	chien	-	-	6	8	D
199	équipement de protection	-	-	7	7	D
225	acarien ou mite ou gale	détergent désinfectant pour locaux (sols, murs, surfaces) domestique et professionnel	-	2	6	C
279	maïs	-	-	4	4	C
280	sarrasin	-	-	4	4	C
283	solvant de nettoyage	-	-	4	4	C
289	alpha amylase	-	-	3	4	C
305	décolorant, éclaircissant capillaire	oxydant et décolorant (cosmétique)	-	3	4	C
313	produit d'entretien	poussière inorganique	-	3	4	D
327	persulfate	décolorant, éclaircissant capillaire	oxydant et décolorant (cosmétique)	3	4	C

Le statut dynamique des motifs a été caractérisé pour l'année 2014 (Figure 41), 36 motifs ont été identifiés comme **persistants**, 15 en **déclin**, 23 en **émergence** et 362 en état de **latence**, dont 13 ont été considérés comme actifs en 2013.

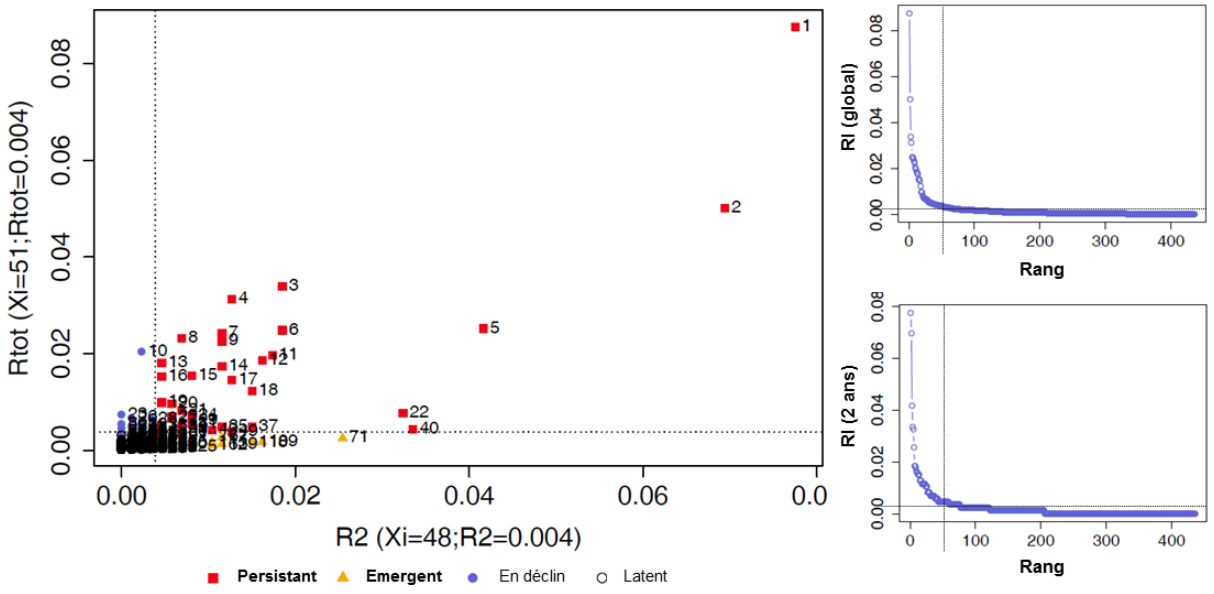


Figure 41 : Statut dynamique des 438 motifs de l'asthme en 2014. Les délimitations des RI a été calculé à 0.004 pour les deux années glissantes et 0.004 pour le RI global. Il y avait 36 motifs persistants, 16 en déclin, 23 en émergence et 363 en latence. ***Les numéros indiqués correspondent au numéro d'identification du Tableau 6.***

Quatre-vingt-sept motifs ont été caractérisés comme actifs et ont été représentés sur l'arbre de la multi-association (Figure 42), dont 19 d'ordre supérieur à I. Les motifs les plus important associés à l'asthme étaient les « poussières de farine » (id: 1), recrutées dans 769 PST, et en association dans 18 autres motifs ; suivi par les « composés d'ammonium quaternaire » (id : 2) dans 440 PST, en association avec 21 autres motifs ; les « cyanates et isocyanates » (id : 3) enregistrés dans 298 PST et formant des associations avec 13 autres motifs ; les « oxydants et colorants (cosmétiques) » (id : 4) présents dans 274 PST et formant 11 autres motifs ; et enfin les « acarien ou mite ou gale » (id : 5) et les « produits d'entretien » (id : 6) enregistrés dans 211 et 218 PST respectivement, et participant à la formation de 23 et 18 autres motifs respectivement. Parmi les 23 motifs en émergence, 14 étaient d'ordre I et 9 d'ordre II. Trois de ces motifs émergents concernaient l'exposition aux « acarien ou mite ou gale » en association avec des « graminées », des « pollens » ou des « détergents et désinfectants ». Deux autres concernaient le « blé », notamment en association avec le « seigle ». Cinq motifs émergents concernaient également les différentes associations entre les « décolorants et éclaircissants capillaire », les « oxydants et décolorants (cosmétiques) » et les « persulfates ». En enfin, les autres motifs concernaient soit des produits d'origine végétale (poussière, sarrasin), soit des composés chimiques (composé d'ammonium quaternaire, méthacrylate).

Le spectrosome des motifs d'expositions professionnelles associés à l'asthme est disponible en Figure 43. Il est à noter que la majorité des motifs mis en évidence par le poids ont été retrouvés sur l'exposome. Ainsi l'asthme est caractérisé par d'importantes expositions particulières mais permettant un grand nombre d'associations et structurant l'information de manière très dense.

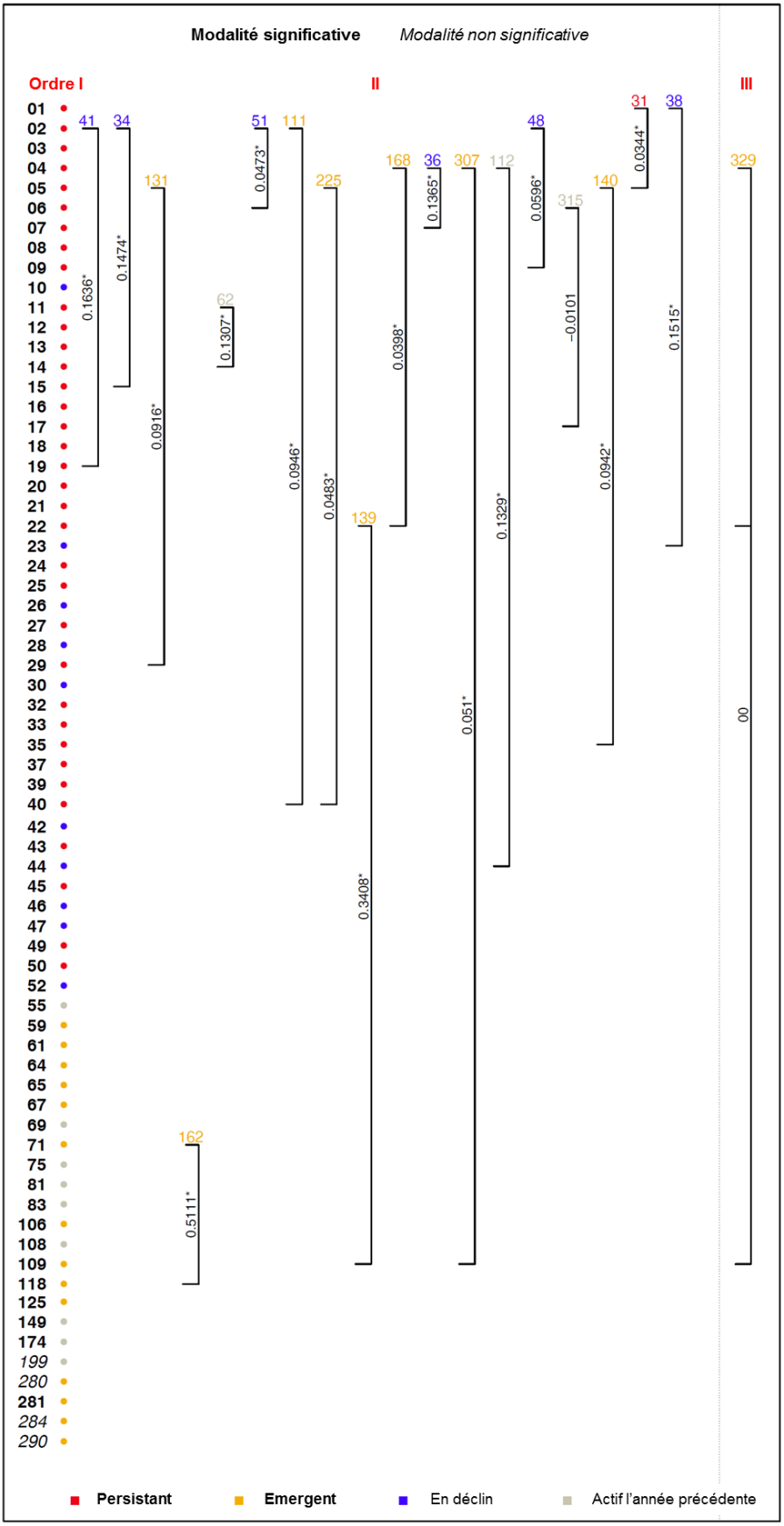


Figure 42 : **Arbre de la multi-association des 87 motifs actifs liés à l'asthme en 2014, dont 68 motifs d'ordre I actifs et 19 motifs d'ordre supérieurs actifs.** ***Les numéros indiqués correspondent au numéro d'identification du Tableau 6.***

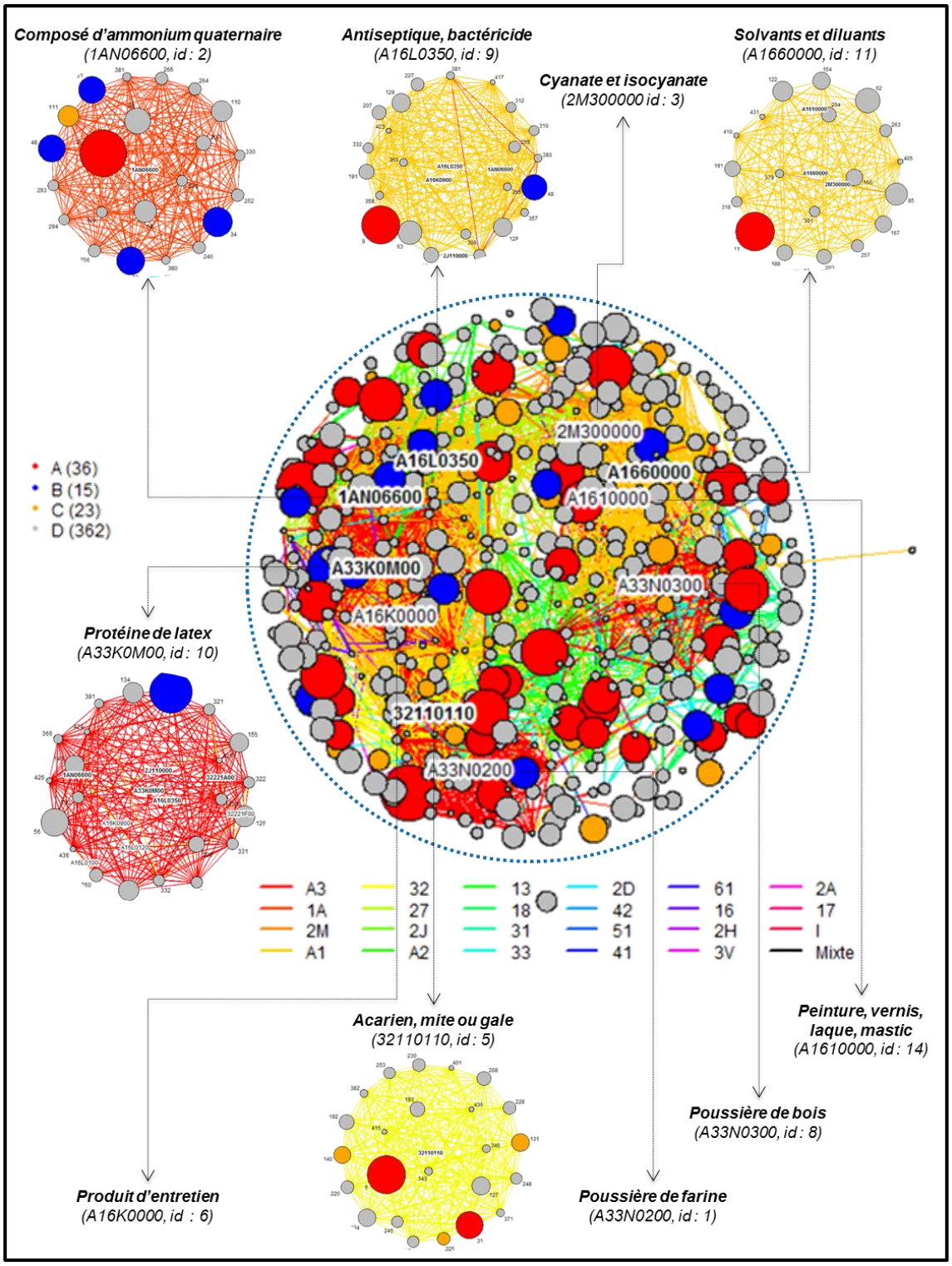


Figure 43 : Spectrosome des motifs de l'asthme en 2014. Au centre du spectrosome des motifs d'expositions professionnelles associés à l'asthme. La taille des nœuds correspond au poids du motif et la couleur correspond au statut dynamique du motif. La couleur des liens correspond aux différentes familles de nuisances identifiées en relation avec l'asthme. Des motifs particuliers ont été identifiés en raison du nombreux nombre de liens auxquels ils participent, de gauche à droite et de haut en bas : Composé d'ammonium quaternaire (2), antiseptique, bactéricide (9), solvants et diluants (11), protéine de latex (10) et acarien ou mite ou gale (5).

Le spectre issu de la Surveillance Observationnelle pour l'asthme est disponible en Figure 44. Dix-sept motifs ont été caractérisés comme spécifiques avec une information tendant à être localisée, tandis que 4 ont été caractérisés comme non localisés. La majorité des motifs intermédiaires ont une valeur de spécificité supérieure à 0.5 et tendent à être localisés dans des associations de modalités spécifiques. C'est le cas par exemple de la « poussière de farine » (id : 1) qui avec un poids de 769 PST distribués dans 68 nœuds, dont 623 PST ont été enregistré de manière identique et ont formé un seul nœud distinct. C'est donc un motif très spécifique, enregistré majoritairement, dans ce cas précis, sous forme de nœud simple. A l'inverse les « solvants et diluants » (id : 11) ont été enregistrés dans 173 PST distribués dans 100 nœuds différents et représentent donc un motif ubiquitaire.

Les trajectoires atypiques de l'asthme sont disponibles en Figure 37. Sont présentés les exemples des motifs « pollens » persistant en 2014 mais avec un recrutement variable depuis 2001 ; le motif des « huile de coupe, fluide d'usinage » persistant depuis 2001 mais instable depuis 2007 et en déclin en 2014 ; les « poussières d'origine végétale » qui après une phase de latence de 2007 à 2013 sont à nouveau en émergence depuis 2013 ; et le motif « peinture, vernis, laque, mastic sans précision en phase solvant » en augmentation depuis 2001 et instable depuis 2007.

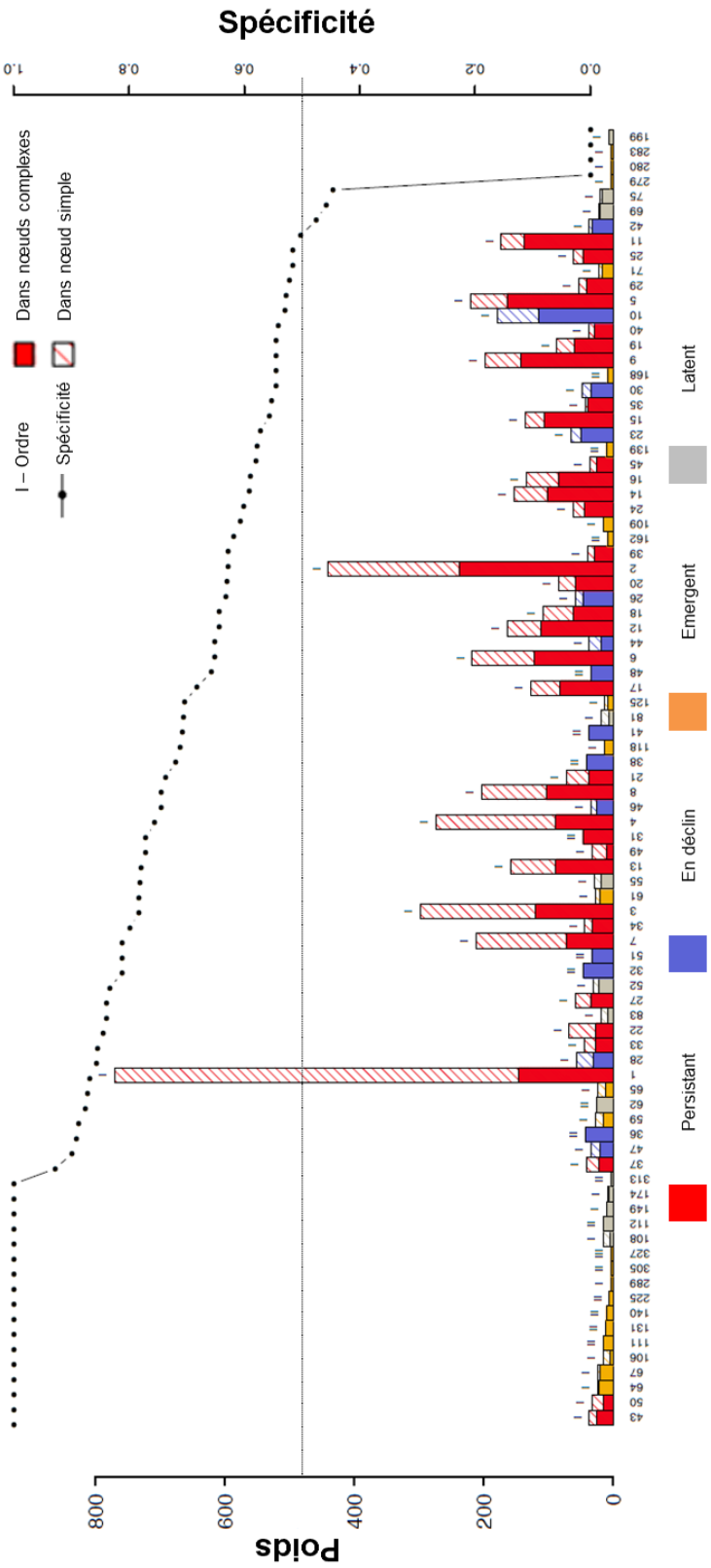


Figure 44 : Spectre de l'asthme en 2014. Le spectre de l'asthme a été caractérisé par 47 motifs actifs (dont 8 persistants, 6 en déclin, 1 en émergence et 3 latents mais actifs en 2013), parmi lesquels 13 étaient des motifs d'ordre I et 5 des motifs d'ordre II. Huit motifs ont eu une localisation de leur information spécifique c'est-à-dire privilégiant une association en particulier, 6 étaient plutôt ubiquitaires avec une information apportée par différents nœuds de manière homogène et 4 étaient des motifs intermédiaires mais dont la spécificité tendaient vers 1 car possédant une grosse partie de leur poids réparti sur un seul nœud simple.

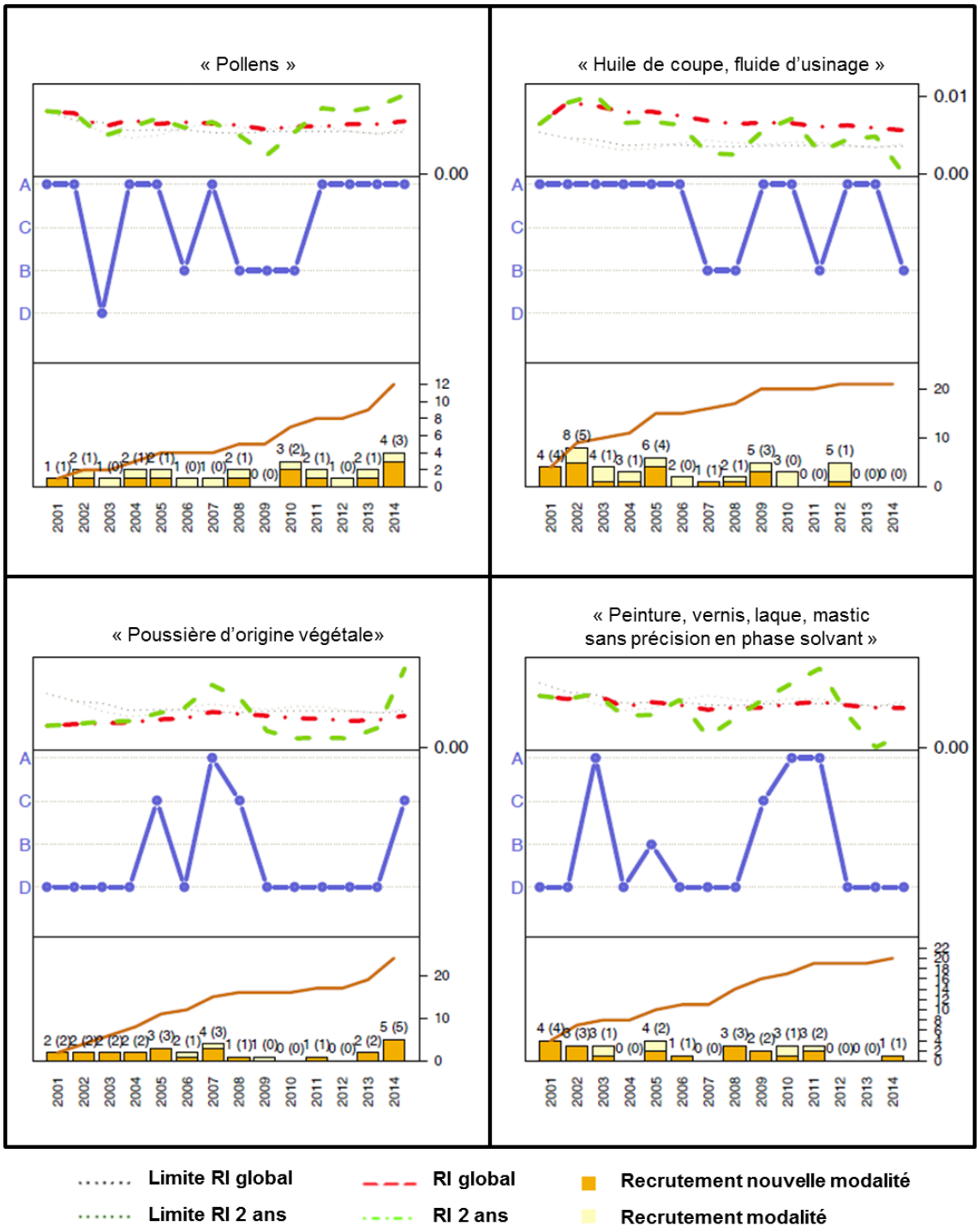


Figure 45 : Exemple de trajectoire atypique de motifs identifiés pour l'asthme de 2001 à 2014 pour chaque catégorie de statut dynamique : persistant, en déclin, émergent et latent. Exemple du motif « Pollens » (persistant), « Huile de coupe, fluide d'usinage » (en déclin), «Poussière d'origine végétale » (en émergence) et la « peinture, vernis, laque, mastic, sans précision (latent). La courbe bleue correspond à l'évolution du statut dynamique des motifs : A – persistant, B – en déclin, C – émergent et D – latent. Les courbes rouge et verte correspondent respectivement aux valeurs de $\boldsymbol{RI}_{\infty}$ et $\boldsymbol{RI}_{\boldsymbol{2}}$ pour le motif à l'année donnée, les courbes pointillées correspondent aux limites $\boldsymbol{\xi}_{\boldsymbol{RI}}$ calculées en fonction des autres recrutements pour l'année donnée. La partie basse correspond à l'évolution du recrutement des motifs dans des activités professionnelles particulières, distinguant les nouvelles activités professionnelles et les activités déjà rencontrées.

La matrice de croisement des motifs en fonction des activités professionnelles n'a pas été incluse en raison de sa taille importante (436 motifs pour 638 activités professionnelles). Cependant, à l'instar du cancer de la vessie, un exemple de suivi de trajectoire est présenté en Figure 46. Il s'agit de l'exemple de l'association d'ordre II du motif de la « poussière de farine » et du motif « acarien ou mite ou gale ».

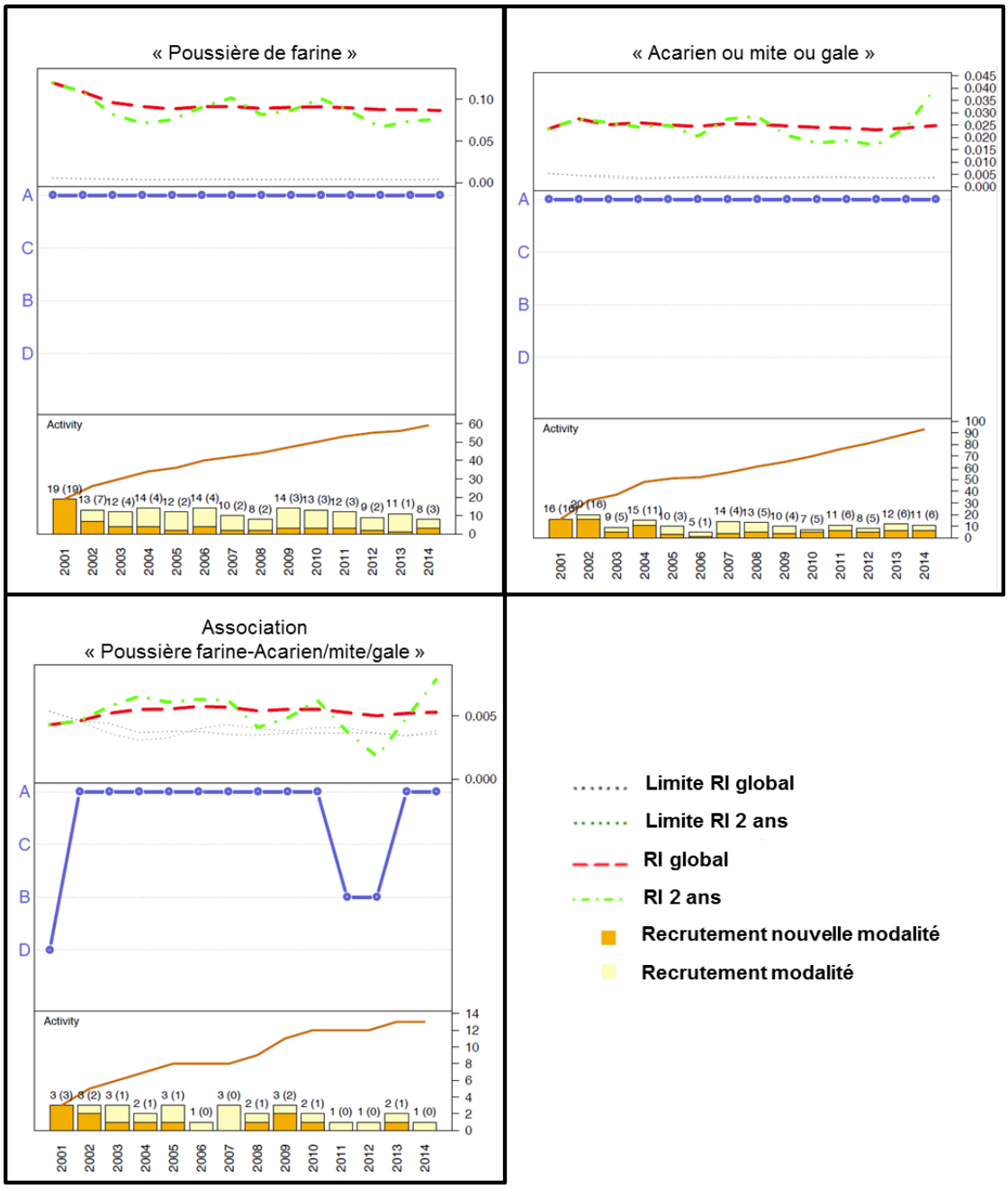


Figure 46 : Exemple de trajectoire dynamique de deux motifs d'ordre I, poussière de farine, et acarien/mite/gale, et leur association d'ordre II, associés à l'asthme, de 2001 à 2014. La courbe bleue correspond à l'évolution du statut dynamique des motifs : A – persistant, B – en déclin, C – émergent et D – latent. Les courbes rouge et verte correspondent respectivement aux valeurs de $\boldsymbol{RI_{\infty}}$ et $\boldsymbol{RI_2}$ pour le motif à l'année donnée, les courbes pointillées correspondent aux limites $\boldsymbol{\xi_{RI}}$ calculées en fonction des autres recrutements pour l'année donnée. La partie basse correspond à l'évolution du recrutement des motifs dans des activités professionnelles particulières, distinguant les nouvelles activités professionnelles et les activités déjà rencontrées.

Ces deux motifs d'ordre I sont des motifs persistants depuis 2001, cependant, leur association n'a pas été aussi stable, car en 2011 et 2012 le motif a été classé dans la catégorie en déclin, son recrutement sur les deux années glissantes étant en baisse par rapport aux autres informations apportées au sein de la base. Cette différence de profil entre les motifs d'ordre I et les associations d'ordre II met en évidence la nécessité de prendre en compte les recrutements multiples au sein des bases de données, notamment au sein du RNV3P car la multi-exposition est très présente dans un contexte professionnel. Enfin, cette association a également été identifiée dans un plus petit nombre d'activités professionnelles que les motifs d'ordre I la composant. Treize activités professionnelles distinctes ont été associées à ce motif d'ordre II, pour 56 et 88 respectivement pour les deux motifs d'ordre I « poussière de farine » et « acarien ou mite ou gale ».

8.2. Conclusion

L'asthme a été caractérisé par un très grand nombre de nuisances associées, ainsi que de motifs d'expositions professionnelles associés. Il s'agit pour la plupart d'association « attendues », comme les associations entre le « blé » et le « seigle » ou bien les « composés d'ammonium quaternaire » et les « détergents ». Cependant cet exemple a permis de mettre en évidence la nécessité de tenir compte de la multi-association, y compris dans le cadre d'une pathologie bien connue et reconnue. Le spectre de l'asthme a été caractérisé par **438 motifs, dont 87 actifs en 2014**. Parmi ces 438, 240 motifs étaient d'ordre supérieur à I, présentant **une multi-exposition plus faible** que pour le cancer de la vessie. Les motifs les plus importants rapportés comme associés à l'asthme sont **la poussière de farine, les composés d'ammonium quaternaire, les acariens, et les produits d'entretien**.

L'application de la méthodologie de la Surveillance Observationnelle à l'asthme a permis de démontrer l'adaptabilité de la méthodologie pour des pathologies ayant des comportements et latences différents (chronique ou aigue). De même, cet exemple permet d'illustrer le caractère unique du spectre obtenu pour chaque analyse, de par le nombre de motifs identifiés (évaluation de l'état de la connaissance), le nombre de motifs d'ordre I ou d'ordre supérieur (mise en évidence de la multi-exposition) et le dynamisme de ces motifs (évolution de la connaissance et émergence des risques).

9. LYMPHOME NON HODGKINIEN

Le lymphome non hodgkinien (LNH) représente un groupe hétérogène de maladies causées par la prolifération de lymphocytes s'accumulant le plus souvent dans les ganglions lymphatiques. Ils représentent 90% des lymphomes et sont subdivisés en plusieurs groupes, les lymphomes malins agressifs et les lymphomes non hodgkiniens indolents. Ces deux catégories de lymphomes regroupent différents sous-types, très différents les uns des autres d'un point de vue cytogénétique, morphologique, immunologique et diagnostic, rendant leur classification très complexe. C'est donc une pathologie complexe et variée à latence longue, chronique et potentiellement soumise à la multi-exposition.

Les taux d'incidence et de mortalité, ajustés sur l'âge, dans la plupart des pays développés (Europe, Amérique du Nord, Australie/Nouvelle Zélande et Japon) sont de 8.6/100 000 et 2.7/100 000 [IARC, 2012]. En France le LNH est rare et ne représente que 3,3% de tous les cancers incidents (non ajusté par âges), mais représente le 8^{e} cancer incident avec un taux ajusté selon l'âge de 12 et 7,6/100 000 chez les hommes et les femmes, respectivement, et le 13^{e} en termes de mortalité, avec un taux normalisé selon l'âge de 3,5 et 1,9/100 000 chez les hommes et les femmes, respectivement [IARC, 2012].

La particularité du LNH et la raison de l'engouement scientifique pour cette pathologie provient de l'augmentation inexpliquée de son incidence d'environ 4% par an dès les années 70 jusqu'à la fin du 20^{e} siècle [Alexander, 2007, De Roos et Bhatti, 2014]. Cette incidence a commencé à décroitre ces dernières années, mais la cause de cette augmentation et les facteurs de risques associés restent largement méconnus. De plus, la magnitude de l'augmentation n'a pu être expliquée par les facteurs de risques connus tel que le statut immunitaire ou les facteurs comportementaux (tabagisme et consommation d'alcool) [De Roos et Bhatti, 2014]. De nos jours, il existe de nombreuses hypothèses à propos d'une relation causale entre le LNH et un contexte professionnel, tel que l'agriculture ou l'industrie [Alavanja, 2014, De Roos et Bhatti, 2014] et les expositions associées. Les hypothèses d'expositions professionnelles les plus discutées à l'heure actuelle concernent les solvants organiques, dont le benzène [Alexander et Wagner, 2010, Cocco, 2010, Vlaanderen, 2011], le trichloroéthylène [Karami, 2013, Purdue, 2011], le tétrachloroéthylène et le toluène ; le diesel et l'essence [Kane et Newton, 2010] ; le formaldéhyde ; les produits agricoles (pesticides, fongicides et herbicides) [Schinasi et Leon, 2014] ; les virus et enfin les radiations ionisantes [Karipidis, 2007].

Le LNH étant un cancer varié, en terme clinique, de latence (distribution de latence, moyenne à longue) et d'étiologie, potentiellement sujet à la multi-exposition, l'apport de la Surveillance Observationnelle avait pour objectif de fournir un spectre des motifs d'expositions professionnelles potentiellement associés en s'appuyant sur des données de qualité obtenues lors de diagnostics et de consultations professionnelles effectués par des médecins experts du réseau RNV3P. Les étiologies étant encore peu certaines, les résultats obtenus seront plus développés que lors des exemples précédents, puis comparés avec la littérature.

Les codes CIM-10 utilisés pour cette analyse sont : C82*, C83* et C85*.

9.1. Analyse

En 2014, au sein des 107 713 PST de la base RNV3P et des 14 055 cancers professionnels, 288 PST concernaient le LNH (Figure 47), dont 251 hommes et 37 femmes, avec une moyenne d'âge de 54.5 et 51.92 ans respectivement. Un nombre total de 178 nuisances distinctes ont été enregistrées depuis 2001, majoritairement des produits chimiques organiques (hydrocarbures aromatiques ou aliphatiques) (n=32) et des produits, matériaux ou des processus industriels (notamment des produits pesticides) (n=69) dans 149 activités professionnelles distinctes, notamment dans les secteurs de l'agriculture et de l'industrie chimique.

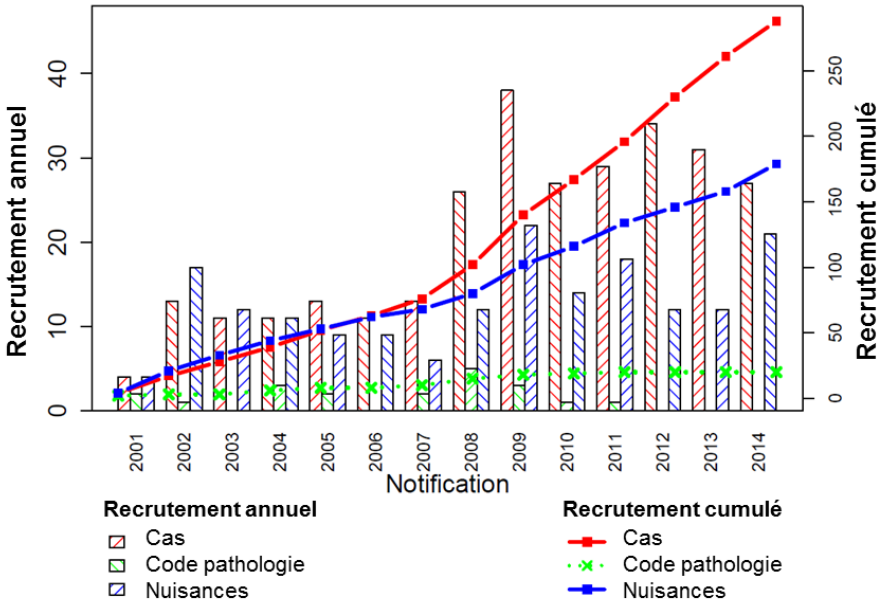


Figure 47 : Recrutement et effectif cumulé des cas, des nuisances distinctes et des codes pathologies distincts pour le LNH de 2001 à 2014.

Parmi les 178 nuisances distinctes, seulement 15 furent considérées comme significatives par le calcul basé sur l'imputabilité, et 200 PST possédaient au moins une de ces nuisances dans leur cortège. A partir de ces 200 PST, 110 cortèges distincts d'expositions professionnelles ont été identifiés, formant les nœuds de l'exposome pour un total de 1 798 liens entre ces nœuds (Figure 48). Comme dans l'exemple du cancer de la vessie, la couleur des liens de l'exposome représente une famille d'expositions au 3ème niveau d'imbrication du thésaurus utilisé, par exemple, la couleur rouge correspond au code « A16 : Matériel et produits divers ».

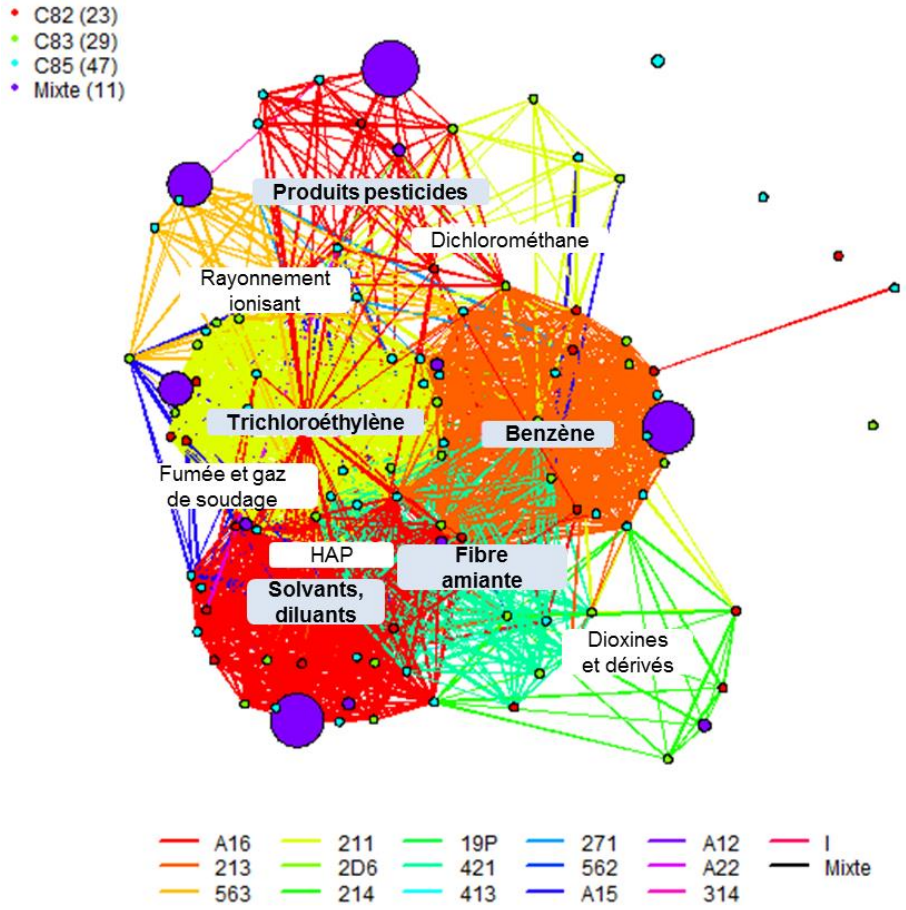


Figure 48 : Exposome du LNH en 2014. La taille des nœuds correspond au poids des nœuds (nombre de PST identiques), la couleur correspond aux différents codes pathologies. La couleur des liens correspond aux différentes familles (3e niveau du niveau imbriqué) de nuisances identifiées en relation avec le LNH et permettent d'identifier les clusters de l'exposome. Les 5 clusters majoritaires identifiés sont : « Produits pesticides » (A16P0000), « Solvants, diluants » (A1660000), « Benzène » (21310100), « Trichloroéthylène » (21170204) et « Fibre d'amiante » (42110000).

La majorité des liens correspondait à des nuisances de type « Matériel et produit divers », notamment l'usage « Solvants, diluants » (A1660000) qui représentait un des plus gros « clusters » de l'exposome. Des nuisances de type « Hydrocarbures et dérivés » étaient également fortement représentés, avec les deux clusters « trichloroéthylène » et « benzène »

(21170204 et 21310100), et un autre cluster correspondant à « l'amiante » (42110000) fut également identifié. Quatre nœuds isolés, c'est-à-dire ne partageant aucune nuisance en commun avec les autres nœuds, ont été mis en évidence : les « traitements phytosanitaires » en tant qu'usage (BM110000), le « mouvement répétitif du membre supérieur » (62200000), les « produits de vulcanisation du caoutchouc » (A1670220) et l'association des usages « d'herbicide phytohormone/d'herbicide triazine » (A16P0550/A16P0560).

Quarante motifs d'expositions professionnelles ont été identifiés, c'est-à-dire 40 nuisances ou associations de nuisances ayant été rapportées comme potentiellement associées au LNH. Un **numéro d'identification** a été attribué à chacun en fonction de leur poids permettant de les identifier dans tous les résultats issus de la Surveillance Observationnelle (Tableau 7). Parmi ces 40 motifs, 21 correspondaient à des motifs d'ordre I (une seule nuisance au sein du cortège), impliquant un total de 19 motifs d'ordre supérieurs à I, correspondant à des combinaisons de motifs d'ordre I : 15 motifs d'ordre II, 4 d'ordre III et 1 motif d'ordre IV.

Tableau 7 : Liste des 40 motifs du LNH identifiés par la Surveillance Observationnelle en 2014.

ID	Ordre I	Ordre II	Ordre III	Nœuds	Poids	Statut*
1	**solvants, diluants**	**-**	**-**	**34**	**58**	**A**
2	**benzène**	**-**	**-**	**33**	**54**	**A**
3	**trichloroéthylène**	**-**	**-**	**30**	**38**	**A**
4	**produit pesticide**	**-**	**-**	**12**	**38**	**A**
5	**radiation ionisante**	**-**	**-**	**10**	**22**	**A**
6	**amiante (fibre)**	**-**	**-**	**18**	**18**	**B**
7	**trichloroéthylène**	**benzène**	**-**	**10**	**11**	**A**
8	**fumée et gaz de soudage**	**-**	**-**	**9**	**9**	**A**
9	**dichlorométhane**	**-**	**-**	**8**	**8**	**B**
10	**dioxine et dérivés**	**-**	**-**	**7**	**8**	**B**
11	**trichloroéthylène**	**solvants, diluants**	**-**	**7**	**8**	**B**
12	**benzène**	**solvants, diluants**	**-**	**7**	**8**	**B**
13	**Hydrocarbures aromatiques polycycliques (HAP)**	**-**	**-**	**6**	**6**	**A**
14	**trichloroéthylène**	**amiante (fibre)**	**-**	**6**	**6**	**B**
15	peinture, vernis, laque, mastic	-	-	5	5	D
16	suie et autre produit dégagé lors de l'utilisation de produit noir	-	-	5	5	D
17	**insecticide**	**-**	**-**	**4**	**5**	**D**
18	amiante (fibre)	solvants, diluants	-	5	5	D
19	fumée et fumée de soudage	solvants, diluants	-	5	5	D
20	**produit pesticide**	**insecticide**	**-**	**4**	**5**	**D**
21	**1,3-butadiene**	**-**	**-**	**4**	**4**	**D**
22	tétrachloroéthylène	-	-	4	4	D
23	formaldéhyde	-	-	4	4	D
24	dérivé organophosphoré	-	-	3	4	D

25	autre produit de soudage	-	-	4	4	D
26	herbicide	-	-	3	4	D
27	trichloroéthylène	tétrachloroéthylène	-	4	4	D
28	produit pesticide	herbicide	-	2	3	D
29	insecticide	herbicide	-	2	3	D
30	produit pesticide	insecticide	herbicide	2	3	D
31	matière plastique, caoutchouc	-	-	1	2	D
32	**traitement phytosanitaire**	**-**	**-**	**1**	**2**	**C**
33	solvants, diluants	matière plastique, caoutchouc	-	1	2	D
34	dérivé organophosphoré	produit pesticide	-	1	2	D
35	dérivé organophosphoré	insecticide	-	1	2	D
36	dérivé organophosphoré	herbicide	-	1	2	D
37	dérivé organophosphoré	produit pesticide	insecticide	1	2	D
38	dérivé organophosphoré	produit pesticide	herbicide	1	2	D
39	dérivé organophosphoré	insecticide	herbicide	1	2	D
40	dérivé organophosphoré	produit pesticide	insecticide	1	2	D

*__En gras__ sont identifiés les motifs dont le statut est considéré comme actif (A : persistant, B : en déclin, C : émergent) ainsi que trois motifs latents (D) actifs en 2013.

Le statut dynamique des motifs a été caractérisés pour l'année 2014 (Figure 49A), 8 motifs ont été identifiés comme **persistants** (id : 1-5, 7-8, 13), 6 en **déclin** (id : 6, 9-12, 14), 1 en **émergence** (id : 32) et 25 en état de **latence**, dont 3 ont été considérés comme actifs en 2013 (id : 17, 20-21). **Dix-huit motifs sont apparus comme actifs** et ont été représentés sur l'arbre de la multi-association (Figure 49B), dont 5 d'ordre supérieur à I. Les motifs les plus important associés au LNH sont les « Solvants, diluants » (id: 1), recrutés dans 58 PST, et en association avec 5 autres motifs d'ordre I (id: 11-12, 18-19, 33) ; suivi par le « benzène » (id : 2) dans 54 PST, en association avec deux autres motifs (id : 7, 12), tous les deux actifs. Les autres motifs important identifiés pour le LNH sont le « trichloroéthylène » (id : 3) et les « produits pesticides » (id : 4), enregistrés chacun dans 38 PST, et formant une association avec 4 et 7 autres motifs respectivement (id : 7, 11, 14, 27 et id : 20, 28, 30, 34, 37-38, 40). Le seul motif émergent identifié correspondait aux « traitements phytosanitaires » (id : 32).

Le spectrosome des motifs d'expositions professionnelles potentiellement associés au LNH est disponible en Figure 50. Contrairement à l'exposome des nœuds de la Figure 48, seuls les 40 motifs identifiés par l'analyse de la SO ont été représentés, permettant la mise en évidence des grandes familles de nuisances associées au LNH ainsi que la visualisation de la structure de ces différentes associations. Les motifs ainsi mis en évidence par le spectrosome n'étaient pas les mêmes que ceux mis en évidence par un classement de poids. En effet, les motifs tels que les « insecticides » (id : 17), « herbicides » (id : 26) et « dérivés organophosphorés » (id : 24) ont été mis en évidence sur le spectrosome en tant que cluster car permettant la formation

de plusieurs autres associations alors que leur poids était inférieur ou égal à 5 seulement. Ceci met en évidence les deux caractéristiques importantes d'un motif, son poids, mais également son implication dans les différentes associations et la structure de celles-ci.

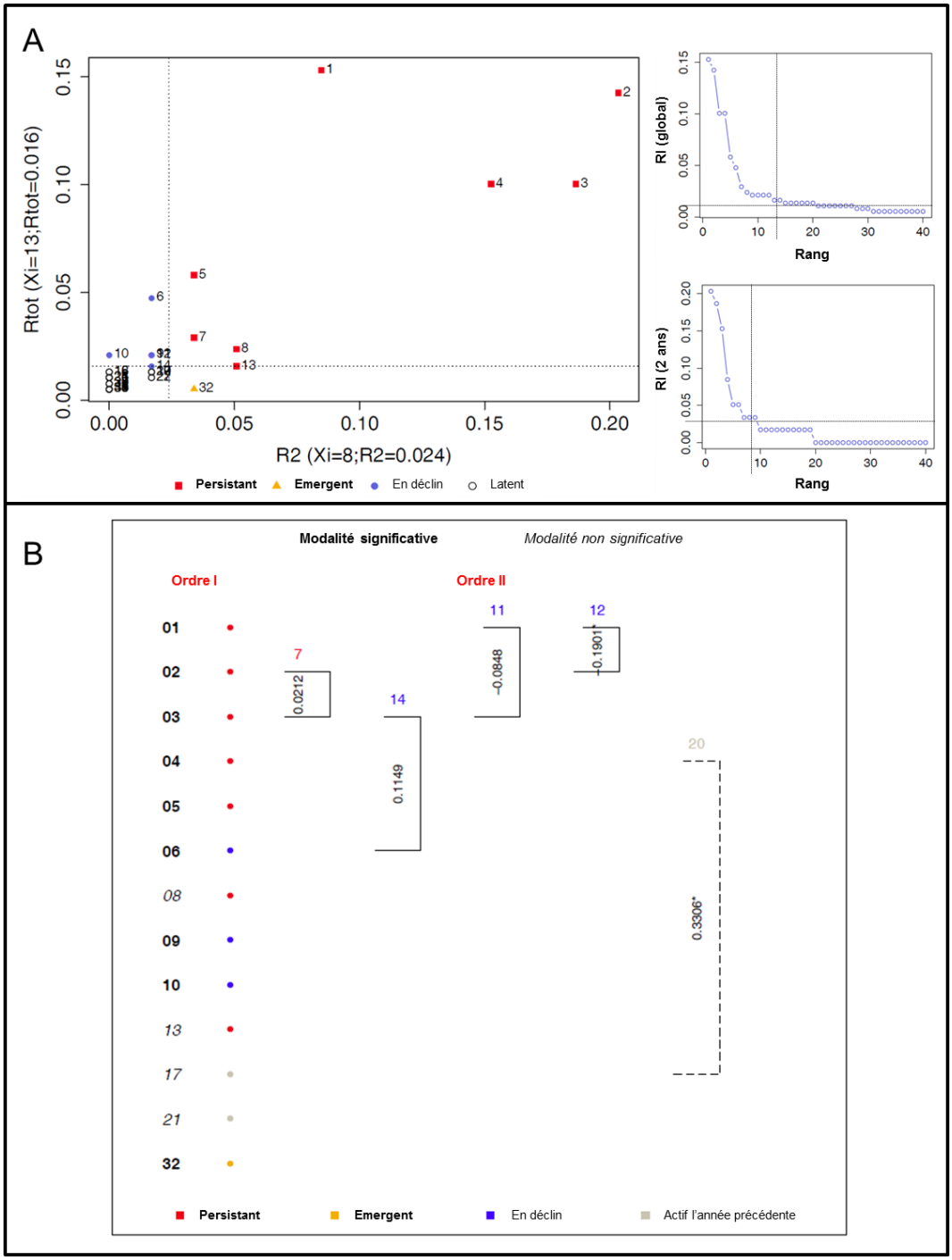


Figure 49 : Caractérisation des motifs issus de l'exposome du LNH en 2014. A) Statut dynamique des 40 motifs en 2014. Les délimitations des RI ont été calculées à 0.024 pour les deux années glissantes et 0.016 pour le RI global. Il y avait 8 motifs persistants, 6 en déclin, 1 en émergence et 25 en latence. B) Arbre de la multi-association des 18 motifs actifs liés au LNH en 2014, avec 13 motifs d'ordre I actifs et 5 motifs d'ordre II actifs. ***Les numéros indiqués correspondent au numéro d'identification du Tableau 7.***

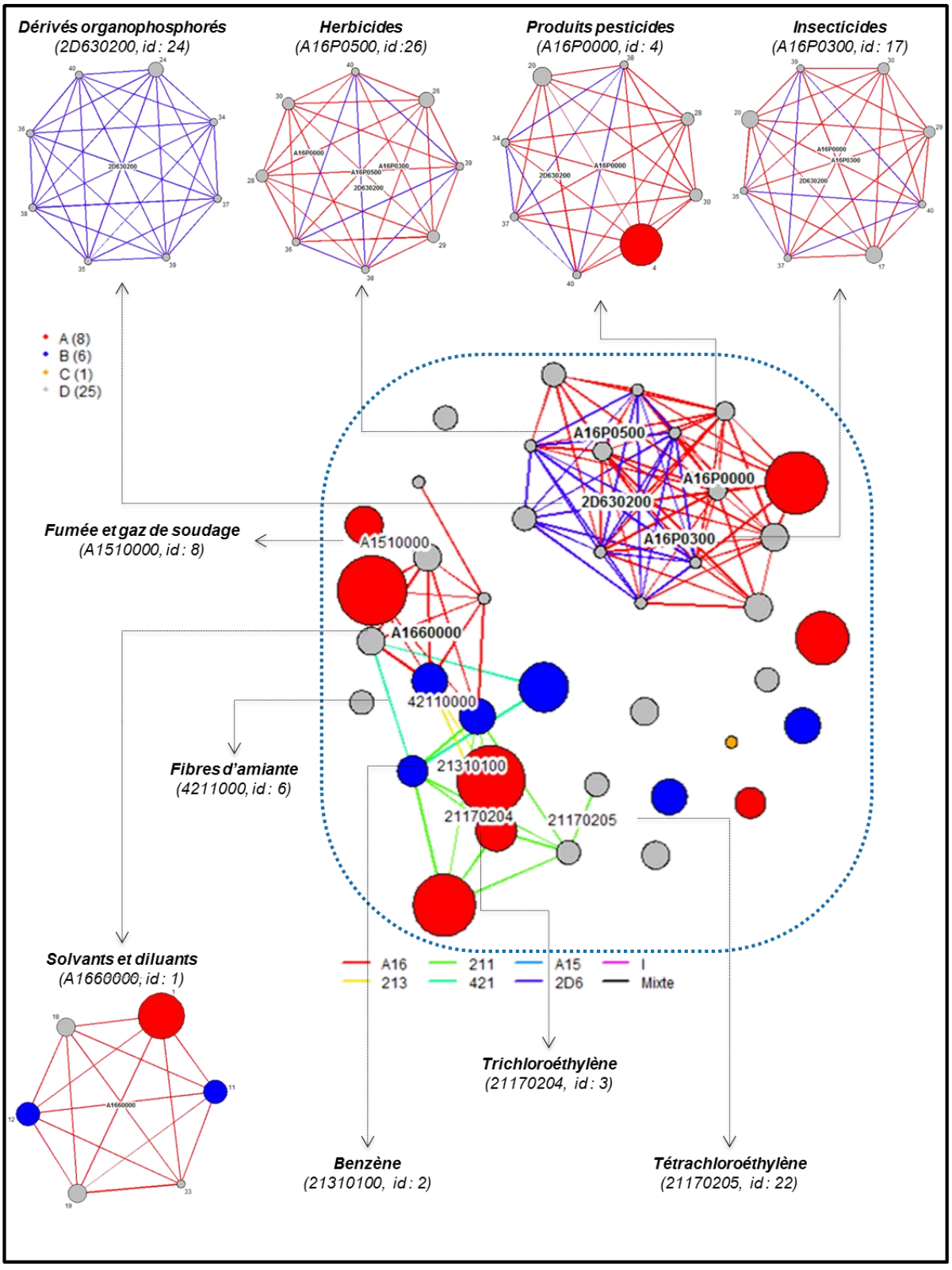


Figure 50 : Spectrosome du LNH en 2014. Au centre du spectrosome des motifs d'expositions professionnelles associés au LNH. La taille des nœuds correspond au poids du motif et la couleur correspond au statut dynamique du motif. La couleur des liens correspond aux différentes familles de nuisances identifiées en relation avec le LNH. Cinq clusters ont été identifiés en raison des nombreuses associations auxquelles ils participent, de gauche à droite et de haut en bas : dérivés organophosphorés (24), herbicides (26), produits pesticides (4), insecticides (17) et solvants, diluants (1).

Le résultat final de la Surveillance Observationnelle pour le LNH correspond au spectre des motifs d'expositions professionnelles en Figure 51. Seuls les motifs actifs ainsi que les 3 motifs latents en 2014 mais actifs en 2013 identifiés **en gras** dans le Tableau 7 ont été présentés sur cette signature de l'exposition du LNH afin d'en faciliter la lisibilité.

Huit motifs ont été caractérisés comme spécifiques (id : 5, 7, 10-12, 17, 20, 32), c'est-à-dire ayant un recrutement plutôt localisé, tandis que 6 ont été caractérisés comme non spécifiques (id : 6, 8-9, 13-14, 21), c'est-à-dire dont l'information a été répartie de manière homogène sur plusieurs nœuds différents. Par exemple, le motif de « l'amiante » (id : 6) a été vu dans 18 nœuds différents avec un poids égal à 1 à chaque fois, c'est donc un motif ubiquitaire et non spécifique d'une association. Concernant les motifs intermédiaires, c'est-à-dire dont la valeur de spécificité est comprise entre 0 et 1, ceux dont la valeur est supérieure à 0.5 tendent vers une localisation de l'information, tandis que les autres vers une ubiquité de leur information. C'était le cas par exemple des 4 premiers motifs associées au LNH : « solvants, diluants », « benzène », « trichloroéthylène » et « produit pesticide » qui tendent tous vers une localisation de l'information avec une forte proportion d'enregistrements en tant que nœud unique, c'est-à-dire une exposition uniquement lors de l'enregistrement du PST.

La matrice primaire secondaire des motifs et des activités professionnelles est disponible à titre d'exemple en Figure 52 et les identifiants des activités professionnelles sont disponibles en Annexe 3. La présence d'un motif dans une activité professionnelle est indiquée par la présence d'un marquage de couleur noire, verte, bleue et rouge indiquant respectivement les apparitions des motifs dans des activités particulières de 2001 à 2011, en 2012, 2013 et 2014. Cette couleur est également utilisée pour indiquer l'année d'apparition des motifs. Par exemple l'activité professionnelle « Professions intermédiaires des sciences physiques et techniques » dans la « Cokéfaction, raffinage, industries nucléaires » (id : 111) est apparue en 2009, associée aux motifs « benzène » (id : 2), « trichloroéthylène » (id : 3), et l'association « trichloroéthylène-benzène » (id : 7) ; tandis que les « traitements phytosanitaires » (id : 32) sont apparus en 2013 chez les « agriculteurs spécialisés dans la culture de céréales » (id : 14).

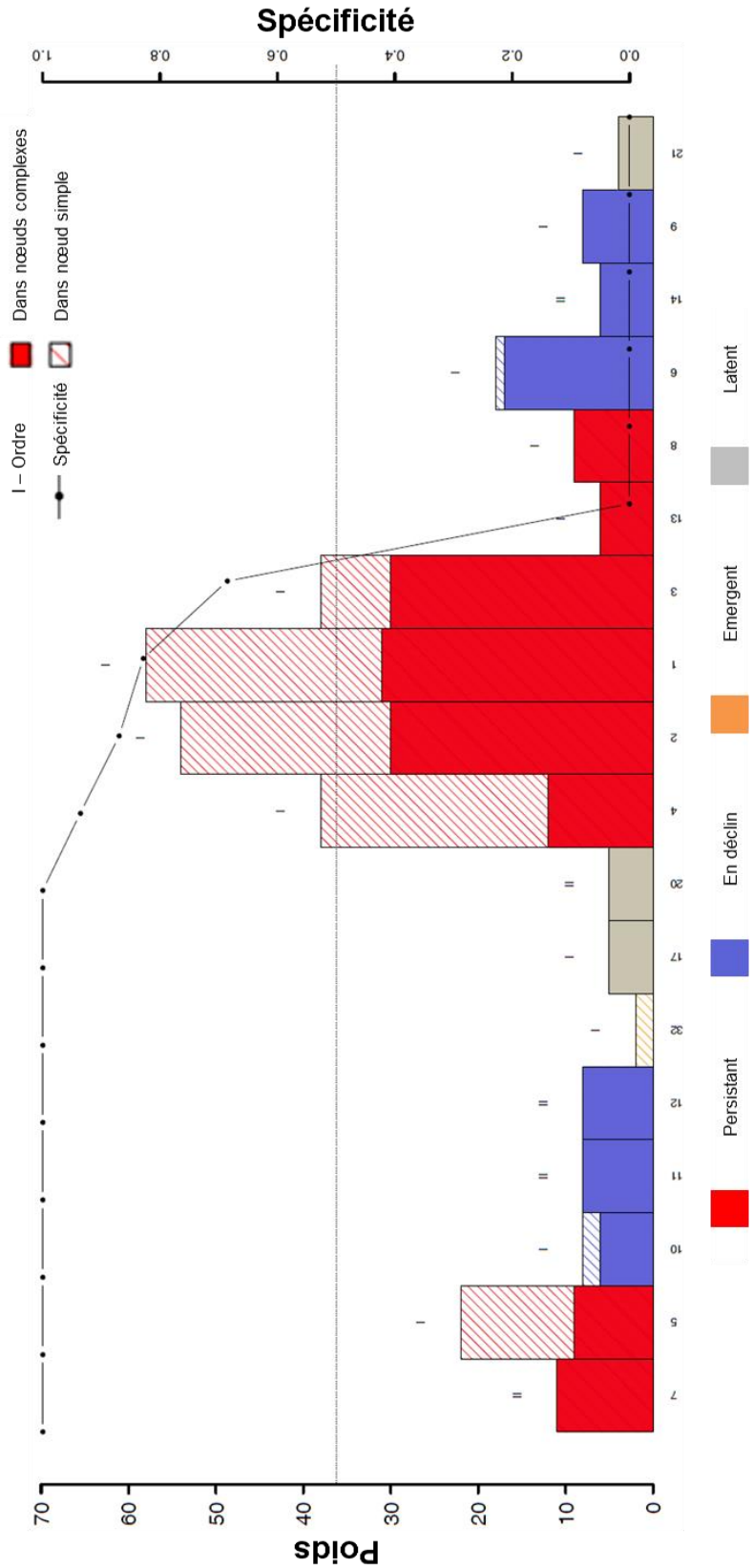


Figure 51 : Spectre du LNH en 2014. Le spectre du LNH a été caractérisé par 18 motifs actifs (dont 8 persistants, 6 en déclin, 1 en émergence et 3 latents mais actifs en 2013), parmi lesquels 13 sont des motifs d'ordre I et 5 des motifs d'ordre II. Huit motifs avaient une localisation de leur information spécifique c'est-à-dire privilégiant une association en particulier, 6 étaient plutôt ubiquitaires avec une information apportée par différents nœuds de manière homogène et 4 étaient des motifs intermédiaires tendant vers une spécificité de par leur forte implication dans un nœud simple.

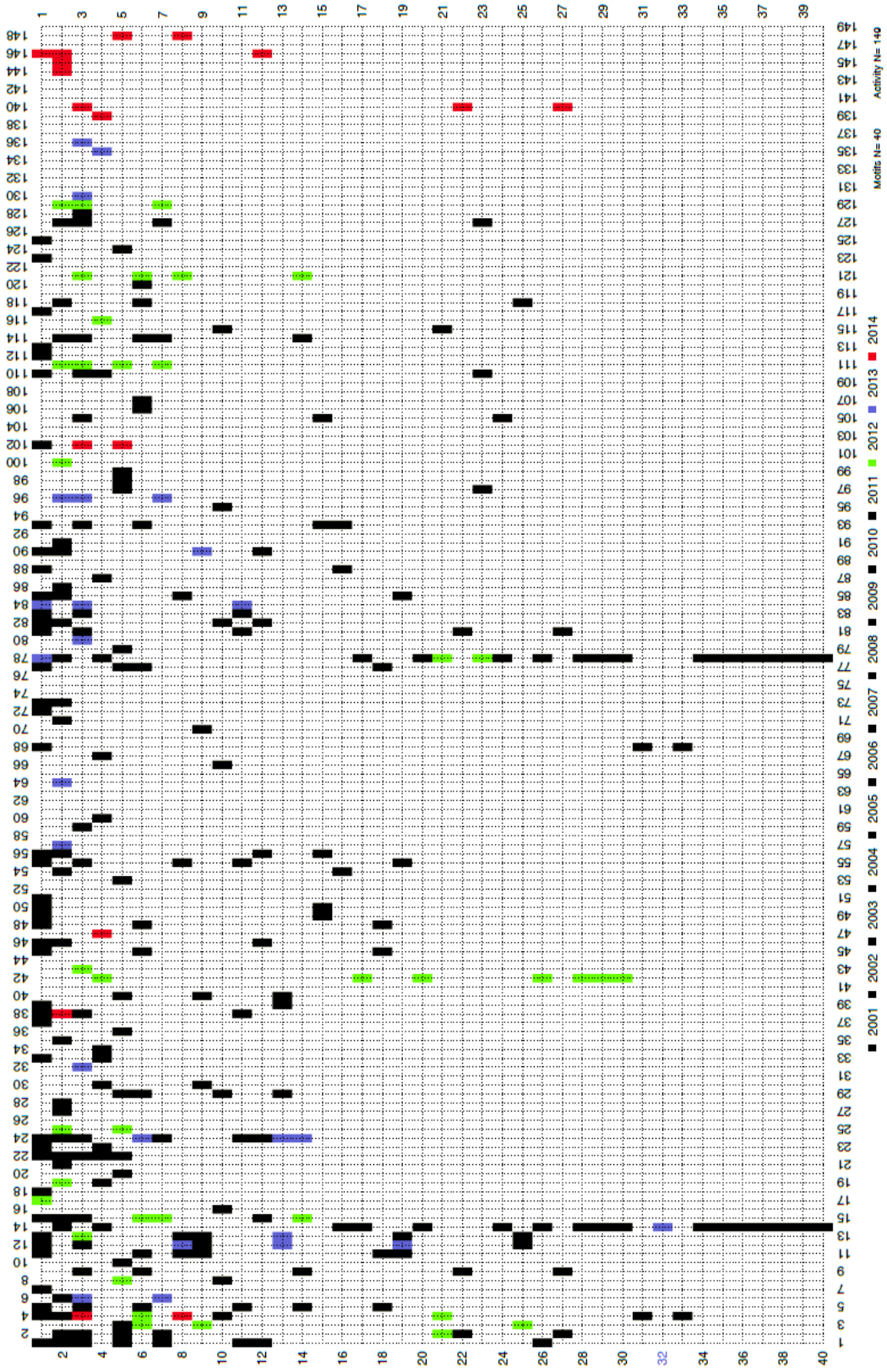


Figure 52 : Matrice primaire-secondaire des motifs et des activités professionnelles associées au LNH en 2014. Chaque motif est représenté sur les lignes horizontales et chaque activité professionnelle sur les lignes verticales. Si un motif est enregistré dans une activité professionnelle particulière au moins une fois, l'intersection est alors marquée d'un rectangle de couleur. La couleur noire correspond aux recrutements ayant eu lieu de 2001 à 2011, le vert au recrutement en 2012, le bleu en 2013 et le rouge en 2014. Les numéros indiqués sur la figure correspondent aux numéros d'identification des motifs et des activités, disponibles respectivement dans le Tableau 7 et Annexe 3.

En Figure 53 sont présentés les exemples atypiques du motif « Trichloroéthylène/benzène » identifié comme motif persistant en 2014 du à son grand nombre de recrutement depuis 2001 et les derniers recrutements en 2013 et 2014 ; le motif de « l'amiante » en tant que motif en déclin car très rapporté également depuis 2001, mais beaucoup moins en 2013/2014 ; le seul motif émergent les « traitements phytosanitaires » dont le premier enregistrement a été effectué en 2013 ; et enfin l'exemple des « peintures, vernis, laque, mastic », motif latent en 2014, avec pourtant certains pics d'enregistrement en 2001/2002 et 2010/2011.

En Figure 54 est présenté un autre exemple, celui des « Solvants, diluants » et de l'« Amiante », permettant l'illustration et l'intérêt de suivre des expositions professionnelles particulières, mais également de suivre l'évolution de leur association au cours du temps. Ces deux motifs d'ordre I « Solvants, diluants » (persistant) et « Amiante » (en déclin) ont permis la formation de leur association pour la première fois en 2004, grâce à l'apparition du motif d'ordre I de « l'amiante ». Cependant, le motif de l'amiante a vu son recrutement baisser en 2006 et 2007, puis remonter de manière stable jusqu'en 2013 pour finalement retomber en déclin en 2014, impactant ainsi la trajectoire du motif d'ordre II qui est resté de manière générale beaucoup plus longtemps en période de latence. Il est également visible que l'association particulière des solvants et de l'amiante n'a été retrouvée que dans 18 activités professionnelles particulières alors que les motifs d'ordre I ont été identifiés dans respectivement 46 et 18 activités professionnelles distinctes, mettant ainsi en évidence un contexte particulier impliquant les solvants et l'amiante dans le secteur de la « défense » (employé des stocks), la « construction de moteurs d'aéronef » (soudeurs, oxycoupeurs), la « fabrication de produits minéraux non métalliques » (conducteurs de machine à fabriquer du ciment), le « transport maritime » (techniciens de matériel d'émission radio, télévision, téléphone) et la « manutention non portuaire » (manutentionnaire).

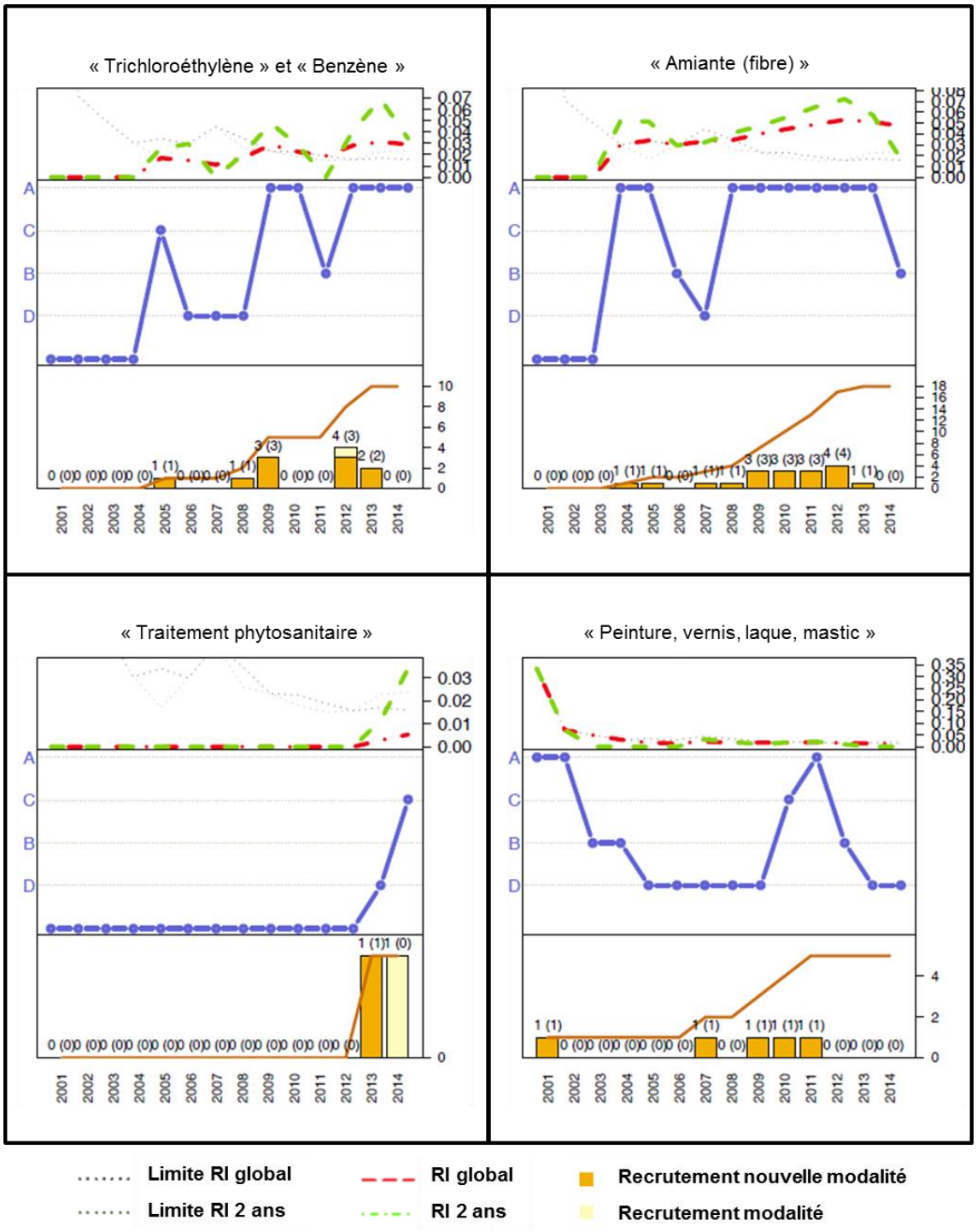


Figure 53 : Exemple de trajectoire atypique de motifs identifiés pour le LNH de 2001 à 2014 pour chaque catégorie de statut dynamique : persistant, en déclin, émergent et latent. Exemple du motif « Trichloroéthylène/Benzène » (persistant), « l'amiante » (en déclin), les « traitements phytosanitaire » (seul motif en émergence en 2014) et le motif « Peinture, vernis, laque, mastic » (latent). La courbe bleue correspond à l'évolution du statut dynamique des motifs : A – persistant, B – en déclin, C – émergent et D – latent. Les courbes rouge et verte correspondent respectivement aux valeurs de $\boldsymbol{RI}_{\infty}$ et $\boldsymbol{RI}_{\boldsymbol{2}}$ pour le motif à l'année donnée, les courbes pointillées correspondent aux limites $\boldsymbol{\xi}_{\boldsymbol{RI}}$ calculées en fonction des autres recrutements pour l'année donnée. La partie basse correspond à l'évolution du recrutement des motifs dans des activités professionnelles particulières, distinguant les nouvelles activités professionnelles et les activités déjà rencontrées.

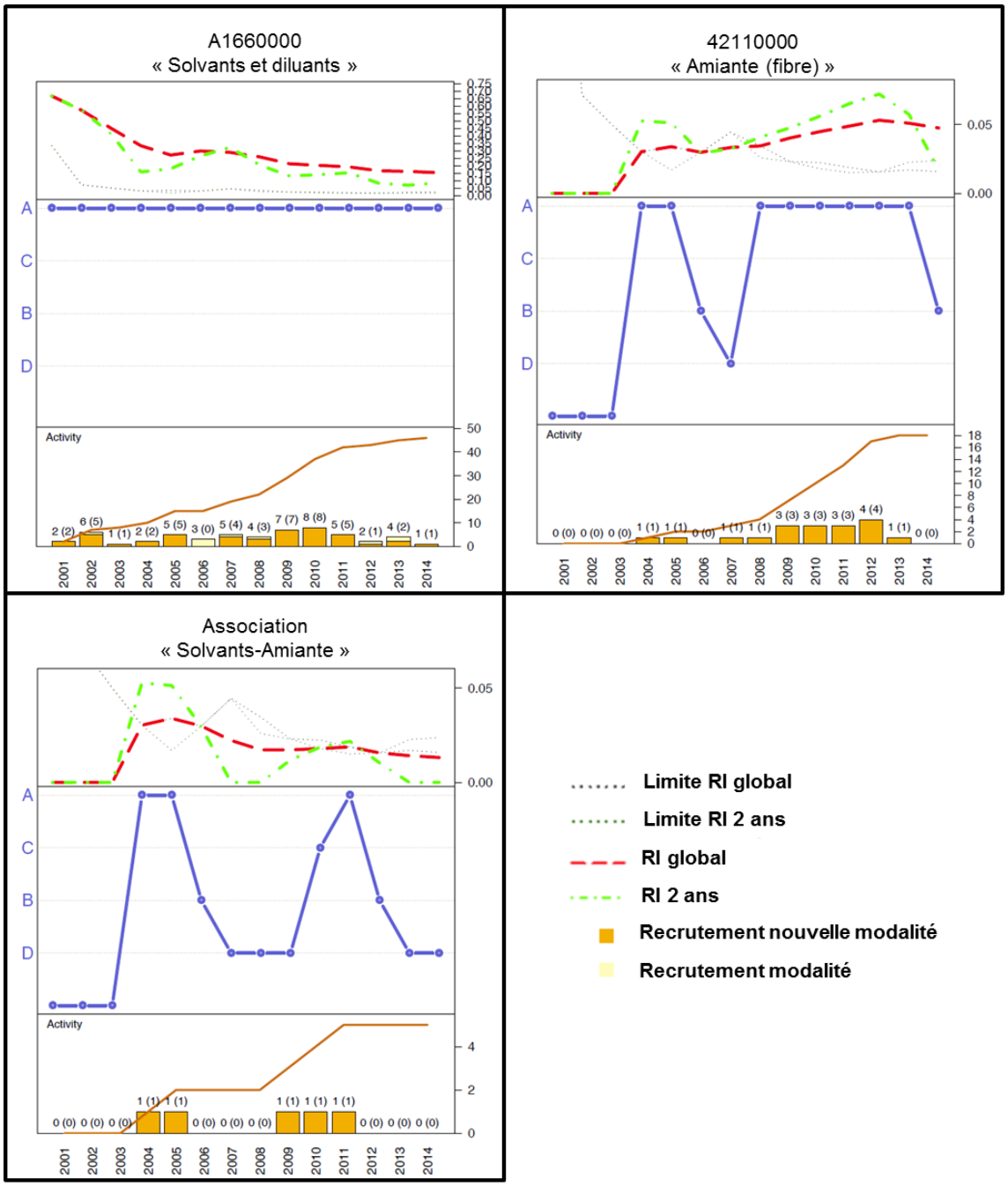


Figure 54 : Exemple de trajectoire dynamique de deux motifs d'ordre I, solvants, diluants et amiante, et leur association d'ordre II, associés au LNH, de 2001 à 2014. La courbe bleue correspond à l'évolution du statut dynamique des motifs : A – persistant, B – en déclin, C – émergent et D – latent. Les courbes rouge et verte correspondent respectivement aux valeurs de $\boldsymbol{RI}_{\infty}$ et $\boldsymbol{RI}_{\boldsymbol{2}}$ pour le motif à l'année donnée, les courbes pointillées correspondent aux limites $\boldsymbol{\xi}_{\boldsymbol{RI}}$ calculées en fonction des autres recrutements pour l'année donnée. La partie basse correspond à l'évolution du recrutement des motifs dans des activités professionnelles particulières, distinguant les nouvelles activités professionnelles et les activités déjà rencontrées.

9.2. Conclusion

Le spectre du lymphome non hodgkinien, cancer rare du système lymphatique, a été caractérisé par **40 motifs d'expositions professionnelles** particuliers dont **18 ont été considérés comme actifs en 2014**. Une multi-exposition importante a été mise en évidence avec un total de 19 motifs d'ordre supérieur à I, dont un d'ordre IV, **soit 47.5% des motifs identifiés concernant la multi-exposition**. Les motifs les plus importants rapportés comme associés au LNH étaient les **solvants, diluants, dont le benzène et le trichloroéthylène, l'amiante et les produits pesticides** (dont herbicides et insecticides à un moindre niveau).

9.3. Comparaison des résultats

Les étiologies du LNH étant toujours très discutées dans la littérature, il était intéressant de pouvoir comparer les hypothèses émises dans la littérature avec les résultats obtenus avec la Surveillance Observationnelle. Un travail listant et classant tous les contextes professionnels (considéré comme une activité professionnelle, ou une exposition professionnelle, ou l'association des deux) associés au risque de LNH mentionnés dans la littérature a été développé et soumis à la revue *American journal of industrial medicine* (en cours de review par les pairs).

9.3.1. Article soumis dans la revue *American Journal of Industrial Medicine*

American Journal of Industrial Medicine

Ranking occupational contexts associated with risk of non-Hodgkin Lymphoma

Journal:	*American Journal of Industrial Medicine*
Manuscript ID	Draft
Wiley - Manuscript type:	Research Article
Keywords:	non-Hodgkin lymphoma, occupational activity, occupational exposure, association risk

SCHOLARONE™
Manuscripts

For Review

John Wiley & Sons

Article soumis dans American Journal of Industrial Medicine

Ranking occupational contexts associated with risk of non-Hodgkin Lymphoma

Delphine Rieutort (phD student) [a,*], Oriane Moyne (MSc) [a], Pierluigi Cocco (Professor) [b], Régis de Gaudemaris (Professor) [a,c], Dominique J. Bicout (phD) [a,d,e,*]

[a] UJF-Grenoble 1 / CNRS / TIMC-IMAG UMR 5525 (EPSP team - Environment and Health Prediction of Populations), Grenoble, F-38041, France

[b] Department of Public Health, Clinical and Molecular Medicine, Occupational Health Section, University of Cagliari, Italy

[c] Occupational and Environmental Diseases Centre, Grenoble Teaching Hospital (CHU Grenoble), Grenoble, France

[d] Biomathematics and Epidemiology EPSP-TIMC, VetAgro Sup, Veterinary Campus of Lyon, F – 69280 Marcy l'Etoile, France

[e] Laue – Langevin Institute, Theory Group, F – 38042 Grenoble cedex 9, France

* Corresponding authors: Faculté de Médecine, Domaine de la Merci, Bat Jean Roget. 38706 La Tronche Cedex, France. Tel: +33 4 76 63 75 06.

E-mail address: delphine.rieutort@imag.fr; bicout@ill.fr (D. Rieutort, DJ Bicout)

ACKNOWLEDGMENTS

We thank our documentarist Sylvette Liaudy for her help with the literature search. DR is a PhD student supported by a grant from the French Ministry of Education and Research through the Ecole Doctorale Ingénierie pour la Santé, la Cognition et l'Environnement (EDISCE) of Grenoble University. We thank Dr Alison Foote (Grenoble Clinical Research Centre) for helpful suggestions in editing the manuscript.

AUTHORS' CONTRIBUTION

Contributed concepts: DJB and DR. Performed the analyses and calculations: DR and OM. Wrote the paper: DJB, DR, PC and RG

Article soumis dans American Journal of Industrial Medicine

SUMMARY

From the literature, we obtain three inventories of occupational contexts associated with risk of NHL (activities, exposures and both), leading to highlight "hot" occupational activities and exposures related to NHL, in both terms of scientific interest and associated risk.

Words: 5669 Char: 33112 (all except tab/fig) ; Fig: 2 Tab: 3

ABSTRACT

Objectives: To inventory from the accumulated knowledge the associations between occupational contexts and the risk of Non-Hodgkin Lymphoma (NHL).

Methods: The occupational contexts associated with the NHL risk are inventoried from the literature review. For each context, items were ranked both in terms of scientific interest and of the strength of association with NHL risk.

Results: Three ranked lists of associations between occupational contexts and NHL risk were constructed. We found that NHL was associated with 31 occupational activities, 91 occupational exposures and 35 occupational activity-exposure combinations. Among them, 5 occupational activities 2 occupational exposures and 3 occupational activity-exposure combinations, mainly involving agricultural or industrial sector and solvents and pesticides products, were highlighted with the highest number of publications and the strongest association with NHL risk.

Conclusions: Within the context of the health risk assessment at the workplace, the results reported are very instructive and useful in both providing a ranked inventory of occupational contexts associated with NHL risk and highlighting "hot" occupational activities and exposures.

KEYWORDS

occupational activity; occupational exposure; association risk

INTRODUCTION

Non-Hodgkin's Lymphomas (NHLs) are a heterogeneous group of diseases caused by a proliferation of malignant lymphocytes that accumulate in lymph nodes. The complexity in characterizing NHL relies both on different etiologies between subtypes (1) and on the various classifications used to distinguish the different subtypes that now include for example chronic lymphocytic leukemia and multiple myeloma (2).

The incidence rates of NHL reported in USA, Australia and Western Europe are about 10/100,000 (2). In France, for instance, NHLs are rare representing only 3.2% of all incident cancers (3), but they are the 6th incident cancer with a standardized incident ratio (SIR) of 12.1 and 8.2 in men and women, respectively, and the 9th in terms of mortality with a standardized mortality ratio (SMR) of 4.0 and 2.3 in men and women, respectively (3). In most developed countries, the NHL incidence was found increasing at a rate of about 4% per year from the 1970s until the end of the 20th century and started to level off in the last few years. Causes of the increase of incidence and associated risk factors remain largely unknown, despite numerous epidemiological studies. For instance, there are not yet conclusive evidences about the causal relationships between agricultural or industrial occupations and NHL risk. In this context, it would be interesting to review the accumulated knowledge about situations of exposure found to be associated with NHL.

To date, most of epidemiological investigations have considered the associations between the NHL and numerous specific occupations and/or specific exposures. Within the occupational framework, the objectives of this paper are twofold, *i)* to use the literature to identify the occupational contexts potentially associated with NHL risk and, in addition for further insights, *ii)* to rank these occupational contexts both in terms of the degree of scientific interest, as measured by the relative number of published studies, and of the strength of association with NHL, based on the distribution of association risks of occupational contexts with NHL. Occupational contexts meant either the occupational activities or the occupational exposures or combinations of both occupational activity and exposure. The main results of the paper are three ranked lists of occupational contexts associated with NHL. Such results may be useful for the assessment of occupational exposure related to health risks.

MATERIAL & METHODS

Literature search and identification of occupational context

Literature search was conducted from June 11th to July 16th 2013 in four electronic databases: PubMed, the Catalog and Index of French language Medical Sites (Catalogue et Index des Sites Médicaux de langue Française, CISMeF), the French Database in Public Health (Base de Données en Santé Publique, BDSP) and the French University System of DOcumentation catalog (catalogue du Système Universitaire de DOcumentation, SUDOC). We used the phrase "lymphome malin non hodgkinien" when searching in French language databases and for PubMed MeSH (Medical Subject Headings) terms combining NHL with occupational exposure or activity terms, such as (work* OR occupation*) AND (Disease) where "Disease" was "non-Hodgkin, Lymphoma" and multiple declinations as provided in the appendix (A1).

All articles published from January 1st 1990 to the end of the search period (July 16th 2013) were considered. The articles were selected in three steps using criteria based on the language, title and abstract (see appendix A2). Briefly, studies had to be written either in English or French, dealing with an epidemiological, etiological or pathological content, i.e., articles had to concern cancer, especially non-Hodgkin lymphoma or hematological diseases family, in an epidemiological or etiological way. Articles describing analyses of heredity, childhood or environment (other than occupational), genetic, molecular or cellular biology were not included in this study. We focused on the NHL (or hematological family) only and excluded studies dealing with other cancers in addition to NHL such as some cohort analyses, to avoid bias in the score calculations.

From all available and retrieved full publications, we extracted all risk estimates (relative risk (RR), odds ratio (OR) and/or SIR/SMR) corresponding to a general occupational context (job, exposure, or both) or to subgroups like gender, NHL subtypes and so on. When dealing with meta-analyses, we either extracted all risk values, when available, or only the meta-RR/OR otherwise to avoid loss of information. When available, the population size (or number of control), number of cases, details about the disease (NHL subtype) and occupational context such as exposure (intensity, duration) and/or activity (sector or occupation) were also recorded. Confidence intervals (CI) were recorded when available and converted from CI

90% into CI 95% if needed. As a result of this process, we obtained a database containing all occupational contexts related to NHL.

To standardize the obtained database of risks for occupational contexts related to NHL, the NHL subtypes were coded using the International Classification of Disease, 10th edition (ICD-10), the activities/sectors were coded using the French national classification of activities (NAF, Nomenclature des Activités Françaises) and the French Health Insurance Fund (CNAM) thesaurus was used to standardize the types of occupational exposures. For example, the exposure "Benzene" was coded as "21311" corresponding to the family of "2-organic compounds", "21-Hydrocarbons", "213-Aromatic hydrocarbons" and finally "2131-Monocyclic aromatic hydrocarbons". Likewise, "Farmers" occupation without precision has been coded in the "A-Agriculture, hunting, forestry" family of the NAF classification whereas a food industry worker has been coded as "DA-Agricultural and food industries". When occupational activities or exposures could not be matched with specific NAF or CNAM codes, more general codes were used or new ones were generated. For example, the "35110-Organic solvents and thinners" code was used when unspecified solvents and the new generated codes included unspecified activity ("ZZ"), scientific activity ("SC"), exposure to unspecified metals ("metals"), chlorinated aliphatic hydrocarbon ("22X20") and unspecified chlorinated hydrocarbon ("22Z20"). In this way we end up with several distinct standardized occupational contexts, named items, with the corresponding distribution of risks.

Some observations from reviews and meta-analyses design were excluded from the standardized database to avoid duplicate information between values from original articles and values used to perform meta-analyses. Moreover, due to the well-known differences in socio-economic conditions between Asia and other developed country as North America, Europe's country and Australia, observations from Asian papers were also excluded because of difficulties in comparing working conditions. Finally, for the sake of statistical power, observations with no reported CI value and items (activity, exposure or both) of sample size smaller than 5 were excluded from the standardized database and not considered in the ranking process.

Ranking procedure

Items were classified within the three categories of the occupational context: activity, occupational exposure or a combination of both activity and exposure. For an occupational context with M different items, let "α" denotes the item index (α = 1, 2, …, M) and $r_{\alpha,i}$ the risk (with 95% confidence interval, ($CI_{inf,\alpha,i}$, $CI_{sup,\alpha,i}$)) for the α item – NHL association with, i = 1, 2, …, n_α where n_α represents the total number of observations for the α item – NHL association viewed and distributed among a total of p_α publications, such $p_\alpha \leq n_\alpha$. The total number of publications and observations (studies) for any occupational context are given by $\sum_{\alpha=1}^{M} p_\alpha = P$ and $\sum_{\alpha=1}^{M} n_\alpha = N$, respectively, with $P \leq N$. Two indices were used to characterize the item – NHL associations within each category of the occupational context.

The Scientific Interest or Popularity (I): I denotes the relative importance of an item in terms of the number of studies (publications) found dealing with NHL among the pool of all considered associations in the occupational context of interest. By definition, the degree of interest of the item α is given by the relation, $I_\alpha = (H_\alpha - H_{min})/(H_{max} - H_{min})$, where $H_\alpha = -(p_\alpha/P) \times \log(p_\alpha/P)$ is the Shannon information, $H_{min} = \log(P)/P$ corresponding to H_α with $p_\alpha = 1$, and $H_{max} = \max(H_\alpha)$. Such a score is also used in biology or ecology for scoring the species diversity (4). By definition, $0 \leq I \leq 1$, with $I = 0$ (for single or very few publications) representing items having very low prominence and rarely studied, while $I \to 1$ (for numerous publications) corresponding to the items more frequently studied.

The Association Strength (S): based on the distributions of risks (all observations), the association between a given item α of occupational context and NHL is rated on a scale $0 \leq S_\alpha \leq 1$; $S = 0$ (low risk) denotes weak or non-associated items to NHL while $S_\alpha \to 1$ (high risk) corresponds to a strong association between the item and NHL. The association strength S_α is assigned to a α item – NHL association by using the logistic transformation as, $\text{logit}(S_\alpha) = b \times (r_{\alpha,95} - r_{th}) \times \log(3)$, where "b" is a stiffness constant, r_{th} the association risk threshold at which $S_\alpha = 0.5$, and $r_{\alpha,95} = 95^{th}$ - percentile of $\{d_\alpha(r)\}$, where $d_\alpha(r)$ is the meta-distribution of risks r (association measures) for the α item – NHL association. Therefore, the strength S_α is equal to 0.5 and 0.75 for associations with risks $r_{\alpha,95} = r_{th}$ and $r_{\alpha,95} = r_{th} + 1/b$, respectively. The values of b = 1 and r_{th} = 2 were used in all calculations reported in the paper. In this case, S_α = 0.25, 0.5 and 0.75 for associations with risks $r_{\alpha,95}$ = 1, 2 and 3, respectively.

As the association measure risk for each i observation in the n_α observations of α item – NHL association is given by a lognormal distribution, LogNorm($\mu_{\alpha,i}$, $\sigma_{\alpha,i}$), with $\mu_{\alpha,i} = \log(r_{\alpha,i})$ and $\sigma_{\alpha,i} = [\log(CI_{sup,\alpha,i}/CI_{inf,\alpha,i})]/2 \times 1.96$, the $d_\alpha(r)$ is obtained by weighted averaging the partial lognormal distributions over all observations n_α as, $d_\alpha(r) = \sum_{i=1}^{n_\alpha} w_{\alpha,i} \times \text{LogNorm}(\mu_{\alpha,i}, \sigma_{\alpha,i})$, where $w_{\alpha,i}$ is the weight of each observation. In absence of additional information, we used in the calculations reported here the following weights, $w_{\alpha,i} = r_{\alpha,i}/\sum_{i=1}^{n_\alpha} r_{\alpha,i}$, that account for relative importance of risks; higher weight being associated to items with higher risk.

Data processing and analyzes

The database was setup on an Excel spreadsheet and all statistical processing, analyzes and modeling were performed using R software (5).

RESULTS

Identification of occupational contexts

The literature search identified a total of 4,157 articles of which 85 met the selection criteria based on the language, title and abstract (see appendix). Sixty-six publications were retained for further analysis after the exclusion of duplicate reports, those with Asian study populations, and lacking of confidence intervals (Figure 1A). The Figure 1B lists the 66 studies, which included 51 case-control studies, 5 cohort studies and 10 meta-analyses (6-71).

The 66 studies included a total of 4,260 observations of which 2,311 (54%) concerned the activities, 2,986 (70%) the kind of exposures, and 1,076 (25%) both occupational activity and exposure (Figure 1B). An observation is defined as a row in the database table containing the complete information on the occupational context item (either activity or exposure or combination of both) and the corresponding risk (association measure with NHL) value with the associated confidence interval. After removal of items with less than 5 observations, we obtained 31 distinct 2-digit codes denoting occupational activities, 91 distinct 5-digit codes corresponding to specific kinds of occupational exposure and 35 distinct occupational activity-exposure combinations. Regarding the NHL disease, only 1,546 observations could be included with accurate information on the subtype including “Follicular non-Hodgkin's

lymphoma (nodular)" (ICD-10 Code family: C82) accounted for 298 observations, "Diffuse non-Hodgkin's lymphoma" (C83) for 316 observations, "Other and unspecified types of non-Hodgkin's lymphoma" (C85) for 608 observations, "Malignant immunoproliferative diseases" (C88.X) for 15 observations and "Chronic lymphocytic leukemia" (C91.1) for 305. We found also a combination of C82 and C91.1 code that accounted for 4 observations.

Occupational activities (2,311 observations): Seven studies contributed to 50% of all observations and the item "agriculture" was found in 33 studies (data not shown). The majority of observations concerned the items "agriculture" (48.3%) and "manufacturing industries" (16.8%) which were also the occupational activities with the most information about subtypes. (Appendix A3). Regarding the distribution of risk, the item "Agriculture, hunting, forestry" accounted for 1,100 observations of which only 12.9% had a risk value significantly greater than one (lower bound of the CI greater than or equal to 1), and the item "Production and distribution of electricity, gas and water" accounted for 190 observations with 9.5% of risks significantly greater than one (data available in appendix A4).

Occupational exposures (2,986 observations): Three studies provided 50% of the information on occupational exposures and exposure to the item "Organic solvents and thinners" was found in 17 studies (data not shown). Exposure to the items "Organic compounds and industrial substances" (family code in CNAM thesaurus) and "Organic molecules and physical agents" prevail with 38.8 and 44.7% of total observations, respectively, and these were the occupational exposures with the most NHL's subtypes information (appendix A6). The exposures "Organic solvents and thinners" and "Benzene" accounted for 272 and 206 observations with 14.7% and 4.4% of risks being significantly greater than one, respectively (data available in appendix A7).

Occupational activity-exposure combinations (1,076 observations): The majority (82.9%) of reports were related to agriculture with 50% of the data from two studies. The most frequently studied item was "agriculture-herbicides", found in 7 studies (data not shown) with 87 observations (data available in appendix A9) and 2.3% of risks significantly greater than one. They were only few combinations with information on NHL's subtypes, mostly in the agricultural sector (appendix A9). Exposure to "Magnetic waves" in the activity "Production and distribution of electricity, gas and water" and exposure to "Organophosphorous insecticides" in the activity "Agriculture, hunting and forestry" were explored 101 and 95

times with 8.9% and 6.3% of risks significantly greater than one, respectively (available in appendix A10).

Ranked lists of occupational contexts

For simplicity, the items were grouped into one of three classes according to the value of the corresponding ranking indices I and S: $0 \leq I, S \leq 0.25$, $0.25 < I, S \leq 0.5$ and $0.5 < I, S \leq 1$ corresponding for I to "Low", "Medium" and "High" prominence or extent studied and for S as "Zero", "Medium" and "Strong" strength of association with NHL respectively. Because of lack of information on NHL's subtypes the ranking results reported below were performed using all observations regardless the subtypes. Additional ranking calculations taking into account the subtypes for each occupational context are given in the appendix A5, A8 and A11.

The majority of occupational activities were found with low or moderate degrees of interest (80.6%) but 15 had high association scores (Figure 2A and D). On the contrary, a majority of kinds of occupational exposures had a low prominence score (61.5%) and among them a small proportion had a high score of association (12.1% of them). Moreover, two occupational exposures were found with high scores both of interest and association (Figure 2B and E). Similar profile is also observed for activity-exposure combinations (Figure 2C and E) with a great number of items with low interest (60%) and medium association scores (71.4%). Details for each group are available in Table 1, Table 2 and Table 3.

Occupational activities

Six occupational activities were frequently mentioned in the literature (19.4%), among them 5 had a high score for the strength of association (Table 1, **Figure 2**A and D), and concerned mostly agricultural and manufacturing sector: "Agricultural and food industries", "Agriculture, hunting, forestry", "Manufacturing industries" (unspecified), "Metallurgy and metalworking", "Production and distribution of electricity, gas and water" and "Property business, renting and business". Eleven occupational activities have been little studied (35.5%) but 6 of them scored high for association score: "Extractive industries", "Hotels and restaurants", "Leather and footwear industries", "Rubber and plastics industries", "Public administration" and "Sciences". Only one sector had the lower degree of interest and had a

weak score of association: "Electrical and electronic equipment manufacturing". Among the 31 occupational activities, 12 involve subtypes (appendix A5).

Table 1: List of the 31 occupational activities associated with NHL ranked according to the French national classification of activities and the degree of interest (I) and strength of association (S). Occupational activities labels are indicated from general family (left) to more specific activity (right). Colors "white", "hatched grey" and "black" correspond to "Low", "Medium" and "High" interest and for S to "Zero", "Medium" and "Strong" strength of association, respectively, with the following classes: 0 ≤ I, S ≤ 0.25, 0.25 < I, S ≤ 0.5 and 0.5 < I, S ≤ 1.

Occupational activity		I	S
Agriculture, hunting, forestry			
Agriculture, hunting, forestry and aquaculture, fishing			
Commerce, car and household goods repairs			
Community, social and personal services			
Construction			
Domestic services			
Education			
Extractive industries			
Extraterritorial activities			
Financial activities			
Healthcare and social working			
Hotels and restaurants			
Manufacturing industries	*(Unspecified)*		
	Agricultural and food industries		
	Chemical industry		
	Coking, refining and nuclear industries		
	Electrical and electronic equipment manufacturing		
	Engine and equipment manufacturing		
	Leather and footwear industries		
	Metallurgy and metalworking		
	Other non-metallic mineral products manufacturing		
	Paper and cardboard industry, publishing and printing		
	Rubber and plastics industries		
	Textile and clothing industry		
	Transport equipment manufacturing		
	Woodworking and wood items manufacturing		
Production and distribution of electricity, gas and water			
Property business, renting and business			
Public administration			
Sciences			
Transports and communications			

Occupational exposures:

Among the 91 listed items of occupational exposures, only 7 were highly mentioned in the literature (7.7%), and only 2 of these, "Carboxylic acids and peracids" and "Organic solvents and thinners", were found with high score of association (Table 2, Figure 2B and E). Most

kinds of occupational exposure appeared to have been little studied (61.5%) nevertheless six stood out due to their high score of association: “Black products”, “Heterocycles”, “Metals”, “Potassium oxides”, “Non-ionizing radiation” and “Radiations”. Finally, 22 occupational exposures were ranked as being in categories of “non-studied” and non-associated (24.2%). Only 29 occupational exposures out of 91 were concerned with subtypes (appendix A8).

Table 2: List of the 91 occupational exposures associated with NHL ranked according to the French Insurance Fund for disease thesaurus and the degree of interest (I) and strength of association (S). Occupational exposures labels are indicated from general family (left) to more specific exposure (right). Colors “white”, “hatched grey” and “black” correspond to “Low”, “Medium” and “High” interest and for S to “Zero”, “Medium” and “Strong” strength of association, respectively, with the following classes: $0 \leq I, S \leq 0.25$, $0.25 < I, S \leq 0.5$ and $0.5 < I, S \leq 1$.

Occupational exposures			I	S
Inorganic compounds	Inorganic arsenic compounds			
	Inorganic cadmium compounds			
	Inorganic chromium compounds (except acids)			
	Potassium oxides			
Unspecified metals				
Hydrocarbons	**(Unspecified)**			
	Aliphatic hydrocarbons			
	Aromatic hydrocarbons	(Unspecified)		
		Benzene		
		Toluene		
		Xylene		
		Styrene		
		Benzene + toluene + xylene (BTX)		
	Heterocycles			
Halogenated derivative of hydrocarbons	**(Unspecified)**			
	Halogenated derivative of saturated aliphatic hydrocarbons	*Chloromethane (methylchloride)*		
		Dichloromethane (methylene chloride)		
		Trichloromethane (chloroform)		
		Tetrachloromethane (carbon tetrachloride)		
		1,2-Dichloroethane		
		1,2-Dichloropropane		
		Bromoethane		
	Halogenated derivative of unsaturated aliphatic hydrocarbon	*Trichloroethylene*		
	Halogenated derivative of alicyclic hydrocarbons	*Lindane*		
		Lindane + acide dichlorophenoxyacetique (2.4-D)		
		Chlordane		
	Halogenated derivative of aromatic hydrocarbons	*Polychlorobiphenyles (PCB)*		
		Dichlorodiphenyltrichloroethane (DDT)		
		DDT (+ 2.4-D)[1]		
		DDT (+ organophosphorus pesticides)[1]		
		DDT (+ 2.4-D or organophosphorus pesticides)[1]		
	Chlorinated aliphatic hydrocarbon			

	Chlorinated hydrocarbon, unspecified			
Phenols and derivative	**(Unspecified)**			
	Halogenated derivative of phenols			
Aldehydes	Aliphatic aldehydes	*Formaldehyde*		
Carboxylic acids and peracids				
Phosphates phosphites pyrophosphates thiophosphates	Organophosphorus pesticides	(Unspecified)		
		Diazinon		
		Malathion		
		Malathion (+ other)[2]		
	Carbamates			
Amides, sulfonamides, phosphoramides, imides				
Amines and imines derivatives				
Nitrate derivatives				
Oxygenated hydrocarbon compounds				
Dusts (except fibers)	Inorganic dust (metal and mineral)			
	Organic dusts	(Unspecified)		
		Wood dusts		
	Vegetable-based dusts			
Fibers	**(Unspecified)**			
	Organic fibers	*Asbestos fibers*		
Fumes	Welding fumes			
Liquids or vapours	Organic liquids or vapours	*Organic solvents and thinners*		
Materials and various products	**(Unspecified)**			
	Adhesives and glues			
	Dyestuffs			
	Inks			
	Foodstuff and feeding products	*Meat*		
		Fish and shellfish		
	Fuel	*Automotive fuel, diesel, gas, oil*		
		Petrol		
		Liquified petroleum gas		
	Oils and greases	(Unspecified)		
		Cutting or machining oils		
		Lubricant mineral oils and greases		
	"Black" products (asphalt, tar, creosote)			
	Cleaning products			
Products for agricultural use or gardening or wood treatment	**(Unspecified)**			
	Insecticides	(Unspecified)		
		Anticholinesterasic carbamate insecticides		
		Organochlorine insecticides		
		Organophosphorus insecticides		
		Pyrethrum and pyrethroids insecticides		
	Herbicides	(Unspecified)		
		Triazine herbicides		
		Triazines herbicides + triazole and derivatives		
	Fungicides	(Unspecified)		
		Dithiocarbamate fungicides		
		Organo metal fungicides (mercury tin ...)		
		Phthalimide fungicides		
Radiations	**(Unspecified)**			
	Ionizing radiations and radioactivity	(Unspecified)		
		Ultraviolet		
		Magnetic waves		
	Non-ionizing radiations	(Unspecified)		

		Electrostatic field		
Thermal ambience				
Vegetable-based products	Vegetable-based insecticides	(Unspecified)		
		Nicotine		
Animals and animal products				

[1] *"in association with (exposures)"*

[2] *Potential combinations of Malathion and: DDT or 2.4-D or Organophosphorus pesticides or Anticholinesterasic carbamate insecticides or Herbicides*

Occupational activity-exposure combinations:

Five distinct occupational activities and 32 distinct occupational exposures were reported as well as 35 occupational activity-exposure combinations, sometimes with multi-exposure (i.e. several exposures in the same occupational activity) (Table 3, Figure 2**Erreur ! Source du renvoi introuvable.**C and E). Most concerned the agriculture sector (82.9%). Nevertheless; most reports concerned exposure in the agriculture sector (7/29 agricultural combinations) and only three of these were also highly scored for the association: "Carboxylic acids and peracids", "Lindane", and "Potassium oxides". The other combination with numerous report and high degree of interest was the exposure to "Trichloroethylene" in "Manufacturing industries", but the strength of association with NHL was moderate. The majority of these combinations were found with a low degree of interest and only two had a high association score: "Fungicides" in "Agricultural, hunting, forestry" and "Non-ionizing radiations" in "Production and distribution of electricity, gas and water". Finally, four combinations were characterized both with less interest and no association with NHL. Only 5 out of 35 combinations were concerned with subtypes (appendix A11).

Table 3: List of the 35 occupational activity-exposure combinations associated with NHL ranked according to the French national classification of activities and the French Insurance Fund for disease thesaurus and the degree of interest (I) and strength of association (S). Occupational activities and exposures labels are indicated from general family (left) to more specific activity/exposure (right). Colors "white", "hatched grey" and "black" correspond to "Low", "Medium" and "High" interest and for S to "Zero", "Medium" and "Strong" strength of association, respectively, with the following classes: $0 \leq I, S \leq 0.25$, $0.25 < I, S \leq 0.5$ and $0.5 < I, S \leq 1$.

Occupational activity	Occupational exposure	I	S
Agriculture, hunting, forestry	Potassium oxides		
	1,2-Dichloropropane		
	Bromoethane		
	Lindane		
	Lindane + 2.4D		
	Chlordane		
	DDT		
	DDT + 2.4-D + Organophosphorus pesticides		

		DDT + 2.4-D			
		DDT + Organophosphorus pesticides			
		Carboxylic acids and peracids			
		Organophosphorus pesticides	Unspecified		
			Diazinon		
			Malathion		
		Carbamates			
		Automotive fuel, diesel, gas, oil			
		Products for agricultural use or gardening or wood treatment			
		Insecticides	Unspecified		
			Anticholinesterasic carbamate insecticides		
			Organochlorine insecticides		
			Organophosphorus insecticides		
		Herbicides	Unspecified		
			Triazine herbicides		
		Fungicides	Unspecified		
			Dithiocarbamate fungicides		
			Organo metal fungicides (mercury tin ...)		
			Phthalimide fungicides		
		Nicotine			
		Animals and animal products			
Manufacturing industries		Trichloroethylene			
	Transport equipment manufacturing	Organic solvents and thinners			
Production and distribution of electricity, gas and water		Ultraviolet			
		Magnetic waves			
		Non-ionizing radiations			
Property business, renting and business		Organic solvents and thinners			

CONCLUDING REMARKS

The aim of this work was twofold: to use the literature for inventorying the occupational contexts (activities, exposures and activity-exposure combinations) associated with a risk of NHL and for further insights rank these occupational contexts.

The results reported in this paper are based on the literature search conducted from June 11th to July 16th 2013 in four electronic databases: PubMed, CISMeF, BDSP and SUDOC. The selection process allowed us to identify 31 distinct 2-digit codes denoting occupational activities, 91 distinct 5-digit codes corresponding to specific kinds of occupational exposure and 35 distinct occupational activity-exposure combinations, mainly involving agricultural or industrial sector and solvents and pesticides products.

The ranking of occupational contexts was performed both in terms of the degree of scientific interest in the literature, as measured by the relative number of reports, and of the strength of association with NHL, based on the distribution of association risks of occupational contexts

with NHL. It appears that 20 activity sectors are strongly associated with NHL of which 6 are frequently reported in literature: "Agricultural and food industries", "Agriculture, hunting, forestry", "Manufacturing industries", "Metallurgy and metalworking", "Production and distribution of electricity, gas and water" and "Property business, renting and business". Only one sector, "Electrical and electronic equipment manufacturing", was found both rarely studied and with no association with NHL. Concerning occupational exposure items, only 2 exposures, "Carboxylic acids and peracids" and "Organic solvents and thinners", out of 91 were ranked as both highly interested and strongly associated with NHL. Among the majority of occupational exposures that have received little attention in the literature (61.5%), 6 ranked high for the association: "Black products", "Heterocycles", "Metal" (unspecified), "Potassium oxides", "Non-ionizing radiation" and "Radiations". This result would indicate the need of further inquiry into these types of exposures. Twenty-two occupational exposures were neither associated with NHL nor deemed worth of interest from researchers. Most of the activity-exposure combinations were related to "Agriculture, hunting and forestry" sector (82.9%) and three exposures were frequently found in this sector: "Carboxylic acids and peracids", "Lindane" and "Potassium oxides".

For the purpose of the objectives of this work, the approach developed and outlined above differs from the classical meta-analysis in two major aspects: the criteria for selecting articles and the way of scoring the outputs. Indeed, the aim of this work was to inventory all occupational contexts related to NHL and to rank these occupational contexts by means of two scores, interest and association. In contrast to meta-analyses design, for the selection of articles, we loosely disregarded the Material and Methods section to allow, for statistical concerns, a larger recruitment of studies or observations dealing with NHL. And, in the place of calculating meta-RR, we developed two scores involving the interest or prominence of studies and the strength of association with the risk of NHL based on the total number of publications and the total number of risks in observations, respectively. The interest index allows scoring the degree of most studied occupational contexts regardless their associated risk values whereas the association strength combines risks no matter its interest. Accordingly, we excluded from our analyses publications broadly dealing with other cancers and not strictly focusing on NHL, like in some cohort studies, because the aims of those publications were to listing diseases appearing an occupational context rather than investigating occupational contexts associated to NHL.

It might be useful to underline that the approach outlined above is quite flexible enough to allow both incorporating more data, studies or articles when available, and including data of different kind like study's designs or kinds of measurement of the exposure (as job-exposure matrix or direct measurement) that can be used, for example, in weighting risks for the calculation of the association strength. In general, the reported results were not stratified in terms of co-variables like gender, exposure intensity, NHL subtypes, …, because either of very low number of observations or information missing. Regarding the important differences of etiologies, it would be worthwhile to re-do such an analysis at the NHL – subtype level with more data available.

Finally, this work is instructive in both providing a ranked inventory of occupational contexts associated with risk of NHL and highlighting "hot" occupational activities and exposures. Such a material obviously does not address the biological plausibility of the associations between occupational context and NHL but turns out already very useful in assessment of the health risk in the workplace.

REFERENCE LIST

(1) Morton LM, Slager SL, Cerhan JR, et al. Etiologic Heterogeneity Among Non-Hodgkin Lymphoma Subtypes: The InterLymph Non-Hodgkin Lymphoma Subtypes Project. J Natl Cancer Inst Monogr 2014;2014:130-144.

(2) IARC, Swerdlow SH, Campo E, et al. WHO classification of tumours of haematopoietic and lymphoid tissues, 2008.

(3) Monnereau A, Danzon A, Delafosse P et al. Lymphome malin non hodgkinien. In: Bélot A, Velten M, Grosclaude P, et al., eds. Estimation nationale de l'incidence et de la mortalité par cancer en France entre 1980 et 2005. Saint-Maurice (Fra): Institut de Veille Sanitaire, 2008:96-9. French

(4) Shannon CE. A Mathematical Theory of Communication. The Bell System Technical Journal 1948;27:379-423.

(5) RCoreTeam. R: A Language and Environment for Statistical Computing. Vienna (Austria): R Foundation for Statistical Computing, 2013.

(6) Amadori D, Nanni O, Falcini F, et al. Chronic lymphocytic leukaemias and non-Hodgkin's lymphomas by histological type in farming-animal breeding workers: a population case-control study based on job titles. Occup Environ Med 1995;52:374-9.

385 (7) Band PR, Le ND, Fang R, et al. Identification of occupational cancer risks in British
386 Columbia: a population-based case-control study of 769 cases of non-Hodgkin's
387 lymphoma analyzed by histopathology subtypes. J Occup Environ Med 2004;46:479-
388 89.

389 (8) Baris D, Zahm SH, Cantor KP, et al. Agricultural use of DDT and risk of non-
390 Hodgkin's lymphoma: pooled analysis of three case-control studies in the United
391 States. Occup Environ Med 1998;55:522-7.

392 (9) Blair A, Cantor KP, Zahm SH. Non-hodgkin's lymphoma and agricultural use of the
393 insecticide lindane. Am J Ind Med 1998;33:82-7.

394 (10) Boffetta P, van der Hel O, Kricker A, et al. Exposure to ultraviolet radiation and risk
395 of malignant lymphoma and multiple myeloma--a multicentre European case-control
396 study. Int J Epidemiol 2008;37:1080-94.

397 (11) Briggs NC, Levine RS, Hall HI, et al. Occupational risk factors for selected cancers
398 among African American and White men in the United States. Am J Public Health
399 2003;93:1748-52.

400 (12) Cantor KP, Blair A, Everett G, et al. Pesticides and other agricultural risk factors for
401 non-Hodgkin's lymphoma among men in Iowa and Minnesota. Cancer Res
402 1992;52:2447-55.

403 (13) Cocco P, t'Mannetje A, Fadda D, et al. Occupational exposure to solvents and risk of
404 lymphoma subtypes: results from the Epilymph case-control study. Occup Environ
405 Med 2010;67:341-7.

406 (14) Cocco P, Satta G, D'Andrea I, et al. Lymphoma risk in livestock farmers: results of the
407 Epilymph study. Int J Cancer 2013;132:2613-8.

408 (15) Cocco P, Satta G, Dubois S, et al. Lymphoma risk and occupational exposure to
409 pesticides: results of the Epilymph study. Occup Environ Med 2013;70:91-8.

410 (16) Costantini AS, Miligi L, Kriebel D, et al. A multicenter case-control study in Italy on
411 hematolymphopoietic neoplasms and occupation. Epidemiology 2001;12:78-87.

412 (17) De Roos AJ, Zahm SH, Cantor KP, et al. Integrative assessment of multiple pesticides
413 as risk factors for non-Hodgkin's lymphoma among men. Occup Environ Med
414 2003;60:E11.

415 (18) Druet-Cabanac M, Moalic-Juge S, Fristch F, et al. Etude de l'exposition
416 professionnelle et environnementale dans la survenue des lymphomes malins non-
417 Hodgkiniens en Limousin. Arch Mal Prof Env 2005:523-31.

418 (19) Dryver E, Brandt L, Kauppinen T, et al. Occupational exposures and non-Hodgkin's
419 lymphoma in Southern Sweden. Int J Occup Environ Health 2004;10:13-21.

(20) Eheman CR, Tolbert PE, Coates RJ, et al. Case-control assessment of the association between non-Hodgkin's lymphoma and occupational radiation with doses assessed using a job exposure matrix. Am J Ind Med 2000;38:19-27.

(21) Fritschi L, Johnson KC, Kliewer EV, et al. Animal-related occupations and the risk of leukemia, myeloma, and non-Hodgkin's lymphoma in Canada. Cancer Causes Control 2002;13:563-71.

(22) Fritschi L, Benke G, Hughes AM, et al. Occupational exposure to pesticides and risk of non-Hodgkin's lymphoma. Am J Epidemiol 2005;162:849-57.

(23) Fritschi L, Siemiatycki J. Lymphoma, myeloma and occupation: results of a case-control study. Int J Cancer 1996;67:498-503.

(24) Fritschi L, Benke G, Hughes AM, et al. Risk of non-Hodgkin lymphoma associated with occupational exposure to solvents, metals, organic dusts and PCBs (Australia). Cancer Causes Control 2005;16:599-607.

(25) Garabedian MJ, Hoppin JA, Tolbert PE, et al. Occupational chlorophenol exposure and non-Hodgkin's lymphoma. J Occup Environ Med 1999;41:267-72.

(26) Hardell L, Eriksson M. A case-control study of non-Hodgkin lymphoma and exposure to pesticides. Cancer 1999;85:1353-60.

(27) Hoar Zahm S, Weisenburger DD, Cantor KP, et al. Role of the herbicide atrazine in the development of non-Hodgkin's lymphoma. Scand J Work Environ Health 1993;19:108-14.

(28) Hohenadel K, Harris SA, McLaughlin JR, et al. Exposure to multiple pesticides and risk of non-Hodgkin lymphoma in men from six Canadian provinces. Int J Environ Res Public Health 2011;8:2320-30.

(29) Karipidis KK, Benke G, Sim MR, et al. Occupational exposure to ionizing and non-ionizing radiation and risk of non-Hodgkin lymphoma. Int Arch Occup Environ Health 2007;80:663-70.

(30) Karipidis KK, Benke G, Sim MR, et al. Non-Hodgkin lymphoma and occupational radiation exposure assessed using local data. Occup Med (Lond) 2009;59:437-9.

(31) Kato I, Watanabe-Meserve H, Koenig KL, et al. Pesticide product use and risk of non-Hodgkin lymphoma in women. Environ Health Perspect 2004;112:1275-81.

(32) Mao Y, Hu J, Ugnat AM, et al. Non-Hodgkin's lymphoma and occupational exposure to chemicals in Canada. Canadian Cancer Registries Epidemiology Research Group. Ann Oncol 2000;11 (Suppl 1):69-73.

(33) Massoudi BL, Talbott EO, Day RD, et al. A case-control study of hematopoietic and lymphoid neoplasms: the role of work in the chemical industry. Am J Ind Med 1997;31:21-7.

(34) Mester B, Nieters A, Deeg E, et al. Occupation and malignant lymphoma: a population based case control study in Germany. Occup Environ Med 2006;63:17-26.

(35) Miligi L, Costantini AS, Benvenuti A, et al. Occupational exposure to solvents and the risk of lymphomas. Epidemiology 2006;17:552-61.

(36) Miligi L, Costantini AS, Bolejack V, et al. Non-Hodgkin's lymphoma, leukemia, and exposures in agriculture: results from the Italian multicenter case-control study. Am J Ind Med 2003;44:627-36.

(37) Mills PK, Yang R, Riordan D. Lymphohematopoietic cancers in the United Farm Workers of America (UFW), 1988-2001. Cancer Causes Control 2005;16:823-30.

(38) Moore T, Brennan P, Becker N, et al. Occupational exposure to meat and risk of lymphoma: a multicenter case-control study from Europe. Int J Cancer 2007;121:2761-6.

(39) Orsi L, Delabre L, Monnereau A, et al. Occupational exposure to pesticides and lymphoid neoplasms among men: results of a French case-control study. Occup Environ Med 2009;66:291-8.

(40) Orsi L, Monnereau A, Dananche B, et al. Occupational exposure to organic solvents and lymphoid neoplasms in men: results of a French case-control study. Occup Environ Med 2010;67:664-72.

(41) Persson B, Fredriksson M, Olsen K, et al. Some occupational exposures as risk factors for malignant lymphomas. Cancer 1993;72:1773-8.

(42) Purdue MP, Severson RK, Colt JS, et al. Degreasing and risk of non-Hodgkin lymphoma. Occup Environ Med 2009;66:557-60.

(43) Purdue MP, Bakke B, Stewart P, et al. A case-control study of occupational exposure to trichloroethylene and non-Hodgkin lymphoma. Environ Health Perspect 2011;119:232-8.

(44) Rafnsson V. Risk of non-Hodgkin's lymphoma and exposure to hexachlorocyclohexane, a nested case-control study. Eur J Cancer 2006;42:2781-5.

(45) Richardson DB, Terschuren C, Hoffmann W. Occupational risk factors for non-Hodgkin's lymphoma: a population-based case-control study in Northern Germany. Am J Ind Med 2008;51:258-68.

(46) Schenk M, Purdue MP, Colt JS, et al. Occupation/industry and risk of non-Hodgkin's lymphoma in the United States. Occup Environ Med 2009;66:23-31.

(47) Seidler A, Becker N, Nieters A, et al. Asbestos exposure and malignant lymphoma: a multicenter case-control study in Germany and Italy. Int Arch Occup Environ Health 2010;83:563-70.

(48) Tatham L, Tolbert P, Kjeldsberg C. Occupational risk factors for subgroups of non-Hodgkin's lymphoma. Epidemiology 1997;8:551-8.

(49) 't Mannetje A, Dryson E, Walls C, et al. High risk occupations for non-Hodgkin's lymphoma in New Zealand: case-control study. Occup Environ Med 2008;65:354-63.

(50) Tranah GJ, Holly EA, Bracci PM. Solvent exposure and non-Hodgkin lymphoma: no risk in a population-based study in the San Francisco Bay Area. Cancer Epidemiol Biomarkers Prev 2009;18:3130-2.

(51) Villeneuve PJ, Agnew DA, Miller AB, et al. Non-Hodgkin's lymphoma among electric utility workers in Ontario: the evaluation of alternate indices of exposure to 60 Hz electric and magnetic fields. Occup Environ Med 2000;57:249-57.

(52) Waddell BL, Zahm SH, Baris D, et al. Agricultural use of organophosphate pesticides and the risk of non-Hodgkin's lymphoma among male farmers (United States). Cancer Causes Control 2001;12:509-17.

(53) Wang R, Zhang Y, Lan Q, et al. Occupational exposure to solvents and risk of non-Hodgkin lymphoma in Connecticut women. Am J Epidemiol 2009;169:176-85.

(54) van Wijngaarden E, Savitz DA. Occupational sunlight exposure and mortality from non-Hodgkin lymphoma among electric utility workers. J Occup Environ Med 2001;43:548-53.

(55) Zheng T, Zahm SH, Cantor KP, et al. Agricultural exposure to carbamate pesticides and risk of non-Hodgkin lymphoma. J Occup Environ Med 2001;43:641-9.

(56) Zheng T, Blair A, Zhang Y, et al. Occupation and risk of non-Hodgkin's lymphoma and chronic lymphocytic leukemia. J Occup Environ Med 2002;44:469-74.

(57) Bloemen LJ, Mandel JS, Bond GG, et al. An update of mortality among chemical workers potentially exposed to the herbicide 2.4-Dichlorophenoxyacetic acid and its derivatives. J Occup Med 1993;35:1208-12.

(58) Cano MI, Pollan M. Non-Hodgkin's lymphomas and occupation in Sweden. Int Arch Occup Environ Health 2001;74:443-9.

(59) Morrison HI, Semenciw RM, Wilkins K, et al. Non-Hodgkin's lymphoma and agricultural practices in the prairie provinces of Canada. Scand J Work Environ Health 1994;20:42-7.

(60) Neasham D, Sifi A, Nielsen KR, et al. Occupation and risk of lymphoma: a multicentre prospective cohort study (EPIC). Occup Environ Med 2011;68:77-81.

(61) Schroeder JC, Savitz DA. Lymphoma and multiple myeloma mortality in relation to magnetic field exposure among electric utility workers. Am J Ind Med 1997;32:392-402.

526 (62) Alexander DD, Wagner ME. Benzene exposure and non-Hodgkin lymphoma: a meta-
527 analysis of epidemiologic studies. J Occup Environ Med 2010;52:169-89.

528 (63) Boffetta P, de Vocht F. Occupation and the risk of non-Hodgkin lymphoma. Cancer
529 Epidemiol Biomarkers Prev 2007;16:369-72.

530 (64) Kane EV, Newton R. Benzene and the risk of non-Hodgkin lymphoma: a review and
531 meta-analysis of the literature. Cancer Epidemiol 2010;34:7-12.

532 (65) Kane EV, Newton R. Occupational exposure to gasoline and the risk of non-Hodgkin
533 lymphoma: a review and meta-analysis of the literature. Cancer Epidemiol
534 2010;34:516-22.

535 (66) Karami S, Bassig B, Stewart PA, et al. Occupational trichloroethylene exposure and
536 risk of lymphatic and haematopoietic cancers: a meta-analysis. Occup Environ Med
537 2013;70:591-9.

538 (67) Keller-Byrne JE, Khuder SA, Schaub EA, et al. A meta-analysis of non-Hodgkin's
539 lymphoma among farmers in the central United States. Am J Ind Med 1997;31:442-4.

540 (68) Khuder SA, Schaub EA, Keller-Byrne JE. Meta-analyses of non-Hodgkin's lymphoma
541 and farming. Scand J Work Environ Health 1998;24:255-61.

542 (69) Mandel JH, Kelsh MA, Mink PJ, et al. Occupational trichloroethylene exposure and
543 non-Hodgkin's lymphoma: a meta-analysis and review. Occup Environ Med
544 2006;63:597-607.

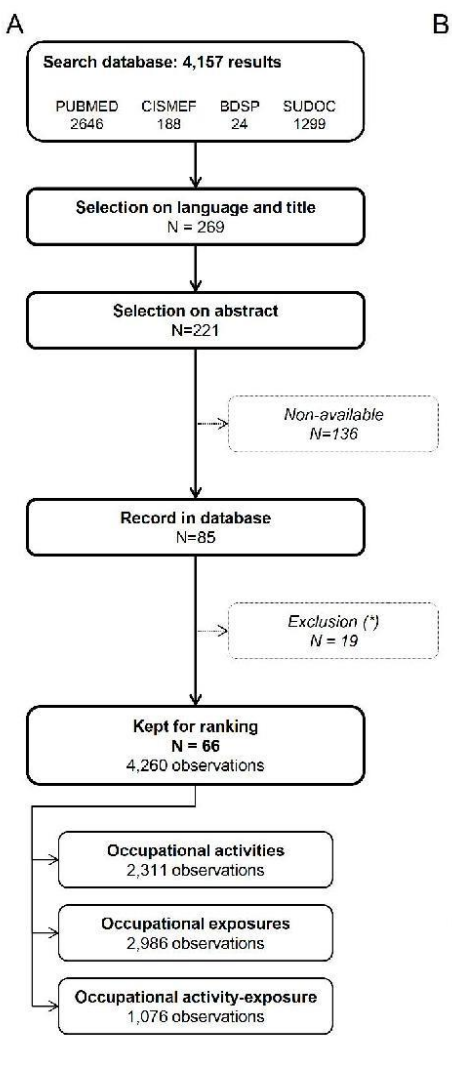
545 (70) Steinmaus C, Smith AH, Jones RM, et al. Meta-analysis of benzene exposure and non-
546 Hodgkin lymphoma: biases could mask an important association. Occup Environ Med
547 2008;65:371-8.

548 (71) Vlaanderen J, Lan Q, Kromhout H, et al. Occupational benzene exposure and the risk
549 of lymphoma subtypes: a meta-analysis of cohort studies incorporating three study
550 quality dimensions. Environ Health Perspect 2011;119:159-167.

551

Article soumis dans American Journal of Industrial Medicine

FIGURES AND CAPTIONS



B

REFERENCE	1st AUTHOR	YEAR	LOCATION
Case-control			
(8)	Amadori	1995	Italy
(7)	Band	2004	Canada
(5)	Baris	1990	USA
(9)	Blair	1998	USA
(10)	Boffetta	2008	Europe
(11)	Briggs	2003	USA
(12)	Cantor	1992	USA
(13)	Cocco	2010	Europe
(14)	Cocco	2013	Europe
(15)	Cocco	2013	Europe
(16)	Costantini	2001	Italy
(17)	De Roos	2003	USA
(18)	Druet-Cabanac	2001	France
(19)	Dryver	2004	Sweden
(20)	Eheman	2000	USA
(21)	Fritschi	1996	Canada
(22)	Fritschi	2002	Canada
(23)	Fritschi	2005	Australia
(24)	Fritschi	2005	Australia
(25)	Garabedian	1999	USA
(26)	Hardell	1999	Sweden
(27)	Hoar Zhang	1993	USA
(28)	Hohenadel	2011	Canada
(29)	Karipidis	2007	Australia
(30)	Karipidis	2009	Australia
(31)	Kato	2004	USA
(32)	Mao	2000	Canada
(33)	Massoudi	1997	USA
(34)	Mester	2006	Germany
(35)	Miligi	2003	Italy
(36)	Miligi	2006	Italy
(37)	Mills	2005	USA
(38)	Moore	2007	Europe
(39)	Orsi	2009	France
(40)	Orsi	2010	France
(41)	Persson	1993	Sweden
(42)	Purdue	2009	USA
(43)	Purdue	2011	USA
(44)	Rafnsson	2006	Island
(45)	Richardson	2008	Germany
(46)	Schenk	2009	USA
(47)	Soldin	2010	Germany, Italy
(48)	Tatham	1997	USA
(49)	tMannetje	2008	New Zealand
(50)	Tranah	2009	USA
(51)	Villeneuve	2000	USA
(52)	Waddell	2001	USA
(53)	Wang	2009	USA
(54)	Wijngaarden	2001	USA
(55)	Zheng	2001	USA
(56)	Zheng	2002	USA
Cohort			
(57)	Bloemen	1993	USA
(58)	Cano	2001	Sweden
(59)	Morrison	1994	Canada
(60)	Ncasham	2011	Europe
(61)	Schroeder	1997	USA
Meta-analysis			
(62)	Alexander	2010	
(63)	Boffetta	2007	
(64)	Kane	2010	
(65)	Kane	2010	
(66)	Karami	2013	
(67)	Keller-Byrne	1997	
(68)	Khuder	1998	
(69)	Mandel	2006	
(70)	Steinmaus	2008	
(71)	Vlaanderen	2011	

Figure 1: Selection of epidemiological studies about non-Hodgkin lymphoma
A) From 4,157 articles in the beginning, 269 publications were selected based on the language and title, then 221 publications have been retained on the abstract but 136 were not available, resulting in a reading of 85 articles. (*) After reading we excluded Asian studies, reviews and publications without CI to obtain the final database with 66 articles representing a total of 4,260 observations including 2,311 observations on the occupational activities, 2,986 for the occupational exposures and 1,076 for the combination of both. B) List of the 66 selected publications, grouped by design of study and first author name.

Article soumis dans American Journal of Industrial Medicine

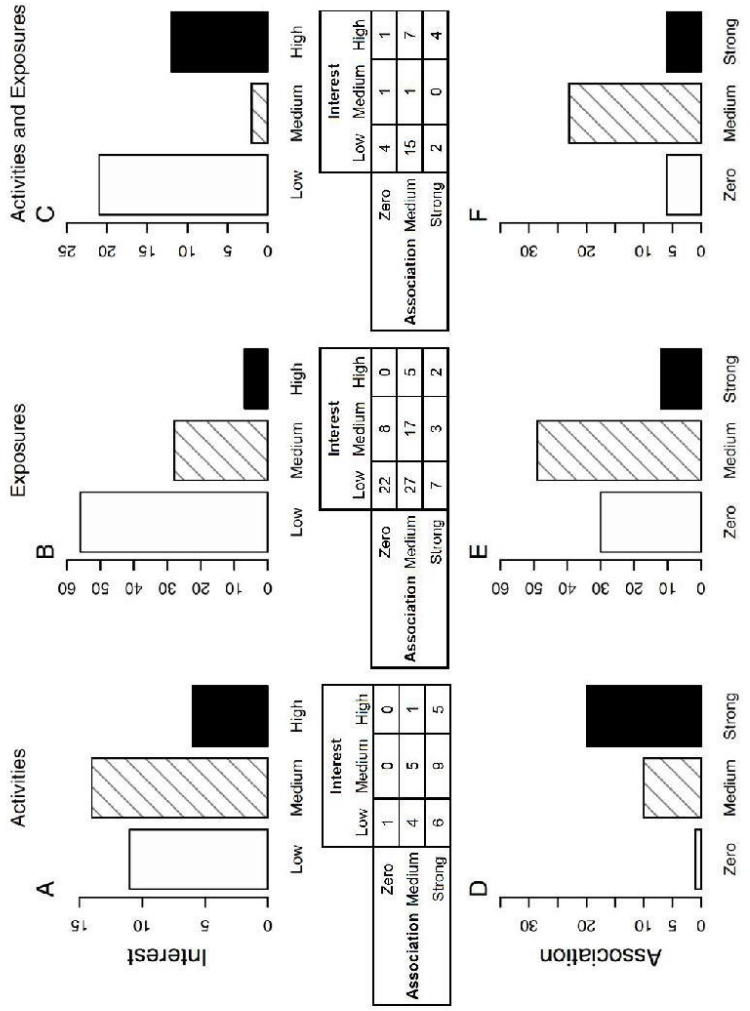


Figure 2: Distributions of score of association and interest

Topline: distribution of degree of interest of Activities (A), Occupational exposures (B) and joint activity-exposure (C). Line below: distribution of strength of association of Activities (D), Occupational exposures (E) and Combinations of both (F). Crossed effective by score is available for each class of occupational context in related table in the middle.

9.3.2. Comparaison des résultats

Brièvement, l'article avait pour premier objectif de **lister à partir de la littérature tous les contextes professionnels potentiellement associés au LNH**, que ce soit des activités professionnelles, des expositions professionnelles, ou bien l'association des deux. Le deuxième objectif était de pouvoir c**lasser ces contextes professionnels selon deux scores**, un premier appelé **score d'intérêt**, permettant de caractériser l'engouement scientifique d'un contexte professionnel par rapport aux autres ; puis un **score d'association**, qui, à partir de la distribution des risques listés dans la littérature, permettait d'attribuer une force d'association entre le contexte professionnel et le LNH. Afin de permettre une standardisation des libellés de la littérature ainsi que la comparaison avec le RNV3P, il a été fait le choix d'utiliser les mêmes thésaurus que le RNV3P lors de la méthodologie de codification et standardisation des données récupérées.

Il a ainsi été mis en évidence dans ce travail **20 secteurs d'activité fortement associés au LNH**, dont 6 sont fréquemment étudiés dans la littérature : « Industries alimentaires et agricoles », « Agriculture, chasse et sylviculture », « Industries manufacturières », « Métallurgie et travail des métaux », « Production et distribution d'électricité, de gaz et d'eau » et « Immobilier, location et services aux entreprises ». Concernant les expositions, seules les **« Acides carboxyliques » et les « solvants organiques et diluants »** ont été mis en évidence avec un fort intérêt et une force d'association importante. Mais certaines expositions **peu étudiées furent mises en évidence par un fort risque d'association,** telles que les « produits noirs », « hétérocycles », « métaux non spécifiés », « oxyde de potassium », « radiations non ionisantes » et « radiations ».

Une majorité des expositions identifiés depuis la littérature ont été retrouvés dans les résultats de la Surveillance Observationnelle appliquée au RNV3P, telles que par exemple les solvants, diluants, notamment le benzène et le trichloroéthylène, l'amiante, les radiations ionisantes, et les produits pesticides, notamment les insecticides et herbicides. Pour le cas des expositions aux solvants et diluants, benzène et trichloroéthylène, largement étudiées dans la littérature, les motifs ont été mis en évidence par leur poids et leur recrutement persistant, mais également par leur importance dans la construction du spectrosome. Cependant, concernant les insecticides et les herbicides, également très étudiés dans la littérature, les motifs issus de la Surveillance Observationnelle ont révélé leur importance au niveau de la complexification du spectrosome et moins au niveau de leur recrutement.

Il est également à noter que certains motifs d'expositions professionnelles n'ont pas été mis en évidence comme étant fortement étudié et associé au LNH par la littérature. C'est le cas du tétrachloroéthylène ou le 1.3-butadiène, pourtant classé cancérigène certain pour l'homme par le CIRC en 2012, mais dont l'étude dans la littérature se fait majoritairement par l'intermédiaire de l'étude d'autres matrices pour lesquels le 1.3-butadiène est retrouvé en tant que composant.

Beaucoup d'expositions issus de la littérature étaient caractérisés par leur molécule propre tandis qu'il s'agissait dans le RNV3P majoritairement de famille d'expositions, notamment les solvants et les produits pesticides. Cette différence est facilement explicable par la différence de recueil d'information entre les études prospectives de la littérature et l'enregistrement observationnel au sein de la base de données du RNV3P. Alors que les études de cohortes et de cas témoins ont potentiellement accès aux différentes expositions et mesures d'expositions des travailleurs, les médecins du RNV3P quant à eux sont souvent dans l'obligation de se baser sur les souvenirs des patients, ce qui peut conduire à des généralisations des nuisances en famille dans le but de ne pas perdre d'information malgré le manque de précision.

Cependant, il est apparu que les résultats de la Surveillance Observationnelle appliquée au réseau RNV3P était cohérente avec la littérature et a également permis de mettre en évidence des motifs d'expositions professionnelles, notamment des associations dans des contextes professionnels spécifiques, telle que les solvants et l'amiante par exemple, qui à ce jour n'avait jamais été mentionné **comme association** propre dans la littérature.

RESUME DES APPLICATIONS SUR LE RNV3P

	Cancer de la vessie	Asthme	LNH
	Forme stable à latence longue, chronique Exposition bien connues	Pathologie aigue à latence courte Expositions bien connues	Complexe et variée à distribution de latence (moyenne à longue), chronique, multi-exposée, Expositions mal connues
RNV3P			
Code	C65*, C66*, C67* et D090	J45* sans J451 et J458	C82*, C83* et C85*
Cas RNV3P	**1 006**	**5 339**	**288**
Nuisances distinctes	331	714	178
Nuisances significatives	**37 (11,18%)**	**236 (33,05%)**	**15 (8,43%)**
Cas conservés	880 (87,48%)	4 959 (92,88%)	200 (69,44%)
Exposome			
Nombre de nœuds	477	1 544	110
Nombre de liens	31 059	63 628	1 798
Motifs			
Nombre de motifs	**166**	**436**	**40**
Ordre I	54 (32,53%)	196 (44,95%)	21 (52,50%)
Ordre II	93	216	15
Ordre III	19	24	4
Ordre IV	0	0	1
Dynamique			
Motifs persistants	26	36	8
Motifs en déclin	8	15	6
Motifs en émergence	7	23	1
Motifs en latence	125	362	25
Motifs en latence actif en 2013	*10*	*13*	*3*
Nombre de motifs actifs	**51 (30,72%)**	**87 (19,86%)**	**18 (45%)**
Caractéristiques du spectre	Près de 70% des motifs en multi-expositions concernant les HAP, l'amiante et les solvants	Un très grand nombre d'expositions simples recensées, formant des associations "attendues" de type poussière organique, acarien, solvants, détergents, bactéricide	Presque 50% de multi-exposition majoritairement de type solvants ou pesticides. Un nombre important de motif ayant une information localisée

Contrairement au système habituel de Surveillance, la méthodologie peut donc être appliquée à différentes pathologies, même si celles-ci possèdent des caractéristiques très différentes, en termes de latence ou d'étiologies. La liste complète des expositions identifiées pour le RNV3P et les 3 pathologies est disponible en Annexe 4.

Partie 4 : Applications à la base belge IDEWE

L'objectif de cette partie est de présenter la généralisation et la portabilité de la méthodologie de la Surveillance Observationnelle à une autre base de données que le RNV3P. Toujours dans un contexte de surveillance de la santé au travail il a été choisi de travailler en collaboration avec des médecins du travail du groupe belge « IDEWE, optimizing work environments » afin d'étudier les pathologies professionnelles enregistrées au sein de leur base de données.

Ces exemples ont également l'intérêt de présenter des analyses sous un angle différent de celui du RNV3P. En effet, leurs problématiques étant différentes, il a été décidé d'étudier, en plus d'une pathologie professionnelle, deux secteurs d'activités particuliers afin de les caractériser en termes de pathologies professionnelles. Les exemples ainsi présentés dans cette partie sont :

- Les ***douleurs à la gorge*** : symptôme (assimilé à une pathologie) initialement choisi pour permettre une étude de faisabilité et ayant donné lieu à des ajustements nécessaire de la base de données permettant l'application de la Surveillance, justifiant ainsi l'intérêt de le présenter dans ces travaux ;
- Les ***aides-soignants*** : il a été effectué l'analyse du secteur d'activité des aides-soignants grâce à la Surveillance Observationnelle, permettant la caractérisation de ce secteur d'activité par des pathologies professionnelles caractéristiques potentiellement associées ;
- Les ***agriculteurs*** : il a également été fait l'analyse de l'activité des agriculteurs afin de les caractériser par les pathologies professionnelles caractéristiques potentiellement associées.

L'intérêt de présenter ces trois exemples est donc d'une part de pouvoir démontrer la **généralisation** possible de la Surveillance Observationnelle à d'autres bases de données que le RNV3P, mais également de démontrer la **flexibilité** de cette surveillance permettant de changer de variable cible à expliquer selon les problématiques posées.

10. PRESENTATION DU GROUPE ET DE LA BASE IDEWE

Le groupe « IDEWE optimizing work environments » est un groupe proposant des services d'aides aux entreprises concernant la surveillance médicale, la gestion des risques encourus par les travailleurs mais également la coordination de la sécurité, la sécurité du travail, l'environnement et l'hygiène industrielle au travail [IDEWE, 2015]. Le groupe possède 150 centres médicaux en Belgique et vient en aide à plus de 35 000 employeurs et 600 000 travailleurs en Belgique pour leur permettre de travailler dans un environnement de travail sûr et sain.

A l'issu de ces prestations, chaque année et pour chaque employé le groupe dispose des informations concernant les pathologies, le secteur d'activité professionnelle et les expositions professionnelles pouvant être rencontrées dans ce secteur d'activité. Il en résulte alors une base de données complète de 67 variables, pour laquelle chaque ligne correspond à un enregistrement (observation). La particularité de cette base de données est que toutes les expositions recensées représentent uniquement des hypothèses émises en rapport avec l'activité professionnelle et ne propose aucune relation de causalité avec la pathologie, permettant ainsi une étude exploratoire avec la Surveillance Observationnelle.

Il s'agit d'une base de données récente, créé en 2010 et contenant déjà près de 400 000 observations.

10.1. Restructuration pour l'application de la Surveillance Observationnelle

Afin de pouvoir appliquer la méthode de la Surveillance Observationnelle à cette base de données, il a été nécessaire de la restructurer car elle était difficilement exploitable en l'état pour une quelconque surveillance. En effet, de par les objectifs du groupe et leur mode de recueil des données, chaque année un nouvel enregistrement est fait pour chaque employé, contenant l'ensemble des pathologies qui ont été constatées sur l'année et l'ensemble des expositions ayant pu être rencontrées sur le lieu de travail, conduisant à l'existence de plusieurs lignes pour un même employé et à des doublons ou des disparitions d'informations dans le cas où les pathologies et expositions n'auraient pas été retrouvées (Figure 55A). De même, la multitude d'information concernant les pathologies et les expositions dans une

même observation rendait difficile une mise en relation causale entre un contexte professionnel et une pathologie.

A

	ID	JAAR	ZIEKTE_TYPE_CODE_LEV7.1	ZIEKTE_TYPE_CODE_LEV7.2	ZIEKTE_TYPE_CODE_LEV7.3	ZIEKTE_TYPE_CODE_LEV7.4	ZIEKTE_TYPE_COD
1	1	2010	780.6	NA	NA	NA	NA
2	1	2011	780.6	NA	NA	NA	NA
3	1	2013	780.6	NA	NA	NA	NA
4	154	2012	719.41	NA	NA	NA	NA
5	166	2010	460	525.9	724.2	NA	NA
6	166	2011	460	525.9	724.2	NA	NA
7	166	2012	460	525.9	724.2	NA	NA
8	166	2013	460	724.2	NA	NA	NA
9	265	2013	727.03	780.2	NA	NA	NA
10	280	2011	460	NA	NA	NA	NA
11	280	2012	460	NA	NA	NA	NA

B

	ID	JAAR	ZIEKTE_TYPE_CODE_LEV7.1	ZIEKTE_TYPE_CODE_LEV7.2	ZIEKTE_TYPE_CODE_LEV7.3	ZIEKTE_TYPE_CODE_LEV7.4	ZIEKTE_TYPE_CODE
1	1	2010	780.6	NA	NA	NA	NA
2	154	2012	719.41	NA	NA	NA	NA
3	166	2010	460	525.9	724.2	NA	NA
4	265	2013	727.03	780.2	NA	NA	NA
5	280	2011	460	NA	NA	NA	NA

Figure 55 : Illustration de la structure initiale et restructurée de la base belge IDEWE. A) Base initiale, illustration de deux patients enregistrés au sein de la base durant 3 et 4 années, avec redondance de l'information pour le patient numéro 1 (rouge) et redondance et perte de l'information pour le patient 166 (bleu). B) Restructuration des données, illustration des deux patients après fusion des informations.

Il a été proposé lors de la collaboration de restructurer la base de données afin de n'obtenir qu'une seule ligne d'observation par employé, contenant la première année d'enregistrement et la conservation de toutes les pathologies et les expositions enregistrées pour ce patient (Figure 55B).

Cette restructuration a permis de palier au problème de perte d'information au fil des années ainsi qu'au problème de doublon, cependant la relation de causalité entre le contexte professionnel et la pathologie restait difficile à établir. Dans le cas où une variable d'imputabilité avait été présente, cette relation causale aurait été plus envisageable. C'est pour cette raison qu'il a été proposé la solution d'ajouter une variable d'imputabilité externe *a posteriori* à chacune des variables associées rencontrées en fonction de sa pertinence et de sa plausibilité par rapport à la variable cible étudiée. Il a alors été nécessaire d'identifier, pour chacun des exemples d'évènement de santé traité, chaque modalité explicative enregistrée afin de leur attribuer une valeur d'imputabilité allant de « -1 » à « 3 », la valeur « -1 » servant à exclure la modalité si non pertinente avec la variable cible étudiée et probablement en relation avec une autre des modalités présente dans le cortège, la valeur 0 pour « pas de relation » jusqu'à « 3 » forte relation entre la modalité et la variable cible. Cette attribution d'imputabilité a été demandée auprès de plusieurs médecins afin de simuler une variabilité

que l'on pourrait rencontrer dans des bases de données puis d'appliquer le test de significativité adéquat des modalités.

Par exemple :

Dans le cas d'une surveillance d'une dermatite de contact (la variable cible), toutes les expositions professionnelles enregistrées pour les employés ayant eu une dermatite de contact vont être extraites (variables explicatives), puis traitée une à une afin d'exclure des expositions n'ayant aucun rapport possible avec une dermatite de contact.

10.2. Déploiement de la Surveillance Observationnelle pour IDEWE

Dans le cadre de l'application de la Surveillance Observationnelle à la base IDEWE et aux 3 exemples choisis, la transposition des termes est la suivante :

- **Variable cible** : variable renseignant la **pathologie ou symptôme** diagnostiqué (douleur de la gorge) ou variable renseignant l'**activité professionnelle** (aides-soignants, agriculteurs) ;
- **Variables explicatives associées** : les variables renseignant les **expositions professionnelles** suspectées (douleurs de la gorge) ou les variables renseignant les **pathologies professionnelles** (aides-soignants, agriculteurs) ;
- **Imputabilité :** Variable renseignant pour chaque exposition l'imputabilité potentielle entre la pathologie et l'exposition (douleurs de la gorge) ou entre l'activité et la pathologie (aides-soignants, agriculteurs), attribuée par les médecins experts.

Les nœuds et motifs seront alors formés par des cortèges d'expositions professionnelles ou des pathologies professionnelles et serviront à la caractérisation de la pathologie professionnelle ou des activités professionnelles étudiées.

10.2.1. Prise en compte de la variable d'imputabilité

La significativité de l'ensemble des modalités *H* a été évaluée grâce au calcul basé sur l'imputabilité *I (voir équation 3.3, p. 54)*, en utilisant l'imputabilité proposée *a posteriori* par les spécialistes, allant de 0 à 3, et en utilisant la valeur « -1 » pour éliminer les modalités non plausibles cliniquement. L'imputabilité numérique a été obtenue de la même manière que pour le RNV3P (*voir 6.3.2, p.116*).

Tableau 8 : Exemple d'une distribution d'imputabilité externe, allant de 0 à 3, renseignée par *n* spécialiste.

Modalité	Freq*	Spec 1**	Spec 2	Spec 3	…	Spec *n*
1	10	0	3	2	2	3
…	…	…	…	…	…	…
H	…	…	…	…	…	…

**Freq : Fréquence. **Spec : Spécialiste*

Deux procédures de récupération des valeurs d'imputabilité fournies par les spécialistes ont été développées. La première est basée sur **le nombre d'occurrences n_i** de chaque valeur d'imputabilité pour une modalité, tandis que la deuxième est basée sur la **génération d'une distribution de taille n_i** de valeurs d'imputabilité à partir de la distribution initiale proposée par les spécialistes.

10.2.1.1. Utilisation du nombre d'occurrences des imputabilités

Pour chaque modalité *h* seule la valeur d'imputabilité *I* ayant le nombre d'occurrences le plus élevé est conservée pour l'analyse. En d'autres termes, la valeur conservée correspond à celle plébiscitée par les spécialistes ayant apporté leur opinion sur la relation entre la modalité regardée et la variable cible. Dans le cas d'une égalité, il est possible de choisir entre un choix dit « ***optimiste*** », privilégiant un faible « risque », et donc une imputabilité faible, ou bien un choix « ***pessimiste*** », privilégiant un « risque » fort et une imputabilité forte.

Par exemple :

Dans l'exemple du Tableau 1 et de la modalité 1, la valeur de l'imputabilité conservée sera 3 dans un cas « pessimiste » privilégiant un risque fort ; et 2 dans un cas « optimiste » privilégiant un risque plus faible. Le calcul de l'imputabilité sera alors appliqué sur une distribution telle que :

	Imputabilité			
	0	**1**	**2**	**3**
Optimiste	0	0	10	0
Pessimiste	0	0	0	10

10.2.1.2. Génération d'une distribution d'imputabilité

Pour chaque modalité *h* la totalité de la distribution d'imputabilité est considérée afin de générer une nouvelle distribution de n_i tirages.

Par exemple :

Dans l'exemple du Tableau 1 et de la modalité 1, pour un nombre d'occurrences $n_i = 10$, le nombre d'occurrences pourrait être réparti de la manière suivante :

	Imputabilité			
	0	**1**	**2**	**3**
Distribution	2	0	4	4

Cette récupération a la particularité de faire légèrement varier le nombre de modalités significatives conservées dû à la génération aléatoire de la distribution. Les modalités étant « limites » peuvent alors être significatives ou non d'une analyse à l'autre et faire varier le nombre de nœuds et de motifs générés. Il s'agit majoritairement de motifs « inactifs », faisant partie de l'ensemble des motifs latents, et n'a donc pas d'impact sur la mise en évidence des motifs représentatifs de l'évènement d'intérêt. Cependant, il est conseillé de n'utiliser cette récupération qu'en cas de très faible information fournie par les spécialistes afin d'apporter un peu plus de variabilité au sein de la distribution.

10.2.2. Détermination des paramètres pour IDEWE

Les paramètres ont été établis afin de tenir compte de la jeunesse de la base mais également de la structure particulière des informations au sein d'une même observation (Tableau 9). Ainsi la longueur choisie pour les nœuds et les motifs a été fixée à $d = 3$ afin de limiter les « faux liens » entre les modalités explicatives et la variable cible. L'intervalle temporel Δt a quant à lui était fixé à 1 an dû à la jeunesse de la base. Le nombre d'occurrences de référence critique sous imputabilité minimale a été augmenté à $n_{\Delta t,cr} = 5$ par rapport au RNV3P pour un $\lambda_{min} = 1.98$ afin de fournir une restriction plus importante de la sélection des modalités.

Tableau 9 : Variables et paramètres considérés pour l'application de la Surveillance Observationnelle à la base IDEWE.

Exemple	Variable cible	Variable associées
Douleur de la gorge	Pathologie/Symptôme	Expositions professionnelles
Aide-soignant	Secteur d'activité	Pathologies/Symptômes
Agriculteur		
Paramètres		
Significativité		
Intervalle temporel Δt	1 an	
Nombre de recrutement de référence nécessaire sous imputabilité nulle ($n_{i,\Delta t,cr}$)	5 ($\lambda_{min} = 1.98$)	
Nombre de recrutement de référence nécessaire sous imputabilité maximale ($n_{i,\Delta t,cr}$)	1 ($\lambda_{max} = 0.05$)	
Paramètre ν	0.5	
Longueur de nœuds *d* maximale	3	
Longueur de motifs *d* maximale	3	

11. DOULEUR A LA GORGE

Ce symptôme a été choisi pour permettre une étude de faisabilité de la transposition de la Surveillance à la base IDEWE, en utilisant le même type de variables cible et explicatives primaires que pour le RNV3P : pathologies et expositions professionnelles. Le symptôme sera alors assimilé à une pathologie.

11.1. Analyse

En 2013, au sein des 183 614 cas de la base IDEWE, 520 concernaient les douleurs à la gorge (Figure 56). Un total de 191 nuisances distinctes a été recensé pour ces cas. La liste a été soumise à un panel de 5 médecins afin d'attribuer une valeur d'imputabilité *a posteriori* à chacune de ces nuisances ou de les écarter si non plausible cliniquement avec les douleurs de la gorge.

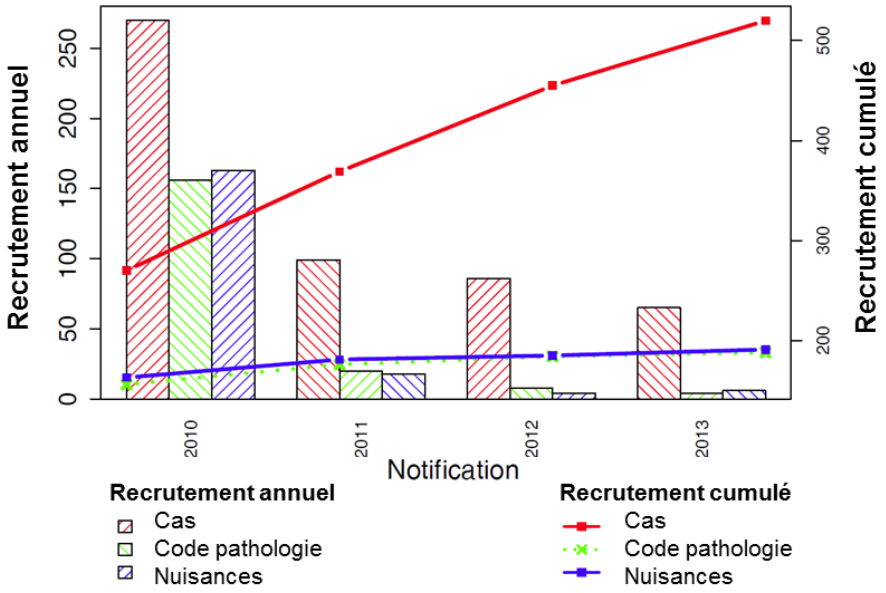


Figure 56 : Recrutement et effectif cumulé des cas, des nuisances distinctes et des codes pathologies distinctes correspondant aux douleurs à la gorge et aux autres pathologies enregistrées au sein des mêmes observations de 2010 à 2013.

Un total de 95 nuisances a été conservé afin d'appliquer le calcul de la significativité par l'imputabilité. La méthode d'attribution de l'imputabilité choisie pour ce cas a été l'attribution de la valeur d'imputabilité plébiscitée par les médecins, possédant le plus grand nombre d'occurrences (*voir 10.2.1.1, p.200*). Dans le Tableau 10 sont présentés les résultats selon une attribution pessimiste, privilégiant les fortes valeurs d'imputabilité en cas d'égalité, et une attribution optimiste, privilégiant une valeur faible d'imputabilité en cas d'égalité.

Dans la version pessimiste de l'analyse, 20 nuisances ont été considérées comme significatives (77%), retrouvées dans 400 cas, permettant la formation de 193 nœuds et 105 motifs ; contre 12 dans la version optimiste de l'analyse, retrouvées dans 375 cas (72%), permettant la formation de 169 nœuds et 78 motifs.

Tableau 10 : Synthèse des résultats de la Surveillance Observationnelle suivant la méthode pessimiste ou optimiste de l'attribution de la valeur d'imputabilité *a posteriori.*

	Pessimiste	**Optimiste**
Cas	520	520
Modalités	191	191
Modalités significatives	**20**	**12**
Cas conservés	400	375
Nœuds	193	169
Motifs	**105**	**78**
A	17	16
B	0	0
C	14	13
D	74	49
D actifs en 2012	*7 actifs*	*1 actif*
Motifs actifs	**38**	**30**

Les exposomes obtenus pour chacune des deux analyses sont disponibles dans la Figure 57. La couleur des liens correspond aux familles des expositions professionnelles au 2ème niveau d'imbrication du thésaurus de la base belge fourni avec les données.

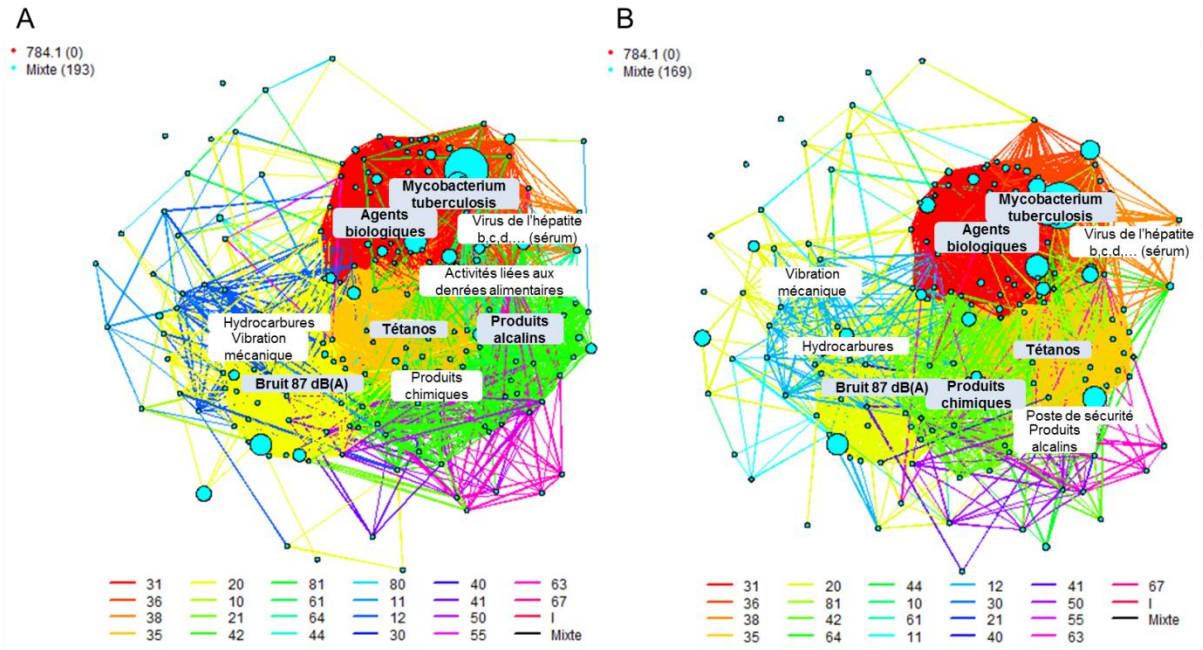


Figure 57 : Exposomes des douleurs de la gorge en 2013. A) Exposome suite à une analyse pessimiste, 193 nœuds, dont 2 isolés, et 4 969 liens. Les 5 clusters majoritaires sont les « agents biologiques », « mycobacterium tuberculosis », « tétanos », « produits alcalins » et « bruit 87dB(A) ». B) Exposome suite à une analyse optimiste, 169 nœuds, dont 2 isolés, et 3 854 liens. Les 5 clusters majoritaires sont les « agents biologiques », » mycobacterium tuberculosis », « tétanos », « produits chimiques » et « bruit 87dB(A) ».

Pour la version pessimiste, 193 nœuds ont été identifiés, dont 2 isolés, connectés par un total de 4 969 liens ; la version optimiste de l'exposome était formée de 169 nœuds, dont 2 isolés, et 3 854 liens. Cinq clusters majoritaires ont été identifiés en gras pour chacun des exposomes, et 4 d'entre eux étaient communs : « agents biologiques », « mycobacterium tuberculosis », « tétanos » et « bruit 87 dB(A) ». Pour la version pessimiste, le cluster des « produits chimiques anorganiques » a également été mis en évidence, tandis que dans la version optimiste, c'est le cluster des « produits chimiques » en général qui a été mis en évidence.

Un total de 105 et 78 motifs ont été identifiés pour les douleurs à la gorge selon la version pessimiste ou optimiste de l'analyse. La liste des différents motifs d'expositions identifiés est disponible dans le Tableau 11, chaque motif étant identifié par un « P » si identifié dans la version pessimiste et « O » dans la version optimiste. Pour la suite des résultats, seuls les **résultats de la version « Pessimiste »** du calcul sera présentée puisqu'elle englobe également les motifs générés par la version « Optimiste ». Parmi ces 105 motifs, 35 correspondaient à des motifs d'ordre I (une seule exposition professionnelle au sein du cortège), impliquant un total de 70 motifs d'ordre supérieur à I, correspondant à des combinaisons de motifs d'ordre I : 48 motifs d'ordre II et 22 motifs d'ordre III.

Tableau 11 : Liste des 105 motifs identifiés par la Surveillance Observationnelle avec calcul « Pessimiste » (P) de l'imputabilité appliquée aux douleurs de la gorge, incluant les 78 motifs obtenus avec la version « Optimiste » (O) du calcul de l'imputabilité. Les libellés des motifs ont été fournis par les acteurs du réseau IDEWE.

ID	Ordre I	Ordre II	Ordre III	Nœuds	Poids	Statut	I
1	agents biologiques (activité non délibérée)	-	-	58	219	A	P et O
2	mycobacterium tuberculosis	-	-	36	194	A	P et O
3	agents biologiques (activité non délibérée)	mycobacterium tuberculosis	-	25	172	A	P et O
4	virus de l'hépatite b,c,d,...(sérum)	-	-	15	140	A	P et O
5	mycobacterium tuberculosis	virus de l'hépatite b,c,d,...(sérum)	-	8	130	A	P et O
6	agents biologiques (activité non délibérée)	virus de l'hépatite b,c,d,...(sérum)	-	5	122	A	P et O
7	agents biologiques (activité non délibérée)	mycobacterium tuberculosis	virus de l'hépatite b,c,d,...(sérum)	1	116	A	P et O
8	tétanos	-	-	50	83	A	P et O
9	bruit 87 dB(a)	-	-	40	61	A	P et O
10	produits alcalins	-	-	28	45	A	P et O
11	agents biologiques (activité non délibérée)	tétanos	-	15	31	A	P et O
12	produits chimiques	-	-	25	26	A	P et O
13	virus de l'hépatite a	-	-	6	20	A	P et O
14	tétanos	mycobacterium tuberculosis	-	3	19	A	P et O
15	vibrations mécaniques	-	-	13	17	A	P et O
16	hydrocarbures	-	-	15	15	A	P et O
17	mycobacterium tuberculosis	virus de l'hépatite a	-	2	15	A	P et O
18	bruit 87 dB(a)	tétanos	-	9	14	C	P et O
19	activité liée aux denrées alimentaires	-	-	12	13	C	P et O
20	manutention manuelle de charges	-	-	11	13	C	P et O
21	agents biologiques (activité non délibérée)	virus de l'hépatite a	-	3	13	C	P et O
22	agents biologiques (activité non délibérée)	produits alcalins	-	5	12	C	P et O
23	tétanos	produits alcalins	-	7	11	C	P
24	agents biologiques (activité non délibérée)	tétanos	mycobacterium tuberculosis	1	11	D	P et O
25	poste de sécurité	-	-	10	10	C	P et O
26	travail sur écran de visualisation	-	-	8	10	C	P et O
27	bruit 87 dB(a)	vibrations mécaniques	-	7	10	C	P et O
28	agents biologiques (activité non délibérée)	mycobacterium tuberculosis	virus de l'hépatite a	1	10	C	P et O

29	**coupure avec des objets tranchants**	**-**	**-**	**5**	**9**	**C**	**P et O**
30	**rayonnements ultraviolets (général)**	**-**	**-**	**6**	**9**	**D**	**P**
31	**risque d'affection pulmonaire**	**-**	**-**	**9**	**9**	**D**	**P et O**
32	bruit 87 dB(a)	agents biologiques (activité non délibérée)	-	4	9	D	P et O
33	mycobacterium tuberculosis	produits alcalins	-	2	8	D	P
34	radiations ionisantes (général)	-	-	5	7	D	P et O
35	**hydrocarbures**	**bruit 87 dB(a)**	**-**	**7**	**7**	**C**	**P et O**
36	tétanos	virus de l'hépatite a	-	3	7	D	P et O
37	virus de l'hépatite b,c,d,...(sérum)	produits alcalins	-	3	7	D	P
38	**produits alcalins**	**manutention manuelle de charges**	**-**	**5**	**7**	**C**	**P et O**
39	**chrome**	**-**	**-**	**3**	**6**	**D**	**P**
40	**chaleur**	**-**	**-**	**6**	**6**	**C**	**P et O**
41	suie, goudron, ...	-	-	6	6	D	P
42	microtraumatismes	-	-	5	6	D	P et O
43	**bruit 87 dB(a)**	**rayonnements ultraviolets (général)**	**-**	**3**	**6**	**D**	**P**
44	mycobacterium tuberculosis	travail sur écran de visualisation	-	4	6	D	P et O
45	risque d'intoxication	-	-	5	5	D	P et O
46	tétanos	virus de l'hépatite b,c,d,...(sérum)	-	3	5	D	P et O
47	bruit 87 dB(a)	produits chimiques	-	5	5	D	P et O
48	agents biologiques (activité non délibérée)	produits chimiques	-	4	5	D	P et O
49	tétanos	produits chimiques	-	5	5	D	P et O
50	produits alcalins	activité liée aux denrées alimentaires	-	5	5	D	P
51	agents biologiques (activité non délibérée)	travail sur écran de visualisation	-	3	5	D	P et O
52	tétanos	mycobacterium tuberculosis	virus de l'hépatite a	1	5	D	P et O
53	coupure avec des objets tranchants	-	-	4	4	D	P et O
54	nuisances causées par contamination de la peau	-	-	3	4	D	P et O
55	monoxyde de carbone	-	-	4	4	D	P
56	plomb	-	-	4	4	D	P et O
57	chlore et ses composes	-	-	4	4	D	P et O
58	dérivés halogénés	-	-	4	4	D	P et O
59	température abaissée	-	-	4	4	D	P
60	contact léger avec agent biologique	-	-	4	4	D	P
61	**chrome**	**bruit 87 dB(a)**	**-**	**1**	**4**	**D**	**P**
62	**chrome**	**rayonnements ultraviolets (général)**	**-**	**1**	**4**	**D**	**P**
63	radiations ionisantes (général)	agents biologiques (activité non délibérée)	-	2	4	D	P et O
64	vibrations mécaniques	tétanos	-	3	4	D	P

65	produits chimiques	produits alcalins	-	4	4	D	P
66	tétanos	activité liée aux denrées alimentaires	-	4	4	D	P et O
67	tétanos	manutention manuelle de charges	-	4	4	D	P et O
68	**chrome**	**bruit 87 dB(a)**	**rayonnements ultraviolets (général)**	**1**	**4**	**D**	**P**
69	bruit 87 dB(a)	agents biologiques (activité non délibérée)	tétanos	1	4	D	P et O
70	agents biologiques (activité non délibérée)	mycobacterium tuberculosis	produits alcalins	1	4	D	P
71	mycobacterium tuberculosis	virus de l'hépatite b,c,d,...(sérum)	produits alcalins	1	4	D	P
72	mycobacterium spp groupe 2	-	-	2	3	D	P et O
73	agents biologiques (activité non délibérée)	mycobacterium spp groupe 2	-	2	3	D	P et O
74	radiations ionisantes (général)	mycobacterium tuberculosis	-	1	3	D	P et O
75	virus de l'hépatite a	virus de l'hépatite b,c,d,...(sérum)	-	2	3	D	P et O
76	mycobacterium tuberculosis	produits chimiques	-	2	3	D	P et O
77	bruit 87 dB(a)	produits alcalins	-	2	3	D	P
78	agents biologiques (activité non délibérée)	manutention manuelle de charges	-	2	3	D	P et O
79	radiations ionisantes (général)	agents biologiques (activité non délibérée)	mycobacterium tuberculosis	1	3	D	P et O
80	tétanos	mycobacterium tuberculosis	virus de l'hépatite b,c,d,...(sérum)	1	3	D	P et O
81	agents biologiques (activité non délibérée)	mycobacterium tuberculosis	travail sur écran de visualisation	1	3	D	P et O
82	faux mouvement en cours d'effort	-	-	1	2	D	P et O
83	contact léger avec agent biologique	-	-	1	2	D	P et O
84	virus	-	-	1	2	D	P et O
85	faux mouvement en cours d'effort	agents biologiques (activité non délibérée)	-	1	2	D	P et O
86	nuisances causées par contamination de la peau	agents biologiques (activité non délibérée)	-	1	2	D	P
87	contact léger avec agent biologique	agents biologiques (activité non délibérée)	-	1	2	D	P et O
88	nuisances causées par contamination de la peau	tétanos	-	1	2	D	P
89	faux mouvement en cours d'effort	mycobacterium tuberculosis	-	1	2	D	P et O
90	contact léger avec agent biologique	mycobacterium tuberculosis	-	1	2	D	P
91	mycobacterium spp groupe 2	mycobacterium tuberculosis	-	1	2	D	P et O
92	agents biologiques (activité non délibérée)	virus	-	1	2	D	P et O
93	mycobacterium tuberculosis	virus	-	1	2	D	P et O

94	virus de l'hépatite b,c,d,...(sérum)	activité liée aux denrées alimentaires	-	1	2	D	P et O
95	bruit 87 dB(a)	vibrations mécaniques	tétanos	1	2	D	P
96	nuisances causées par contamination de la peau	agents biologiques (activité non délibérée)	tétanos	1	2	D	P
97	faux mouvement en cours d'effort	agents biologiques (activité non délibérée)	mycobacterium tuberculosis	1	2	D	P et O
98	contact léger avec agent biologique	agents biologiques (activité non délibérée)	mycobacterium tuberculosis	1	2	D	P
99	agents biologiques (activité non délibérée)	mycobacterium spp groupe 2	mycobacterium tuberculosis	1	2	D	P et O
100	agents biologiques (activité non délibérée)	mycobacterium tuberculosis	virus	1	2	D	P et O
101	agents biologiques (activité non délibérée)	virus de l'hépatite a	virus de l'hépatite b,c,d,...(sérum)	1	2	D	P et O
102	agents biologiques (activité non délibérée)	mycobacterium tuberculosis	produits chimiques	1	2	D	P et O
103	bruit 87 dB(a)	tétanos	produits alcalins	1	2	D	P
104	agents biologiques (activité non délibérée)	virus de l'hépatite b,c,d,...(sérum)	produits alcalins	1	2	D	P
105	agents biologiques (activité non délibérée)	produits alcalins	manutention manuelle de charges	1	2	D	P et O

***En gras** sont identifiés les motifs dont le statut est considéré comme actif (A : persistant, B : en déclin, C : émergent) ainsi que 7 motifs latents (D) actifs en 2012.

Le statut dynamique des motifs a été caractérisé pour l'année 2013 (Figure 58A), 17 motifs ont été identifiés comme **persistants**, aucun en **déclin**, 14 en **émergence** et 74 en état de **latence**, dont 7 actifs en 2012 (id : 30-31, 39, 43, 61-62, 68). **Trente-huit motifs ont été caractérisés actifs** en 2013 et sont représentés dans l'arbre de la multi-association (Figure 58B), dont 19 d'ordre supérieur à I. Les motifs les plus importants rapportés comme associés aux douleurs de la gorge sont les « agents biologiques » (id : 1), recruté dans 219 cas, et en association avec 31 autres motifs ; suivi par « mycobacterium tuberculosis » (id : 2) dans 194 cas en association avec 26 autres motifs, ainsi que leur association (id : 3) avec 172 cas. Les autres motifs importants identifiés pour les douleurs de la gorge sont le « virus de l'hépatite b,c,d » (id : 4) avec 140 cas et l'association de ce motif avec « agents biologiques » (id : 5) ou « mycobacterium tuberculosis » (id : 6) dans 130 et 122 cas respectivement. Aucun motif en déclin n'a été identifié.

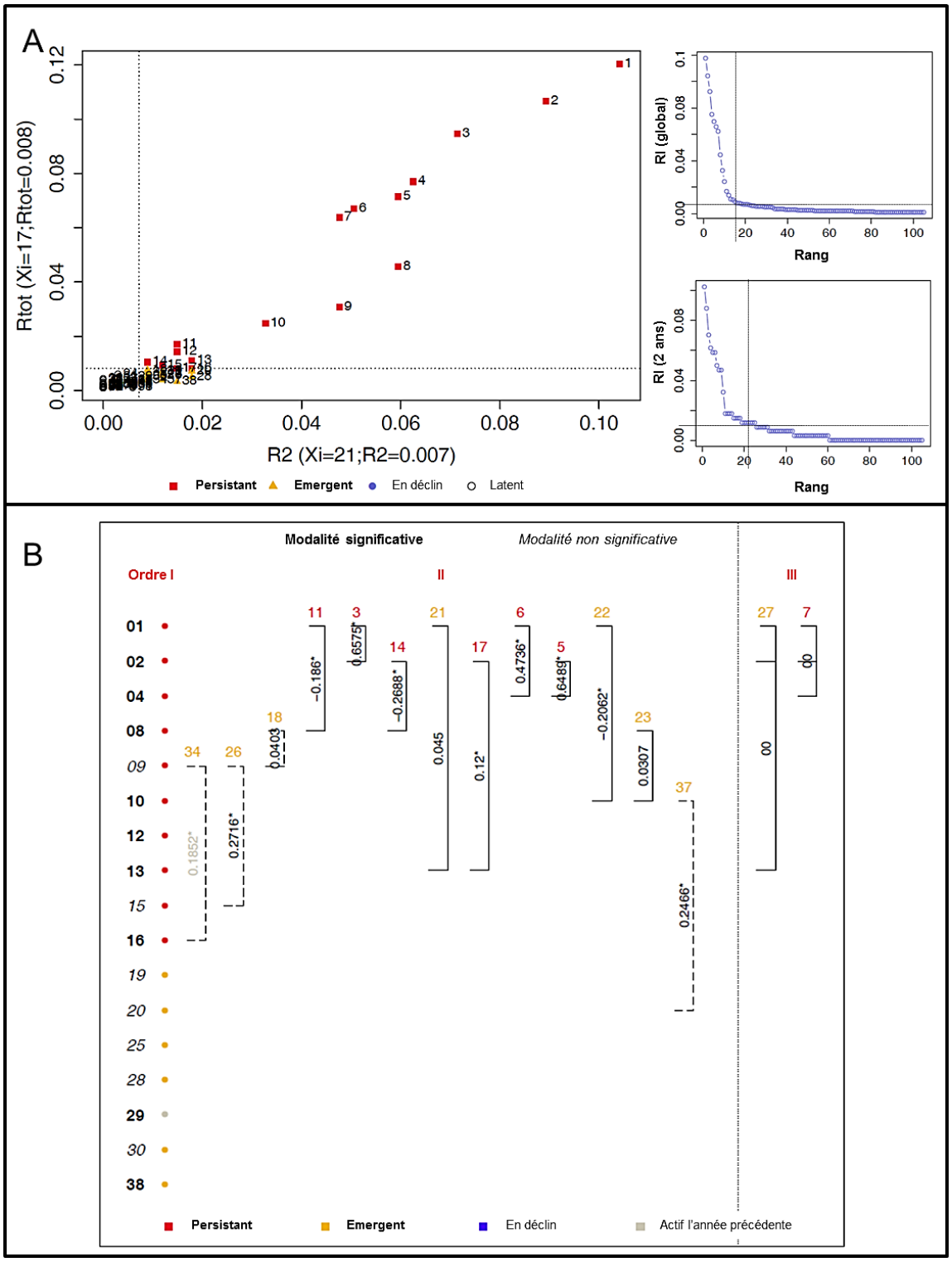


Figure 58 : Caractérisation des motifs issus de l'exposome des douleurs à la gorge en 2013. A) Statut dynamique des 105 motifs en 2013. Les délimitations des RI ont été calculées à 0.006 pour les deux années glissantes et 0.008 pour le RI global. Il y avait 17 motifs persistants, aucun en déclin, 14 en émergence et 74 en latence. B) Arbre de la multi-association des 38 motifs actifs liés aux douleurs à la gorge en 2013, avec 19 motifs d'ordre I actifs et 19 motifs d'ordre II actifs.

Le spectrosome des motifs d'expositions professionnelles rapportées comme potentiellement associées aux douleurs de la gorge est disponible en Figure 59. Les deux plus gros motifs d'ordre I en termes de poids, « agents biologiques » et « mycobacterium tuberculosis », font également partie des motifs formant les plus gros clusters. Viennent ensuite les motifs « produits alcalins », « produits chimiques » et « tétanos », rapportés dans moins de cas, mais impliqué dans de nombreux autres motifs d'ordre supérieur à I, formant ainsi des clusters importants.

Le spectre issu de la Surveillance Observationnelle pour les douleurs à la gorge est disponible en Figure 60. Douze des motifs actifs ont été caractérisés comme spécifiques avec une information tendant à être localisée comme le motif associant « agents biologiques », « mycobacterium tuberculosis » et « Virus de l'hépatite b,c,d,...(sérum) » (id : 7), ayant un poids de 116 pour un seul nœud enregistré ; tandis que 21 ont été caractérisés comme non localisés comme par exemple la « chaleur » (id : 40) retrouvée dans 10 nœuds différents, chacun ayant un poids égal à 1. Le symptôme ciblé des douleurs de la gorge est caractérisé par une tendance à la spécificité pour les motifs les plus importants. Ainsi, les 6 premiers motifs en termes de poids présentent tous une spécificité tendant vers 1.

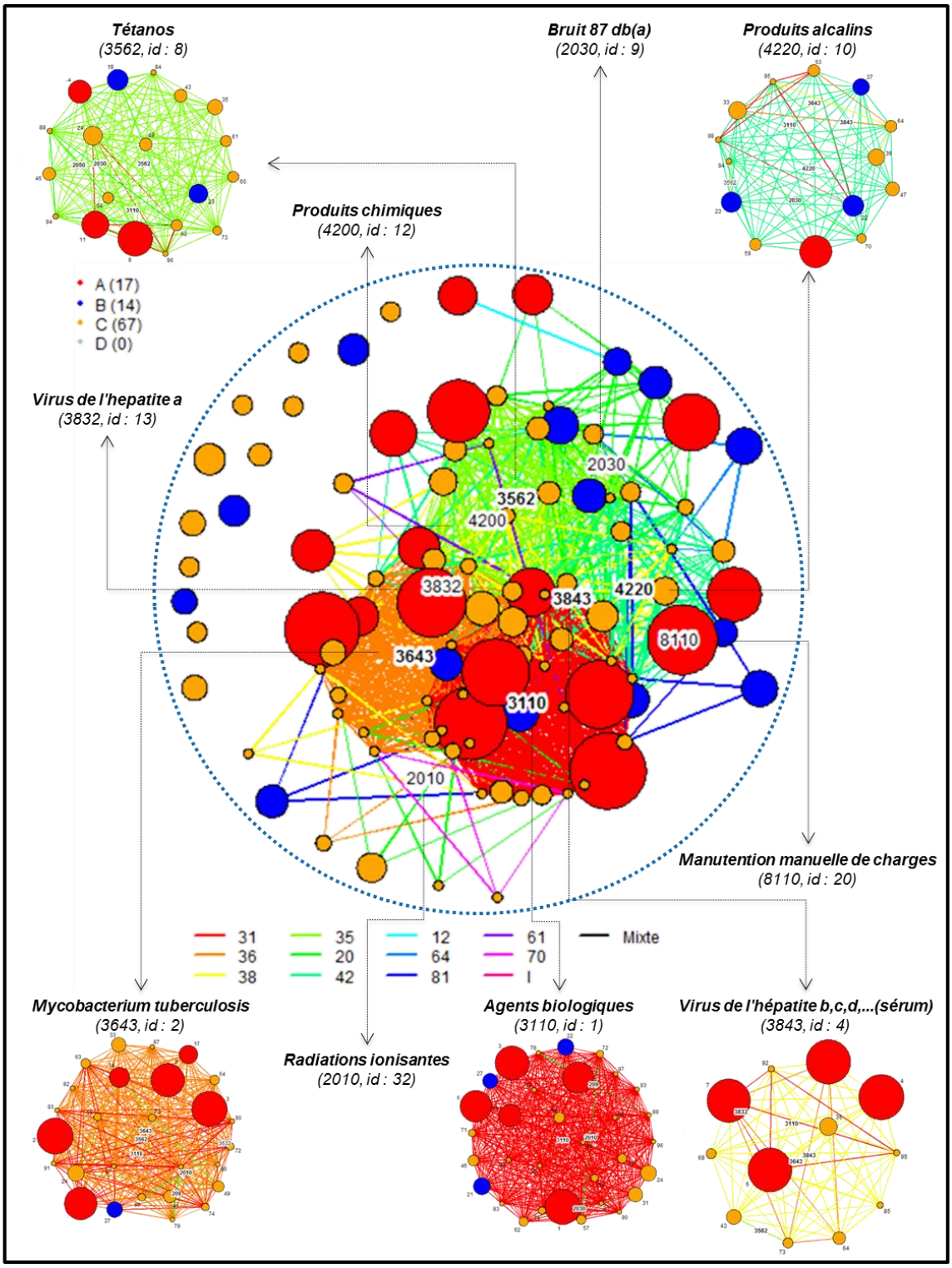


Figure 59 : Spectrosome des motifs des douleurs à la gorge en 2013. Au centre le spectrosome des motifs d'expositions professionnelles rapportés comme associés aux douleurs à la gorge. La taille des nœuds correspond au poids du motif et la couleur correspond au statut dynamique du motif. La couleur des liens correspond aux différentes familles de nuisances identifiées en relation avec les douleurs à la gorge. Cinq clusters ont été identifiés en raison du nombreux nombre de liens auxquels ils participent, de gauche à droite et de haut en bas : tétanos (8), bruit 87 dB(a) (9), produits alcalins (10), mycobacterium tuberculosis (2), agents biologiques (1) et virus de l'hépatite b,c,d…. (sérum) (4).

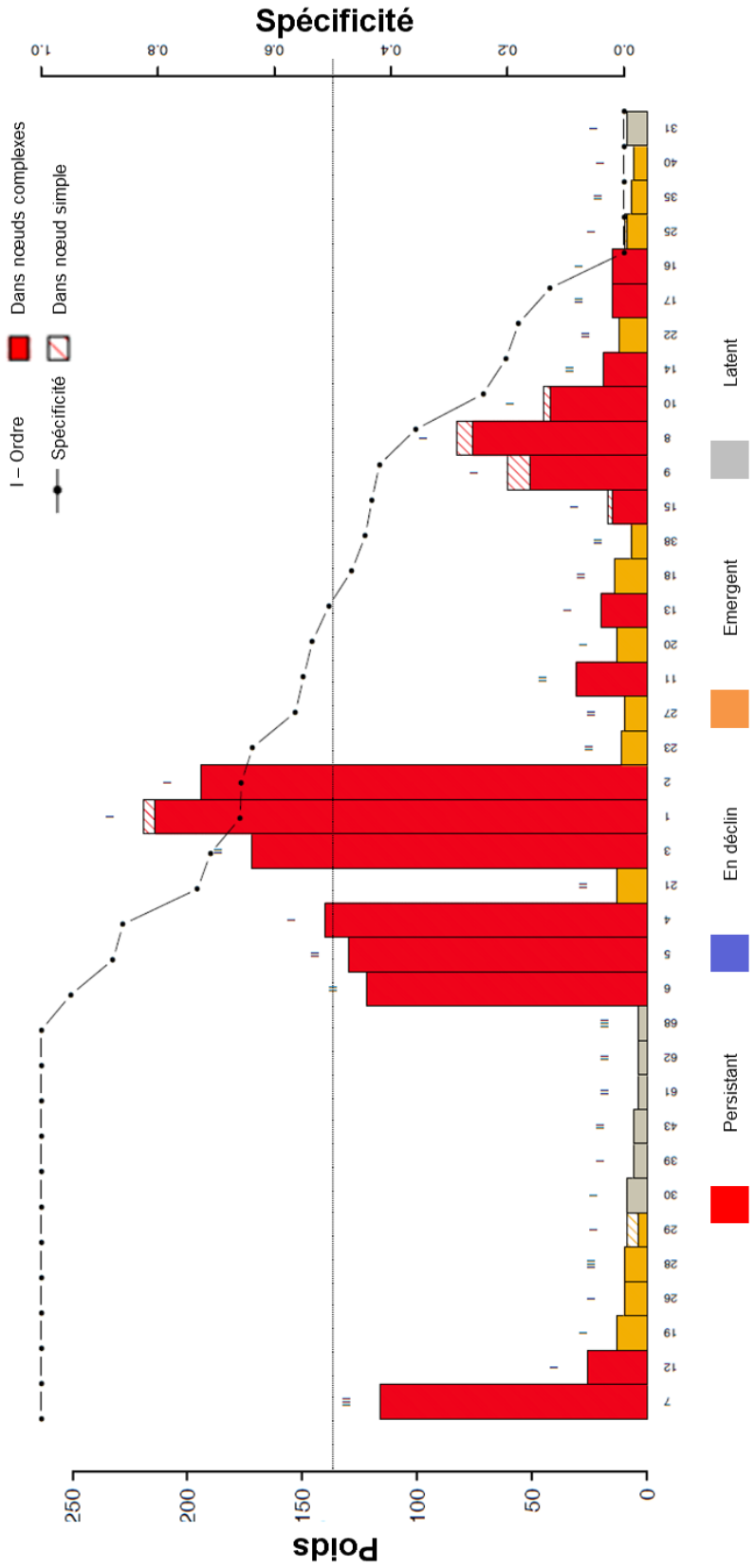


Figure 60 : Spectre des douleurs de la gorge en 2013. Le spectre des douleurs à la gorge a été caractérisé par 38 motifs actifs (dont 17 persistants, 0 en déclin, 14 en émergence et 7 latents mais actifs en 2012), parmi lesquels 19 étaient des motifs d'ordre I et 19 des motifs d'ordre supérieur. Douze motifs avaient une localisation de leur information spécifique c'est-à-dire privilégiant une association en particulier, 21 étaient ubiquitaires avec une information apportée par différents nœuds de manière homogène et les autres étaient des motifs intermédiaires.

Les trajectoires atypiques des douleurs à la gorge sont disponibles en Figure 61. Sont présentés les exemples des motifs « Agents biologiques » persistant depuis 2010 ; le motif « Bruit 87 dB(a) » et « tétanos » apparu en 2010 en latence et actuellement en émergence le nombre de recrutement n'étant pas suffisant pour une classification en persistant ; et le motif « Rayonnements ultraviolets (général) » actuellement en latence après une émergence en 2012. Il n'y avait aucun motif en phase de déclin en 2013.

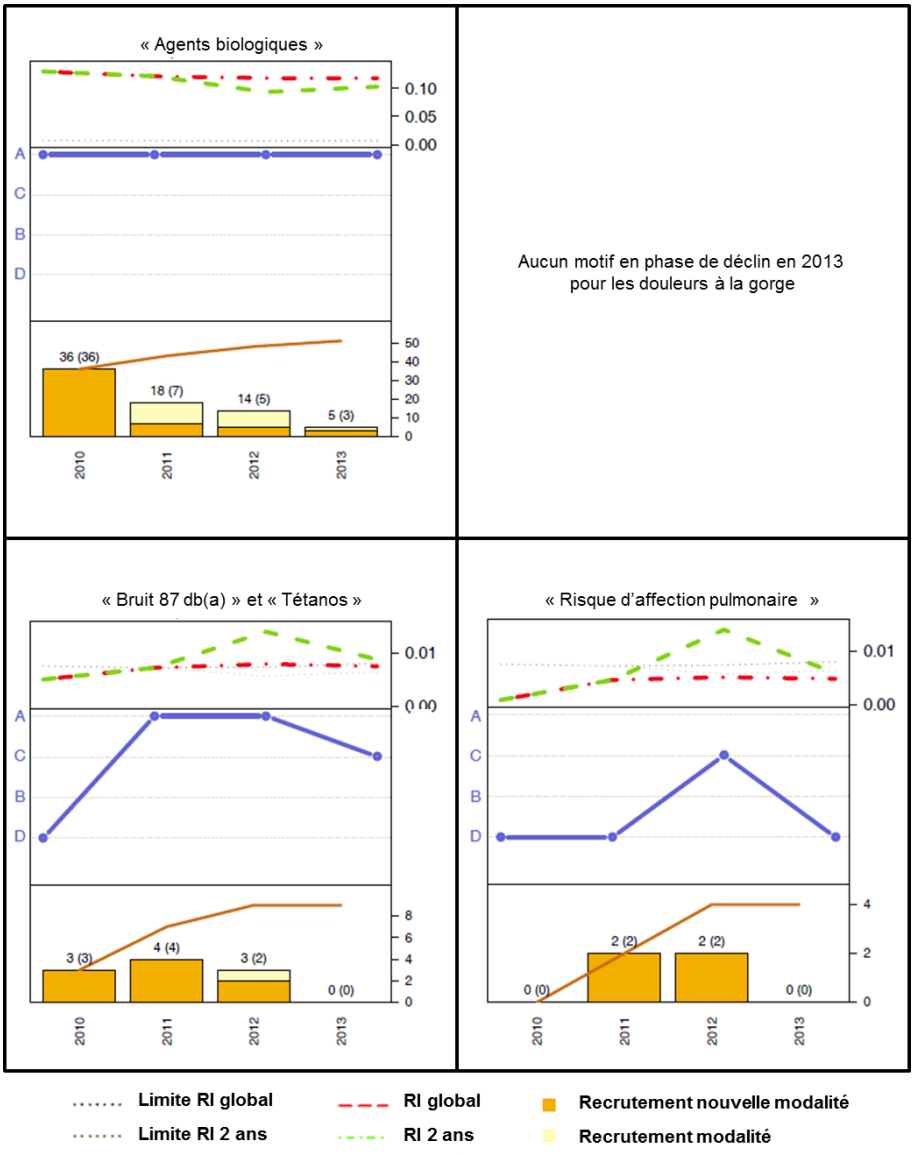


Figure 61 : Exemple de trajectoire atypique de motifs identifiés pour les douleurs à la gorge de 2010 à 2013 pour chaque catégorie de statut dynamique : persistant, en déclin, émergent et latent. Exemple du motif « Agent biologique » (persistant), « Bruit 87 dB(a) » et « tétanos » (en émergence) et « rayonnements ultraviolets » (latent). Aucun motif en déclin n'a été identifié. La courbe bleue correspond à l'évolution du statut dynamique des motifs : A – persistant, B – en déclin, C – émergent et D – latent. Les courbes rouge et verte correspondent respectivement aux valeurs de $\boldsymbol{RI_{\infty}}$ et $\boldsymbol{RI_2}$ pour le motif à l'année donnée, les courbes pointillées correspondent aux limites $\boldsymbol{\xi_{RI}}$ calculées en fonction des autres recrutements pour l'année donnée. La partie basse correspond à l'évolution du recrutement des motifs dans des activités professionnelles particulières, distinguant les nouvelles activités professionnelles et les activités déjà rencontrées.

11.2. Conclusion

Le symptôme des douleurs à la gorge a été choisie pour une étude de faisabilité de la généralisation de la méthodologie à une autre base de données, dont la structure et les paramètres d'études étaient différents. Cet exemple illustre avec succès qu'il est possible d'appliquer la méthodologie de la Surveillance Observationnelle à d'autres bases de données, et que la barrière de la langue peut être dépassée grâce à la standardisation des informations sous forme de thésaurus. Cependant, cette étude met en évidence que la manière d'enregistrer les informations a un rôle important dans le résultat fourni. En effet, il a été fait le choix d'ajouter des valeurs d'imputabilités *a posteriori* pour être capable de mettre en évidence des potentielles relations entre la pathologie et des situations professionnelles précises car le calcul de significativité basé sur le Risque Relatif aurait pu être biaisé par les multiples variables explicatives enregistrées pour chaque observation et qui n'ont pas obligatoirement un lien causal avec la pathologie étudiée. En effet, les expositions professionnelles enregistrées sont des **hypothèses d'expositions pouvant être rencontrées dans l'activité professionnelle de l'employé**, et ne correspondent pas obligatoirement à une relation « causale » avec la pathologie diagnostiquée comme dans le cas du RNV3P.

La Surveillance Observationnelle a permis de mettre en évidence d'après les informations de la base IDEWE un total de **105 motifs, dont 38 actifs en 2013, et une proportion importante de multi exposition de 66.66%.** Les douleurs à la gorge sont régulièrement enregistrées dans des observations avec des **agents biologiques, mycobacterium tuberculosis, le tétanos et des produits alcalins ou biologiques**. Au vu de la structure de la base, **il est alors nécessaire de se demander si ces expositions sont directement liées aux douleurs de la gorge**, ou si elles sont liées à une autre pathologie qui est fréquemment associée aux douleurs de la gorge, notamment le cas de l'exposition « bruit 87dB(A) ».

12. AIDES-SOIGNANTS

Cette application a pour objectif de présenter un exemple de Surveillance Observationnelle appliquée à une autre variable cible que les maladies professionnelles : les activités professionnelles, démontrant ainsi la généralisation possible de la méthodologie de la Surveillance Observationnelle. Le premier exemple concerne les aides-soignants.

12.1. Analyse

En 2013, au sein des 183 614 cas de la base IDEWE, 2 224 concernaient les aides-soignants (Figure 62), dont 111 hommes et 2 094 femmes, avec une moyenne d'âge de 32.02 et 36.63 ans respectivement. Une liste de 721 pathologies distinctes a été recensée pour ces cas, dont beaucoup ont été supprimées car avaient une fréquence égale à 1 et apportaient trop de bruit pour l'analyse. La liste des 233 pathologies restantes a été soumise à un panel de 3 médecins afin d'attribuer une valeur d'imputabilité à chacune de ces pathologies, ou de les écarter si non pertinente avec l'activité professionnelle étudiée, et 167 ont été conservées pour l'analyse.

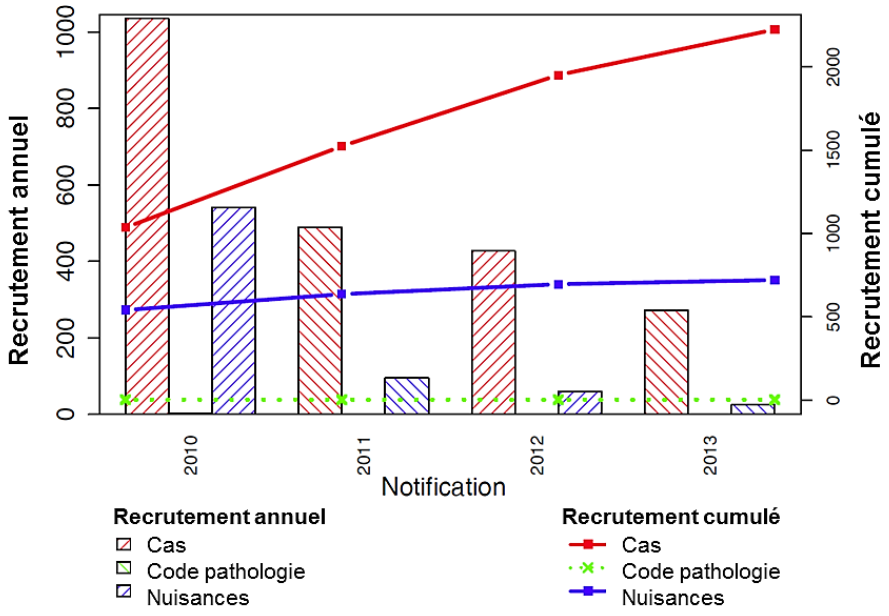


Figure 62 : Recrutement et effectif cumulé des cas, des nuisances distinctes et du code d'activité professionnelle correspondant aux aides-soignants de 2010 à 2013.

Le nombre de médecins étant assez faible pour l'attribution de l'imputabilité, il a été choisi de récupérer les imputabilités renseignées *a posteriori* avec la méthode de la distribution aléatoire (*voir 10.2.1.2, p.200*). Pour chaque pathologie conservée, une distribution

d'imputabilité de taille correspondant au nombre d'occurrences n_i a été générée, à partir de la distribution initiale proposée par les médecins. Une fois les valeurs d'imputabilités générées, le calcul de la significativité par l'imputabilité a été appliqué et a permis de mettre en évidence un total de 57 pathologies significatives (sur les 167 conservées et les 721 enregistrées initialement) et 1 561 observations contenaient au moins une de ces pathologies significatives dans son cortège. A partir de ces 1 561 observations, 880 cortèges distincts de pathologies ont été identifiés, formant les nœuds de l'exposome pour un total de 28 524 liens entre ces nœuds (Figure 63).

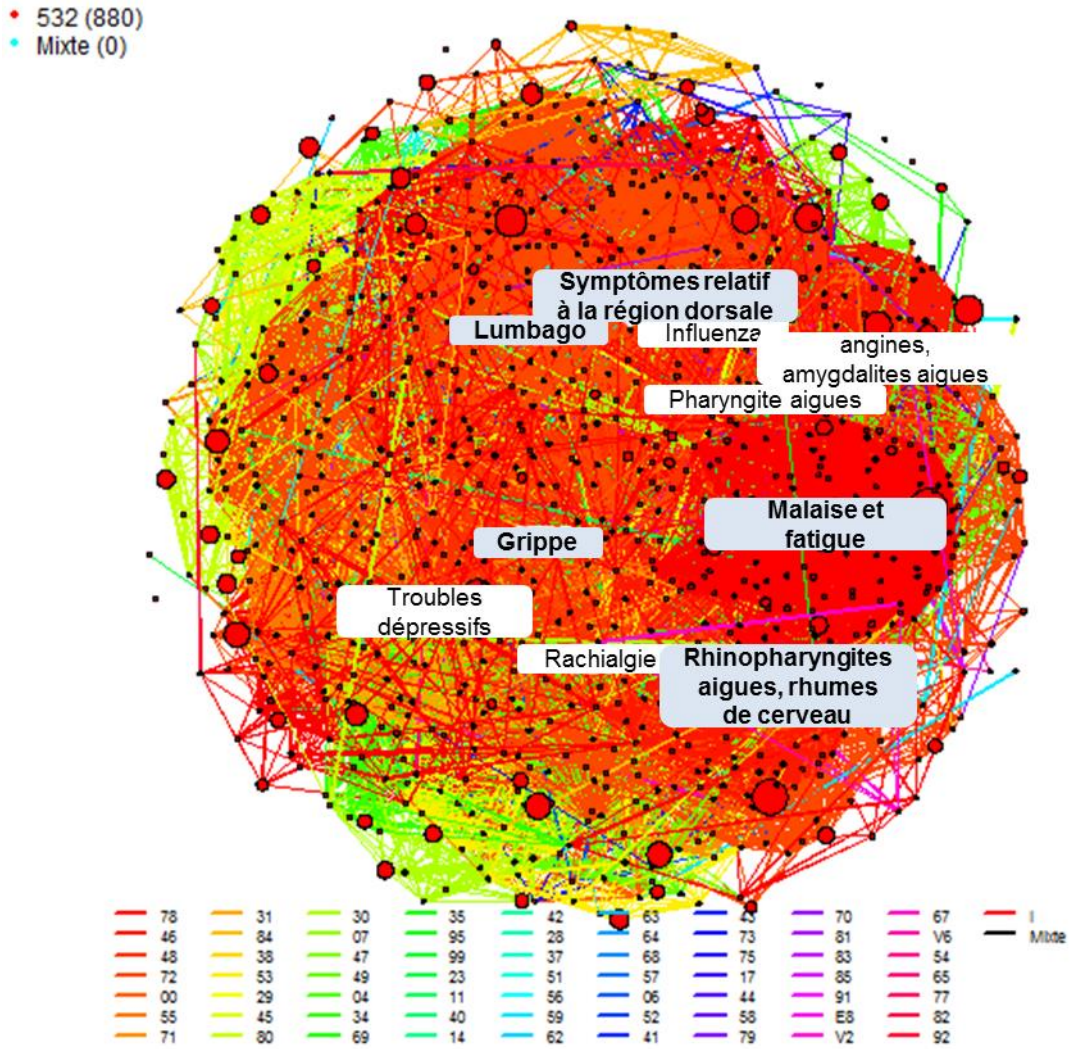


Figure 63 : Exposome des aides-soignants en 2013. Formé par 880 nœuds et 28 584 liens. La taille des nœuds correspond au poids des nœuds (nombre de PST identiques). La couleur des liens correspond aux différentes familles (2e digit) des pathologies identifiées en relation avec les aides-soignants et permettent d'identifier les clusters de l'exposome. Cinq clusters majoritaires ont été identifiés : « Symptômes relatif la région dorsale », « Lumbago », « Malaise et fatigue », « Grippe » et « Rhinopharyngites aigues, rhumes de cerveau ».

La liste des 145 motifs identifiés pour les aides-soignants ainsi que leur numéro d'identification est disponible dans le tableau en Annexe 5, la liste partielle des 44 motifs **actifs** est disponible dans le Tableau 12. Parmi les 145 motifs, 94 correspondaient à des motifs d'ordre I et 51 à des motifs d'ordre II.

Tableau 12 : Liste des 44 motifs actifs des aides-soignants identifiés par la Surveillance Observationnelle en 2013. Les libellés des motifs ont été fournis par les acteurs du réseau IDEWE.

ID	Ordre I	Ordre II	Nœuds	Poids	Statut
1	malaise et fatigue	-	119	239	A
2	rhinopharyngites aigues, rhumes de cerveau	-	92	164	A
3	grippe, avec manifestation, autre	-	63	120	A
4	lumbago	-	67	116	A
5	symptôme relatif à la région dorsale, autre	-	59	86	A
6	influenza	-	39	74	A
7	angines, amygdalites aigues	-	36	69	A
8	rachialgie, sans autre précision	-	37	62	A
9	troubles dépressifs, non classés ailleurs	-	43	57	A
10	gastro-entérite et colite non infectieuse, autre et sans autre précision	-	33	52	A
11	colite, entérite et gastro-entérite infectieuses	-	25	50	A
12	affection de la région dorsale, autre, sans autre précision	-	31	46	A
13	infection aigue de voie respiratoire supérieure, sans autre précision	-	32	45	A
14	pharyngites aigues	-	35	43	B
15	hernie discale, sans autre précision, sans myélopathie	-	28	41	B
16	épicondylite, sans autre précision	-	32	41	A
17	douleur articulaire, d'épaule	-	22	39	A
18	asthme non spécifié, sans autre précision	-	29	37	A
19	bronchite aigue	-	24	33	B
20	bronchites, sans autre précision	-	27	32	B
21	symptôme concernant les systèmes nerveux et musculo-squelettique, autre	-	21	32	A
22	cervicalgie	-	24	31	B
23	réaction aigue au stress, sans autre précision	-	24	28	A
24	dyspepsie et trouble fonctionnel de l'estomac, autre	-	21	27	A
25	trouble bipolaire, sans autre précision	-	19	26	B
26	diarrhée	-	20	26	B
27	sinusite chronique, sans autre précision	-	18	24	A
28	capsulite adhésive de l'épaule	-	20	24	A
29	otite moyenne, sans autre précision	-	14	23	B
30	bronchite chronique obstructive, avec exacerbation (aigue)	-	19	23	A
31	réaction aigue au stress, autre	-	16	22	B
32	migraine	-	18	22	B
33	migraine, sans autre précision	-	18	22	B
34	colite, entérite et gastro-entérite présumées infectieuses	-	16	21	A
35	varice de membre inférieur, asymptomatique	-	14	21	A

36	gastrite et gastroduodénite, sans autre précision, sans mention d'hémorragie	-	13	21	A
37	libération du canal carpien	-	15	20	B
38	malaise et fatigue, sans autre précision	-	16	19	D
39	arthroscopie, genou	-	12	19	C
40	infection virale sans autre précision	-	12	18	C
41	hernie thoracique noyau pulpeux ou lombaire sans myélopathie	-	13	18	D
42	traitement chirurgical d'une hernie discale	-	11	16	C
51	opération du système musculo-squelettique, autre	-	9	12	C
59	lumbago	symptôme relatif à la région dorsale, autre	9	9	D

Le statut dynamique des motifs a été caractérisé pour l'année 2013 (Figure 64), 25 motifs étaient **persistants**, 12 en **déclin**, 4 en **émergence** et 104 en état de **latence**, dont 3 actifs en 2012. **Quarante-quatre motifs ont donc été caractérisés comme actifs en 2013**, mais seulement un motif actif était d'ordre II : « lumbago/symptôme relatif à la région dorsale, autre », c'est pourquoi l'arbre de la multi-association n'a pas été représenté.

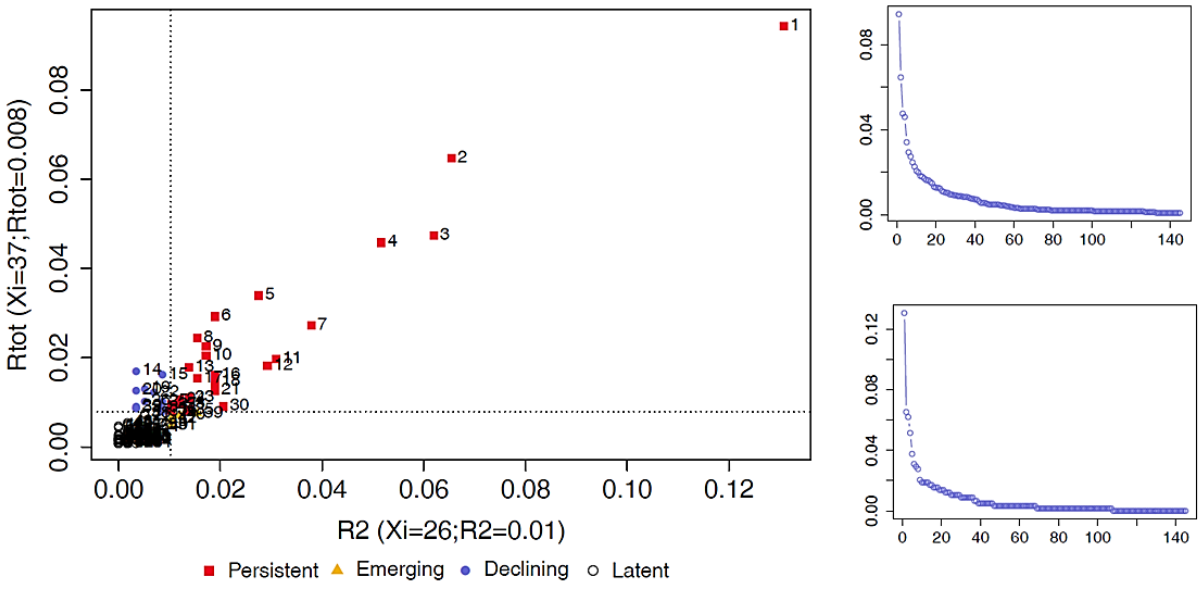


Figure 64 : Caractérisation du statut dynamique des motifs issus de l'exposome des aides-soignants en 2013. Les délimitations des RI ont été calculé à 0.01 et 0.008 pour les deux années glissantes et de manière globale respectivement. Il y avait 25 motifs persistants, 12 en déclin, 4 en émergence et 104 latents donc 3 actifs en 2012.

Les motifs les plus importants rapportés comme potentiellement associées à la profession des aides-soignants étaient : « malaise et fatigue » (id : 1) d'un poids de 239, enregistré en association avec 17 autres motifs ; « rhinopharyngites aigues, rhumes de cerveau » (id : 2) enregistré dans 164 observations et participant à la formation de 11 autres motifs ; « grippe » (id : 3) d'un poids de 120 observations et identifié dans 7 autres motifs ; « lumbago » (id : 4) dans 116 observations et participant à la formation de 10 motifs ; et « symptôme relatif à la région dorsale, autre » (id : 5) enregistré dans 86 observations et identifié dans 6 autres motifs. Les motifs identifiés en émergence étaient : « arthroscopie, genou » (id : 39), « infection virale sans autre précision » (id : 40), « traitement chirurgical d'une hernie discale » (id : 42) et « opération du système musculo-squelettique, autre » (id : 51).

Le spectrosome des motifs des pathologies rapportées comme associées à la profession des aides-soignants est disponible en Figure 65. Cinq clusters majoritaires ont été mis en évidence de par les nombreux liens qu'ils forment au sein du spectrosome : « malaise et fatigue » (id : 1), « lumbago » (id : 4), « symptômes relatif à la région dorsale » (id : 5), « rhinopharyngites aigues, rhumes du cerveau » (id : 2) et « grippe » (id : 3).

Le spectre issu de la Surveillance Observationnelle des aides-soignants est disponible en Figure 66. Dix-sept motifs ont été caractérisés comme spécifiques, avec une information localisée dans des associations spécifiques de pathologies ou alors en motif d'ordre I, et seulement un motif a été caractérisé comme ubiquitaire : « lumbago/symptôme relatif à la région dorsale, autre ».

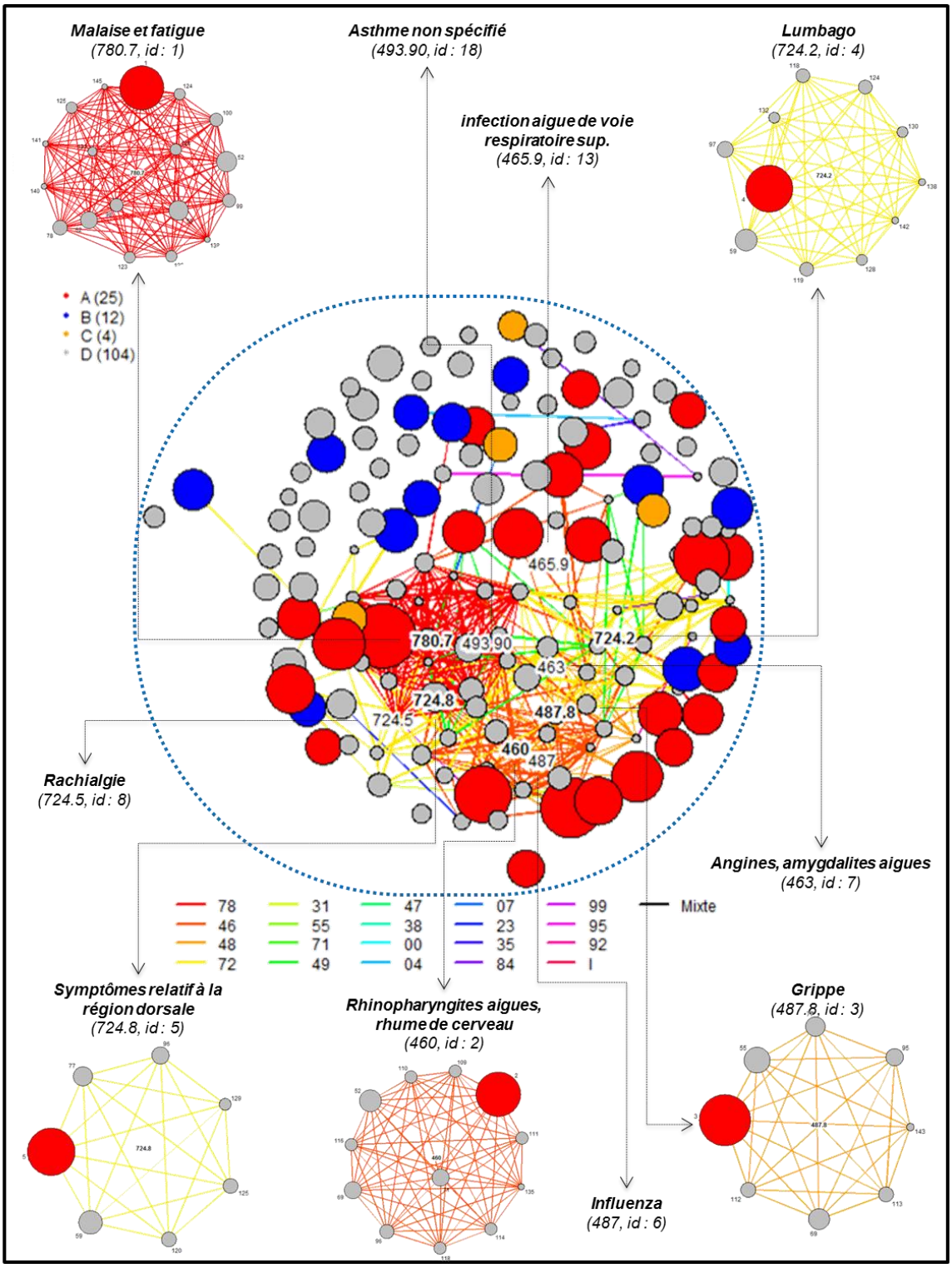


Figure 65 : Spectrosome des motifs des aides-soignants en 2013. Au centre le spectrosome des motifs des pathologies associées aux aides-soignants. La taille des nœuds correspond au poids du motif et la couleur correspond au statut dynamique du motif. La couleur des liens correspond aux différentes familles de nuisances identifiées en relation avec le cancer de la vessie. Cinq clusters ont été identifiés en raison du nombreux nombre de liens auxquels ils participent, de gauche à droite et de haut en bas : malaise et fatigue (1), lumbago (4), symptômes relatif à la région dorsale (5), rhinopharyngites aigues, rhumes du cerveau (2) et grippe (3).

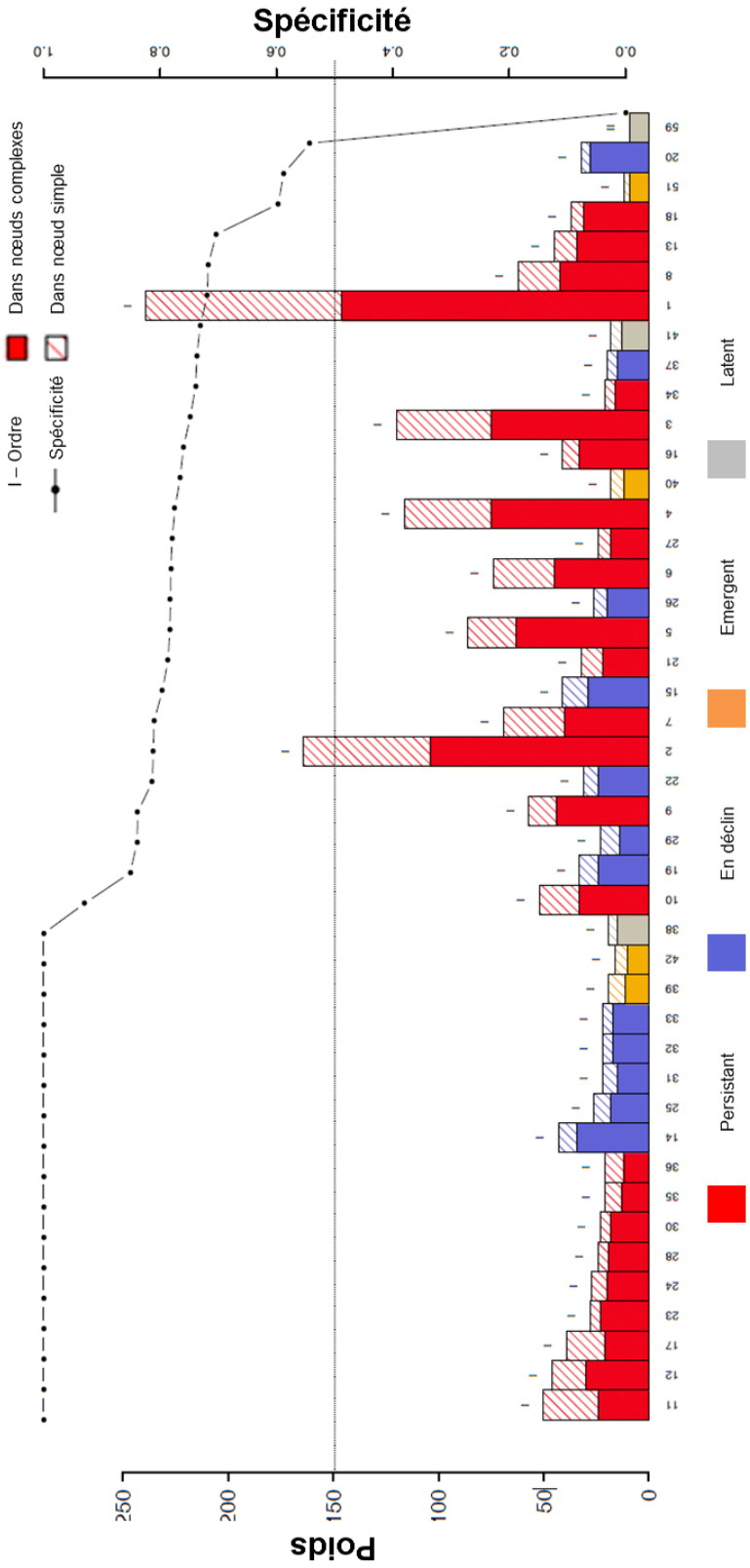


Figure 66 : Spectre des aides-soignants en 2013. Le spectre des aides-soignants a été caractérisé par 44 motifs actifs (dont 25 persistants, 12 en déclin, 4 en émergence et 3 latents mais actifs en 2012), parmi lesquels 43 étaient des motifs d'ordre I et 1 seul motif d'ordre II. Dix-sept motifs ont eu une localisation de leur information spécifique c'est-à-dire privilégiant une association en particulier, 1 seul était ubiquitaire avec une information apportée par différents nœuds de manière homogène et les autres étaient des motifs intermédiaires, dont les motifs 1 et 2 qui tendaient vers une spécificité avec une valeur supérieure à 0.5.

Les trajectoires atypiques des motifs rapportés comme associés à la profession des aides-soignants est disponible en Figure 67. Sont présentés les exemples des motifs « Malaise et fatigue » persistant depuis 2010 ; le motif « Pharyngites aigues » persistant en 2010 et 2011 et en déclin depuis ; le « traitement chirurgicale d'une hernie discale » qui après un déclin en 2011 et 2012 est de nouveau en émergence en 2013 ; et enfin le motif « Malaise et fatigue, autre » qui après une phase de déclin est devenu un motif latent.

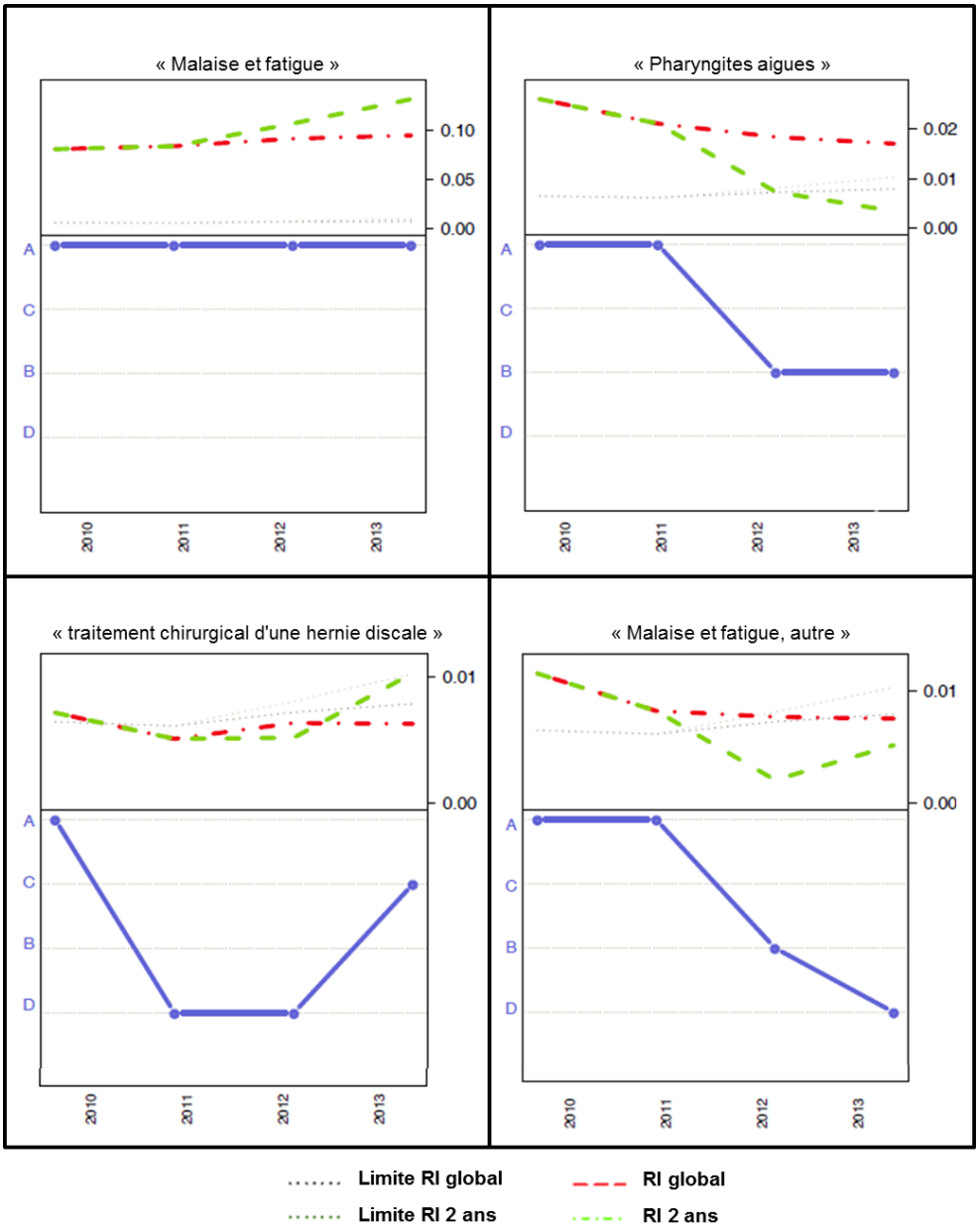


Figure 67 : Exemple de trajectoire atypique de motifs identifiés pour les aides-soignants de 2001 à 2013 pour chaque catégorie de statut dynamique : persistant, en déclin, émergent et latent. Exemple du motif « Malaise et fatigue » (persistant), « Pharyngites aigues » (en déclin), le « traitement chirurgicale d'une hernie discale » (émergence) et le motif « Malaise et fatigue, autre » (latent). La courbe bleue correspond à l'évolution du statut dynamique des motifs : A – persistant, B – en déclin, C – émergent et D – latent. Les courbes rouge et verte correspondent respectivement aux valeurs de $\boldsymbol{RI_{\infty}}$ et $\boldsymbol{RI_2}$ pour le motif à l'année donnée, les courbes pointillées correspondent aux limites $\boldsymbol{\xi_{RI}}$ calculées en fonction des autres recrutements pour l'année donnée.

12.2. Conclusion

L'exemple de l'activité professionnelle des aides-soignants a permis de proposer une application pour laquelle la variable cible de la surveillance n'était pas la même que pour le RNV3P, afin de présenter l'adaptabilité de cette méthode de surveillance. Il est à noter qu'en raison du faible nombre de médecins participant au renseignement de l'information, il a été nécessaire de générer de manière aléatoire la distribution de l'imputabilité afin de pouvoir appliquer un calcul de significativité adéquat.

La Surveillance Observationnelle a permis de mettre en évidence d'après les informations de la base IDEWE que sur les 167 pathologies identifiées au début, 57 ont été considérées comme significatives par la Surveillance Observationnelle, formant **145 motifs dont 44 actifs et 4 en émergence en 2013**. Cette profession est fortement caractérisée par une mono-pathologie (64.83%), car peu d'associations d'expositions ont été retrouvées lors de l'analyse (peu de motifs d'ordre supérieur à I), et une seule seulement était active (« lumbago/symptôme relatif à la région dorsale, autre »). Les aides-soignants semblent être touchés par **des douleurs liées à la zone dorsale, des malaises et des pathologies liées au pharynx**. Les informations issues de la Surveillance Observationnelle appliquée sur les activités professionnelles de la base IDEWE semblent plus « fiables » que celle appliquée à une pathologie (ou symptôme) spécifique, notamment parce que les motifs ici sont formés par les pathologies avérées et diagnostiquées par le médecin et non par des hypothèses comme dans le cas des expositions professionnelles et des douleurs de la gorge.

13. AGRICULTEURS

Cette deuxième application a pour objectif de présenter un autre exemple de Surveillance Observationnelle appliquée à une autre activité professionnelle : les agriculteurs.

13.1. Analyse

En 2013, au sein des 183 614 cas de la base IDEWE, 1 157 concernaient les agriculteurs (Figure 68), dont 1 035 hommes et 119 femmes, avec une moyenne d'âge de 41.44 et 41.48 ans respectivement. Une liste de 488 pathologies distinctes a été recensée pour ces cas, dont une partie a été supprimée à cause de leur fréquence égale à 1. La liste des 188 pathologies restantes a été soumise à un panel de 3 médecins afin d'attribuer une valeur d'imputabilité à chacune de ces pathologies ou de les écarter si non pertinente avec l'activité professionnelle étudiée et 145 ont été conservées pour l'analyse.

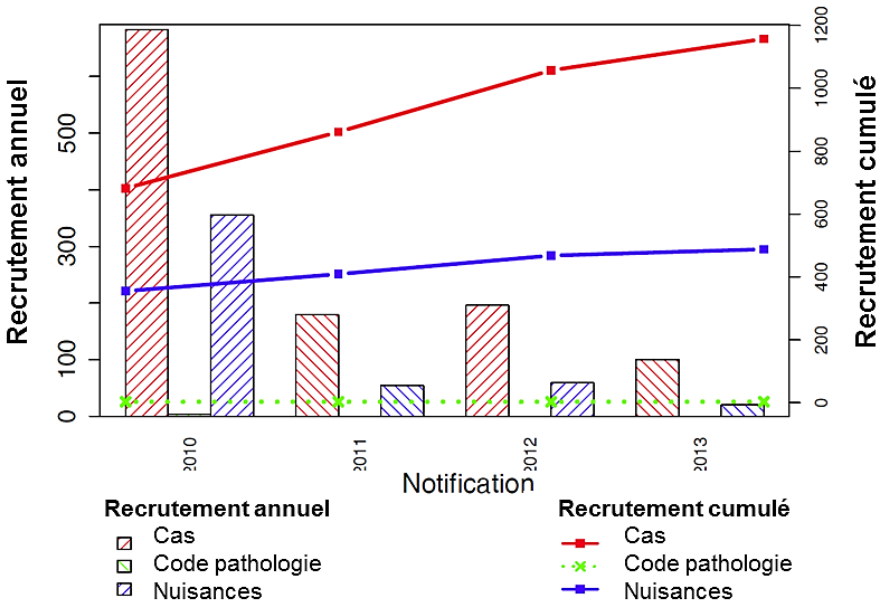


Figure 68 : Recrutement et effectif cumulé des cas, des nuisances distinctes et du code d'activité professionnel correspondant aux agriculteurs de 2010 à 2013.

Le nombre de médecins étant assez faible pour l'attribution de l'imputabilité, il a été choisi de générer, comme pour l'exemple des aides-soignants, une distribution d'imputabilité dont la taille correspond au nombre d'observations enregistrées, à partir de la distribution initiale proposée par les médecins. Le calcul de la significativité a permis de mettre en évidence un total de 39 pathologies significatives (sur les 145 conservées et les 488 enregistrées

initialement) et 696 observations contenaient au moins une de ces pathologies significatives dans son cortège. A partir de ces 696 observations, 363 cortèges distincts de pathologies ont été identifiés, formant les nœuds de l'exposome pour un total de 6 438 liens entre ces nœuds (Figure 69).

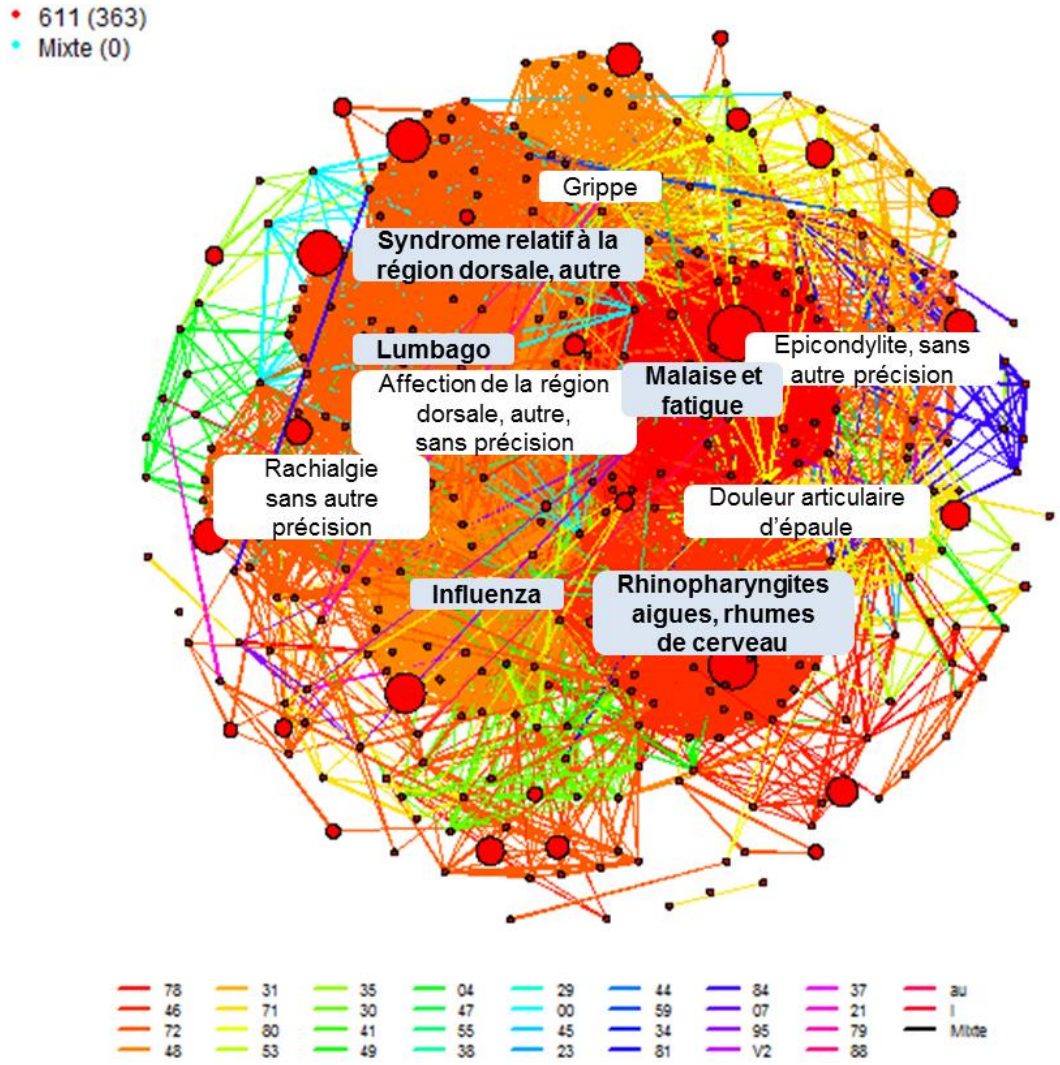


Figure 69 : Exposome des agriculteurs en 2013. Formé par 363 nœuds et 6 438 liens. La taille des nœuds correspond au poids des nœuds (nombre de PST identiques). La couleur des liens correspond aux différentes familles (2^{e} digit) des pathologies identifiées en relation avec les agriculteurs et permettent d'identifier les clusters de l'exposome. Cinq clusters majoritaires ont été identifiés : « Syndrome relatif à la région dorsale, autre », « Lumbago », « Malaise et fatigue », « Influenza » et « Rhinopharyngites aigues, rhumes de cerveau ».

La liste des 66 motifs identifiés pour les agriculteurs ainsi que leur numéro d'identification est disponible dans le tableau en Annexe 6, la liste partielle des 28 motifs **actifs** est disponible dans le Tableau 13. Parmi ces 66 motifs, 49 correspondaient à des motifs d'ordre I et 17 à des motifs d'ordre supérieurs : 16 d'ordre II et 1 d'ordre III.

Tableau 13 : Liste des 28 motifs <u>actifs</u> des agriculteurs identifiés par la Surveillance Observationnelle en 2013. Les libellés des motifs ont été fournis par les acteurs du réseau IDEWE.

ID	Ordre I	Ordre II	Nœuds	Poids	Statut
1	malaise et fatigue	-	57	139	A
2	rhinopharyngites aigues, rhumes de cerveau	-	46	91	A
3	lumbago	-	40	82	A
4	symptôme relatif à la région dorsale, autre	-	38	68	A
5	influenza	-	34	61	A
6	grippe, avec manifestation, autre	-	24	40	A
7	rachialgie, sans autre précision	-	21	35	A
8	affection de la région dorsale, autre, sans autre précision	-	23	32	A
9	épicondylite, sans autre précision	-	18	29	A
10	douleur articulaire, d'épaule	-	16	26	A
11	pharyngites aigues	-	11	21	A
12	arthroscopie, genou	-	12	20	B
13	troubles dépressifs, non classés ailleurs	-	10	19	B
14	dyspepsie et trouble fonctionnel de l'estomac, autre	-	12	17	A
15	hernie discale, sans autre précision, sans myélopathie	-	11	16	A
16	capsulite adhésive de l'épaule	-	8	16	A
17	réaction aigue au stress, sans autre précision	-	13	15	A
18	hernie thoracique noyau pulpeux ou lombaire sans myélopathie	-	8	11	A
19	colite, entérite et gastro-entérite infectieuses	-	9	9	B
20	maladie des insertions ligamentaires et tendineuses, sans autre précision	-	5	9	A
21	entorse et rupture de muscle ou tendon, sans autre précision	-	9	9	B
22	libération du canal carpien	-	8	8	D
23	syndrome du canal carpien	-	5	8	C
27	troubles de disques intervertébraux	-	5	7	D
29	lumbago	affection de la région dorsale, autre, sans autre précision	6	7	D
35	douleur articulaire, de région pelvienne et cuisse	-	6	6	D
40	hernie discale cervicale, sans myélopathie	-	5	5	D
44	douleur de membre, sans autre précision	-	3	5	C

Le statut dynamique des motifs a été caractérisé pour l'année 2013 (Figure 64), 17 motifs étaient **persistants**, 4 en **déclin**, 2 en **émergence** et 43 en état de **latence**, dont 5 actifs en 2012. Vingt-huit motifs ont donc été caractérisés comme actifs en 2013, mais seulement un motif actif était d'ordre II : « lumbago/affection de la région dorsale, autre, sans autre précision », c'est pourquoi l'arbre de la multi-association n'a pas été représenté.

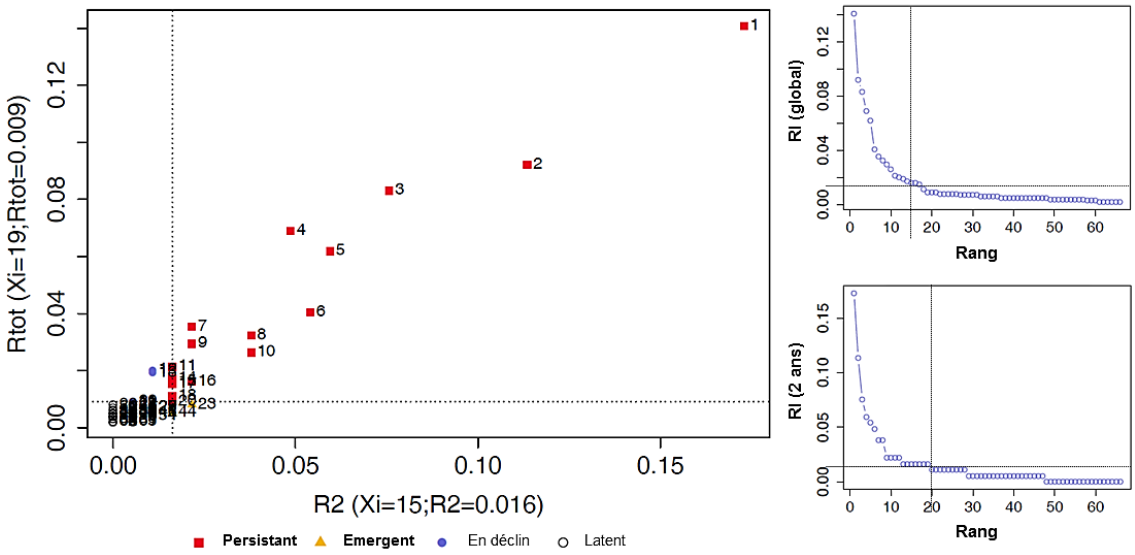


Figure 70 : Caractérisation du statut dynamique des motifs issus de l'exposome des agriculteurs en 2013. Les délimitations des RI ont été calculé à 0.016 et 0.009 pour les deux années glissantes et de manière globale respectivement. Il y avait 17 motifs persistants, 4 en déclin, 2 en émergence et 43 latents donc 5 actifs en 2012.

Les motifs les plus importants rapportés comme potentiellement associées à la profession des agriculteurs étaient : « malaise et fatigue » (id : 1) d'un poids de 139, enregistré en association avec 5 autres motifs ; « rhinopharyngites aigues, rhumes de cerveau » (id : 2) enregistré dans 91 observations et participant à la formation de 2 autres motifs ; « lumbago » (id : 3) d'un poids de 82 observations et identifié dans 7 autres motifs ; « symptôme relatif à la région dorsale, autre » (id : 4) dans 68 observations et participant à la formation de 6 motifs ; et « influenza » (id : 5) enregistré dans 61 observations et identifié dans 3 autres motifs. Les motifs identifiés en émergence étaient : « syndrome du canal carpien » (id : 23) et « douleur de membre, sans autre précision » (id : 44).

Le spectrosome des motifs des pathologies rapportées comme associées à la profession des agriculteurs est disponible en Figure 71. Cinq clusters majoritaires ont été mis en évidence de par les nombreux liens qu'ils forment au sein du spectrosome : « lumbago » (id : 3), « malaise et fatigue » (id : 1), « symptômes relatif à la région dorsale, autre » (id : 4), « douleurs articulaires de l'épaule » (id : 10) et « influenza » (id : 5).

Le spectre issu de la Surveillance Observationnelle des agriculteurs est disponible en Figure 72. Treize motifs ont été caractérisés comme spécifiques, avec une information localisée dans des associations spécifiques de pathologies comme le motif 7 « rachialgie, sans autre précision » enregistré 35 fois dans 21 nœuds différents, et 5 motifs étaient ubiquitaires comme le motif 9 « colite, entérite et gastro-entérite infectieuses » enregistré 9 fois dans 9 associations différentes.

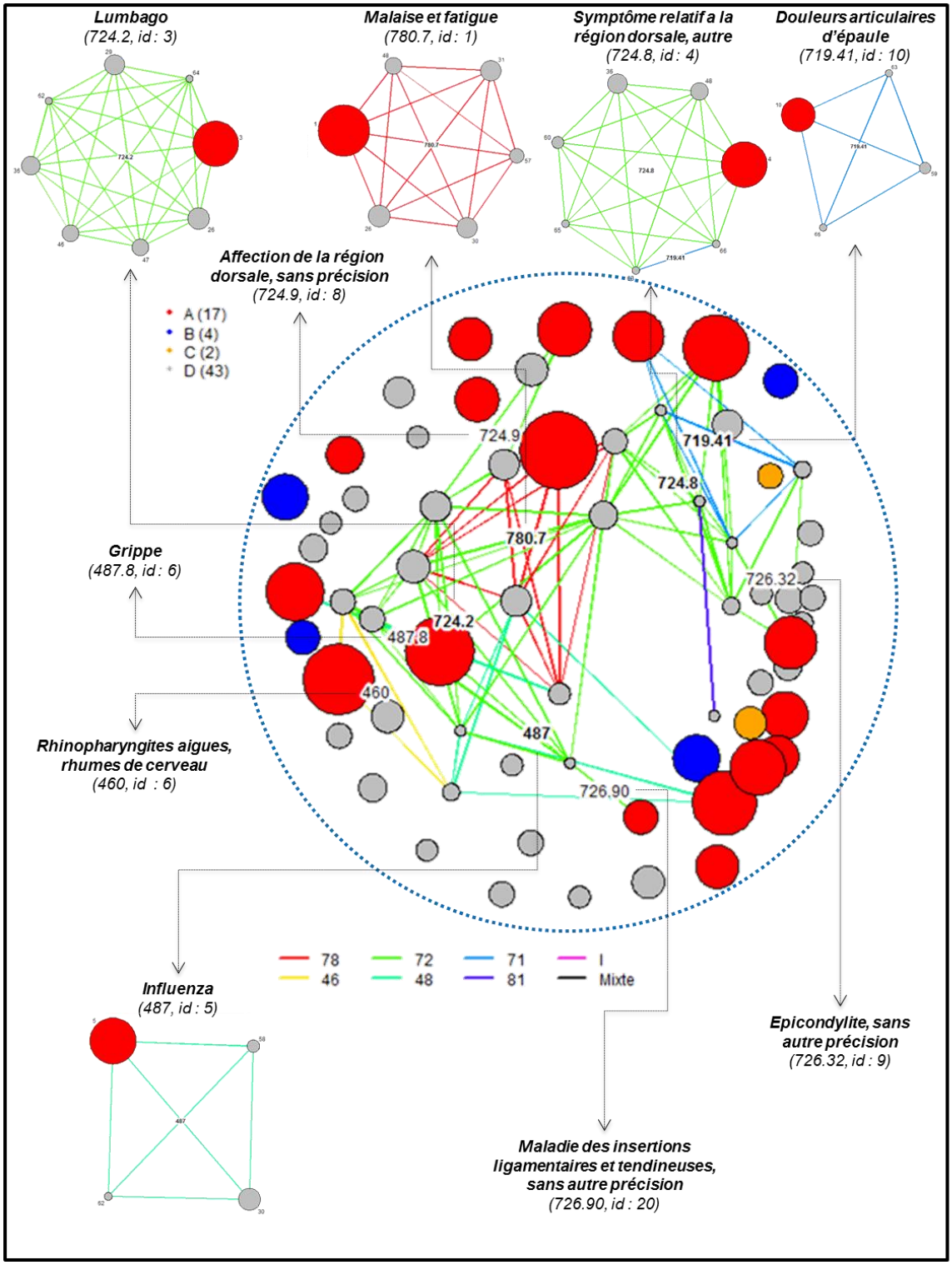


Figure 71 : Spectrosome des motifs des agriculteurs en 2013. Au centre le spectrosome des motifs des pathologies associées aux agriculteurs. La taille des nœuds correspond au poids du motif et la couleur correspond au statut dynamique du motif. La couleur des liens correspond aux différentes familles de nuisances identifiées en relation avec le cancer de la vessie. Cinq clusters ont été identifiés en raison du nombreux nombre de liens auxquels ils participent, de gauche à droite et de haut en bas : lumbago (3), malaise et fatigue (1), symptômes relatif à la région dorsale, autre (4), douleurs articulaires de l'épaule (10) et influenza (5).

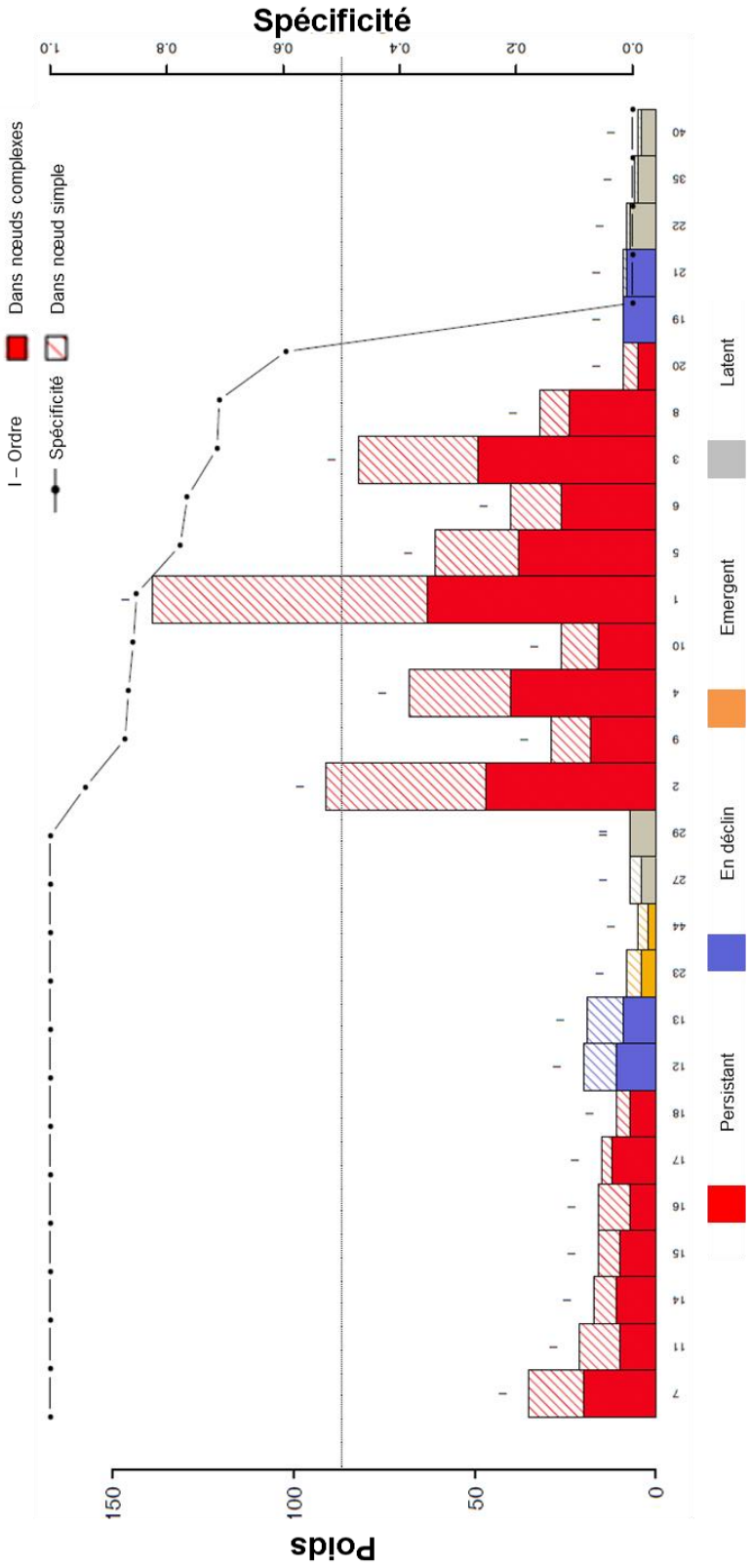


Figure 72 : Spectre des agriculteurs en 2013. Le spectre des agriculteurs a été caractérisé par 28 motifs actifs (dont 17 persistants, 4 en déclin, 2 en émergence et 5 latents mais actifs en 2012), parmi lesquels 27 sont des motifs d'ordre I et 1 seul motif d'ordre supérieur. Treize motifs ont une localisation de leur information spécifique c'est-à-dire privilégiant une association en particulier, 5 sont ubiquitaire avec une information apportée par différents nœuds de manière homogène et les autres étant des motifs intermédiaires, dont les motifs 1 et 2 tendent vers une spécificité avec une valeur supérieure à 0.5.

Les trajectoires atypiques des motifs rapportés comme associés à la profession des agriculteurs sont disponibles en Figure 73. Sont présentés les exemples des motifs « Malaise et fatigue » persistant depuis 2010 ; le motif « Pharyngites aigues » persistant en 2010 et 2011 et en déclin depuis ; le « traitement chirurgicale d'une hernie discale » qui après un déclin en 2011 et 2012 est de nouveau en émergence en 2013 ; et enfin le motif « Malaise et fatigue, autre » qui après une phase de déclin est devenu un motif latent.

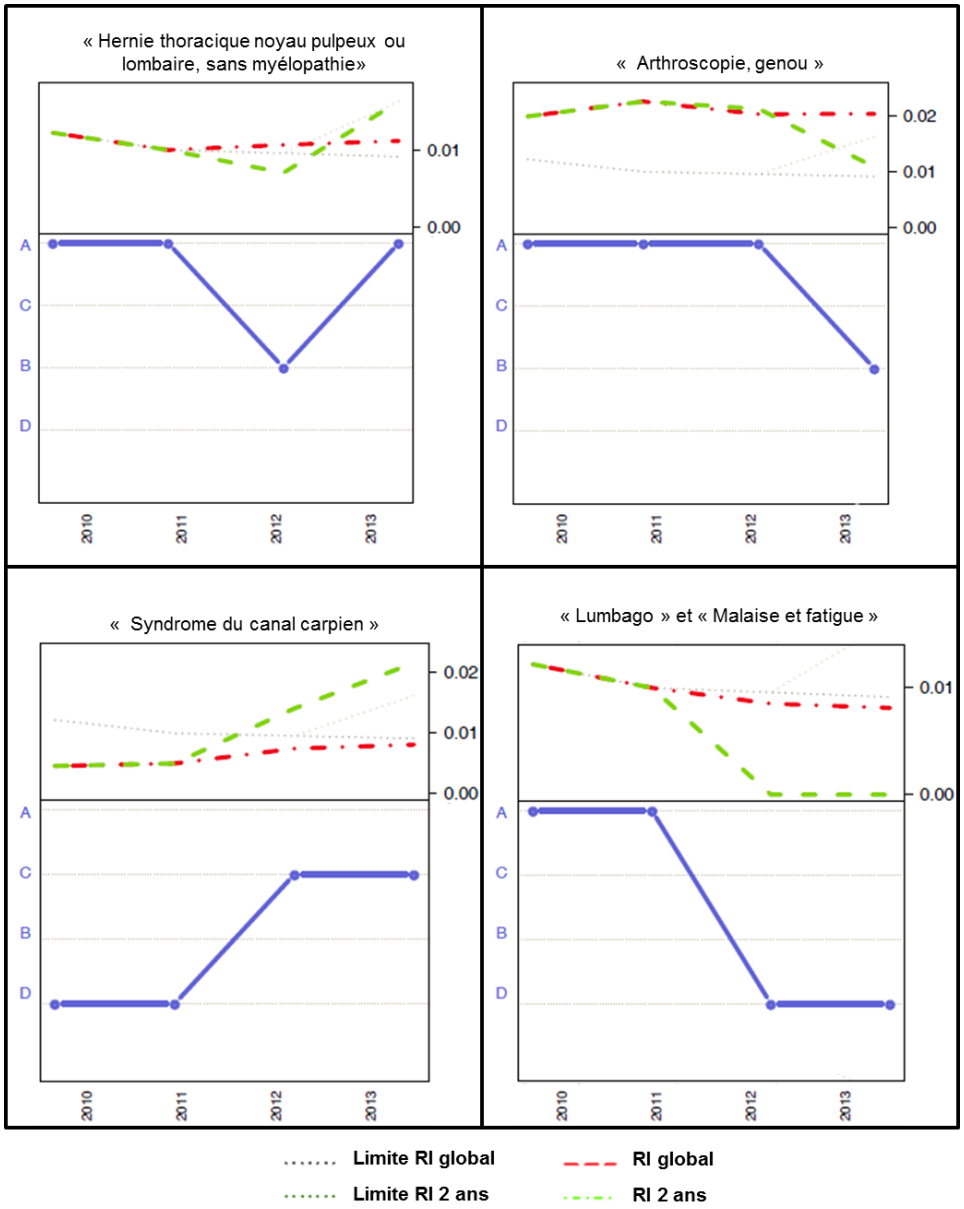


Figure 73 : Exemple de trajectoire atypique de motifs identifiés pour les agriculteurs de 2010 à 2013 pour chaque catégorie de statut dynamique : persistant, en déclin, émergent et latent. Exemple du motif « Hernie thoracique noyaux pulpeux ou lombaire, sans myélopathie » (persistant), « Arthroscopie, genou » (en déclin), le « syndrome du canal carpien » (émergence) et le motif « Lumbago » et « Malaise et fatigue » (latent). La courbe bleue correspond à l'évolution du statut dynamique des motifs : A – persistant, B – en déclin, C – émergent et D – latent. Les courbes rouge et verte correspondent respectivement aux valeurs de $\boldsymbol{RI}_{\infty}$ et $\boldsymbol{RI}_{\boldsymbol{2}}$ pour le motif à l'année donnée, les courbes pointillées correspondent aux limites $\boldsymbol{\xi}_{\boldsymbol{RI}}$ calculées en fonction des autres recrutements pour l'année donnée.

13.2. Conclusion

A l'instar des aides-soignants, en raison du faible nombre de médecins participant au renseignement de l'information, il a été nécessaire de générer de manière aléatoire la distribution de l'imputabilité afin de pouvoir appliquer le calcul de significativité.

La Surveillance Observationnelle a permis de mettre en évidence d'après les informations de la base IDEWE que les agriculteurs sont également très touchés par **des douleurs liées à la zone dorsale, des malaises et des pathologies liées au pharynx**. Il y a de nombreux motifs en commun entre les deux activités professionnelles étudiées. Sur les 145 pathologies identifiées au début, 39 ont été considérées comme significatives par la Surveillance Observationnelle, formant **66 motifs dont 28 actifs et 2 en émergence en 2013**. Ce secteur d'activité est également caractérisé par une **mono-pathologie** (74.24%), peu d'associations ayant été retrouvées lors de l'analyse, et une seule seulement étant active (« lumbago/affection de la région dorsale, autre, sans autre précision »).

RESUME DES APPLICATIONS SUR LA BASE IDEWE

	Douleurs de la gorge	Aides-soignants	Agriculteurs
	Symptôme	Activité professionnelle	Activité professionnelle
IDEWE			
Code	784.1	531	611
Cas IDEWE	**520**	**2 224**	**1 157**
Modalités distinctes	191 (nuisances)	167 (pathologies)	145 (pathologies)
Modalités significatives	**20 (10,47%)**	**57 (34,13%)**	**39 (26,9%)**
Cas conservés	400 (76,92%)	1 561 (70,19%)	696 (60,16%)
Exposome			
Nombre de nœuds	193	880	363
Nombre de liens	4 969	28 584	6 438
Motifs			
Nombre de motifs	**105**	**145**	**66**
Ordre I	35 (33,33%)	94 (64,83%)	49 (74,24%)
Ordre II	48	51	16
Ordre III	22	-	1
Dynamique			
Motifs persistants	17	25	17
Motifs en déclin	0	12	4
Motifs en émergence	14	4	2
Motifs en latence	74	104	43
Motifs en latence actif en 2012	7	3	5
Nombre de motifs actifs	**38 (36,19%)**	**44 (30,35%)**	**28 (42,42%)**
Caractéristiques du spectre	Pas de motifs en déclin, peu de multi-exposition, quid de la plausibilité des motifs identifiés	Peu de multi-associations de pathologies recensées. Angle de réflexion plus pertinent par rapport à la structure de la base	

Il a donc été présenté dans ces différents chapitres les différentes possibilités d'applications de la Surveillance Observationnelle, adaptable à différentes bases de données, mais également à différents contextes et différentes problématiques grâce à la possibilité de choisir des variables cibles différentes pour la même base de données.

Partie 5 : Discussion et Perspectives

Les objectifs de ces travaux étaient de développer une nouvelle méthodologie de Surveillance qui permettrait d'une part **l'utilisation optimale** de bases de données observationnelles et d'autre part la prise en compte des **origines multifactorielles** de certains évènements de santé. L'objectif sous-jacent était de **développer et fournir un outil** qui permettrait de rendre cette méthodologie directement utilisable en routine et pouvant s'adapter aux différentes bases de données et problématiques ; avec pour finalité de pouvoir proposer **une surveillance systématique**.

Ces travaux ont abouti à la **conceptualisation** de la Surveillance dite « Observationnelle » et de ses différents indicateurs associés, ainsi qu'à la **création d'un véritable outil utilisable en pratique**. Les différentes applications sur les deux bases de données **RNV3P et IDEWE** ont permis de présenter **l'intérêt et la flexibilité (adaptabilité)** de la méthodologie par rapport aux bases de données, mais également en présentant des exemples pour lesquels les variables cibles et explicatives primaires étaient différentes (IDEWE). Ces exemples ont permis de démontrer que la **prise en compte dynamique des origines multifactorielles** pouvaient être un outil de surveillance grâce aux **signaux d'alertes statistiques** pouvant être émis suite à l'émergence ou au déclin des différents motifs mis en évidence.

Ainsi, au même titre qu'une surveillance classique, la Surveillance Observationnelle permet de fournir, à partir d'un recueil de données observationnelles, l'analyse et l'émission de signaux d'alerte statistiques. Les **hypothèses de travail** ainsi générées pouvant permettre une **amélioration des connaissances** vis-à-vis d'un évènement de santé d'intérêt.

SYNTHESE ET DISCUSSION

Impact des paramètres de la Surveillance Observationnelle

Lors des exemples d'applications de la Surveillance Observationnelle au RNV3P et IDEWE, les paramètres présentés dans la partie II (*voir 3.3, p. 49*) ont été fixés puis conservés entre chaque analyse afin de permettre une simplification du message ainsi qu'une comparaison des résultats. Cependant, il est nécessaire de rappeler que ces paramètres ont un impact direct sur les résultats obtenus et peuvent être modifiés.

✓ *Les paramètres Δt, λ_{min} et ν*

Seul les paramètres Δt et λ_{min} ont été modifié d'une base à l'autre :

- RNV3P [$\Delta t = 4$ et $\lambda_{min} = 1.37$]
- IDEWE [$\Delta t = 5$ et $\lambda_{min} = 1.98$]

Pour rappel, le paramètre Δt correspond à l'intervalle de temps de référence permettant le calcul du nombre de recrutement de référence $n_{i,\Delta t}$ de chacune des modalités et de la significativité de chacun de ces modalités (*voir 3.3.1, p.52*) ; le paramètre λ_{min} correspond au nombre d'occurrences minimum critique $n_{\Delta t,cr}$ pour qu'une modalité enregistrée avec une imputabilité minimale soit significative (*voir 3.3.1.2, p.53*). Il est donc possible de pouvoir faire varier ces différents paramètres afin de préserver un maximum d'hypothèses de travail, ou au contraire, durcir la sélection afin de ne sélectionner que les cas dont la représentativité ne fait aucun doute. De même, toujours afin de permettre une **inclusion ou exclusion fine des modalités**, il est possible de faire varier le paramètre *v* ayant ainsi un impact direct sur la relation entre λ et I (*voir 3.3.1.2.1, p.55*).

Dans le Tableau 14 sont présentés quelques exemples de résultats obtenus pour le cancer de la vessie (*voir Chapitre 7, p.121*) selon les différentes variations des paramètres précédemment cités, en ordre d'apparition : le nombre de modalité considérée significative grâce au calcul avec l'imputabilité *H*, le nombre de nœuds formés *V* grâce aux PST conservés et enfin le nombre de motifs représentatifs identifiés pour le cancer de la vessie *M*.

Tableau 14 : Résumé des différents résultats de la Surveillance Observationnelle appliquée au cancer de la vessie du RNV3P selon les variations des paramètres Δt, λ_{min} et ν.

λ_{min}	ν	$\Delta t = 1$	$\Delta t = 2$	$\Delta t = 4$
$\lambda_{min} = 4$	$\nu = 0.1$	H : 32 V : 469 M : 164	H : 46 V : 491 M : 169	H : 83 V : 531 M : 184
	$\nu = 0.5$	H : 28 V : 459 M : 164	**H : 37** **V : 477** **M : 166**	H : 51 V : 500 M : 173
	$\nu = 2$	H : 24 V : 441 M : 161	H : 30 V : 462 M : 164	H : 51 V : 500 M : 173
	$\nu = 10$	H : 23 V : 436 M : 160	*H : 28* *V : 459* *M : 164*	H : 34 V : 472 M : 165
$\lambda_{min} = 6$	$\nu = 0.1$	H : 32 V : 469 M : 164	H : 41 V : 487 M : 169	H : 70 V : 522 M : 180
	$\nu = 0.5$	H : 27 V : 456 M : 163	H : 31 V : 463 M : 164	H : 42 V : 489 M : 171
	$\nu = 2$	H : 21 V : 417 M : 155	H : 28 V : 459 M : 164	H : 42 V : 489 M : 171
	$\nu = 10$	H : 20 V : 408 M : 153	H : 24 V : 441 M : 161	H : 30 V : 462 M : 164
$\lambda_{min} = 10$	$\nu = 0.1$	*H : 28* *V : 459* *M : 164*	H : 37 V : 477 M : 166	H : 51 V : 500 M : 174
	$\nu = 0.5$	H : 23 V : 436 M : 160	H : 28 V : 459 M : 164	H : 32 V : 465 M : 164
	$\nu = 2$	H : 15 V : 257 M : 110	H : 23 V : 436 M : 160	H : 32 V : 465 M : 164
	$\nu = 10$	H : 14 V : 231 M : 101	H : 21 V : 417 M : 155	H : 27 V : 453 M : 164

Le minimum de modalités considérées comme significatives était de 14 pour les conditions les plus strictes (couleur orange dans le tableau) [$\Delta t = 1,\ \lambda_{min} = 10\ et\ \nu = 10$], nécessitant un recrutement de référence critique de 10 modalités pour l'imputabilité minimale sur un an et privilégiant les imputabilités fortes ; le nombre maximum était de 83 (couleur verte) pour les conditions les plus souples [$\Delta t = 4,\ \lambda_{min} = 4\ et\ \nu = 0.1$], nécessitant un minimum critique de 4 occurrences dans le cas de l'imputabilité minimale, sur 4 ans, et permettant la conservation des plus petites valeurs d'imputabilité. Pour rappel, les résultats de l'analyse présentée en *partie III* pour le cancer de la vessie sont indiqués en couleur bleue. Les 14 nuisances significatives retenues lors de l'analyse avec les conditions les plus strictes (orange) faisaient également partie des 37 nuisances significatives présentées dans l'analyse du cancer de la vessie (bleu), elles-mêmes présentes dans les 83 nuisances significatives dans l'analyse

la plus souple (vert). L'analyse la plus stricte a permis d'identifier 101 motifs dont 24 actifs : 13 persistants, 4 en déclin, 6 en émergence et 1 latent mais actif en 2013. L'analyse la plus souple a permis l'identification de 184 motifs dont 47 actifs : 27 persistants, 7 en déclin, 5 en émergence et 8 latents mais actifs en 2013. Les différences observées entre les signaux proviennent de la souplesse ou de la rigidité des seuils de significativité utilisés pour la sélection des expositions. Ainsi des expositions régulièrement rapportées avec une imputabilité faible peuvent contribuer ou non aux formations des motifs selon la stringence des paramètres. Des motifs émergents pourraient alors être non détectés en cas d'analyses trop rigides.

De manière générale, la moyenne de modalités significatives conservées était de 23.92 pour l'intervalle de référence $\Delta t = 1$, 31.17 pour $\Delta t = 2$ et 45.42 pour $\Delta t = 4$. La moyenne du nombre de nœuds était 411.5, 460.7 et 487.3 pour $\Delta t = 1, 2\ et\ 3$ respectivement (Figure 74). La moyenne du nombre de motifs quant à elle était de 151.6, 163.8 et 170.6 pour $\Delta t = 1, 2\ et\ 3$ respectivement. Plus l'intervalle de temps de référence est large, **plus le nombre de modalités significatives augmente**, permettant ainsi la formation **d'un plus grand nombre de nœuds et de motifs**.

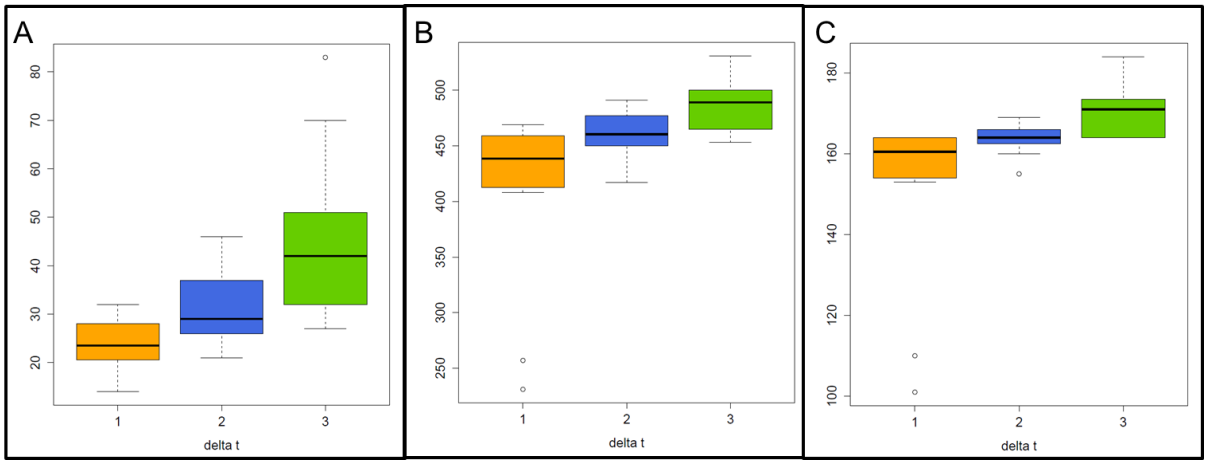


Figure 74 : Répartition de la distribution des résultats obtenus pour chaque Δt pour le cancer de la vessie selon les variations des paramètres λ_{min} et ν. A) Moyenne du nombre de modalités significatives pour $\Delta t = 1, 2\ et\ 3$. B) Moyenne du nombre de nœuds formés pour $\Delta t = 1, 2\ et\ 3$. C) Moyenne du nombre de motifs représentatifs pour $\Delta t = 1, 2\ et\ 3$.

Des « équivalences » sont apparues entre certaines configurations, par exemple la condition $[\Delta t = 1,\ \lambda_{min} = 10\ et\ \nu = 0.1]$ et $[\Delta t = 2,\ \lambda_{min} = 4\ et\ \nu = 0.5]$. Il est donc possible de pouvoir adapter l'analyse aux caractéristiques de la base de données, par exemple à la plage temporelle considérée (RNV3P : 14 ans, IDEWE : 4 ans) ou à la structure des informations des variables explicatives primaires (RNV3P : potentiellement en relation ; IDEWE :

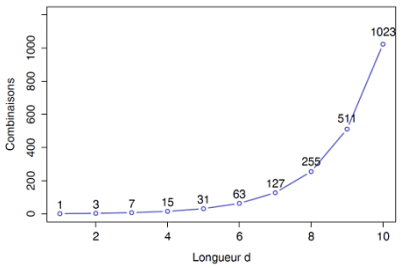
potentiellement sur le lieu de l'activité), tout en conservant la possibilité de moduler la stringence de l'analyse.

✓ ***La longueur* d *des cortèges***

La longueur maximale d des cortèges, correspondant au nombre maximal de modalités des variables explicatives primaires au sein d'une observation, possède également un impact important concernant la formation des motifs. Considérant l'équation (3.7) des combinaisons, le nombre de motifs générés sera proportionnel à la longueur des cortèges : plus le cortège sera long, plus le nombre de combinaisons possible sera grand (*voir 3.3.2.3, p.59*).

Par exemple :

Pour une longueur de nœuds maximal de $d = 3$ le nombre de combinaisons sera de 7, pour une longueur de $d = 5$ il sera de 31 et pour une longueur de $d = 10$ il sera de 1023.



Il est important de souligner que l'un des objectifs de la Surveillance Observationnelle est de **pouvoir prendre en compte les origines multifactorielles** des évènements d'intérêts surveillés. Il a donc été établi pour ces premières études que la longueur d des motifs devait au moins être égale à 2, **l'exposome et le spectrosome n'ayant de sens que si des nœuds sont susceptibles de partager des modalités en commun**. Or si les nœuds sont de longueurs 1, ils sont tous distincts et ne peuvent partager aucune information en commun. De même, il devient contre-informatif d'avoir des cortèges très longs.

Choix des variables sélectionnées pour la Surveillance Observationnelle

✓ ***IDEWE***

L'exemple d'application de la Surveillance Observationnelle aux douleurs de la gorge de la base IDEWE (*voir Chapitre 11, p.203*) a permis d'étudier la faisabilité de l'application de la méthodologie à une autre base de données que le RNV3P. Les variables utilisées pour l'analyse étaient donc les mêmes que pour les exemples d'applications du RNV3P, à savoir la pathologie professionnelle (ou symptôme) en tant que variable cible et les expositions professionnelles « associées » en tant que variables explicatives primaires. Cependant, il est apparu que la méthode de recueil des variables au sein des bases de données peut avoir un

impact important sur la plausibilité des résultats. En effet, les expositions professionnelles enregistrées au sein de la base IDEWE ne représentent pas des hypothèses d'expositions professionnelles en relation avec une pathologie diagnostiquée comme le RNV3P, mais représentent un ensemble d'expositions pouvant être rencontrées dans l'exercice habituel de l'activité de l'employé, sans préciser si cette exposition a été avérée ou non. Cette différence a mené à la mise en évidence de l'exposition professionnelle « Bruit 87dB(A) » en tant que motif majeur d'exposition pour les douleurs de la gorge, ce qui semble être difficilement crédible d'un point de vue clinique. Il est donc primordial, comme à l'instar des autres systèmes et méthodes de surveillance, d'établir des problématiques scientifiques solides et adaptées, permettant ainsi de pouvoir recueillir les données en adéquation avec ces problématiques.

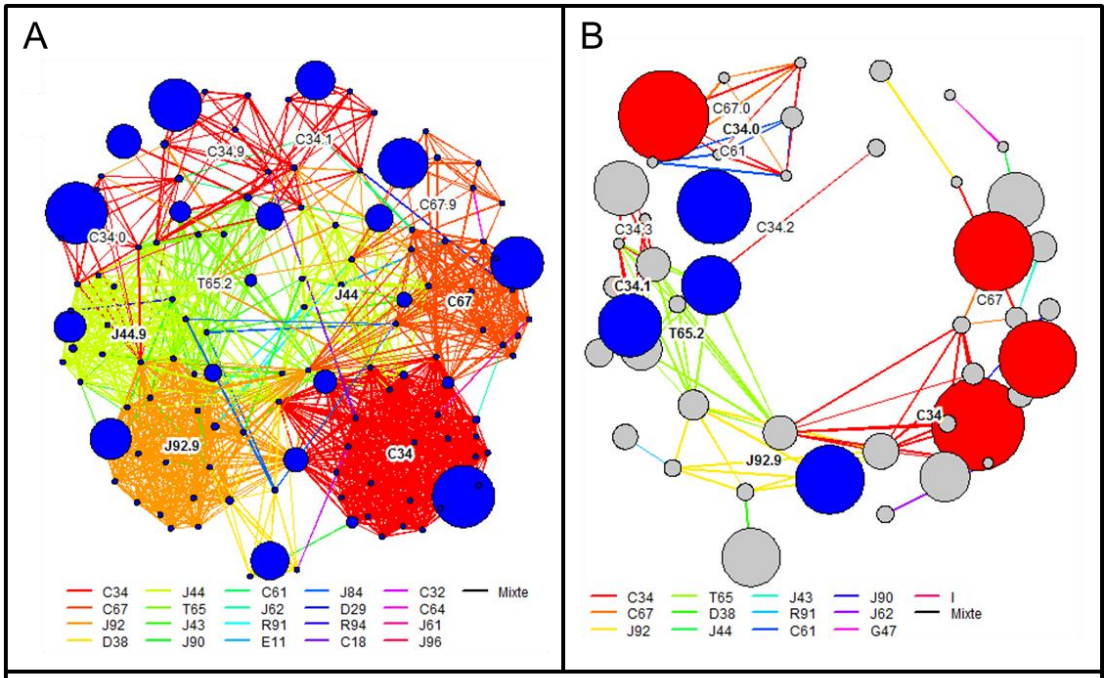
Dans le cas de la base IDEWE, la méthode de la Surveillance Observationnelle semble plus intéressante en termes de pertinence scientifique lorsque la variable cible correspond à une activité professionnelle, telle que les aides-soignants ou les agriculteurs (*Chapitre 12 et 13, p.217 et 227*), et les variables explicatives primaires aux différentes pathologies professionnelles enregistrées. Il a ainsi pu être mis en évidence un **spectre des pathologies professionnelles les plus fréquemment enregistrées pour des employés dont les activités professionnelles étaient aides-soignants ou agriculteurs**. Ce spectre des pathologies professionnelles peut alors servir à améliorer la connaissance en apportant de nouvelles hypothèses de travail ainsi qu'un caractère dynamique à cette connaissance avec la mise en évidence de motifs (pathologies professionnelles en relation avec l'activité professionnelle) persistants et/ou émergents.

- ✓ ***RNV3P***

Dans le cadre des applications sur la base du RNV3P (*Partie 3 : Applications au RNV3P, p.111*) ; la variable cible a été établie en fonction de l'objectif même du réseau qui est la surveillance des pathologies professionnelles. Cependant, il aurait été possible d'inverser ce choix, et d'analyser en tant que variable cible les expositions professionnelles et en tant que variables explicatives primaires les pathologies professionnelles, afin d'obtenir le **spectre des pathologies professionnelles associées à une nuisance particulière**. Au sein du RNV3P, en plus de la pathologie dit « principale » (utilisée jusqu'à présent dans les exemples d'applications), 3 autres pathologies peuvent être renseignées au sein du PST. Il alors possible de créer un exposome et un spectrosome de la nuisance choisie en fonction des différentes

pathologies enregistrées. Dans ce contexte, il n'est toutefois pas possible d'utiliser la variable de l'imputabilité pour le calcul de la significativité car celle-ci permet d'établir un lien entre la pathologie principale et les expositions ; les autres pathologies seraient alors écartées par leur manque d'imputabilité. Le calcul de la significativité par le Risque Relatif (voir *3.3.1.1, p.53, équation 3.2*) permet donc de palier à ce manque d'information.

La Figure 75 propose **l'exposome et le spectrosome des pathologies professionnelles associées à l'exemple des Hydrocarbures Aromatiques Polycycliques** (HAP).



C

Code	Libellé	Réseau
C34	Tumeur maligne des bronches et du poumon	Exposome, Spectrosome
C34.0	Tumeur maligne bronche souche	Exposome, Spectrosome
C34.1	Tumeur maligne lobe supérieur, bronches ou poumons	Exposome, Spectrosome
C34.2	Tumeur maligne lobe moyen, bronches ou poumon	Spectrosome
C34.3	Tumeur maligne lobe inférieur, bronches ou poumon	Spectrosome
C34.9	Tumeur maligne des bronches et du poumon, sans précision	Exposome
C61	Tumeur maligne de la prostate	Spectrosome
C67	Tumeur maligne de la vessie	Exposome, Spectrosome
C67.0	Tumeur maligne, trigone de la vessie	Spectrosome
C67.9	Tumeur maligne de la vessie, sans précision	Exposome
J44	Autres maladies pulmonaires obstructives chroniques	Exposome
J44.9	Maladie pulmonaire obstructive chronique, sans précision	Exposome
J92.9	Plaque pleurale sans asbestose	Exposome, Spectrosome
T65.2	Effet toxique de substances, tabac et nicotine	Exposome, Spectrosome

Figure 75 : Illustration de la Surveillance Observationnelle appliquée au RNV3P à une exposition professionnelle : les Hydrocarbures Aromatiques Polycycliques. A) Exposome des pathologies professionnelles associées à l'exposition aux HAP, 127 nœuds et 1310 liens. B) Spectrosome des motifs de pathologies professionnelles associées aux HAP, 45 nœuds (motifs), 90 liens. C) Correspondances des codes de la CIM-10 utilisés pour l'exposome et le spectrosome.

Parmi les observations du RNV3P, 1 672 PST possédaient l'exposition aux HAP (la famille générale, sans précision) et un total de 221 pathologies ont été identifiées pour ces patients, parmi lesquelles 12 ont été considérées significatives grâce au calcul par le risque relatif et retrouvées dans 1 426 PST. A partir de ces cas, un total de 127 cortèges distincts de pathologies professionnelles a été identifié, formant les nœuds de l'exposome (Figure 75A), connectés entre eux par 1 310 liens. Ces nœuds ont permis la construction d'un total de 45 motifs représentatifs des HAP, dont 19 étaient d'ordre II et 2 d'ordre III, visibles sur le spectrosome en Figure 75B. Concernant le statut dynamique, 4 étaient persistants, 4 en déclin et 37 en état de latence. Aucun motif en émergence n'a été identifié. Il s'agissait majoritairement de pathologies des voies respiratoires ou de la vessie (Figure 75C).

On peut également imaginer diverses autres configurations autres que pathologie-exposition et exposition-pathologie (exemple des HAP) pour cette base de données professionnelle :

- Variable cible : **Pathologie** ; Variable explicative primaire : **Activité**

Cette configuration permettrait l'obtention d'un spectre des motifs d'activités professionnelles caractéristiques d'une pathologie professionnelle donnée et permettrait ainsi de cibler sur les activités professionnelles les plus à risques concernant cette activité professionnelle.

- Variable cible : **Exposition** ; Variable explicative primaire : **Activité**

Cette configuration permettrait l'obtention d'un spectre des activités professionnelles pour lesquelles l'exposition professionnelle a été signalée, permettant ainsi de révéler des activités professionnelles à risque encore peu reconnues pour certains types d'expositions.

- Variable cible : **Activité** ; Variable explicative primaire : **Pathologie**

Cette configuration permettrait cette fois-ci d'obtenir pour une activité professionnelle le spectre des pathologies professionnelles associées.

- Variable cible : **Activité** ; Variable explicative primaire : **Exposition**

Cette configuration permettrait d'obtenir pour une activité professionnelle le spectre global des expositions associées, pouvant mettre en évidence des activités professionnelles ayant une forte tendance à la mono-exposition ou au contraire à la multi-exposition.

La matrice primaire-secondaire

La matrice primaire secondaire (*voir 3.5.7, p.75*) n'a pas été développée de manière approfondie dans ce manuscrit, notamment par un souci de lisibilité. Les matrices primaires-secondaires établies pour les exemples du RNV3P concernaient les motifs d'expositions professionnelles en fonction des activités professionnelles recensées en très grand nombre rendant difficile la lecture du graphique et seul l'exemple du LNH a pu être inclus (à noter que l'exploitation de ces matrices ne nécessite pas les graphiques). Cependant, la matrice est disponible sous format tableur afin de pouvoir **investiguer en profondeur des contextes professionnels particuliers** alliant une activité professionnelle et une exposition professionnelle. Cet outil permettrait, à l'instar des matrices emploi-expositions, de faire un nouveau pas vers la compréhension des risques liés aux activités professionnelles et pourrait permettre la prise de décision pour la prévention des travailleurs : connaissant le secteur d'activité et les expositions, il est possible de connaître le panel des pathologies associées possibles.

PERSPECTIVES

Troisième axe de la SO : Prospection

Parmi les objectifs de la Surveillance on trouve d'abord la surveillance des évènements (en général et/ou choisis) et, ce faisant, la détection des phénomènes ou évènements émergents, mais un autre objectif est également de pouvoir effectuer de la prospection ou projection : tenant compte de l'état actuel des connaissances et de la surveillance sur un évènement à un temps donné *t*, il est alors possible d'effectuer des projections à un temps $t' > t$ pour faire de la prévention.

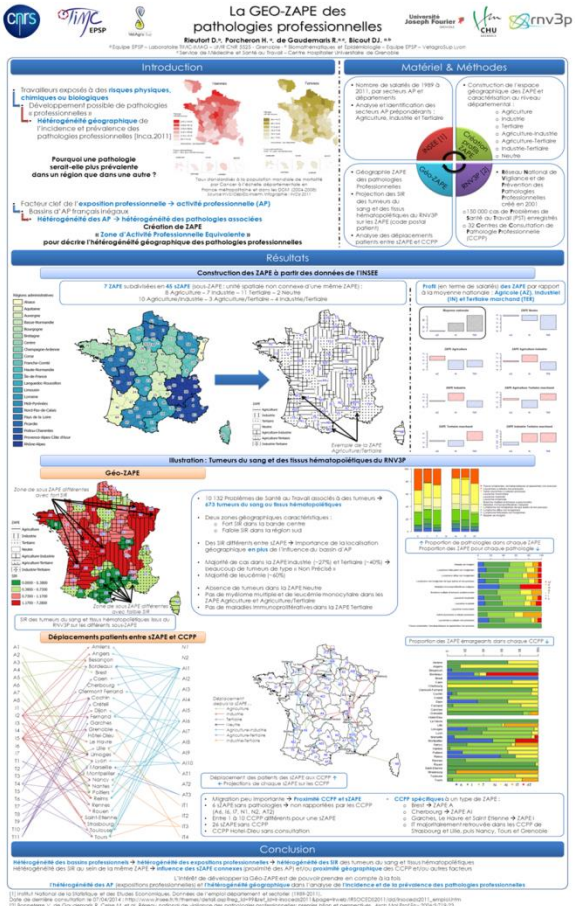
C'est dans cette optique que les précédents travaux de Laurie Faisandier [Faisandier, 2011a], appliqué à la base du RNV3P, avaient initié la réflexion sur le développement des Fonctions Expertes, définies comme :

« Un modèle probabiliste qui prend comme données d'entrée une pathologie et une activité professionnelle associée, et fournit en sortie une séquence de nuisances (1 à 5). Cette fonction génère ainsi des PST reflétant les pratiques de codages et les règles d'associations entre une pathologie et ses expositions professionnelles multiples, induisant une certaine connaissance. »

Le but étant de pouvoir, à partir des connaissances enregistrées au sein de la base, générer de nouveaux PST, et par conséquent des nouveaux nœuds de l'exposome. Durant les travaux actuels, des pistes avaient été envisagées concernant le développement des fonctions expertes et la prise en compte globale des *Cursus Laboris*, défini comme l'historique complet de l'environnement professionnel du patient. Une étude concernant la construction de *Cursus Laboris* de patients atteints d'Hémopathies et d'adénocarcinomes des sinus enregistrées dans le CCPP de Grenoble a été menée en 2013 [Kosicki, 2013], et avait permis d'identifier des profils différents d'activités professionnelles selon les pathologies. Cette étude avait alors renforcé l'importance de la prise en compte du contexte professionnel « activité-exposition » pour une pathologie donnée. Cependant, le développement n'a pas été approfondi car il paraissait nécessaire de développer une surveillance solide avant de pouvoir prétendre à effectuer de la prévention et intégrer des informations telles que les *Cursus Laboris* au sein des Fonctions Expertes. Il est à noter cependant qu'il serait envisageable d'utiliser les motifs générés lors du processus de la SO afin d'effectuer de la prospection. En effet, les motifs représentatifs sont issus de toutes les combinaisons possibles formées à partir des nœuds (*voir 3.3.2.3, p.59*), ce concept pourrait alors s'apparenter à la construction des fonctions expertes.

La géo-ZAPE

Lors des analyses effectuées sur le RNV3P, les expositions professionnelles ont été utilisées en tant que variables explicatives primaires afin de pouvoir caractériser les différentes pathologies choisies, ces expositions professionnelles étant intrinsèquement liées aux activités professionnelles des individus. Or, il a déjà été mis en évidence l'existence de l'hétérogénéité spatiale en terme d'activités professionnelles, et donc d'expositions professionnelles puis de pathologies professionnelles. Il serait alors naturel et pertinent d'incorporer l'aspect géographique à l'analyse de la SO. Différentes variables géographiques sont disponibles au sein du RNV3P mais n'ont pas été utilisées dans ce développement (CCPP d'enregistrement, code postal du patient, code postal entreprise, etc.). Afin d'améliorer la SO, une réflexion concernant l'incorporation de cet aspect géographique avait été amorcé en 2013 : la « géo-ZAPE » pour la « géographie des Zones d'Activités Professionnelles Equivalentes » [Porcheron, 2013]. Cette étude avait consisté en la construction de l'espace (ou paysage) des activités professionnelles équivalentes à l'échelle nationale dans lesquels les analyses pourraient être effectuées. Cette approche pourrait résulter en des analyses plus précises de secteurs d'activités grâce à une échelle géographique adaptée.



<u>Rieutort D</u>, Porcheron H, de Gaudemaris R, Bicout D

La Géo-ZAPE des pathologies professionnelles.

Journée de la Recherche Médicale - Action de Recherche Clinique Innovations Diagnostiques et Thérapeutiques ; 13 juin 2014 - UFR de Médecine et Pharmacie, Université Joseph Fourier, Grenoble. Poster. Résumé p. 93. Actes : 115 pages.

Voir Valorisation des travaux

Extension à l'analyse des bases de données ornithologiques

Une utilisation de la SO qui pourrait également être envisagée est l'exportation de cette méthodologie à des bases de données, ou simplement des recueils de données, qui ne sont pas obligatoirement structurés comme des bases de données observationnelles. Une étude de faisabilité a ainsi été mise en place afin d'étudier la possibilité d'appliquer la méthodologie à un contexte de zoonose, sur des données de terrain ornithologiques, dont la structure ne correspond pas à une base de données classique. Ce travail exploratoire se place dans le contexte de l'étude de l'influenza Aviaire dans la Dombes en France.

En 2006, la Dombes avait été la scène d'une épizootie mineure d'Influenza Aviaire Hautement Pathogène H5N1. Cet épisode avait soulevé la question des contacts entre avifaune sauvage et volailles domestiques. Des études mises en place par l'Institut Technique de l'Aviculture, l'équipe EPSP-TIMC/VetagroSup et la Société d'Etude et de Recherche sur les Oiseaux et leur Ecosystème avaient pour objectif d'évaluer les contacts directs et indirects entre les volailles sur les parcours et l'avifaune sauvage [Doctrinal, 2009]. La zone d'étude de la Dombes, caractérisée par de nombreux étangs artificiels (plus de 1 000) rassemble plus de 20 000 oiseaux d'eau de décembre à février. C'est également une zone d'arrêt d'oiseaux migrateurs et un site de reproduction d'environ 130 espèces d'oiseaux qui héberge une avifaune sédentaire importante. La période d'étude s'étendait du 15 janvier au 6 décembre 2008 et les données recueillies comprenaient des informations sur la topologie de chaque site, avec le détail de chaque unité écologique (UE, Tableau 15) présente et le nombre de contacts des 81 espèces d'oiseaux différentes étudiées, le nombre de contacts étant défini comme le nombre de fois où l'espèce a été vue ou entendue dans une unité écologique.

Tableau 15 : Libellés des 12 unités écologiques étudiées dans l'étude en Dombes en 2008.

Code	Libellé	Code	Libellé
CBO	Culture blé orge	HU	Habitat humain urbain
CH	Chaume	LA	Labour
ET	Etang	MR	Marais et roselière
FO	Forêt	PP	Prairie et pâture
HA	Haie	PV	Parcours de volaille
HR	Habitat urbain rural	RR	Rivière et ripisylve

Dans cette configuration, la **variable cible correspond aux espèces d'oiseaux** contactées et les **variables explicatives primaires correspondent à chacune des unités écologiques** rencontrées. Le but étant de pouvoir mettre en évidence des profils de fréquentation et potentiellement des groupes caractéristiques d'oiseaux dont le profil de fréquentation serait proche ou identique. Un premier essai préliminaire a été effectué et l'exposome associé est disponible en Figure 76. Chaque nœud correspond à une espèce d'oiseau particulière, la couleur des nœuds correspond à l'ordre de chacune de ces espèces, la taille des nœuds au nombre de contact total de l'espèce, et enfin la couleur des liens aux milieux fréquentés de manière commune par deux (ou plus) espèces d'oiseaux. Il apparaît déjà de manière visuelle la présence de groupe spécifique, et donc de cluster. Des espèces semblent être présentes uniquement sur les milieux « Etang » (ET en bas à droite de l'exposome) et « Rivière et ripisylve » (RR à gauche de l'exposome). Une espèce, la chouette effraie, semble également n'être présente que sur une unité écologique particulière : les « habitats urbains » (HU).

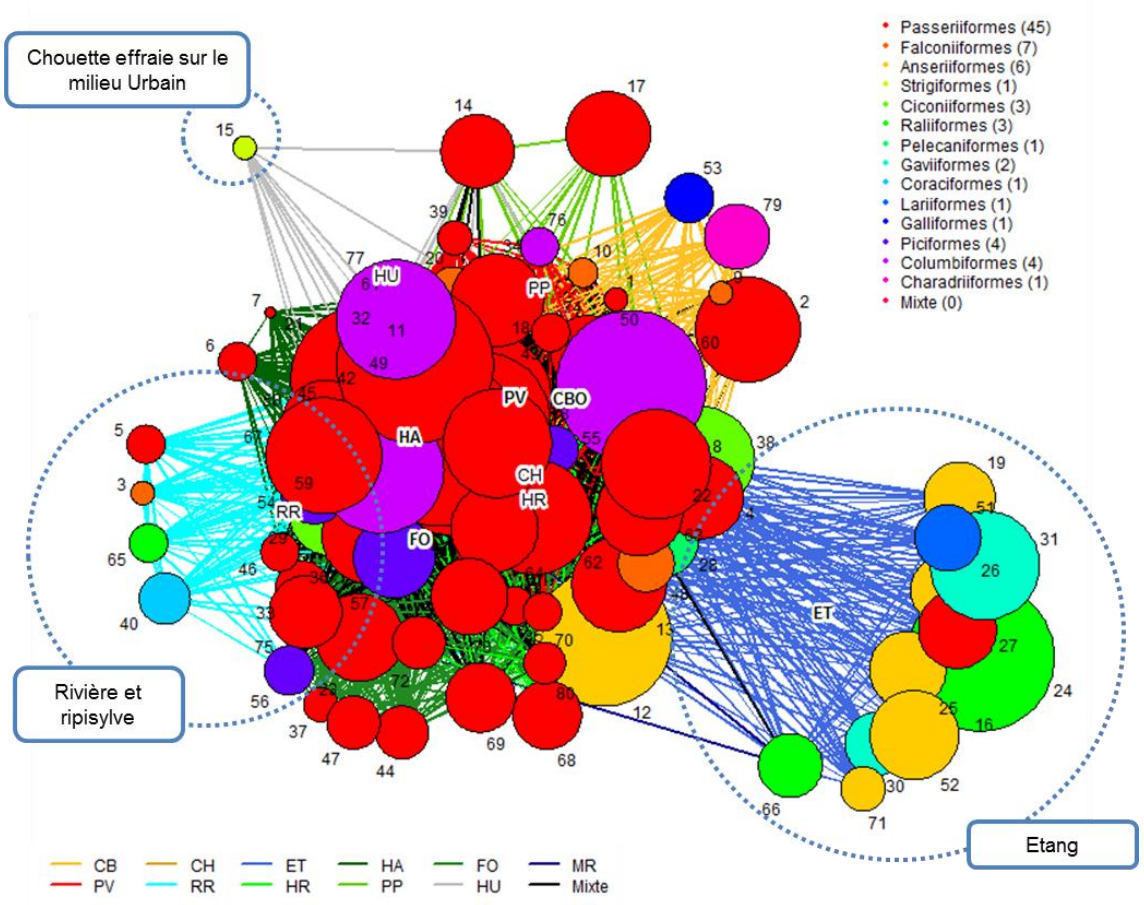


Figure 76 : Illustration de la Surveillance Observationnelle dans le contexte de l'influenza aviaire dans la région de la Dombes. Exposome de la fréquentation des oiseaux sur les différents milieux : 80 nœuds et 1 788 liens. Chaque nœud correspond à une espèce, la taille est proportionnelle au nombre de contact de l'espèce, la couleur des nœuds à l'ordre de l'espèce et la couleur des liens aux milieux fréquentés en commun par les oiseaux.

Cette approche préliminaire laisse **entrevoir les perspectives prometteuses de l'utilisation de la Surveillance Observationnelle à d'autres types de données** et d'autres contextes sanitaires autre que les pathologies professionnelles développées dans ces travaux. Il serait alors également possible d'établir l'exposome des unités écologiques fréquentées par les oiseaux.

Amélioration de l'outil OSA

Outre la théorie et son utilisation, des améliorations sont également envisageables pour cette version de l'outil.

La première, et constante amélioration, concerne la propreté du script et **l'optimisation du programme** afin de réduire le temps d'analyse. En effet, actuellement le temps de l'analyse est encore impacté par le nombre d'observations au sein de l'échantillon ainsi que la longueur maximale des nœuds et des motifs, deux paramètres augmentant de manière exponentielle le nombre possible de combinaisons.

Une autre amélioration serait l'intégration de la méthodologie et de l'outil au sein **d'un package R** afin de permettre son utilisation de manière plus large et dans un processus de surveillance systématique.

Une troisième amélioration possible concerne l'ergonomie de l'interface, permettant une utilisation plus fluide, mais également la gestion de **l'ouverture de l'interface sans avoir besoin de recourir au logiciel R**, comme l'utilisation des extensions « batch » par exemple.

Une autre amélioration potentielle serait également la **migration de l'outil dans une autre plateforme** de développement telle que Java par exemple, afin de permettre l'amélioration globale des points cités précédemment : ergonomie, aspect et utilisation sans le logiciel **R**, mais également son intégration au sein de système déjà développé, comme celui du réseau RNV3P par exemple.

CONCLUSION

Dans un contexte où la santé des populations et son interaction avec l'environnement est au cœur des préoccupations l'objectif de ce travail était de proposer une nouvelle Surveillance permettant l'utilisation optimale des bases de données massives en santé, ainsi que la prise en compte des origines multifactorielles des évènements d'intérêts.

Les travaux de ce manuscrit ont permis de présenter le développement et la conceptualisation de la Surveillance Observationnelle, de la théorie à la mise en pratique grâce à la création d'un outil « OSA » permettant son utilisation en routine. Au travers des différentes applications, il a été mis en évidence l'adaptabilité de la Surveillance Observationnelle aux différentes bases de données, mais également aux différentes variables potentiellement utilisables pour cette analyse. La surveillance Observationnelle semble alors prometteuse en tant qu'outil pour la veille sanitaire des évènements d'intérêts dont les origines sont multifactorielles, grâce à la génération de nouvelles hypothèses de travail à destination des acteurs des réseaux de surveillance.

Des développements concernant la partie prospection et l'adaptabilité à d'autres types de recueils de données sont en cours de réflexion, de même que l'amélioration de l'outil « OSA » permettant ainsi une intégration future dans des systèmes de surveillance déjà mis en place.

Références

Alavanja MC, Hofmann JN, Lynch CF, Hines CJ, Barry KH, Barker J, et al.
Non-hodgkin lymphoma risk and insecticide, fungicide and fumigant use in the agricultural health study.
PLoS One. 2014;9(10):e109332.

Alexander DD, Mink PJ, Adami HO, Chang ET, Cole P, Mandel JS, et al.
The non-Hodgkin lymphomas: a review of the epidemiologic literature.
Int J Cancer. 2007;120(Suppl 12):1-39.

Alexander DD, Wagner ME.
Benzene exposure and non-Hodgkin lymphoma: a meta-analysis of epidemiologic studies.
J Occup Environ Med. 2010;52(2):169-89.

ANSES.
Réseau National de Vigilance et de Prévention des Pathologies Professionnelles. Rapport scientifique.
Maison-Alfort: ANSES; 2011. 282 p.

Arnaudo B, Léonard M, Sandret N, Cavet M, Coutrot T, Rivalin R.
L'évolution des risques professionnels dans le secteur privé entre 1994 et 2010: premiers résultats de l'enquête SUMER.
Dares Analyses. 2012(23):1-10.

Astagneau P.
Définitions et concepts.
In: Astagneau P, Ancelle T. Surveillance épidémiologique Principes, méthodes et applications en santé publique.
Paris: Médecine Sciences Publications, Lavoisier; 2011. p. 7-12.

Bonneterre V, Faisandier L, Bicout D, Bernardet C, Piollat J, Ameille J, et al.
Programmed health surveillance and detection of emerging diseases in occupational health: contribution of the French national occupational disease surveillance and prevention network (RNV3P).
Occup Environ Med. 2009;67(3):178-86.

Brown SC, Alberts R, Schoenberg M.
Cancer incidence and mortality among workers exposed to benzidine.
Am J Ind Med. 2011;54(4):300-6.

Brown T, Slack R, Rushton L.
Occupational cancer in Britain. Urinary tract cancers: bladder and kidney.
Br J Cancer. 2012;107(Suppl 1):S76-84.

Case RA, Hosker ME, McDonald DB, Pearson JT.
Tumours of the urinary bladder in workmen engaged in the manufacture and use of certain dyestuff intermediates in the British chemical industry. I. The role of aniline, benzidine, alpha-naphthylamine, and beta-naphthylamine.
Br J Ind Med. 1954;11(2):75-104.

Cocco P, t'Mannetje A, Fadda D, Melis M, Becker N, de Sanjose S, et al.
Occupational exposure to solvents and risk of lymphoma subtypes: results from the Epilymph case-control study.
Occup Environ Med. 2010;67(5):341-7.

Colt JS, Karagas MR, Schwenn M, Baris D, Johnson A, Stewart P, et al.
Occupation and bladder cancer in a population-based case-control study in Northern New England.
Occup Environ Med. 2011;68(4):239-49.

Coquin Y.
Surveillance et décision en santé publique.
In: Astagneau P, Ancelle T. Surveillance épidémiologique: principes, méthodes et applications en santé publique.
Paris: Médecine Sciences Publications, Lavoisier; 2011. p. 287-91.

Cramer H.
Mathematical methods of statistics. Princeton Mathematical Series.
Princeton (United States): Princeton University Press; 1946. 575 p.

De Roos AJ, Bhatti P.
Lymphohematopoietic Malignancies.
In: Anttila S, Boffetta P, editors. Occupational Cancers.
London: Springer; 2014. p. 497-529.

Doctrinal D, Ruette S, Hars J, Artois M, Bicout D.
Spatial and temporal analysis of the highly pathogenic asian influenza H5N1 outbreak in the Dombes area, France, in 2006.
Wildfowl. 2009;2:202-14.

Dryson E, t Mannetje A, Walls C, McLean D, McKenzie F, Maule M, et al.
Case-control study of high risk occupations for bladder cancer in New Zealand.
Int J Cancer. 2008;122(6):1340-6.

Faisandier L, Bonneterre V, De Gaudemaris R, Bicout DJ.
Elaboration d'une méthode statistique pour la détection d'évènements émergents : application au réseau national de vigilance et de prévention des pathologies professionnelles (RNV3P).
Epidémiol et santé anim. 2007;51:111-8.

Faisandier L.
De la surveillance à la détection des Problèmes de Santé au Travail potentiellement émergents dans un contexte de multi-exposition : Développement de méthodes appliquées au Réseau National de Vigilance et de Prévention des Pathologies Professionnelles.
Thèse de Sciences: Modèles, Méthodes et Algorithmes pour la Biologie, la Santé et l'Environnement. Grenoble: Université de Grenoble; 2011. 268 p.

Faisandier L, Bonneterre V, De Gaudemaris R, Bicout DJ.
Occupational exposome: a network-based approach for characterizing Occupational Health Problems.
J Biomed Inform. 2011;44(4):545-52.

Finger V, Sari-Minodier I, Charpin D.
Epidemiologie.
In: Bessot JC, Pauli G, Vandeplas O. L'Asthme professionnel 2ème édition.
Paris: Margaux Orange; 2012. p. 13-26.

Goh KI, Cusick ME, Valle D, Childs B, Vidal M, Barabasi AL.
The human disease network.
Proc Natl Acad Sci U S A. 2007;104(21):8685-90.

Handcock MS, Hunter DR, Butts CT, Goodreau SM, Krivitsky PN, Bender-deMoll S, et al.
Package Statnet. Software Tools for the Statistical Analysis of Network Data. R package version 3.1-0.
Irvine, CA (USA): The Statnet Project; 2015.

IARC.
World cancer report 2008.
Lyon: International Agency for Research on Cancer; 2008. 260 p.

IARC.
Some Aromatic Amines, Organic Dyes and Relative Exposures. IARC Monographs on the Evaluation of Carcinogenic Risks to Humans. Volume 99.
Lyon: International Agency for Research on Cancer; 2010. 692 p.

IARC.
Globocan [dernière consultation: 15 septembre 2015].
Lyon: IARC; 2012. Disponible: *http://globocan.iarc.fr*.

IDEWE.
Présentation du groupe IDEWE [dernière consultation: 14 septembre 2015].
Louvain (Belgique): IDEWE; 2015. Disponible: *http://www.idewe.be*.

INCa.
La situation du cancer en France en 2012.
Paris: Institut National du Cancer; 2012 décembre 2012. 352 p.

INCa.
Plan cancer 2014-2019. Guérir et prévenir les cancers : donnons les mêmes chances à tous, partout en France.
Paris: Ministère des Affaires sociales, de la Santé et des Droits des femmes; 2014. 152 p.

INRS.
Les maladies professionnelles. Guide d'accès aux tableaux du régime général et du régime agricole de la Sécurité sociale. Mise à jour mai 2012. ED 835.
Paris: INRS; 2012. 361 p.

INSERM, InVS.
Présentation du réseau Sentinelles [dernière consultation: 14 septembre 2015].
Paris; 2015. Disponible: *https://websenti.u707.jussieu.fr/sentiweb/?page=presentation*.

InVS. [a]
Centres Nationaux de Référence [dernière consultation: 14 septembre 2015].
Saint-Maurice; 2012. Disponible: *http://www.invs.sante.fr/Espace-professionnels/Centres-nationaux-de-reference/Missions*.

InVS. [b]
Maladies à déclaration obligatoire [dernière consultation: 14 septembre 2015].
Saint-Maurice; 2012. Disponible: *http://www.invs.sante.fr/Dossiers-thematiques/Maladies-infectieuses/Maladies-a-declaration-obligatoire/31-maladies-a-declaration-obligatoire*.

InVS.
Surveillance et signalement des infections nosocomiales et associées aux soins dans le cadre du RAISIN [dernière consultation: 14 septembre 2015].
Saint-Maurice; 2013. Disponible: *http://www.cclin-arlin.fr/RAISIN.html*.

InVS.
Surveillance Syndromique - SurSaUD[dernière consultation: 14 septembre 2015].
Saint-Maurice; 2015. Disponible: *http://www.invs.sante.fr/Espace-professionnels/Surveillance-syndromique-SurSaUD-R*.

Kane EV, Newton R.
Occupational exposure to gasoline and the risk of non-Hodgkin lymphoma: a review and meta-analysis of the literature.
Cancer Epidemiol. 2010;34(5):516-22.

Karami S, Bassig B, Stewart PA, Lee KM, Rothman N, Moore LE, et al.
Occupational trichloroethylene exposure and risk of lymphatic and haematopoietic cancers: a meta-analysis.
Occup Environ Med. 2013;70(8):591-9.

Karipidis KK, Benke G, Sim MR, Kauppinen T, Kricker A, Hughes AM, et al.
Occupational exposure to ionizing and non-ionizing radiation and risk of non-Hodgkin lymphoma.
Int Arch Occup Environ Health. 2007;80(8):663-70.

Kiriluk KJ, Prasad SM, Patel AR, Steinberg GD, Smith ND.
Bladder cancer risk from occupational and environmental exposures.
Urol Oncol. 2012;30(2):199-211.

Kogevinas M, Garcia-Closas M.
Bladder Cancer.
In: Anttila S, Boffetta P, editors. Occupational Cancers.
London: Springer; 2014. p. 461-80.

Kosicki A.
Caractérisation des Cursus Laboris des Pathologies Professionnelles au sein du RNV3P.
Rapport Master 1. Ecole doctorale EDISCE. Université Joseph Fourier, 2013.

Massardier-Pilonchery A, Charbotel B, Normand JC, Bergeret A.
Cancers professionnels. 16-532-A-10.
In: Encyclopédie médico-chirurchicale - Pathologie professionnelle et de l'environnement. Paris: Elsevier Masson; 2013. p. 1-12.

ONCFS.
Le réseau SAGIR [dernière consultation: 14 septembre 2015].
Paris; 2014. Disponible: *http://www.oncfs.gouv.fr/Reseau-SAGIR-ru105*.

Porcheron H.
Caractérisation de la variable imputabilité des bases de données observationnelles, exemple du réseau RNV3P.
Rapport Master 1. Ecole doctorale EDISCE. Université Joseph Fourier, 2013.

Purdue MP, Bakke B, Stewart P, De Roos AJ, Schenk M, Lynch CF, et al.
A case-control study of occupational exposure to trichloroethylene and non-Hodgkin lymphoma.
Environ Health Perspect. 2011;119(2):232-8.

R Core Team.
R: A language and environment for statistical computing. Version 3.1-0.
Vienna (Austria): R Foundation for Statistical Computing; 2013.

Rappaport SM.
Implications of the exposome for exposure science.
J Expo Sci Environ Epidemiol. 2011;21(1):5-9.

Reulen RC, Kellen E, Buntinx F, Brinkman M, Zeegers MP.
A meta-analysis on the association between bladder cancer and occupation.
Scand J Urol Nephrol Suppl. 2008(218):64-78.

Salines G, Eilstein D.
Santé Environnement.
In: Astagneau P, Ancelle T. Surveillance épidémiologique: principes, méthodes et applications en santé publique.
Paris: Médecine Sciences Publications, Lavoisier; 2011. p. 166-76.

Schinasi L, Leon ME.
Non-Hodgkin lymphoma and occupational exposure to agricultural pesticide chemical groups and active ingredients: a systematic review and meta-analysis.
Int J Environ Res Public Health. 2014;11(4):4449-527.

Shannon CE.
A Mathematical Theory of Communication.
Bell Syst Tech J. 1948;27(3):379-423.

Toren K, Blanc PD.
Asthma caused by occupational exposures is common - a systematic analysis of estimates of the population-attributable fraction.
BMC Pulm Med. 2009;9(1):7.

Vlaanderen J, Lan Q, Kromhout H, Rothman N, Vermeulen R.
Occupational benzene exposure and the risk of lymphoma subtypes: a meta-analysis of cohort studies incorporating three study quality dimensions.
Environ Health Perspect. 2011;119(2):159-67.

Vrijheid M, Robinson O, Basagaña Flores X, Bustamante Pineda M, Casas M, Estivill X, et al.
The human early-life exposome (HELIX): project rationale and design.
Environ Health Perspect. 2014;122(6):535-44.

WHO.
Asthma, fact sheet n°307 [dernière consultation: 14 septembre 2015].
Geneva, WHO; 2013. Disponible: *http://www.who.int/mediacentre/factsheets/fs307/en/*.

Wild CP.
Complementing the genome with an "exposome": the outstanding challenge of environmental exposure measurement in molecular epidemiology.
Cancer Epidemiol Biomarkers Prev. 2005;14(8):1847-50.

Winslow CE.
The Untilled Fields of Public Health.
Science. 1920;51(1306):23-33.

Valorisation des travaux

Articles Scientifiques

Rieutort D, de Gaudemaris R, Bicout D
[Development of the Observational Surveillance.]
En cours d'écriture

Rieutort D, Moyne O, Cocco P, de Gaudemaris R, Bicout D
Ranking occupational contexts associated with risk of non-Hodgkin Lymphoma
(Soumis dans Am J Ind Med).

Rieutort D, De Gaudemaris R, Bicout DJ.
Surveillance Observationnelle : Approche par Exposome.
Epidémiol santé anim. 2012;61:127-40.

Communications orales

Rieutort D, de Gaudemaris R, Bicout D
Approche par exposome appliquée aux données du RNV3P. [Conférence téléphonique]
Groupe de travail « Méthodologie », ANSES RNV3P. Grenoble; 27 juillet 2015.

Delaunay M, Petit P, Lutier S, Rieutort D (par ordre alphabétique).
Présentation des travaux des doctorants de l'équipe EPSP.
Evaluation de l'HCERES du laboratoire TIMC-IMAG. Grenoble; 16 janvier 2015.

Rieutort D, de Gaudemaris R, Bicout D
Databased Observational Surveillance on Occupational Asthma.
ICOH. Aware, beware, take care! New insights in occupational health surveillance. Bologna (Italy); October 15-16th, 2014.

Bicout D, Rieutort D, de Gaudemaris R
Workshop on the Observational Surveillance.
COST Meeting MODERNET. Zaragoza (Spain); April 9-11th, 2014.

Rieutort D, de Gaudemaris R, Bicout D
Développement d'une application sous R pour la Surveillance des Problèmes de Santé au Travail.
Deuxième rencontre R. Statistiques appliquées. Lyon; 27-28 juin 2013.
Diaporama 27 pages, disponible : http://r2013-lyon.sciencesconf.org/1886

Rieutort D, de Gaudemaris R, Bicout D
Observational Surveillance of Occupational Health Problems.
COST Meeting MODERNET. Bucharest (Romania); May 29-31th, 2013.

Rieutort D, de Gaudemaris R, Bicout D
La Surveillance Observationnelle, approche par exposome.
Groupe de travail « Méthodologie », ANSES RNV3P. Maison-Alfort; 26 septembre 2012.

Rieutort D, de Gaudemaris R, Bicout D

La Surveillance Observationnelle, approche par exposome.

Journées scientifiques de l'AEEMA. Thème : Surveillance syndromique - Communications libres. Maison-Alfort; 30 mai -1er juin 2012.

> ***Communication avec publication :*** Rieutort D, De Gaudemaris R, Bicout DJ. Surveillance Observationnelle : Suveillance Observationnelle : Approche par Exposome. Epidémiol santé anim. 2012;61:127-40.

Communications affichées

Rieutort D, de Gaudemaris R, Bicout D

Databased Observational Surveillance of occupational non Hodgkin lymphoma.

European Congress of Epidemiology 2015. June 25-27th, 2015 – Maastricht, The Netherlands.

Rieutort D, Porcheron H, de Gaudemaris R, Bicout D

La Geo-ZAPE des pathologies professionnelles.

Journée de la Recherche Médicale - Action de Recherche Clinique Innovations Diagnostiques et Thérapeutiques ; 13 juin 2014 - UFR de Médecine et Pharmacie, Université Joseph Fourier, Grenoble. Poster. Résumé p. 93. Actes : 115 pages.

Rieutort D, de Gaudemaris R, Bicout D

De la Consultation Professionnelle à la Surveillance et la Détection d'Evènements Emergents. Journée de l'EDISCE - 2e année de doctorat. Poster.

Bicout D, Rieutort D, de Gaudemaris R

Exposomes pour la Surveillance des Problèmes de Santé au Travail.

Journées interuniversitaires, Universités de Grenoble - 26 septembre 2013 – Grenoble.

Rieutort D, de Gaudemaris R, Bicout D

La Surveillance Observationnelle des Problèmes de Santé au Travail.

Journée de la Recherche Médicale - Action de Recherche Clinique Innovations Diagnostiques et Thérapeutiques - 7 juin 2013 - UFR de Médecine et Pharmacie, Université Joseph Fourier, Grenoble. Poster. Résumé p. 95. Actes : 120 pages.

Épidémiol. et santé anim., 2012, **61**, 127-140

SURVEILLANCE OBSERVATIONNELLE : APPROCHE PAR EXPOSOME*

Delphine Rieutort[1], **Régis de Gaudemaris**[1,2] et **Dominique J. Bicout**[3]

RÉSUMÉ

La surveillance observationnelle, fondée sur l'exposome, est un nouveau concept permettant de prendre en compte de manière optimale toutes les informations issues de bases observationnelles. La surveillance observationnelle au sein du Réseau national de vigilance et de prévention des pathologies professionnelles (RNV3P) s'organise en trois étapes : la construction de l'exposome et des groupes d'expositions et la génération d'une matrice activités-groupes d'expositions. Le concept est illustré par l'exemple des lymphomes non hodgkiniens (LNH). Plusieurs groupes d'expositions concordants avec la littérature ont été mis en évidence : solvants et diluants organiques (dont benzène et trichloroéthylène), produits agricoles et radiations ionisantes. La surveillance observationnelle a permis une analyse complémentaire permettant d'apporter et de conforter des hypothèses de travail concernant les maladies professionnelles.

Mots-clés : exposome, surveillance, RNV3P, exposition professionnelle, lymphome non hodgkinien, données observationnelles, maladie professionnelle.

SUMMARY

Observational surveillance, based on the exposome, is a new concept designed to take into account optimally all the information obtained from observational databases. Observational surveillance in the National Network for Vigilance and Prevention of Occupational Diseases (RNV3P) is organized in three steps: the construction of exposome, that of exposure groups and the generation of Matrix Activity-Exposure groups. The concept is illustrated by the example of non-Hodgkin lymphoma (NHL). Several exposure groups consistent with the literature were identified: organic solvents and thinners (including benzene and trichloroethylene), agricultural products and ionizing radiation. This observational surveillance has allowed an additional analysis to improve or strengthen working hypotheses concerning occupational diseases.

Keywords: Exposome, Surveillance, RNV3P, Occupational exposure, Non Hodgkin Lymphoma, Observational data, Occupational disease.

* Texte de la communication orale présentée au cours des Journées scientifiques AEEMA, 1er juin 2012

1 Laboratoire Environnement et prédiction de la santé des populations – TIMC, UMR CNRS 5525 Université Joseph Fourier, Grenoble, Domaine de la Merci, 38706 La Tronche Cedex, France
2 Service de médecine et santé au travail, CHU Grenoble, BP 217, 38043 Grenoble cedex 09 France
3 Biomathématiques et épidémiologie, EPSP – TIMC, UMR CNRS 5525 Université Joseph Fourier, Grenoble, VetAgro Sup Lyon, 69280 Marcy l'Etoile, France
(delphine.rieutort@imag.fr, dominique.bicout@vetagro-sup.fr, bicout@ill.fr)

I - INTRODUCTION

A la suite de récentes crises sanitaires, les processus de surveillance se sont multipliés. Ces réseaux de surveillance ont donné naissance à des bases de données enregistrant diverses informations concernant tous types d'évènements sanitaires dans un but de prévention et de volonté de détecter précocement des menaces pour la santé des populations. On ne parle alors plus de veille sanitaire mais de surveillance épidémiologique (ou surveillance en santé publique), car celle-ci permet une démarche pro-active ayant pour but de générer des alertes. Cette surveillance épidémiologique est un processus de collecte systématique et continu d'évènements de santé connus, que l'on suit dans le temps en utilisant des indicateurs épidémiologiques courants. La surveillance peut se concevoir sous plusieurs aspects (tableau 1) : la surveillance programmée, qui consiste en un recueil d'informations sur une unique maladie ; la surveillance évènementielle, qui consiste en un recueil d'informations sur un évènement particulier ; et la surveillance syndromique qui consiste en un recueil de diverses informations ou syndromes permettant de générer des alertes précoces sur divers évènements non définis. La surveillance observationnelle est un nouveau concept permettant l'exploitation optimale des bases de données de déclarations ou observationnelles. Ces bases de données recueillent des données sur de nombreux évènements de santé, ainsi que de multiples informations qui leur sont associées. La surveillance observationnelle est donc une surveillance complexe permettant une prise en charge globale de toutes les caractéristiques liées aux évènements de santé enregistrés dans les bases observationnelles par les différents acteurs des réseaux

Tableau 1

Les différents types de surveillance

Type de surveillance	Evènement suivi	Exemple
Programmée	Une maladie particulière	Grippe
Evènementielle	Un évènement particulier	Hiroshima
Syndromique	Divers syndromes issus de différentes bases de données	Registre des urgences ou pompiers, SAGIR
Observationnelle	Plusieurs évènements de santé et toutes les caractéristiques associées	RNV3P, SAGIR

Ces bases de données de déclarations existent dans différents domaines. Pour n'en citer que deux, il existe le réseau SAGIR [SAGIR, 2012], créé en 1986, recueillant les signalements volontaires des animaux morts ou moribonds effectués par des chasseurs, techniciens de forêt, ou encore par le public ; et le Réseau national de vigilance et de prévention des pathologies professionnelles (RNV3P) [Anses, 2012] ; créé en 2001 dans un but de surveillance, d'émergence et de prospection des associations pathologies-expositions professionnelles, il recueille de manière systématique et standardisée toutes les informations issues des consultations professionnelles effectuées par les médecins experts du réseau.

Les objectifs de ces travaux sont de développer et de poser les bases d'une surveillance observationnelle fondée sur l'exposome ayant pour but de prendre en compte et d'utiliser de manière optimale ces bases de données observationnelles.

La surveillance observationnelle telle que nous la concevons au sein du RNV3P se décline en trois étapes principales : la construction de l'exposome, la construction des groupes d'expositions (GEXPO) et la génération d'une matrice activités-groupes d'expositions (MAGE) [Rieutort, 2012]. Seules les étapes concernant l'exposome et les groupes d'expositions sont présentés ici. L'exposome et les groupes d'expositions associés ont été développés précédemment par l'équipe EPSP [Faisandier, 2011 ; Faisandier *et al.*, 2011]. Le concept d'exposome est similaire à celui de la science des réseaux complexes [Barabási, 2012]. Des approches semblables sont

retrouvées dans la littérature comme l'exposome à la Wild [Wild, 2005] représentant la synthèse de toutes les expositions individuelles et environnementales rencontrées au cours de la durée de vie d'un individu, ou le « diseasome » illustrant par exemple l'environnement et les facteurs sociaux qui pourraient avoir un rôle potentiel dans les origines de l'obésité [Goh *et al.*, 2007 ; Barabási, 2007]. Ces approches ont en commun la volonté de rechercher les caractéristiques similaires permettant la compréhension et l'identification des facteurs clefs dans l'apparition et le développement de maladies.

Afin de présenter cette surveillance observationnelle, l'exemple des lymphomes non hodgkiniens (LNH) issus de la base RNV3P sera utilisé.

II - MATÉRIEL ET MÉTHODES

1. RÉSEAU RNV3P

Le RNV3P est un réseau d'experts mis en place en 2001 dans un but de veille épidémiologique programmée (suivi au cours du temps de l'évolution des couples pathologies-nuisances) ; de veille clinique ; et de recherche afin d'élaborer des outils de mise en évidence d'associations potentiellement émergentes [Faisandier, 2011].

Les médecins-experts du RNV3P recueillent de manière systématique et standardisée toutes les informations issues des consultations professionnelles effectuées dans les Centres de consultations des pathologies professionnelles (CCPP) des Centres hospitaliers universitaires (CHU) de France. Durant ces consultations, les médecins experts retracent l'historique professionnel du patient, et déterminent les activités professionnelles et nuisances potentiellement responsables de la pathologie diagnostiquée. Ces informations sont alors structurées sous forme de problème de santé au travail (PST).

Chaque PST est caractérisé par une pathologie p diagnostiquée par un médecin-expert, une ou plusieurs nuisances (substances, mouvement, conditions de travail) h_1,h_2,h_3,h_4,h_5 (5 maximum), potentiellement associées à cette pathologie et l'activité professionnelle jugée comme responsable par le médecin expert, définie par le secteur d'activité s et le métier m du patient (figure 1).

La base est actuellement composée de 81 132 PST. Les pathologies sont codées avec les codes issus de la CIM-10 et les nuisances avec la CNAM-TS.

2. SURVEILLANCE OBSERVATIONNELLE

La surveillance observationnelle s'organise selon trois étapes successives : la construction de l'exposome, des groupes d'expositions et la génération d'une matrice activité-groupes d'expositions qui ne sera pas présentée ici [Rieutort, 2012].

Figure 1

Structure d'un problème de santé au travail (PST) au sein de la base observationnelle du RNV3P

2.1. EXPOSOME

L'exposome est un réseau de PST (nœuds) connectés entre eux s'ils partagent un ou plusieurs éléments de l'exposition professionnelle [Faisandier, 2011]. Son principe et sa construction ont été décrits lors de précédent travaux [Faisandier *et al.*, 2011]. Comme tout réseau quelconque, l'exposome peut être caractérisé par les paramètres suivants :

- *Degré de connectivité d'un nœud :* nombre de nœuds auquel il est connecté ;
- *Coefficient de clustering :* probabilité que les voisins d'un nœud soient également connectés entre eux ;
- *Densité :* rapport entre le nombre de liens total du réseau et le nombre de liens maximum possible. Une densité égale à 1 correspond à un réseau complet où chaque nœud est connecté à tous les autres nœuds du réseau ;
- *Assortativité :* coefficient de corrélation de Pearson appliqué à la distribution des nœuds du réseau. Il traduit la tendance des nœuds à se lier avec des nœuds ayant les mêmes degrés de connectivité.

La mise en réseau de l'information permet, en plus d'une analyse exploratoire et analytique des PST, une restructuration des données pour leur exploitation dans la suite de la démarche de la surveillance observationnelle.

2.2. GROUPES D'EXPOSITIONS

Les groupes d'expositions sont des ensembles de nœuds ou de PST distincts partageant exactement D éléments identiques de l'exposition professionnelle, c'est-à-dire D nuisances ou associations de nuisances identiques, avec $1 \leq D \leq 4$. Le nombre minimal de nœuds (seuil) permettant la formation des groupes d'expositions varie selon le niveau de précision de l'étude. Pour ces travaux nous avons choisi un seuil égal à 2, un groupe d'expositions est formé si deux nœuds au minimum partagent D nuisances ou associations de nuisances similaires. On parle de groupes d'expositions simples lorsqu'il n'y a qu'une exposition particulière identique et de groupes d'expositions multiples lorsqu'il y a association de deux à quatre nuisances identiques (tableau 2). Un groupe d'expositions multiple est toujours composé de groupes d'expositions simples.

Tableau 2

Les différents niveaux d'ordres des groupes d'expositions

Ordre	Nuisance	Type d'exposition	Evènement	Description
Ordre 1	1	**Mono-exposition**	A	Déjà présent
			B	Nouveau groupe
Ordre 2	2	**Multi-exposition**	C	Nouveau groupe à partir de A
Ordre 3	3		D	Nouveau groupe à partir de A et B
Ordre 4	4		E	Nouveau groupe à partir de B

La construction de ces groupes d'expositions se fait par la formation d'une matrice d'exposition B. Soit H l'ensemble des nuisances distinctes α et V l'ensemble des nœuds i du réseau, alors la matrice d'exposition $B = H \times V$. Si la nuisance ou l'association de nuisances α sont présentes dans un nœud i, alors l'intersection $B_{\alpha,i}$ est égale à 1, sinon 0. Si pour un nœud α la $\sum B_i > D - 1$ alors un groupe d'expositions d'ordre 1 est créé (tableau 2), prenant le nom de la nuisance α. Pour les groupes d'expositions multiples, si pour plusieurs nœuds α la somme des produits de leur B_i, est supérieure à $D - 1$, des groupes d'expositions d'ordre 2, 3 ou 4 sont formés (tableau 2).

Les groupes d'expositions peuvent se former au cours du temps de manière différente. Il y a cinq possibilités de formation : ce sont les évènements d'intérêts des groupes d'expositions (tableau 2). Deux sont liés à la mono-exposition et trois sont liés à la multi-exposition.

A partir des groupes d'expositions, nous pouvons définir des indicateurs pour la surveillance observationnelle : *Groupes d'expositions :*

- « Prévalence » : Nombre de groupes d'expositions par an suivant leur ordre ;
- « Incidence » : Nombre de nouveaux groupes d'expositions par an suivant leur ordre ;
- Proportion P_g : Nombre de nœuds ou PST ayant le groupe d'exposition g dans son cortège de nuisances.

Evènements d'intérêts : « Incidence » des évènements d'intérêts, A à E (tableau 2).

3. LOGICIELS ET STATISTIQUES

Le logiciel open source R [RDevelopmentCoreTeam, 2011] a été utilisé en tant que logiciel de programmation, afin de mettre en place les composants nécessaires à la mise en place du processus de surveillance : exposome et groupes d'expositions ; et de statistiques pour les tests de corrélation, de significativité et les graphiques.

Le logiciel Gephi [Bastian *et al.*, 2009] a été utilisé pour la création de l'exposome, notamment les outils de spatialisation, de classement et de chronologie.

Le logiciel Excel a servi de support à la base de données et aux résultats obtenus grâce au logiciel R.

III - RÉSULTATS

Le LNH est un cancer rare du système lymphatique (3% de tous les cancers confondus), causé par une prolifération de lymphocytes qui s'accumulent le plus souvent dans les ganglions lymphatiques mais qui peut toucher de nombreuses autres structures du système lymphatique comme la moelle osseuse, la rate, le thymus ou les vaisseaux sanguins. Les LNH représentent 90% des lymphomes [Neasham *et al.,* 2011] et sont subdivisés en plusieurs groupes, les lymphomes malins agressifs et les lymphomes non hodgkiniens indolents. Ces deux catégories de lymphomes regroupent encore de nombreux sous-types, très différents les uns des autres d'un point de vue cytogénétique, morphologique, immunologique et diagnostic, rendant leur classification très complexe [Zhang *et al.,* 2011]. Leurs étiologies sont multiples et encore mal définies pour certaines. Un des facteurs majeurs est le facteur immunologique, les personnes souffrant d'immunodéficience auraient un risque plus élevé de développer un LNH, notamment les porteurs du virus de l'immunodéficience humaine (VIH). Des facteurs comportementaux ou liés au style de vie sont également suggérés, tels que le tabagisme ou la consommation d'alcool [Zhang *et al.,* 2011]. Les expositions professionnelles sont les plus discutées malgré de nombreuses études effectuées depuis plusieurs dizaines d'années. Les hypothèses les plus controversées à l'heure actuelle concernent les solvants organiques [Boffetta *et al.,* 2007 ; Cocco *et al.,* 2010 ; Hartge *et al.,* 2007 ; Karunanayake *et al.,* 2009 ; Purdue *et al.,* 2011 ; Neasham *et al.,* 2011 ; Wang *et al.,* 2009 ; Zhang *et al.*, 2011], dont le benzène [Alexander *et al.,* 2010 ; Hohenadel *et al.,* 2011 ; Smith *et al.,* 2007 ; Vlaanderen *et al.,* 2011], le trichloroéthylène, le tétrachloroéthylène et le toluène ; le diesel [Hohenadel *et al.,* 2011 ; Karunanayake *et al.,* 2011] et l'essence [Cocco *et al.,* 2011 ; Neasham *et al.,* 2011] ; le formaldéhyde [Hohenadel *et al.,* 2011 ; Wang *et al.,* 2009] ; les produits agricoles (pesticides, fongicides et herbicides) [Boffetta *et al.,* 2007 ; Braüner *et al.,* 2012 ; Hartge *et al.,* 2007 ; Hohenadel *et al.,* 2011 ; Karunanayake *et al.*, 2009 ; McDuffie *et al.*, 2009 ; Orsi *et al.,* 2009] ; les virus [Boffetta *et al.,* 2007 ; Braüner *et al.,* 2012 ; Hartge *et al.,* 2007 ; Neasham *et al.,* 2011] et enfin les radiations ionisantes [Hartge *et al.,* 2007 ; Karunanayake *et al.*, 2009].

Dans un souci d'amélioration de nos connaissances à propos de cette pathologie, il est intéressant de mener une étude de surveillance observationnelle, à partir du RNV3P, afin d'analyser les données et les informations que peut nous apporter une base observationnelle.

1. STATISTIQUES DESCRIPTIVES DE L'ÉCHANTILLON

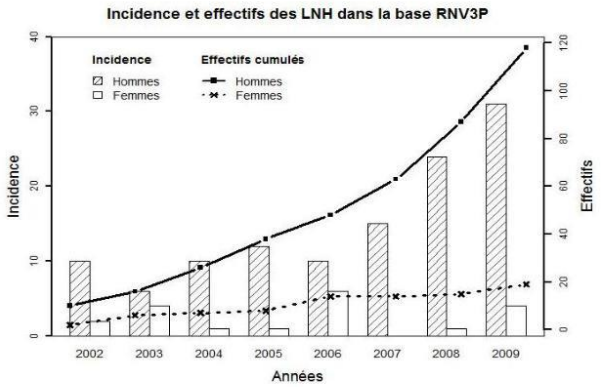
L'échantillon était composé de 137 PST de LNH enregistrés dans la base de 2002 à 2009, dont 16 lymphomes folliculaires non hodgkiniens (Classification CIM-10 : C82), 21 lymphomes diffus non hodgkiniens (C83) et 100 lymphomes non hodgkiniens de types autres et non précisés (C85), sur un total de 11 505 cancers, soit 1,2% de tous les cancers enregistrés dans la base.

Parmi ces 137 PST, 118 concernaient des hommes et 19 des femmes (figure 2). Le LNH touche 1,5 fois plus d'hommes que de femmes [Zhang *et al.*, 2011], nous avons cependant observé une surreprésentation des hommes dans notre échantillon, l'effectif attendu étant de 82,2 hommes et 54,8 femmes pour un échantillon théorique de 137 patients (*p=0,0001*). La moyenne d'âge des hommes et des femmes n'étaient pas significativement différentes (50,65 et 50,13 ans respectivement).

Le taux de recrutement des PST par an était de 17,125, dont 14,75 pour les hommes et 2,375 pour les femmes (figure 2), avec une forte augmentation de cette incidence en 2008 et 2009 avec respectivement 25 et 35 nouveaux PST enregistrés, sexe confondu.

Figure 2

Incidence et effectifs des cas de LNH dans la base RNV3P, de 2002 à 2009, par sexe.



Il y a eu 51% des PST enregistrés en mono-exposition, 17% en multi-exposition avec deux nuisances renseignées, 17% avec trois nuisances renseignées, 14% avec quatre nuisances et 4% avec cinq nuisances.

Le nombre de nuisances distinctes est passé de 18 en 2002 à 103 en 2009 avec une incidence de 8 à 20 nuisances par an, le maximum étant en 2009 et la moyenne étant à 12,875. Les nuisances identifiées concernaient 6 des 14 grands groupes de nuisances répertoriées par la CNAM-TS. La majorité de nuisances associées au LNH dans la base étaient des composés organiques et des substances industrielles (33 et 45% respectivement).

2. L'EXPOSOME

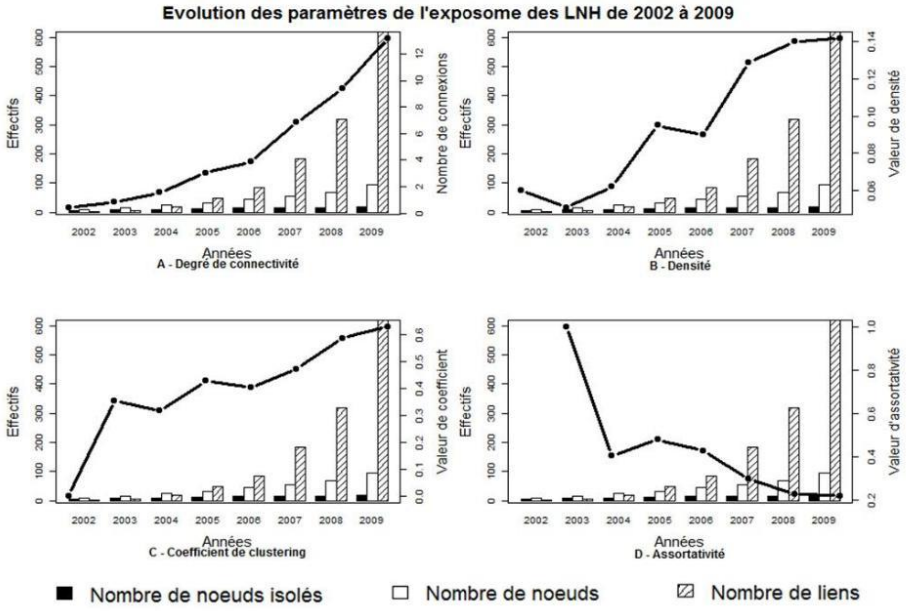
2.1. PARAMETRES DE L'EXPOSOME

L'exposome des LNH est passé de neuf nœuds et deux liens en 2002 à 94 nœuds et 621 liens en 2009 (figure 3). L'augmentation du nombre de nœuds et de liens est très prononcée en 2008, où l'on passe de 54 à 68 nœuds et de 185 liens à 319 et en 2009 où l'on passe de 68 à 94 nœuds et de 319 à 621 liens.

Figure 3

Evolution des paramètres de l'exposome des LNH de 2002 à 2009

(A) Degré de connectivité moyen des nœuds. (B) Densité de l'exposome. (C) Coefficient de clustering moyen. (D) Assortativité des nœuds. Chaque courbe illustre le paramètre présenté en A, B, C et D, sans unité, en fonction du nombre de nœuds, de nœuds isolés et de liens.



Chaque graphique de la figure 3 représente un paramètre spécifique de l'exposome en fonction du nombre de nœuds, de nœuds isolés et de liens de l'exposome. L'unité des paramètres est arbitraire. Le degré de connectivité moyen des nœuds correspond au nombre de liens moyen des nœuds (figure 3.A), il semble augmenter en même temps que le nombre de liens augmente. En 2009 le degré de connectivité moyen des nœuds est de 13,213 et le degré de connectivité maximal pour cet échantillon est de 44. La valeur de la densité de l'exposome (figure 3.B) varie au cours des années et semble se stabiliser à partir de 2008 à 0,14, son minimum étant de 0,051 en 2003. Le coefficient moyen de clustering (figure 3.C) augmente rapidement entre 2002 et 2003, passant de 0 à 0,353 ; il augmente ensuite de manière plus progressive jusqu'à atteindre sa valeur maximale de 0,631 en 2009. Enfin l'assortativité, qui correspond à la tendance des nœuds à se lier à des nœuds ayant un degré de connectivité similaire (figure 3.D), chute brutalement en 2004, passant de 1 à 0,02 (2002 n'est pas représentée car nous ne disposons pas d'assez de valeurs pour appliquer le test de Pearson).

2.2. Dynamique de l'exposome

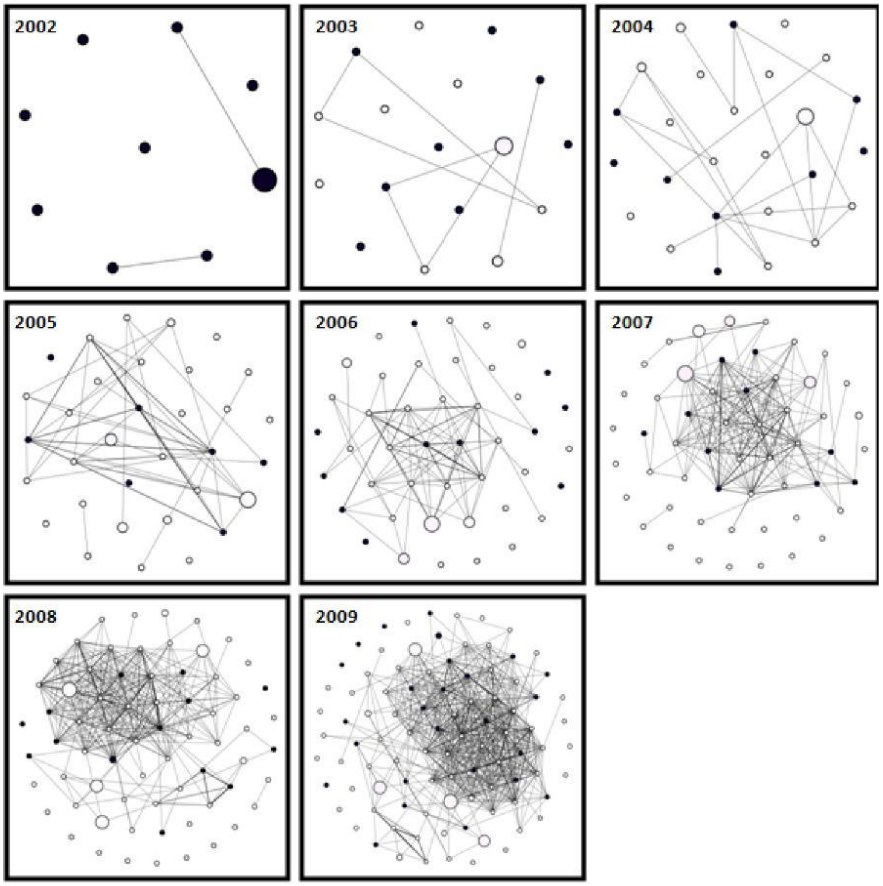
La dynamique temporelle de l'exposome est illustrée par la figure 4. Les nœuds noirs correspondent aux nouveaux nœuds et les nœuds blancs à ceux déjà existants. Leur taille est proportionnelle au nombre d'individus identiques dans la base (poids des nœuds), le minimum est 1 et le maximum est de 13. L'épaisseur des liens est proportionnelle au nombre de nuisances identiques partagées entre deux nœuds.

On peut voir sur cette figure que l'exposome se complexifie à partir de 2007 où il y a une explosion de liens permettant la formation de clusters (agrégats) denses. En 2009, il semble émerger trois clusters majoritaires, imbriqués les uns dans les autres, ainsi qu'un autre plus petit dans le bas-gauche de l'exposome.

Figure 4

Dynamique temporelle de l'exposome des LNH issus de la base RNV3P, de 2002 à 2009

Les nœuds noirs correspondent aux nœuds nouvellement formés et les nœuds blancs aux nœuds existants.
La taille des nœuds est proportionnelle au nombre d'individus identiques dans la base,
l'épaisseur des liens est proportionnelle au nombre de nuisances partagées entre deux nœuds.



En 2002, il y a 5 nœuds isolés sur 9 nœuds, en 2003, 9 sur 17, en 2004, 10 sur 26, en 2005, 11 sur 33, en 2006, 16 sur 44, en 2007, 14 sur 54, en 2008, 15 sur 68 et enfin, en 2009, 18 sur 94. On observe donc une stabilisation du nombre de nœuds isolés à partir de 2006.

4. LES GROUPES D'EXPOSITIONS

Nous avons effectué l'étude des groupes d'expositions associés aux LNH pour la plage temporelle de 2007 à 2009, en prenant en compte tous les groupes d'expositions déjà formés en 2006 (évènement d'intérêt de type A).

Les groupes d'expositions caractérisant les LNH sont au nombre de 64 (tableau 3), dont 33 d'ordre 1, 23 d'ordre 2, sept d'ordre 3 et un d'ordre 4. Nous avons observé qu'il y a 33 groupes d'expositions d'ordre 1, donc 33 nuisances distinctes utilisées pour la formation des groupes d'expositions sur les 103 identifiées dans l'échantillon, soit 32% des nuisances identifiées qui ont été utilisées. Les groupes d'expositions les plus renseignés lors

de l'enregistrement des PST (>10%) dans la base étaient les solvants et diluants organiques, le benzène, le trichloroéthylène, les fumées de soudage, les produits agricoles, le formaldéhyde, l'amiante, les associations benzène et solvants, ainsi que trichloroéthylène et solvants, et enfin les radiations ionisantes.

En 2006, 22 groupes d'expositions étaient déjà formés sur les 64 présents en 2009. Il y a eu une augmentation de 22 groupes sur cinq ans (2002 à 2006) et de 42 sur trois ans (2007 à 2009). En 2009, il n'y a eu aucun groupe d'expositions d'ordre 3 ou 4 de formé. Ces 42 groupes sont apparus de manière différente : il n'y a eu qu'un seul évènement de type E en 2009, l'évènement de formation qui semble le plus récurrent est le type C. Le tableau 3 récapitule le nombre de nouveaux groupes d'expositions par année, par ordre et par évènements d'intérêts.

Tableau 3

Récapitulatif du nombre de nouveaux groupes d'expositions de 2006 à 2009, par ordre et par type d'évènements d'intérêts

	Ordre 1	Ordre 2	Ordre 3	Ordre 4		Ordre 1	Ordre 2	Ordre 3	Ordre 4
2006	**17**	**5**	**-**	**-**	**2008**	**3**	**7**	**5**	**1**
A	17	5	-	-	**A**	-	-	-	-
B	-	-	-	-	**B**	3	-	-	-
C	-	-	-	-	**C**	-	3	2	-
D	-	-	-	-	**D**	-	4	3	1
E	-	-	-	-	**E**	-	-	-	-
2007	**5**	**5**	**2**	**-**	**2009**	**8**	**6**	**-**	**-**
A	-	-	-	-	**A**	-	-	-	-
B	5	-	-	-	**B**	8	-	-	-
C	-	3	1	-	**C**	-	3	-	-
D	-	2	1	-	**D**	-	2	-	-
E	-	-	-	-	**E**	-	**1**	-	-

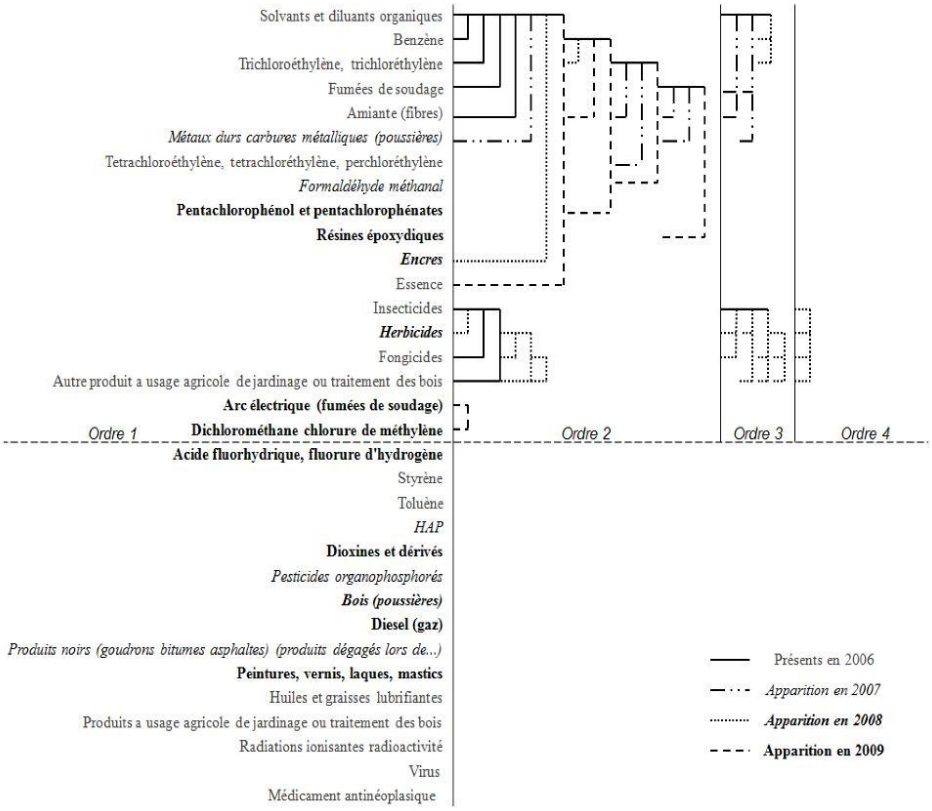
La représentation de la structure des groupes d'expositions est présentée en figure 5. Chaque groupe d'expositions d'ordre 1 est indiqué sur la gauche de la figure, les branches illustrent les associations retrouvées entre ces groupes d'expositions d'ordre 1. C'est en 2008 (en ***gras italique***) que sont apparus le plus grand nombre de groupes d'expositions d'ordre supérieur à 1 grâce à la formation du groupe d'ordre 1 « Herbicides ». Ce groupe d'expositions d'ordre 1 a permis la formation de neuf groupes d'expositions spécifiquement liés à l'agriculture sur une totalité de 11 groupes concernant les produits agricoles. Dix groupes d'expositions d'ordre supérieur à 1 concernent les solvants et diluants organiques d'une manière générale, cinq le benzène et le trichloroéthylène et un pour le tétrachloroéthylène. Les groupes d'expositions d'ordre 1, arcs électriques et dichlorométhane, apparus en 2009, ont permis la formation simultanée d'un groupe d'expositions d'ordre supérieur à 1 (évènement d'intérêt E), ce groupe est isolé des autres.

Il y a seize groupes concernant des solvants et onze concernant les produits agricoles sur un total de 31 groupes d'expositions d'ordre supérieur à 1. Il y a quinze groupes d'expositions d'ordre 1 non associés à d'autres sur les 33 groupes d'expositions d'ordre 1 existants.

Figure 5

Structure des groupes d'expositions des LNH de 2007 à 2009

Les libellés et les branches sont codés selon l'année de leur apparition.
Sous les pointillés sont présentés les groupes d'expositions ne participant à aucune association.



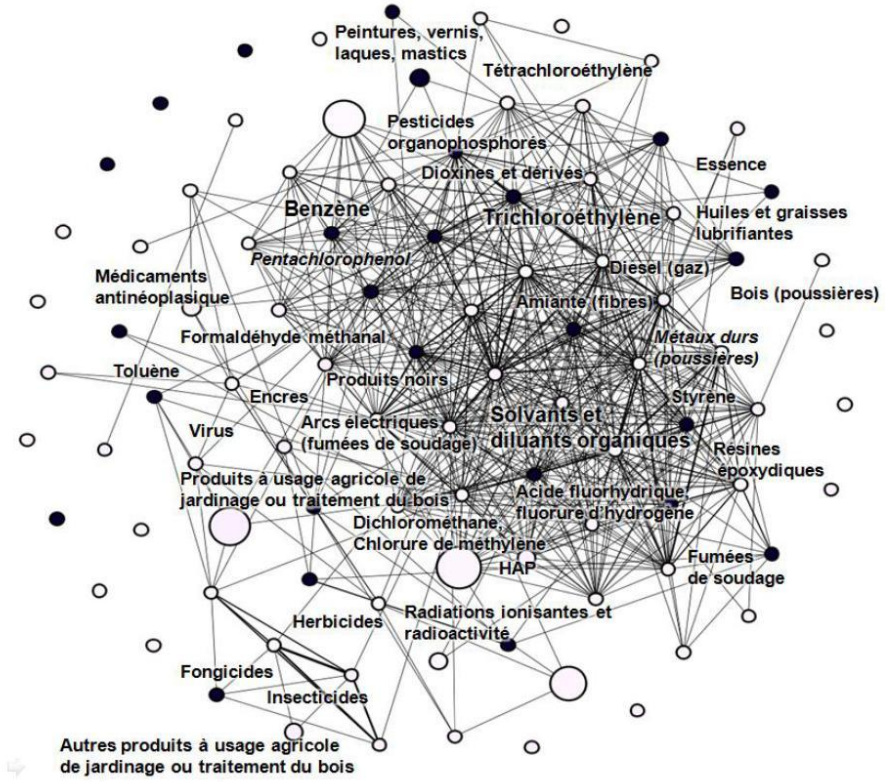
L'exposome de la figure 6 correspond aux données des LNH de 2002 à 2009 et à l'exposome de 2009 de la figure 4. Il concerne 137 PST de LNH pour 94 nœuds, dont 18 isolés et 621 liens. Seuls les groupes d'expositions précédemment énoncés (figure 5) sont libellés sur l'exposome, soit 33 nuisances sur les 103 identifiées au départ. La taille des nœuds correspond au nombre d'individus identiques et l'épaisseur des liens au nombre de nuisances partagées entre deux individus.

On observe une imbrication de trois clusters majoritaires : les solvants et diluants organiques, le benzène et le trichloréthylène, qui correspondent aux groupes d'expositions formant le plus d'associations (figure 5). Il y a également une zone correspondant essentiellement aux produits agricoles sur le côté bas-gauche de l'exposome mais qui ne forme pas de cluster. On observe une sorte de hiérarchie dans la structure de l'exposome : toutes les expositions concernant les usines et manufactures sont très imbriquées, tandis que les expositions aux radiations ionisantes, aux virus, médicaments et aux produits agricoles sont plus éloignées du cluster central.

Figure 6

Exposome des LNH en 2009 et groupes d'expositions associés

Seuls les groupes d'expositions identifiés sont libellés sur cet exposome, soit 33 nuisances sur les 103 recensées. Les libellés dont la taille de police est supérieure aux autres correspondent aux clusters majoritaires de l'exposome.



IV - DISCUSSION

A la suite d'une analyse de la bibliographie, nous avons pu constater que 55% des groupes d'expositions d'ordre 1 identifiés pour les LNH correspondaient à des hypothèses émises par la littérature. Cela soulève l'hypothèse qu'il existe une réelle association entre ces expositions et la pathologie, notamment avec les solvants et diluants organiques, le benzène, le trichloroéthylène, les produits agricoles et les radiations ionisantes. Les groupes d'expositions restants (45%) n'ont pas été retrouvés dans la littérature et 8 expositions émises par la littérature n'ont pas été identifiés comme groupes d'expositions. Cependant, il nous faut préciser que certaines de ces expositions pourraient être potentiellement assimilées aux groupes d'expositions après recodage. En effet, les libellés des groupes d'expositions sont issus de la classification de la CNAM-TS et ont été conservés tels quels, incluant un niveau de précision pouvant diverger par rapport aux appellations de la littérature. De plus, il n'est pas exclu que les expositions issues de la

littérature aient été enregistrées dans la base en tant que nuisances mais n'aient pas été considérées comme groupe d'expositions car n'étant pas identifiées dans au moins deux nœuds du réseau. Cette constatation a été vérifiée et trois des expositions de la littérature ont été retrouvées parmi les 103 nuisances associées aux LNH (PCB, pesticides organophosphorés et composés organo-chlorés). Le suivi de ces expositions est intéressant puisqu'en cas de nouvelles associations identiques enregistrées dans la base, elles seraient susceptibles de former des nouveaux groupes d'expositions. Ces groupes d'expositions et expositions sont donc des éléments de discussion et de réflexion majeurs quant à l'étiologie de la pathologie.

L'exemple des lymphomes non hodgkiniens a permis d'illustrer la démarche de la surveillance observationnelle qui se décline en plusieurs étapes : la construction de l'exposome, la construction des groupes d'expositions et la génération de la matrice activité-groupe d'expositions [Rieutort, 2012].

Nous avons déjà pu observer grâce aux groupes d'expositions et à une comparaison avec la bibliographie que 55% des groupes d'expositions identifiées via cette surveillance observationnelle étaient en accord avec les hypothèses de la littérature, soient 18 des 103 nuisances identifiées pour les LNH. Nous avons également pu observer que certaines des hypothèses de la littérature ne se sont pas démarquées en tant que groupes d'expositions ; cependant, leur suivi est très intéressant puisqu'il est susceptible de nous apporter des informations quant à l'émergence de certaines associations. Il est également nécessaire de rappeler que les articles concernant le LNH traitent essentiellement de mono-exposition, à l'exception de deux articles sur les pesticides et les solvants. Nous avons donc comparé uniquement les groupes d'expositions d'ordre 1 à cette bibliographie, bien que la surveillance observationnelle permette de traiter la multi-exposition. Enfin, pour cet exemple, nous avons choisi de former les groupes d'expositions pour un seuil égal à 2, c'est-à-dire que deux nœuds minimum de la matrice d'exposition devaient comporter la nuisance ou l'association de nuisances en question. Si nous avions choisi un seuil plus élevé, nous aurions pu avoir un nombre de groupes d'expositions plus faible, notamment les groupes d'expositions d'ordre supérieur à 1.

L'exemple choisi concernait une unique pathologie, or la structure d'un PST dans la base du RNV3P est composée des nuisances mais également d'une pathologie et d'une activité professionnelle. Ainsi, il est possible dans la même étude de combiner plusieurs pathologies différentes et d'observer les interactions entre ces pathologies.

De même, il est possible d'utiliser les groupes d'expositions pour explorer la dimension activité professionnelle associée aux LNH [Rieutort, 2012]. Cet aspect qui n'a pas été présenté ici permet d'identifier les secteurs d'activités les plus exposés et les plus susceptibles d'être en lien avec une pathologie donnée.

La surveillance observationnelle, telle que nous l'avons développée dans ces travaux, présente les résultats préliminaires de la démarche. La suite des travaux consistera en l'approfondissement de la démarche et l'étude de ses limites et de ses avantages.

V - CONCLUSION

Ce travail a eu pour but de présenter les premiers développements de la surveillance observationnelle et de ses indicateurs. Le développement des groupes d'expositions a permis d'apporter des dimensions supplémentaires à l'approche par exposome, comme l'évolution des associations des nuisances au cours du temps.

La surveillance observationnelle représente donc une analyse complémentaire permettant d'apporter ou de conforter des hypothèses de travail concernant les pathologies professionnelles enregistrées dans la base du RNV3P. L'automatisation complète de cette surveillance observationnelle est un point essentiel de cette démarche afin de pouvoir l'intégrer en routine au sein du RNV3P.

La surveillance observationnelle n'en est qu'à ses débuts et les possibilités semblent encore nombreuses. En effet, il est possible

d'appliquer cette surveillance observationnelle à des réseaux tels que SAGIR, pour lequel la réflexion de l'exploration des données via l'exposome avait été initiée lors du travail de thèse de Petit sur la surveillance syndromique, [Petit, 2011]. L'aspect d'émergence, un des paradigmes du RNV3P, déjà initié lors des travaux de thèse de Laurie Faisandier [Faisandier, 2011], sera également intégré à la surveillance observationnelle lors de prochains travaux que nous effectuerons.

BIBLIOGRAPHIE

Alexander D.D., Wagner M.E. - Benzene exposure and non-Hodgkin lymphoma: a meta-analysis of epidemiologic studies. *J. Occup. Environ. Med.,* 2010, **52**, 169-189.

Agence nationale de sécurité sanitaire de l'alimentation, de l'environnement et du travail – Présentation du RNV3P. 2012. Consulté le 12 juin 2012. Accessible au : http://www.anses.fr/ET/PPN5BDA.htm

Barabási A.L. – Network medicine—from obesity to the "diseasome*". N. Engl. J. Med.,* 2007, **4**, 404-407.

Barabási A.L. – La théorie de la complexité. *La Recherche*, 2012, **464**, 36-41.

Bastian M., Heymann S., Jacomy M., - Gephi: an open source software for exploring manipulating networks. International AAAI Conference on Weblogs and Social Media, 2009.

Boffetta P., de Vocht F. - Occupation and the risk of non-Hodgkin lymphoma. *Cancer Epidemiol. Biomarkers Prev.*, 2007, **16**, 369-372.

Braüner E.V. - Composés organochlorés dans le tissu adipeux et risque de lymphome non Hodgkinien. *Environ. Risques Santé*, 2012, **11**, 102-103.

Cocco P., t'Mannetje A., Fadda D., Melis M., Becker N., de Sanjose S., *et al.* - Occupational exposure to solvents and risk of lymphoma subtypes: results from the Epilymph case-control study. *Occup. Environ. Med.*, 2010, **67**, 341-347.

Faisandier L. - De la surveillance à la détection des Problèmes de Santé au Travail potentiellement émergents dans un contexte de multi-exposition : Développement de méthodes appliquées au Réseau National de Vigilance et de Prévention des Pathologies Professionnelles, 266 pages, Thèse de Sciences. Université de Grenoble, Grenoble, 2011.

Faisandier L., Bonneterre V., de Gaudemaris R., Bicout D.J. - Occupational exposome: a network-based approach for characterizing Occupational Health Problems. *J. Biomed. Inform.*, 2011, **44**, 545-552.

Goh K.I., Cusick M.E., Valle D., Childs B., Vidal M., Barabási A.L. – The human disease network. *Proc. Natl. Acad. Sci. U S A, 2007,* **104**, 8685-8690.

Hartge P., Smith M.T. - Environmental and behavioral factors and the risk of non-Hodgkin lymphoma. *Cancer Epidemiol. Biomarkers Prev.*, 2007, **16**, 367-368.

Hohenadel K., Harris S.A., McLaughlin J.R., Spinelli J.J., Pahwa P., Dosman J.A., *et al.* - Exposure to multiple pesticides and risk of non-Hodgkin lymphoma in men from six Canadian provinces. *Int. J. Environ. Res. Public Health*, 2011, **8**, 2320-2330.

Karunanayake C.P., Singh G.V., Spinelli J.J., Mc Laughlin J.R., Dosman J.A., McDuffie H.H. *et al.* - Occupational exposures and Hodgkin Lymphoma: Canadian case-control study. *J. Occup. Environ. Med.,* 2009, **51**, 1447-1454.

McDuffie H.H., Pahwa P., Karunanayake C.P., Spinelli J.J., Dosman J.A. - Clustering of cancer among families of cases with Hodgkin Lymphoma (HL), Multiple Myeloma (MM), Non-Hodgkin's Lymphoma (NHL), Soft Tissue Sarcoma (STS) and control subjects. *BMC Cancer*, 2009, **9**, 70.

Neasham D., Sifi A., Nielsen K.R., Overvad K., Raaschou-Nielsen O., Tjonneland A. *et al.* - Occupation and risk of lymphoma: a multicentre prospective cohort study (EPIC). *Occup. Environ. Med.*, 2011, **68**, 77-81.

Orsi L., Delabre L., Monnereau A., Delval P., Berthou C., Fenaux P., *et al.* - Occupational exposure to pesticides and lymphoid neoplasms among men: results of a French case-control study. *Occup. Environ. Med.*, 2009, **66**, 291-298.

Petit E. - Modélisation de données de surveillance épidémiologique de la faune sauvage en vue de la détection de problèmes sanitaires inhabituels, 310 pages, Thèse de Sciences, Université de Grenoble, Grenoble, 2011.

Purdue M.P., Severson R.K., Colt J.S., Stewart P., de Roos A.J., Cerhan J.R., *et al.* - A case-control study of occupational exposure to trichloroethylene and non-Hodgkin lymphoma. *Environ. Health Perspect.*, 2011, **119**, 232-238.

RDevelopementCoreTeam - R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, 2011

Rieutort D. – La Surveillance Observationnelle des Problèmes de Santé au Travail. Mémoire master 2 MRESTE R, 40 pages, Université de Grenoble, Grenoble, 2012.

SAGIR – Le Réseau SAGIR. 2012. Consulté le 12 juin 2012. Accessible au http://www.oncfs.gouv.fr/Reseau-SAGIR-ru105

Smith M.T., Jones R.M., Smith A.H. - Benzene exposure and risk of non-Hodgkin lymphoma. *Cancer Epidemiol. Biomarkers Prev.*, 2007, **16**, 385-391.

Vlaanderen J., Lan Q., Kromhout H., Rothman N., Vermeulen R. - Occupational benzene exposure and the risk of lymphoma subtypes: a meta-analysis of cohort studies incorporating three study quality dimensions. *Environ. Health Perspect.*, 2011, **119**, 159-167.

Wang R., Zhang Y., Lan Q., Holford T.R., Leaderer B., Zahm S.H., *et al.* - Occupational exposure to solvents and risk of non-Hodgkin lymphoma in Connecticut women. *Am. J. Epidemiol.*, 2009, **169**, 176-185.

Wild C.P. - Complementing the genome with an « exposome »: the outstanding challenge of environmental exposure measurement in molecular epidemiology. *Cancer Epidemiol. Biomarkers Prev.,* 2005, **8**, 1847-1850.

Zhang Y., Dai Y., Zheng T., Ma S. - Risk Factors of Non-Hodgkin Lymphoma. *Expert Opin. Med. Diagn.*, 2011, **5**, 539-550.

La Surveillance Observationnelle des Problèmes de Santé au Travail

Rieutort D.[a] *(PhD student)*, de Gaudemaris R.[a,c], Bicout DJ. [a,b]
[a] Equipe EPSP - Laboratoire TIMC-IMAG - UMR CNR 5525 - Grenoble - [b] Biomathématiques et Epidémiologie - Equipe EPSP - VetagroSup Lyon
[c] Service de Médecine et Santé au Travail - Centre Hospitalier Universitaire de Grenoble

Introduction

La surveillance des maladies professionnelles est un outil important pour identifier et prévenir des potentielles menaces pour la santé des travailleurs. C'est ainsi que depuis 2001, le Réseau National de Vigilance et de Prévention des Pathologies professionnelles recueillent dans une base de données tous les Problèmes de Santé au Travail (PST) dans 32 CHU de France. Afin d'améliorer notre compréhension et notre capacité d'analyse des PST ou associations pathologie-expositions professionnelles, nous avons développé le concept d'exposome professionnel [1] permettant d'investiguer les caractéristiques dynamiques des PST. La Surveillance Observationnelle des PST, que nous sommes en train de développer, repose sur **la construction puis la description des exposomes professionnels et des spectres d'expositions professionnelles.**

Matériel & Méthodes

Structure d'un Problème de Santé au travail (PST)

La surveillance observationnelle a pour objectif **d'étudier les relations maladies-expositions professionnelles dans leur dynamique, de repérer les changements de tendance et détecter les évènements émergents.** Cette surveillance se développe à partir de l'analyse à la fois de l'exposome professionnel et le spectre d'expositions professionnelles

- **L'Exposome professionnel [1]**
  Réseau relationnel de PST partageant un ou plusieurs éléments de leur expositions professionnelles.

Code	Libellés
21311	Benzène
21312	Toluène
21320	Hydrocarbures aromatiques polycycliques
214B0	Dioxines et dérivés
22122	Chlorure de méthylène
22224	Trichloroéthylène
22225	Tétrachloroéthylène
29111	Formaldéhyde
2G320	Pesticides organophosphorés
32110	Amiante (fibres)
334C0	Produits noirs (goudron bitume, asphalte)
34100	Fumées de soudage
35110	Solvants et diluants organiques
36200	Peintures, vernis, laques, mastics
36500	Caoutchouc, polymère de plastique
36910	Encres
36B12	Essence
36H00	Produits de nettoyage
36R00	Produits à usage agricole ou traitement du bois
36R20	Insecticides
36R40	Herbicides
36R50	Fongicides
36RZ0	Autres produits agricoles ou traitement du bois
47100	Radiation ionisante et radioactivité

- **Le Spectre des expositions professionnelles**
  Ensemble des groupes d'expositions de l'exposome ayant un nombre significatif d'enregistrements.

Spectre des Expositions Professionnelles des LNH en 2011

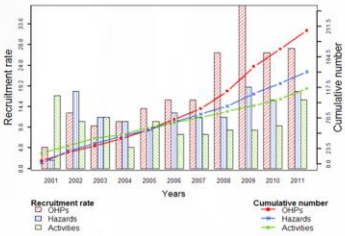
Le groupe d'exposition le plus représenté dans le LNH est « Solvants et diluants organiques ». Ce groupe est associé à 6 autres expositions :
benzène, trichloroéthylène, amiante, fumées de soudage, caoutchouc et encres.

Illustration : Le lymphome non-hodgkinien

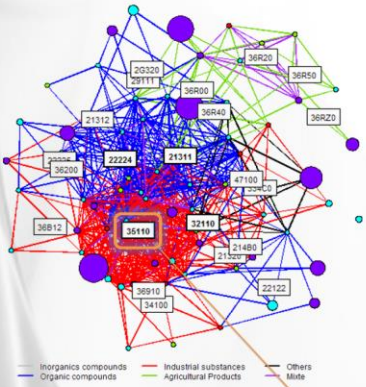
Cancer rare (3% des cancers) et représentant 90% des lymphomes, le lymphome non hodgkinien (NHL) est dû à une prolifération maligne des lymphocytes. Depuis les années 70, on a assisté à une augmentation de l'incidence des LNH qui a été en grande partie attribuée à la propagation du VIH ; les personnes immunodéprimées étant plus susceptibles de développer un LNH. Cependant, malgré l'efficacité des campagnes de prévention contre le VIH, l'augmentation continue de l'incidence des LNH laisse à penser que d'autres facteurs comme les facteurs environnementaux pourraient avoir un lien avec cette pathologie. **La part de l'exposition professionnelle reste méconnue mais concentre de plus en plus d'intérêt...**

Echantillon des LNH de 2001 à 2011

En 2011, la base RNV3P recense 205 PST de LNH, avec une augmentation du taux de recrutement en 2008. Le nombre de nuisances distinctes associées est de 142, et celui des activités professionnelles distinctes de 116.



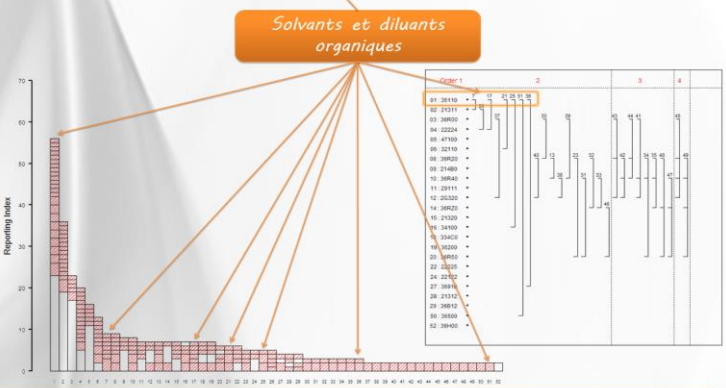
Exposome professionnel des LNH en 2011



L'exposome des LNH comporte 72 nœuds, dont 1 isolés, et 628 liens.

Quatre clusters majeurs sont identifiables :
- **Substances industrielles (en rouge) :** « Solvants et diluants organiques » et « Amiante »
- **Produits organiques (en bleu) :** « benzène » et « trichloroéthylène »

Dans le sud de l'exposome on trouve le cluster en formation des produits agricoles (en vert).



Conclusion

Les expositions professionnelles associées au LNH issues de la base RNV3P sont cohérentes avec la littérature, avec comme groupe d'expositions identifiés : « **Solvants et diluants organiques** », **dont le « Benzène » et « Trichloroéthylène », « Produits agricoles » et « Radiations ionisantes** ». Le groupe d'exposition des « Solvants et diluants organiques » est associé à 6 autres groupes d'expositions identifiés pour le LNH. C'est donc un groupe que l'on peut considérer comme caractéristique de la pathologie, de part sa fréquence et le nombre des associations.
Dans cette illustration, la Surveillance Observationnelle a permis de mettre en évidence **les expositions majoritaires des LNH et la multi-exposition importante associée à cette pathologie.** D'autres indicateurs (non montrés ici) de cette approche permettent d'étudier la dynamique et l'émergence des associations LNH - expositions.

[1] Faisandier L, Bonneterre V, De Gaudemaris R, Bicout DJ. Occupational exposome: a network-based approach for characterizing Occupational Health Problems. J Biomed Inform. 2011 Aug;44(4):545-52.
[2] Rieutort D, De Gaudemaris R, Bicout DJ. Surveillance Observationnelle : Approche par Exposome. Epidémiol et santé anim. 2012;61:127-140.

De la Consultation Professionnelle à la Surveillance et la Détection des Evènements Emergents

Rieutort D.[a], **de Gaudemaris R.**[a,c], **Bicout DJ.**[a,b]

[a] Equipe EPSP – Laboratoire TIMC-IMAG – UMR CNR 5525 - Grenoble - [b] Biomathématiques et Epidémiologie – Equipe EPSP – VetagroSup Lyon
[c] Service de Médecine et Santé au Travail – Centre Hospitalier Universitaire de Grenoble

CONTEXTE

Quel est le point commun de tous ces **travailleurs** ?

L'exposition à des risques **physiques, chimiques ou biologiques**

- **Concerne 26 millions** de travailleurs (52% des 15 ans et plus)
- Risques potentiellement responsables de **maladies « professionnelles »**

« **Conséquence de l'exposition plus ou moins prolongée à un risque lors de l'exercice habituel de la profession** »

Diagnostic des maladies professionnelles dans les **Centres de Consultation de Pathologies Professionnelles** des Centres Hospitaliers Universitaires

Réseau RNV3P (créé en 2001)
- Réseau National de **Vigilance** et de **Prévention** des Pathologies Professionnelles
- Base de données ~**150 000 cas de Problèmes de Santé au Travail (PST)**

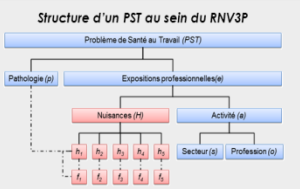
Données « **observationnelles** » ne se prêtant pas à des analyses épidémiologiques conventionnelles :
- **Manque de dénominateur**
- **Maladies multi factorielles** (ex : plusieurs expositions)

Surveillance, identification menaces, prévention ?

Développement d'**outils d'analyse adaptés** à des données « **observationnelles** »

MATERIEL & METHODES

Structure d'un PST au sein du RNV3P



DIME
« Data-based Investigation and Modeling of Events »
Surveillance : Surveillance structurée d'évènements d'intérêts
Emergence : Identification/détection d'évènements d'intérêts émergents
Prospection : Inférence/prédiction d'évènements d'intérêts

Développement de la **Surveillance Observationnelle**
utilisation optimale des bases de données

PST
- Description analytique de la base de données
- Identification des **expositions significatives**

Exposome
- Restructuration par **Exposome [1]** : « **Réseau relationnel de PST significatifs** partageant au moins un élément de l'exposition professionnelle »

Motifs
- Construction des **motifs d'expositions significatifs**

Analyse spectrale
- **Structure des combinaisons**
- **Facteur d'impact :** caractéristique de recrutement du motif : **Persistance (A), Emergence (C), Déclin (B), Latence (D)**
- **Spécificité :** Spécifique d'une association (1) ou Ubiquitaire sur différentes associations (0)

ILLUSTRATION : Lymphome non hodgkinien

- Prolifération maligne des lymphocytes B ou T
- **3% des cancers**, 90% des lymphomes
- Augmentation de l'incidence attribuée au VIH
- Augmentation persistante malgré campagnes de prévention VIH
- Possibilité **d'autres facteurs** : génétiques, environnementaux ou **professionnels**

La part totale de l'exposition professionnelle reste méconnue et suscite de plus en plus d'intérêt…

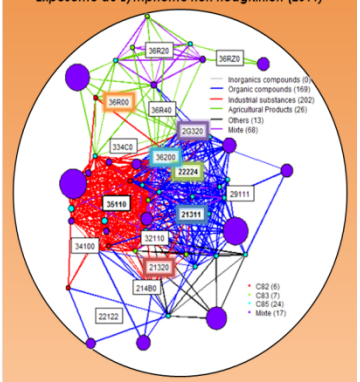
205 PST dans la base RNV3P

142 nuisances distinctes – **10 significatives**

610 combinaisons possibles – **48 significatives**

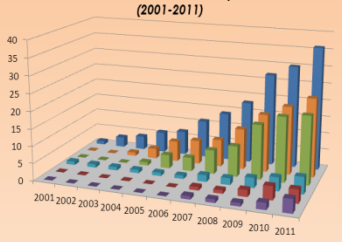
20 Simples				**28 Associations**			
A : 10	B : 2	C : 3	D : 5	A : 3	B : 1	C : 10	D : 14

Exposome du Lymphome non hodgkinien (2011)



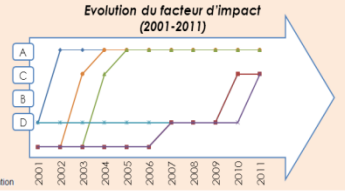
Exemple d'expositions identifiées sur l'Exposome :
Benzène (*21311*) – Trichloroéthylène (*22224*) – HAP (*21320*) – Peintures, laques, mastics (*36200*) – Produits agricoles et traitement bois (*36R00*) – Pesticides organophosphorés (*2G320*)

Recrutement cumulé des expositions (2001-2011)

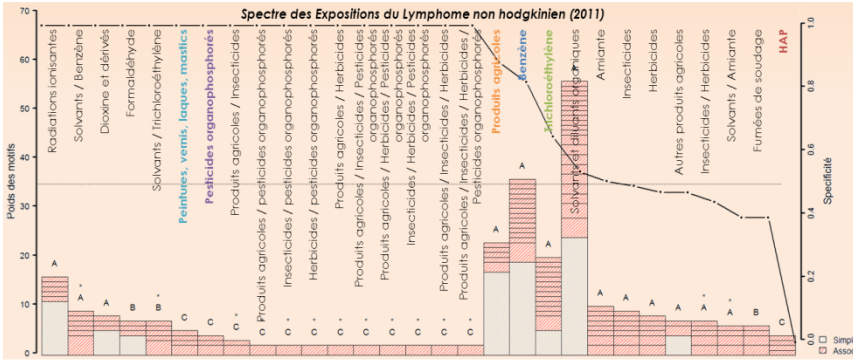


6 profils d'apparition différents :
- **Benzène :** Recrutement depuis 2002 , **persistant**
- **Trichloroéthylène :** **Emergence** en 2004, **persistant** depuis
- **HAP :** **Latence** depuis 2007, **émergence** en 2010, non stabilisé en 2011
- **Peintures, laques :** **Latence** depuis 2001, **émergence** en 2010, non stabilisé en 2011
- **Produits agricoles :** **Emergence** en 2003, **persistant** depuis
- **Pesticides organophosphorés :** **Latence** depuis 2007, **émergence** en 2011

Evolution du facteur d'impact (2001-2011)



Spectre des Expositions du Lymphome non hodgkinien (2011)



CONCLUSION

Surveillance Observationnelle : Utilisation optimale des bases de données observationnelles et données massives
- Création d'une application **OSA** (« Observational Surveillance Analysis ») → **routine**
- **Analyse poussée** de la structure des associations en complément des analyses classiques
- Détection **d'évènements émergents → alerte**

LNH : 20 expositions potentielles pour 48 combinaisons
- Pathologie caractérisée par **multi-expositions** (majorité d'associations)
- Expositions aux **produits chimiques** (ex. benzène) et **produits agricoles** majoritaires
- De nouvelles expositions apparaissent régulièrement → **état de la connaissance incomplet**

EN COURS

- **3 articles en cours** sur la Surveillance Observationnelle
- Amélioration de **l'application OSA**
- **Généralisation** à d'autres bases (collaboration via réseau international MODERNET)
- Développement des autres étapes de **DIME**

[1] Falsandler L, Bonneterre V, De Gaudemaris R, Bicout DJ. Occupational exposome: a network-based approach for characterizing Occupational Health Problems. J Biomed Inform. 2011 Aug;44(4):545-52.

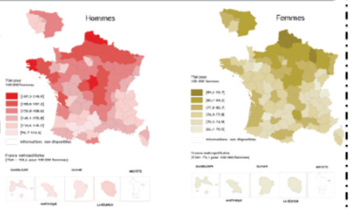
La GEO-ZAPE des pathologies professionnelles

Rieutort D.[a], Porcheron H. [a], de Gaudemaris R.[a,c], Bicout DJ. [a,b]

[a] Equipe EPSP – Laboratoire TIMC-IMAG – UMR CNR 5525 - Grenoble - [b] Biomathématiques et Epidémiologie – Equipe EPSP – VetagroSup Lyon
[c] Service de Médecine et Santé au Travail – Centre Hospitalier Universitaire de Grenoble

Introduction

- Travailleurs exposés à des **risques physiques, chimiques ou biologiques**
  - Développement possible de pathologies « professionnelles »
    - **Hétérogénéité géographique** de l'incidence et prévalence des pathologies professionnelles [Inca,2011]



Taux standardisés à la population mondiale de mortalité par Cancer à l'échelle départementale en France métropolitaine et dans les DOM (2004-2008)
Source InVS/CépiDc-Inserm, infographie : INCa 2011

Pourquoi une pathologie serait-elle plus prévalente dans un région que dans une autre ?

- Facteur clef de l'**exposition professionnelle → activité professionnelle (AP)**
  - Bassins d'AP français inégaux
    - **Hétérogénéité des AP → hétérogénéité des pathologies associées**

Création de ZAPE
« Zone d'Activité Professionnelle Equivalente »
pour décrire l'hétérogénéité géographique des pathologies professionnelles

Matériel & Méthodes

- Nombre de salariés de 1989 à 2011, par secteurs AP et départements
- Analyse et identification des secteurs AP prépondérants : Agriculture, Industrie et Tertiaire

INSEE [1] → Création profils ZAPE

- Construction de l'espace géographique des ZAPE et caractérisation au niveau départemental :
  - Agriculture
  - Industrie
  - Tertiaire
  - Agriculture-Industrie
  - Agriculture-Tertiaire
  - Industrie-Tertiaire
  - Neutre

Géo-ZAPE

- Géographie ZAPE des pathologies Professionnelles
- Projection des SIR des tumeurs du sang et des tissus hématopoïétiques du RNV3P sur les ZAPE (code postal patient)
- Analyse des déplacements patients entre sZAPE et CCPP

RNV3P [2]

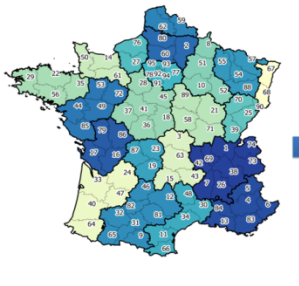
- **R**éseau **N**ational de **V**igilance et de **P**révention des **P**athologies **P**rofessionnelles créé en 2001
  - 150 000 cas de **P**roblèmes de **S**anté au **T**ravail (PST) enregistrés
  - 32 **C**entres de **C**onsultation de **P**athologie **P**rofessionnelle (CCPP)

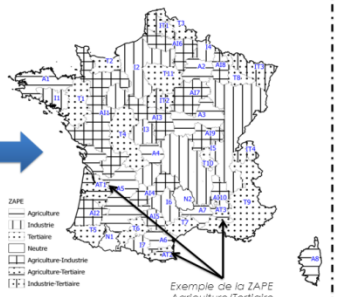
Résultats

Construction des ZAPE à partir des données de l'INSEE

7 ZAPE subdivisées en **45 sZAPE** (sous-ZAPE : unité spatiale non connexe d'une même ZAPE) :
8 Agriculture – 7 Industrie – 11 Tertiaire – 2 Neutre
10 Agriculture/Industrie – 3 Agriculture/Tertiaire – 4 Industrie/Tertiaire

Régions administratives
Alsace
Aquitaine
Auvergne
Basse-Normandie
Bourgogne
Bretagne
Centre
Champagne-Ardenne
Corse
Franche-Comté
Haute-Normandie
Île-de-France
Languedoc-Roussillon
Limousin
Lorraine
Midi-Pyrénées
Nord-Pas-de-Calais
Pays de la Loire
Picardie
Poitou-Charentes
Provence-Alpes-Côte d'Azur
Rhône-Alpes





Exemple de la ZAPE Agriculture/Tertiaire

Profil (en terme de salariés) **des ZAPE** par rapport à la moyenne nationale : **Agricole (AZ), Industriel (IN) et Tertiaire marchand (TER)**

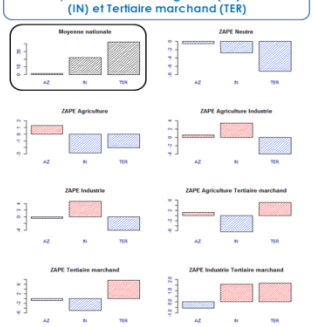
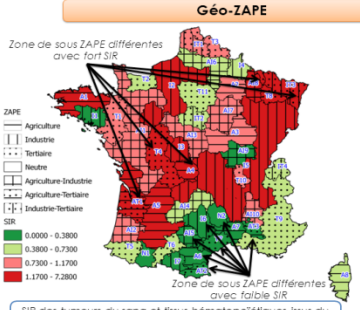


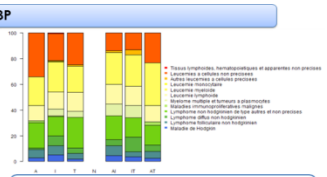
Illustration : Tumeurs du sang et des tissus hématopoïétiques du RNV3P

Géo-ZAPE

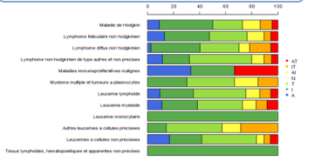


SIR des tumeurs du sang et tissus hématopoïétiques issus du RNV3P sur les différents sous-ZAPE

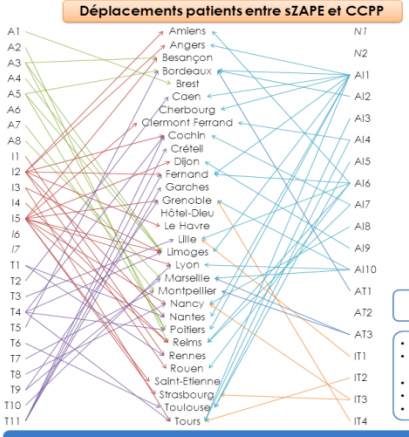
- 10 132 Problèmes de Santé au Travail associés à des tumeurs → **673 tumeurs du sang ou tissus hématopoïétiques**
- Deux zones géographiques caractéristiques :
  - Fort SIR dans la bande centre
  - Faible SIR dans la région sud
- Des SIR différents entre sZAPE → importance de la localisation géographique **en plus** de l'influence du bassin d'AP
- Majorité de cas dans la ZAPE Industrie (~27%) et Tertiaire (~40%) → beaucoup de tumeurs de type « Non Précisé »
- Majorité de leucémie (~60%)
- Absence de tumeurs dans la ZAPE Neutre
- Pas de myélome multiple et de leucémie monocytaire dans les ZAPE Agriculture et Agriculture/Tertiaire
- Pas de maladies immunoprolifératives dans la ZAPE Tertiaire

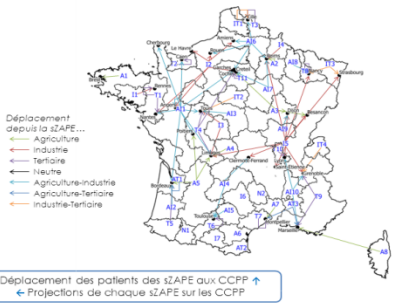


↑ Proportion de pathologies dans chaque ZAPE
Proportion des ZAPE pour chaque pathologie ↓



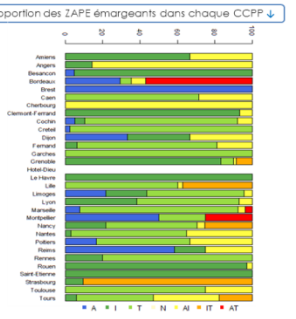
Déplacements patients entre sZAPE et CCPP





Déplacement des patients des sZAPE aux CCPP ↑
← Projections de chaque sZAPE sur les CCPP

Proportion des ZAPE émargeants dans chaque CCPP ↓



- Migration peu importante → **Proximité CCPP et sZAPE**
- 6 sZAPE sans pathologies → non rapportées par les CCPP (A6, I6, I7, N1, N2, AT2)
- Entre 1 à 10 CCPP différents pour une sZAPE
- 26 sZAPE sans CCPP
- CCPP Hotel-Dieu sans consultation

- **CCPP spécifiques** à un type de ZAPE :
  - Brest → ZAPE A
  - Cherbourg → ZAPE AI
  - Garches, Le Havre et Saint Etienne → ZAPE I
  - IT majoritairement retrouvée dans les CCPP de Strasbourg et Lille, puis Nancy, Tours et Grenoble

Conclusion

Hétérogénéité des bassins professionnels → hétérogénéité des expositions professionnelles → hétérogénéité des SIR des tumeurs du sang et tissus hématopoïétiques
Hétérogénéité des SIR au sein de la même ZAPE → **influence des sZAPE connexes** (proximité des AP) et/ou **proximité géographique** des CCPP et/ou autres facteurs

L'intérêt de développer la Géo-ZAPE est de pouvoir prendre en compte à la fois
l'hétérogénéité des AP (expositions professionnelles) et **l'hétérogénéité géographique** dans l'analyse de **l'incidence et de la prévalence des pathologies professionnelles**

[1] Institut National de la Statistique et des Etudes Economiques. Données de l'emploi départements et sectoriel (1989-2011).
Date de dernière consultation le 07/04/2014 : http://www.insee.fr/fr/themes/detail.asp?reg_id=99&ref_id=ir-irsoceds2011&page=irweb/IRSOCEDS2011/dd/irsoceds2011_emploi.htm
[2] Bonneterre V, de Gaudemaris R, Celse M, et al. Réseau national de vigilance des pathologies professionnelles: premier bilan et perspectives. Arch Mal Prof Env 2006;5:719-23.

Databased Observational Surveillance of occupational non Hodgkin lymphoma

Delphine Rieutort[a,*], **de Gaudemaris Régis**[a,b], **Dominique J Bicout**[a,c,*]

[a] UJF-Grenoble 1 / CNRS / TIMC-IMAG UMR 5525 (EPSP team - Environment and Health Prediction of Populations), Grenoble, F-38041, France

[b] Occupational and Health Medicine Department, Grenoble Teaching Hospital, Grenoble, France

[c] Biomathematics and Epidemiology, VetAgro Sup, Veterinary Campus of Lyon, Marcy l'Etoile, France

INTRODUCTION:

- Malignant proliferation of B/T lymphocytes representing 3% of all cancers and non-Hodgkin Lymphoma (NHL), 90% of lymphoma
- Persistent increase of NHL incidence from 70's to late 90's mainly attributed to HIV, despite prevention campaign, and leveling off in the last few years
- Associated risk factors largely unknown because assumptions about immunological and behavioral factors are not completely sufficient
- Occupational exposures such as solvents or agricultural products have been suggested but very widespread multi-exposure at work prevents to conclude from epidemiological studies

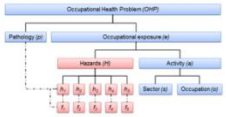
OBJECTIVES:

Use the observational database of the RNV3P network (French National Network of Surveillance and Prevention of Occupational Diseases) to investigate the relationship structure of NHL with occupational (multi)exposures and draw the spectrum of NHL's occupational (multi)exposures.

MATERIAL:

- French National Expert Network of Surveillance and Prevention of Occupational Diseases (RNV3P)

- Since 2001, systematic and standardized collect of Occupational Health Problems (OHP) by Expert Physicians
- About 200,000 observations including 50% of OHPs, from 32 French University Hospital in the RNV3P



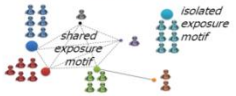
- For instance:
  - 24% of diseases of the respiratory system (ICD 10: J00-J99)
  - 23% of mental and behavioral disorders (F00-F99)
  - 15% of diseases of the musculoskeletal system and connective tissue (M00-M99)
  - 14% of neoplasms (C00-D49)

METHODS:

Observational surveillance*[1] *based on:

- **Occupational Exposome**[2]

Relational network of OHPs with at least 1 significant motif and sharing at least 1 element of the exposure



- **Occupational spectrum**

Structured list of motifs from the exposome ranked according to dynamic, abundance, and diversity

Aims:

- Monitoring the dynamic of pathology-occupational exposure associations as dynamic **motifs**
- Analyzing structural change in our association
- Highlighting appearance or disappearance of associations

Indicators:

- Occupational exposome of motifs
- Spectrum of occupational exposures motifs
- Multi-exposures tree
- Dynamical changes:
  - A: Persistent
  - B: Declining
  - C: Emerging
  - D: Latency

RESULTS:

288 OHPs in RNV3P database

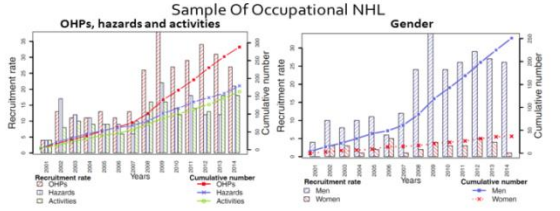
178 distinct hazards – **15 significant hazards**

743 motif combinations – **40 significant motifs**

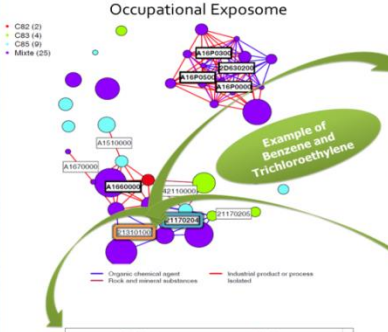
21 Singles | **19 Associations**

A : 6 | B : 3 | C : 2 | D : 10 | A : 1 | B : 2 | C : 0 | D : 16

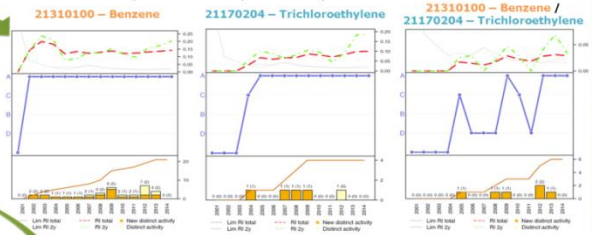
Sample Of Occupational NHL



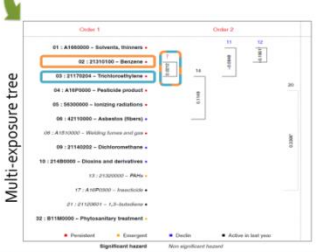
Occupational Exposome



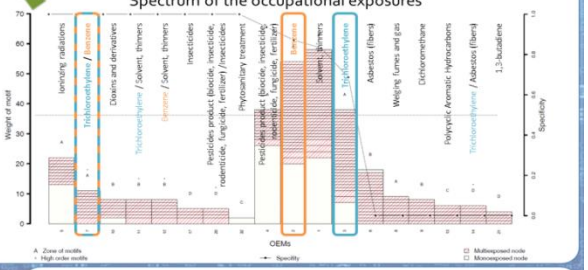
Dynamic of occupational exposures recruitment



Multi-exposure tree



Spectrum of the occupational exposures



CONCLUSION:

Observational Surveillance of NHL from the RNV3P has allowed to:

- highlight "new" occupational (multi)exposures, in addition to those already known, as 1,3-butadiene or dichloromethane, which are not very studied currently[3]
- study the dynamic structure of 40 occupational exposures, including 6 single persistent exposures and one multi-exposure: benzene and trichloroethylene association

PERSPECTIVES:

- It would be interesting to conduct further epidemiological investigations on highlighted associations
- This analysis could be completed by the study of workplaces related to each exposure motif
- Moreover this approach is not specific to occupational databases and has been already applied on other databases, not only in the occupational disease field.

BIBLIOGRAPHY

1 **Rieutort D.** The Observational Surveillance. PhD Project, 2015.

2 **Faisandier L, Bonneterre V, De Gaudemaris R, Bicout DJ.** Occupational exposome: a network-based approach for characterizing Occupational Health Problems. J Biomed Inform. 2011 Aug;44(4):545-52.

3 **Rieutort D, Moyne O, Cocco P, de Gaudemaris R, Bicout DJ.** Ranking occupational contexts associated with risk of non-Hodgkin Lymphoma. (reviewed in BMJ Open)

European Congress of Epidemiology – Maastricht – June 25-27th, 2015

ANNEXES

Annexe 1 : Liste des 166 motifs du cancer de la vessie identifiés par la Surveillance Observationnelle en 2014. 287
Annexe 2 : Liste des 436 motifs de l'asthme identifiés par la Surveillance Observationnelle en 2014. 291
Annexe 3 : Liste des 149 activités identifiées sur la matrice primaire-secondaire du LNH 300
Annexe 4 : Liste de tous les codes et libellés des motifs recensées pour les trois exemples d'application du RNV3P 304
Annexe 5 : Liste des 145 motifs identifiés pour les aides-soignants par la Surveillance Observationnelle en 2013. 308
Annexe 6 : Liste des 66 motifs identifiés pour les aides-soignants par la Surveillance Observationnelle en 2013. 311

Annexe 1 : Liste des 166 motifs du cancer de la vessie identifiés par la Surveillance Observationnelle en 2014.

ID	Ordre I	Ordre II	Ordre III	Nœuds	Poids	Statut*
1	Hydrocarbure aromatique polycyclique (HAP)	-	-	128	285	A
2	amiante (fibre)	-	-	137	160	A
3	solvant, diluant	-	-	86	106	A
4	diesel gaz d'échappement	-	-	64	103	A
5	trichloroéthylène	-	-	83	96	A
6	amine aromatique et dérivé	-	-	41	85	A
7	peinture, vernis, laque, mastic	-	-	51	83	A
8	suie et autre produit dégagé lors de l'utilisation de produit noir	-	-	52	81	B
9	amine aromatique	-	-	41	78	A
10	HAP	amiante (fibre)	-	45	59	A
11	fumée et gaz de soudage	-	-	41	55	A
12	huile de coupe, fluide d'usinage	-	-	34	50	A
13	matière colorante	-	-	30	48	A
14	huile et graisse lubrifiante	-	-	33	37	A
15	encre	-	-	25	33	A
16	huile, graisse et fluide d'usinage industriel	-	-	28	31	A
17	houille, charbon	-	-	20	29	A
18	HAP	diesel gaz d'échappement	-	17	28	A
19	amiante (fibre)	solvant, diluant	-	23	26	A
20	HAP	suie et autre produit dégagé lors de l'utilisation de produit noir	-	18	26	B
21	trichloroéthylène	HAP	-	24	25	A
22	trichloroéthylène	amiante (fibre)	-	24	24	A
23	huile et graisse lubrifiante minérale	-	-	20	22	A
24	gaz d'échappement	-	-	20	22	A
25	peinture, vernis, laque, mastic	solvant, diluant	-	18	22	B
26	amiante (fibre)	fumée et gaz de soudage	-	19	21	A
27	amiante (fibre)	diesel gaz d'échappement	-	18	21	A
28	HAP	amine aromatique et dérivé	-	8	19	B
29	tabac	-	-	7	17	B
30	produit pesticide (biocide, insecticide, rodenticide, fongicide, engrais)	-	-	10	16	B
31	HAP	amine aromatique	-	10	15	A
32	amiante (fibre)	peinture, vernis, laque, mastic	-	15	15	A
33	amiante (fibre)	suie et autre produit dégagé lors de l'utilisation de produit noir	-	13	15	B
34	HAP	fumée et gaz de soudage	-	12	14	B
35	poussière de bois	-	-	13	13	C
36	solvant, diluant	encre	-	8	13	D
37	autre HAP	-	-	8	12	D
38	amiante (fibre)	huile et graisse lubrifiante	-	12	12	C
39	huile et graisse lubrifiante	diesel gaz d'échappement	-	11	12	D
40	adhésif, colle	-	-	11	11	D
41	HAP	peinture, vernis, laque, mastic	-	10	11	D
42	HAP	solvant, diluant	-	10	11	D

43	amiante (fibre)	huile, graisse et fluide d'usinage industriel	-	10	11	D
44	**amiante (fibre)**	**huile de coupe, fluide d'usinage**	**-**	**11**	**11**	**D**
45	HAP	amiante (fibre)	diesel gaz d'échappement	8	11	D
46	benzène	-	-	10	10	D
47	produit noir	-	-	8	10	D
48	suie de combustion du charbon	-	-	9	10	D
49	trichloroéthylène	solvant, diluant	-	10	10	D
50	HAP	huile et graisse lubrifiante	-	9	10	D
51	**amiante (fibre)**	**huile et graisse lubrifiante minérale**	**-**	**10**	**10**	**C**
52	hydrocarbure aromatique et dérivé	-	-	5	9	D
53	dérivé nitrosé d'amine aliphatique	-	-	7	9	D
54	silice cristalline	-	-	9	9	D
55	poussière de métaux durs carbures métalliques	-	-	8	9	D
56	HAP	houille, charbon	-	8	9	D
57	amine aromatique et dérivé	amiante (fibre)	-	9	9	D
58	amine aromatique et dérivé	solvant, diluant	-	6	9	D
59	amine aromatique et dérivé	matière colorante	-	3	9	D
60	HAP	huile, graisse et fluide d'usinage industriel	-	9	9	D
61	trichloroéthylène	huile de coupe, fluide d'usinage	-	9	9	D
62	HAP	huile de coupe, fluide d'usinage	-	8	9	D
63	**solvant, diluant**	**huile et graisse lubrifiante**	**-**	**8**	**9**	**C**
64	diesel gaz d'échappement	suie et autre produit dégagé lors de l'utilisation de produit noir	-	9	9	D
65	**trichloroéthylène**	**HAP**	**amiante (fibre)**	**9**	**9**	**C**
66	plomb	-	-	8	8	D
67	tétrachloroéthylène	-	-	8	8	D
68	trichloroéthylène	peinture, vernis, laque, mastic	-	8	8	D
69	amine aromatique et dérivé	peinture, vernis, laque, mastic	-	3	8	D
70	amine aromatique	solvant, diluant	-	8	8	D
71	**amiante (fibre)**	**gaz d'échappement**	**-**	**8**	**8**	**D**
72	trichloroéthylène	diesel gaz d'échappement	-	7	8	D
73	trichloroéthylène	suie et autre produit dégagé lors de l'utilisation de produit noir	-	7	8	D
74	huile et graisse lubrifiante	suie et autre produit dégagé lors de l'utilisation de produit noir	-	7	8	D
75	solvant, diluant	huile, graisse et fluide d'usinage industriel	-	6	7	D
76	solvant, diluant	gaz d'échappement	-	7	7	D
77	acide chlorhydrique	-	-	6	6	D
78	peinture, vernis, laque, mastic sans précision en phase aqueuse	-	-	6	6	D
79	goudron	-	-	6	6	C
80	HAP	silice cristalline	-	6	6	D
81	amine aromatique	amiante (fibre)	-	6	6	D
82	**trichloroéthylène**	**fumée et gaz de soudage**	**-**	**5**	**6**	**D**
83	fumée et gaz de soudage	huile et graisse lubrifiante	-	6	6	D
84	trichloroéthylène	huile et graisse lubrifiante minérale	-	6	6	D
85	huile, graisse et fluide d'usinage industriel	diesel gaz d'échappement	-	6	6	D
86	huile et graisse lubrifiante minérale	diesel gaz d'échappement	-	6	6	D
87	huile de coupe, fluide d'usinage	suie et autre produit dégagé lors de l'utilisation de produit noir	-	5	6	D

88	HAP	amiante (fibre)	solvant, diluant	5	6	D
89	amiante (fibre)	solvant, diluant	gaz d'échappement	6	6	D
90	ammoniac (NH3)	-	-	5	5	D
91	styrène	-	-	5	5	D
92	silice	-	-	5	5	D
93	peinture, vernis, laque, mastic sans précision en phase solvant	-	-	5	5	D
94	teinture pour bois	-	-	5	5	D
95	huile de coupe pleine, fluide d'usinage pleine	-	-	5	5	D
96	HAP	dérivé nitrosé d'amine aliphatique	-	4	5	D
97	plomb	amiante (fibre)	-	5	5	D
98	benzène	amiante (fibre)	-	5	5	D
99	peinture, vernis, laque, mastic	adhésif, colle	-	5	5	D
100	adhésif, colle	solvant, diluant	-	5	5	D
101	trichloroéthylène	encre	-	5	5	D
102	amine aromatique et dérivé	encre	-	3	5	D
103	amine aromatique	encre	-	5	5	D
104	trichloroéthylène	huile et graisse lubrifiante	-	5	5	D
105	peinture, vernis, laque, mastic	huile et graisse lubrifiante	-	5	5	D
106	fumée et gaz de soudage	diesel gaz d'échappement	-	5	5	D
107	solvant, diluant	diesel gaz d'échappement	-	5	5	D
108	huile de coupe, fluide d'usinage	diesel gaz d'échappement	-	5	5	D
109	**amiante (fibre)**	**goudron**	**-**	**5**	**5**	**C**
110	amine aromatique et dérivé	suie et autre produit dégagé lors de l'utilisation de produit noir	-	5	5	D
111	fumée et gaz de soudage	suie et autre produit dégagé lors de l'utilisation de produit noir	-	5	5	D
112	solvant, diluant	poussière de bois	-	5	5	D
113	amiante (fibre)	peinture, vernis, laque, mastic	solvant, diluant	5	5	D
114	amiante (fibre)	solvant, diluant	huile, graisse et fluide d'usinage industriel	4	5	D
115	HAP	huile et graisse lubrifiante	diesel gaz d'échappement	5	5	D
116	amiante (fibre)	huile et graisse lubrifiante	diesel gaz d'échappement	5	5	D
117	HAP	amiante (fibre)	suie et autre produit dégagé lors de l'utilisation de produit noir	4	5	D
118	HAP	huile et graisse lubrifiante	suie et autre produit dégagé lors de l'utilisation de produit noir	4	5	D
119	toluène	-	-	4	4	D
120	acétone	-	-	4	4	D
121	4,4'-diaminobiphenyle	-	-	4	4	D
122	autre poussière métallique	-	-	4	4	D
123	résine époxydique	-	-	4	4	D
124	essence	-	-	4	4	D
125	bitume	-	-	4	4	D
126	produit photographique	-	-	4	4	D
127	fumée d'explosif	-	-	4	4	D
128	poussière inorganique	-	-	4	4	D
129	acide chlorhydrique	trichloroéthylène	-	4	4	D
130	trichloroéthylène	tétrachloroéthylène	-	4	4	D
131	trichloroéthylène	amine aromatique	-	4	4	D
132	trichloroéthylène	silice cristalline	-	4	4	D
133	silice cristalline	amiante (fibre)	-	4	4	D
134	amine aromatique	peinture, vernis, laque, mastic	-	4	4	D
135	fumée et gaz de soudage	solvant, diluant	-	4	4	D
136	amiante (fibre)	matière colorante	-	4	4	D
137	peinture, vernis, laque, mastic	matière colorante	-	4	4	D

138	solvant, diluant	matière colorante	-	4	4	D
139	HAP	gaz d'échappement	-	4	4	D
140	houille, charbon	diesel gaz d'échappement	-	3	4	D
141	peinture, vernis, laque, mastic	diesel gaz d'échappement	-	4	4	D
142	peinture, vernis, laque, mastic	suie et autre produit dégagé lors de l'utilisation de produit noir	-	4	4	D
143	solvant, diluant	suie et autre produit dégagé lors de l'utilisation de produit noir	-	4	4	D
144	huile, graisse et fluide d'usinage industriel	suie et autre produit dégagé lors de l'utilisation de produit noir	-	4	4	D
145	HAP	suie de combustion du charbon	-	4	4	D
146	amiante (fibre)	suie de combustion du charbon	-	4	4	D
147	peinture, vernis, laque, mastic	poussière de bois	-	4	4	D
148	trichloroéthylène	HAP	silice cristalline	4	4	D
149	HAP	amiante (fibre)	fumée et gaz de soudage	4	4	D
150	HAP	amiante (fibre)	huile, graisse et fluide d'usinage industriel	4	4	D
151	trichloroéthylène	HAP	huile de coupe, fluide d'usinage	4	4	D
152	trichloroéthylène	amiante (fibre)	huile de coupe, fluide d'usinage	4	4	D
153	trichloroéthylène	amiante (fibre)	huile et graisse lubrifiante	4	4	D
154	HAP	amiante (fibre)	huile et graisse lubrifiante	4	4	D
155	4,4'-méthylene-bis-orthochloraniline	-	-	1	3	D
156	fumée de tabac	-	-	2	3	D
157	dérivé nitrosé d'amine aliphatique	amine aromatique et dérivé	-	2	3	D
158	autre HAP	amiante (fibre)	-	2	3	D
159	HAP	poussière de métaux durs carbures métalliques	-	2	3	D
160	HAP	dérivé nitrosé d'amine aliphatique	amine aromatique et dérivé	2	3	D
161	amine aromatique et dérivé	solvant, diluant	encre	2	3	D
162	produit capillaire	-	-	1	2	D
163	dérivé nitrosé d'amine aliphatique	huile de coupe, fluide d'usinage	-	1	2	D
164	huile et graisse lubrifiante	gaz d'échappement	-	1	2	D
165	matière colorante	produit capillaire	-	1	2	D
166	diesel gaz d'échappement	fumée de tabac	-	1	2	D

*En gras sont identifiés les motifs dont le statut est considéré comme actif (A : persistant, B : en déclin, C : émergent) ainsi que 10 motifs latents (D) actifs en 2013.

Annexe 2 : Liste des 436 motifs de l'asthme identifiés par la Surveillance Observationnelle en 2014.

ID	Ordre I	Ordre II	Ordre III	Nœuds	Poids	Statut*
1	poussière de farine	-	-	68	769	A
2	composé d'ammonium quaternaire	-	-	117	440	A
3	cyanate et isocyanate	-	-	79	298	A
4	oxydant et décolorant (cosmétique)	-	-	35	274	A
5	acarien ou mite ou gale	-	-	99	221	A
6	produit d'entretien	-	-	74	218	A
7	produit capillaire	-	-	37	212	A
8	poussière de bois	-	-	76	203	A
9	antiseptique, bactéricide pour la peau (sauf SHA solutions hydro-alcooliques)	-	-	92	198	A
10	protéine de latex	-	-	67	179	B
11	solvant, diluant	-	-	100	173	A
12	formaldéhyde	-	-	88	163	A
13	plastique thermodurcissable polyuréthane	-	-	73	159	A
14	peinture, vernis, laque, mastic	-	-	66	153	A
15	amine aliphatique	-	-	64	136	A
16	lessive détergent	-	-	51	134	A
17	poussière inorganique	-	-	66	128	A
18	fumée et gaz de soudage	-	-	44	108	A
19	glutaraldéhyde	-	-	32	87	A
20	poussière organique	-	-	46	85	A
21	chlore	-	-	29	72	A
22	persulfate	-	-	23	68	A
23	enzyme	-	-	15	65	B
24	résine époxydique	-	-	35	62	A
25	adhésif, colle	-	-	37	61	A
26	chrome	-	-	43	59	B
27	champignon, moisissure	-	-	33	58	A
28	colophane	-	-	28	57	B
29	graminée céréale	-	-	33	53	A
30	huile de coupe, fluide d'usinage	-	-	25	48	B
31	acarien ou mite ou gale	poussière de farine	-	14	47	A
32	composé d'ammonium quaternaire	amine aliphatique	-	16	46	B
33	méthacrylate de méthyle	-	-	26	45	A
34	pollens	-	-	31	44	A
35	amiante (fibre)	-	-	38	43	A
36	produit capillaire	oxydant et décolorant (cosmétique)	-	9	43	B
37	poussière de chantier	-	-	22	42	A
38	enzyme	poussière de farine	-	7	41	B
39	poussière de ciment, aluminosilicate de calcium	-	-	26	40	A
40	détergent désinfectant pour locaux (sols, murs, surfaces) domestique et professionnel	-	-	18	38	A
41	composé d'ammonium quaternaire	glutaraldéhyde	-	17	38	B
42	nickel	-	-	29	37	B
43	caoutchouc, latex naturel	-	-	26	37	A
44	teinture (cosmétique)	-	-	9	37	B
45	rat	-	-	16	36	A
46	encre	-	-	23	35	B
47	adhésif, colle cyanoacrylate	-	-	20	34	B
48	composé d'ammonium quaternaire	antiseptique, bactéricide pour la peau (sauf SHA solutions hydro-alcooliques)	-	14	34	B
49	trichlorure d'azote	-	-	8	33	A
50	poussière de papier carton	-	-	16	33	A
51	composé d'ammonium quaternaire	produit d'entretien	-	12	33	B

52	**peinture, vernis, laque, mastic sans précision en phase solvant**	**-**	**-**	**23**	**31**	**D**
53	poussière d'acier inox poussière d'acier ou fonte poussière de métaux durs carbures métalliques	-	-	18	30	D
54	cobalt	-	-	23	29	D
55	**ambiance thermique froide**	**-**	**-**	**18**	**29**	**D**
56	composé d'ammonium quaternaire	protéine de latex	-	12	29	D
57	souris	-	-	14	28	D
58	poussière d'origine animale	-	-	17	28	D
59	**poussière d'origine végétale**	**-**	**-**	**15**	**28**	**C**
60	sulfate et sulfite	-	-	16	27	D
61	**silice**	**-**	**-**	**21**	**27**	**C**
62	**peinture, vernis, laque, mastic**	**solvant, diluant**	**-**	**14**	**26**	**D**
63	lessive détergent	antiseptique, bactéricide pour la peau (sauf SHA solutions hydro-alcooliques)	-	11	26	D
64	**acide chlorhydrique**	**-**	**-**	**23**	**24**	**C**
65	**méthacrylate**	**-**	**-**	**12**	**24**	**C**
66	autre cyanate et isocyanate	-	-	6	24	D
67	**ambiance humide**	**-**	**-**	**21**	**24**	**C**
68	gaz d'échappement	-	-	14	23	D
69	**décapant**	**-**	**-**	**19**	**23**	**D**
70	parfum et substance aromatique	-	-	16	23	D
71	**blé**	**-**	**-**	**13**	**22**	**C**
72	acrylate	-	-	13	21	D
73	peinture, vernis, laque, mastic polyuréthanne phase solvant	-	-	11	21	D
74	coton (fibre)	-	-	8	21	D
75	**ef3 physique**	**-**	**-**	**15**	**20**	**D**
76	matière plastique, caoutchouc	-	-	11	20	D
77	plume, duvet	-	-	16	20	D
78	latex d'hévéa	-	-	10	20	D
79	hydroxyde de sodium	-	-	17	19	D
80	ammoniac (NH3)	-	-	15	19	D
81	**sel de sodium de la N-chloro-p-toluènesulfonamide**	**-**	**-**	**5**	**19**	**D**
82	éthanolamine	-	-	11	19	D
83	**fourrure, feutre**	**-**	**-**	**9**	**19**	**D**
84	fibre végétale textile	-	-	8	19	D
85	cyanate et isocyanate	solvant, diluant	-	11	19	D
86	tétrachloroéthylène	-	-	12	18	D
87	toluène-2,4-diisocyanate	-	-	11	18	D
88	silice cristalline	-	-	17	18	D
89	laine verre	-	-	15	18	D
90	résine	-	-	14	18	D
91	produit pesticide (biocide, insecticide, rodenticide, fongicide, engrais)	-	-	11	18	D
92	protéine animale	-	-	9	18	D
93	rat	souris	-	7	18	D
94	composé d'ammonium quaternaire	lessive détergent	-	7	18	D
95	mammifère et cétacé	-	-	11	17	D
96	Rongeur	-	-	6	17	D
97	autre poussière métallique	-	-	15	17	D
98	plancher maison	-	-	9	17	D
99	matière colorante	-	-	13	17	D
100	produit pour l'alimentation	-	-	11	16	D
101	insecticide	-	-	14	16	D
102	chrome	nickel	-	13	16	D
103	cyanate et isocyanate	peinture, vernis, laque, mastic	-	5	16	D
104	bois exotique sans précision	-	-	7	15	D
105	fleur et plante verte	-	-	10	15	D
106	**tabac**	**-**	**-**	**6**	**15**	**C**
107	peinture, vernis, laque, mastic sans précision en phase aqueuse	-	-	12	15	D

108	**produit de décomposition thermique de matière plastique**	**-**	**-**	**6**	**15**	**D**
109	**produit de décomposition thermique de matière plastique**	**-**	**-**	**12**	**15**	**C**
110	composé d'ammonium quaternaire	formaldéhyde	-	9	15	D
111	**composé d'ammonium quaternaire**	**détergent désinfectant pour locaux (sols, murs, surfaces) domestique et professionnel**	**-**	**7**	**15**	**C**
112	**oxydant et décolorant (cosmétique)**	**teinture (cosmétique)**	**-**	**4**	**15**	**D**
113	peinture, vernis, laque, mastic	poussière de bois	-	11	15	D
114	acide peracétique	-	-	12	14	D
115	amine aromatique et dérivé	-	-	11	14	D
116	phénylènediamine	-	-	7	14	D
117	insecte	-	-	13	14	D
118	**seigle**	**-**	**-**	**9**	**14**	**C**
119	dégraissant	-	-	14	14	D
120	enzyme protéolytique	-	-	9	14	D
121	formaldéhyde	glutaraldéhyde	-	6	14	D
122	solvant, diluant	plastique thermodurcissable polyuréthane	-	10	14	D
123	hypochlorite	-	-	11	13	D
124	ammoniaque (NH4OH)	-	-	10	13	D
125	**arbre et plante**	**-**	**-**	**8**	**13**	**C**
126	thermoplastique polyester	-	-	11	13	D
127	acarien ou mite ou gale	produit d'entretien	-	7	13	D
128	produit d'entretien	antiseptique, bactéricide pour la peau (sauf SHA solutions hydro-alcooliques)	-	8	13	D
129	antiseptique, bactéricide pour la peau (sauf SHA solutions hydro-alcooliques)	protéine de latex	-	6	13	D
130	cyanate et isocyanate	poussière de bois	-	6	13	D
131	**acarien ou mite ou gale**	**graminée céréale**	**-**	**8**	**12**	**C**
132	cyanate et isocyanate	plastique thermodurcissable polyuréthane	-	8	12	D
133	peinture, vernis, laque, mastic	plastique thermodurcissable polyuréthane	-	8	12	D
134	acarien ou mite ou gale	protéine de latex	-	8	12	D
135	produit d'entretien	protéine de latex	-	8	12	D
136	aldéhyde aliphatique	-	-	8	11	D
137	diphénylméthanediisocyanate	-	-	9	11	D
138	amine aliphatique	huile de coupe, fluide d'usinage	-	7	11	D
139	**persulfate**	**produit de décomposition thermique de matière plastique**	**-**	**8**	**11**	**C**
140	**acarien ou mite ou gale**	**pollens**	**-**	**10**	**11**	**C**
141	lessive détergent	protéine de latex	-	7	11	D
142	adhésif, colle	poussière de bois	-	8	11	D
143	plastique thermodurcissable polyuréthane	poussière de bois	-	6	11	D
144	sulfite	-	-	6	10	D
145	acétone	-	-	10	10	D
146	benzisothiazoline-3-one	-	-	10	10	D
147	autre amine aliphatique	-	-	6	10	D
148	hexaméthylènediisocyanate	-	-	8	10	D
149	**chat**	**-**	**-**	**8**	**10**	**D**
150	adhésif, colle urée formol	-	-	8	10	D
151	diesel gaz d'échappement	-	-	9	10	D
152	produit cosmétique divers	-	-	3	10	D
153	œuf	-	-	3	10	D
154	solvant, diluant	poussière inorganique	-	9	10	D
155	amine aliphatique	protéine de latex	-	5	10	D
156	chromate	-	-	6	9	D
157	anhydride phtalique	-	-	6	9	D
158	aspergillus	-	-	7	9	D
159	crustacé	-	-	3	9	D
160	huile et graisse lubrifiante	-	-	8	9	D
161	produit pharmaceutique	-	-	5	9	D
162	**blé**	**seigle**	**-**	**4**	**9**	**C**

163	formaldéhyde	adhésif, colle	-	7	9	D
164	peinture, vernis, laque, mastic	adhésif, colle	-	6	9	D
165	cyanate et isocyanate	résine époxydique	-	4	9	D
166	adhésif, colle	solvant, diluant	-	5	9	D
167	solvant, diluant	encre	-	6	9	D
168	**persulfate**	**oxydant et décolorant (cosmétique)**	**-**	**6**	**9**	**C**
169	formaldéhyde	poussière de bois	-	7	9	D
170	hydrure de sulfure	-	-	5	8	D
171	hydrocarbure alicyclique insaturé - terpène	-	-	7	8	D
172	hydroquinone	-	-	7	8	D
173	levure	-	-	5	8	D
174	**chien**	**-**	**-**	**6**	**8**	**D**
175	armoise	-	-	8	8	D
176	fumée et gaz de soudage d'aluminium	-	-	3	8	D
177	peinture, vernis, laque, mastic sans précision en poudre	-	-	7	8	D
178	produit de décomposition thermique du PVC	-	-	7	8	D
179	produit dégagé lors de l'utilisation normale du caoutchouc	-	-	5	8	D
180	huile, graisse et fluide d'usinage industriel	-	-	8	8	D
181	insecticide organophosphoré	-	-	6	8	D
182	fumée d'explosif	-	-	7	8	D
183	déchet de poussière organique	-	-	2	8	D
184	cobalt	chrome	-	7	8	D
185	formaldéhyde	amine aliphatique	-	4	8	D
186	amine aliphatique	cyanate et isocyanate	-	6	8	D
187	peinture, vernis, laque, mastic sans précision en phase solvant	peinture, vernis, laque, mastic sans précision en phase aqueuse	-	8	8	D
188	peinture, vernis, laque, mastic sans précision en phase solvant	solvant, diluant	-	7	8	D
189	fumée et gaz de soudage	huile de coupe, fluide d'usinage	-	4	8	D
190	amine aliphatique	antiseptique, bactéricide pour la peau (sauf SHA solutions hydro-alcooliques)	-	6	8	D
191	solvant, diluant	antiseptique, bactéricide pour la peau (sauf SHA solutions hydro-alcooliques)	-	8	8	D
192	acarien ou mite ou gale	enzyme	-	2	8	D
193	acarien ou mite ou gale	poussière organique	-	4	8	D
194	produit d'entretien	poussière organique	-	3	8	D
195	fumée et gaz de soudage	poussière inorganique	-	5	8	D
196	poussière organique	poussière inorganique	-	5	8	D
197	acide carboxylique aliphatique et dérivé	-	-	7	7	D
198	poussière d'alliage léger (aluminium...)	-	-	4	7	D
199	**équipement de protection**	**-**	**-**	**7**	**7**	**D**
200	composé d'ammonium quaternaire	acide peracétique	-	6	7	D
201	cyanate et isocyanate	toluène-2,4-diisocyanate	-	5	7	D
202	fumée et gaz de soudage	peinture, vernis, laque, mastic	-	7	7	D
203	fumée et gaz de soudage	solvant, diluant	-	4	7	D
204	résine époxydique	solvant, diluant	-	5	7	D
205	formaldéhyde	plastique thermodurcissable polyuréthane	-	6	7	D
206	peinture, vernis, laque, mastic sans précision en phase solvant	plastique thermodurcissable polyuréthane	-	7	7	D
207	formaldéhyde	antiseptique, bactéricide pour la peau (sauf SHA solutions hydro-alcooliques)	-	7	7	D
208	acarien ou mite ou gale	poussière inorganique	-	6	7	D
209	oxydant et décolorant (cosmétique)	protéine de latex	-	3	7	D
210	graminée céréale	poussière de farine	-	5	7	D
211	acide phosphorique	-	-	6	6	D

212	acide sulfurique	-	-	6	6	D
213	toluène	-	-	6	6	D
214	thiazole, isothiazole et dérivé	-	-	6	6	D
215	acide acétique	-	-	6	6	D
216	plante alimentaire, légume ou fruit	-	-	6	6	D
217	courant d'air	-	-	5	6	D
218	produit de vulcanisation du caoutchouc	-	-	6	6	D
219	chlore	trichlorure d'azote	-	2	6	D
220	acarien ou mite ou gale	insecte	-	6	6	D
221	ambiance thermique froide	ef3 physique	-	4	6	D
222	amiante (fibre)	fumée et gaz de soudage	-	5	6	D
223	résine époxydique	plastique thermodurcissable polyuréthane	-	6	6	D
224	amine aliphatique	produit d'entretien	-	5	6	D
225	**acarien ou mite ou gale**	**détergent désinfectant pour locaux (sols, murs, surfaces) domestique et professionnel**	**-**	**2**	**6**	**C**
226	lessive détergent	poussière inorganique	-	2	6	D
227	antiseptique, bactéricide pour la peau (sauf SHA solutions hydro-alcooliques)	poussière inorganique	-	6	6	D
228	acarien ou mite ou gale	poussière d'origine animale	-	6	6	D
229	pollens	poussière de farine	-	5	6	D
230	acarien ou mite ou gale	enzyme	poussière de farine	1	6	D
231	oxyde de fer	-	-	5	5	D
232	tungstène	-	-	3	5	D
233	peroxyde d'hydrogène	-	-	4	5	D
234	autre acide et peracide carboxylique	-	-	5	5	D
235	amine alicyclique et dérivé	-	-	5	5	D
236	oiseau	-	-	5	5	D
237	ambiance thermique chaude	-	-	5	5	D
238	port de charge	-	-	5	5	D
239	produit de décomposition thermique d'autre produit divers	-	-	5	5	D
240	carbone (fibre)	-	-	5	5	D
241	déchet de fibre végétale	-	-	5	5	D
242	cobalt	tungstène	-	3	5	D
243	chrome	méthacrylate de méthyle	-	4	5	D
244	formaldéhyde	cyanate et isocyanate	-	5	5	D
245	composé d'ammonium quaternaire	acarien ou mite ou gale	-	4	5	D
246	champignon, moisissure	acarien ou mite ou gale	-	5	5	D
247	chat	chien	-	3	5	D
248	acarien ou mite ou gale	rat	-	5	5	D
249	cobalt	poussière d'acier inox poussière d'acier ou fonte poussière de métaux durs carbures métalliques	-	2	5	D
250	poussière d'acier inox poussière d'acier ou fonte poussière de métaux durs carbures métalliques	fumée et gaz de soudage	-	2	5	D
251	peinture, vernis, laque, mastic sans précision en phase aqueuse	plastique thermodurcissable polyuréthane	-	5	5	D
252	composé d'ammonium quaternaire	décapant	-	4	5	D
253	acarien ou mite ou gale	lessive détergent	-	3	5	D
254	phénylènediamine	oxydant et décolorant (cosmétique)	-	2	5	D
255	plastique thermodurcissable polyuréthane	antiseptique, bactéricide pour la peau (sauf SHA solutions hydro-alcooliques)	-	5	5	D
256	composé d'ammonium quaternaire	enzyme protéolytique	-	5	5	D
257	solvant, diluant	poussière organique	-	4	5	D
258	peinture, vernis, laque, mastic	poussière inorganique	-	4	5	D
259	fumée et gaz de soudage	colophane	-	4	5	D
260	produit capillaire	protéine de latex	-	3	5	D

261	champignon, moisissure	poussière de farine	-	5	5	D
262	œuf	poussière de farine	-	2	5	D
263	solvant, diluant	poussière de bois	-	5	5	D
264	composé d'ammonium quaternaire	formaldéhyde	glutaraldéhyde	3	5	D
265	composé d'ammonium quaternaire	amine aliphatique	antiseptique, bactéricide pour la peau (sauf SHA solutions hydro-alcooliques)	3	5	D
266	acide fluorhydrique	-	-	4	4	D
267	acide nitrique	-	-	4	4	D
268	autre composé d'ammonium quaternaire	-	-	4	4	D
269	styrène	-	-	4	4	D
270	aldéhyde aliphatique saturé	-	-	4	4	D
271	autre anhydride d'acide carboxylique	-	-	4	4	D
272	phtalate	-	-	4	4	D
273	microbiologique	-	-	4	4	D
274	alternaria	-	-	4	4	D
275	autres acariens, mites et gales	-	-	4	4	D
276	bœuf ou veau ou vache	-	-	4	4	D
277	hamster	-	-	4	4	D
278	sapin	-	-	3	4	D
279	**mais**	**-**	**-**	**4**	**4**	**C**
280	**sarrasin**	**-**	**-**	**4**	**4**	**C**
281	soja	-	-	4	4	D
282	laine de roche	-	-	4	4	D
283	**solvant de nettoyage**	**-**	**-**	**4**	**4**	**C**
284	thermoplastique polyacrylique ou polymétacrylique	-	-	4	4	D
285	plastique thermodurcissable urée-formaldéhyde	-	-	4	4	D
286	détergent désinfectant pour matériel médical, instruments	-	-	3	4	D
287	nettoyant désinfectant alimentaire	-	-	4	4	D
288	fongicide	-	-	4	4	D
289	**alpha amylase**	**-**	**-**	**3**	**4**	**C**
290	acide chlorhydrique	chrome	-	4	4	D
291	cobalt	nickel	-	4	4	D
292	composé d'ammonium quaternaire	autre amine aliphatique	-	4	4	D
293	composé d'ammonium quaternaire	éthanolamine	-	3	4	D
294	formaldéhyde	champignon, moisissure	-	4	4	D
295	hamster	rat	-	4	4	D
296	silice cristalline	amiante (fibre)	-	4	4	D
297	cyanate et isocyanate	fumée et gaz de soudage	-	3	4	D
298	poussière d'alliage léger (aluminium...)	fumée et gaz de soudage d'aluminium	-	1	4	D
299	amiante (fibre)	solvant, diluant	-	4	4	D
300	adhésif, colle	plastique thermodurcissable polyuréthane	-	4	4	D
301	glutaraldéhyde	produit d'entretien	-	1	4	D
302	composé d'ammonium quaternaire	détergent désinfectant pour matériel médical, instruments	-	3	4	D
303	produit d'entretien	lessive détergent	-	4	4	D
304	décapant	lessive détergent	-	3	4	D
305	**produit de décomposition thermique de matière plastique**	**oxydant et décolorant (cosmétique)**	**-**	**3**	**4**	**C**
306	produit capillaire	parfum et substance aromatique	-	3	4	D
307	glutaraldéhyde	antiseptique, bactéricide pour la peau (sauf SHA solutions hydro-alcooliques)	-	4	4	D
308	dégraissant	antiseptique, bactéricide pour la peau (sauf SHA solutions hydro-alcooliques)	-	4	4	D
309	glutaraldéhyde	enzyme protéolytique	-	4	4	D
310	antiseptique, bactéricide pour la peau (sauf SHA solutions hydro-	produit pharmaceutique	-	1	4	D

	alcooliques)					
311	gaz d'échappement	poussière organique	-	4	4	D
312	poussière de ciment, aluminosilicate de calcium	poussière inorganique	-	3	4	D
313	**produit d'entretien**	**poussière inorganique**	**-**	**3**	**4**	**D**
314	résine époxydique	carbone (fibre)	-	4	4	D
315	graminée céréale	pollens	-	3	4	D
316	solvant, diluant	colophane	-	4	4	D
317	produit d'entretien	latex d'hévéa	-	2	4	D
318	formaldéhyde	protéine de latex	-	2	4	D
319	rat	protéine de latex	-	2	4	D
320	souris	protéine de latex	-	2	4	D
321	levure	poussière de farine	-	1	4	D
322	protéine de latex	poussière de farine	-	3	4	D
323	amiante (fibre)	poussière de bois	-	3	4	D
324	poussière inorganique	poussière de bois	-	4	4	D
325	peinture, vernis, laque, mastic sans précision en phase solvant	peinture, vernis, laque, mastic sans précision en phase aqueuse	plastique thermodurcissable polyuréthane	4	4	D
326	composé d'ammonium quaternaire	glutaraldéhyde	produit d'entretien	1	4	D
327	**persulfate**	**produit de décomposition thermique de matière plastique**	**oxydant et décolorant (cosmétique)**	**3**	**4**	**C**
328	composé d'ammonium quaternaire	amine aliphatique	protéine de latex	2	4	D
329	rat	souris	protéine de latex	2	4	D
330	lessive détergent	antiseptique, bactéricide pour la peau (sauf SHA solutions hydro-alcooliques)	protéine de latex	2	4	D
331	peinture, vernis, laque, mastic	adhésif, colle	poussière de bois	3	4	D
332	hêtre	-	-	2	3	D
333	fumée et gaz de soudage d'inox	-	-	2	3	D
334	composé chimique carcinogène	-	-	1	3	D
335	hydrolase	-	-	2	3	D
336	aromate épice	-	-	1	3	D
337	chlore	acide chlorhydrique	-	2	3	D
338	hydrure de sulfure	ammoniac (NH3)	-	1	3	D
339	nickel	méthacrylate de méthyle	-	2	3	D
340	chrome	cyanate et isocyanate	-	2	3	D
341	acarien ou mite ou gale	mammifère et cétacé	-	2	3	D
342	arbre et plante	fleur et plante verte	-	2	3	D
343	tungstène	poussière d'acier inox poussière d'acier ou fonte poussière de métaux durs carbures métalliques	-	1	3	D
344	acarien ou mite ou gale	plancher maison	-	1	3	D
345	ef3 physique	poussière de chantier	-	2	3	D
346	autre poussière métallique	fumée et gaz de soudage	-	2	3	D
347	nickel	fumée et gaz de soudage d'inox	-	2	3	D
348	hexaméthylènediisocyanate	peinture, vernis, laque, mastic	-	2	3	D
349	poussière de ciment, aluminosilicate de calcium	peinture, vernis, laque, mastic	-	2	3	D
350	méthacrylate de méthyle	résine époxydique	-	2	3	D
351	poussière d'acier inox poussière d'acier ou fonte poussière de métaux durs carbures métalliques	huile de coupe, fluide d'usinage	-	2	3	D
352	chlore	produit d'entretien	-	2	3	D
353	hydroxyde de sodium	produit d'entretien	-	2	3	D
354	oxydant et décolorant (cosmétique)	parfum et substance aromatique	-	2	3	D
355	chlore	antiseptique, bactéricide pour la peau (sauf SHA solutions hydro-alcooliques)	-	1	3	D
356	peinture, vernis, laque, mastic	antiseptique, bactéricide pour la peau (sauf SHA solutions hydro-alcooliques)	-	2	3	D
357	cyanate et isocyanate	insecticide organophosphoré	-	2	3	D
358	seigle	alpha amylase	-	2	3	D

359	silice	poussière inorganique	-	2	3	D
360	silice cristalline	poussière inorganique	-	2	3	D
361	antiseptique, bactéricide pour la peau (sauf SHA solutions hydro-alcooliques)	plume, duvet	-	2	3	D
362	plume, duvet	fourrure, feutre	-	2	3	D
363	antiseptique, bactéricide pour la peau (sauf SHA solutions hydro-alcooliques)	poussière d'origine animale	-	2	3	D
364	glutaraldéhyde	protéine de latex	-	2	3	D
365	produit d'entretien	poussière de farine	-	1	3	D
366	hydrolase	poussière de farine	-	2	3	D
367	poussière organique	poussière de farine	-	2	3	D
368	aromate épice	poussière de farine	-	1	3	D
369	acarien ou mite ou gale	poussière de bois	-	1	3	D
370	hêtre	poussière de bois	-	2	3	D
371	sapin	poussière de bois	-	2	3	D
372	bois exotique sans précision	poussière de bois	-	2	3	D
373	résine époxydique	poussière de bois	-	2	3	D
374	colophane	poussière de bois	-	2	3	D
375	chrome	nickel	méthacrylate de méthyle	2	3	D
376	cobalt	tungstène	poussière d'acier inox poussière d'acier ou fonte poussière de métaux durs carbures métalliques	1	3	D
377	cyanate et isocyanate	solvant, diluant	plastique thermodurcissable polyuréthane	2	3	D
378	composé d'ammonium quaternaire	lessive détergent	antiseptique, bactéricide pour la peau (sauf SHA solutions hydro-alcooliques)	2	3	D
379	composé d'ammonium quaternaire	antiseptique, bactéricide pour la peau (sauf SHA solutions hydro-alcooliques)	protéine de latex	1	3	D
380	acarien ou mite ou gale	pollens	poussière de farine	2	3	D
381	hydroxyde de calcium	-	-	1	2	D
382	caféier	-	-	1	2	D
383	ail	-	-	1	2	D
384	oignon	-	-	1	2	D
385	adhésif, colle polyuréthane	-	-	1	2	D
386	autre huile de coupage	-	-	1	2	D
387	fromage	-	-	1	2	D
388	huile essentielle	-	-	1	2	D
389	hydroxyde de sodium	ammoniaque (NH4OH)	-	1	2	D
390	péroxyde d'hydrogène	acide peracétique	-	1	2	D
391	acrylate	méthacrylate	-	1	2	D
392	hydroxyde de sodium	sulfate et sulfite	-	1	2	D
393	chlore	sel de sodium de la N-chloro-p-toluènesulfonamide	-	1	2	D
394	champignon, moisissure	aspergillus	-	1	2	D
395	champignon, moisissure	caféier	-	1	2	D
396	hêtre	sapin	-	1	2	D
397	ammoniac (NH3)	graminée céréale	-	1	2	D
398	ail	oignon	-	1	2	D
399	acarien ou mite ou gale	Rongeur	-	1	2	D
400	champignon, moisissure	courant d'air	-	1	2	D
401	hydroxyde de calcium	poussière de ciment, aluminosilicate de calcium	-	1	2	D
402	amine aromatique et dérivé	résine époxydique	-	1	2	D
403	formaldéhyde	solvant, diluant	-	1	2	D
404	éthanolamine	huile de coupe, fluide d'usinage	-	1	2	D
405	fumée et gaz de soudage	autre huile de coupage	-	1	2	D
406	fumée et gaz de soudage	huile et graisse lubrifiante	-	1	2	D
407	peinture, vernis, laque, mastic	diesel gaz d'échappement	-	1	2	D
408	solvant, diluant	diesel gaz d'échappement	-	1	2	D
409	ammoniaque (NH4OH)	produit d'entretien	-	1	2	D
410	plancher maison	produit d'entretien	-	1	2	D
411	sulfite	lessive détergent	-	1	2	D
412	formaldéhyde	produit capillaire	-	1	2	D
413	acarien ou mite ou gale	produit capillaire	-	1	2	D

414	sulfate et sulfite	oxydant et décolorant (cosmétique)	-	1	2	D
415	cyanate et isocyanate	antiseptique, bactéricide pour la peau (sauf SHA solutions hydro-alcooliques)	-	1	2	D
416	graminée céréale	enzyme protéolytique	-	1	2	D
417	blé	alpha amylase	-	1	2	D
418	enzyme	œuf	-	1	2	D
419	adhésif, colle cyanoacrylate	colophane	-	1	2	D
420	glutaraldéhyde	latex d'hévéa	-	1	2	D
421	antiseptique, bactéricide pour la peau (sauf SHA solutions hydro-alcooliques)	latex d'hévéa	-	1	2	D
422	sel de sodium de la N-chloro-p-toluènesulfonamide	protéine de latex	-	1	2	D
423	Rongeur	protéine de latex	-	1	2	D
424	poussière de ciment, aluminosilicate de calcium	protéine de latex	-	1	2	D
425	fumée d'explosif	poussière de farine	-	1	2	D
426	protéine animale	poussière de farine	-	1	2	D
427	poussière d'origine végétale	poussière de farine	-	1	2	D
428	produit pesticide (biocide, insecticide, rodenticide, fongicide, engrais)	poussière de bois	-	1	2	D
429	peinture, vernis, laque, mastic	solvant, diluant	diesel gaz d'échappement	1	2	D
430	hydroxyde de sodium	ammoniaque (NH4OH)	produit d'entretien	1	2	D
431	produit capillaire	oxydant et décolorant (cosmétique)	parfum et substance aromatique	1	2	D
432	blé	seigle	alpha amylase	1	2	D
433	acarien ou mite ou gale	produit d'entretien	poussière organique	1	2	D
434	produit capillaire	oxydant et décolorant (cosmétique)	protéine de latex	1	2	D
435	enzyme	œuf	poussière de farine	1	2	D
436	hêtre	sapin	poussière de bois	1	2	D

*En gras sont identifiés les motifs dont le statut est considéré comme actif (A : persistant, B : en déclin, C : émergent) ainsi que 13 motifs latents (D) actifs en 2013.

Annexe 3 : Liste des 149 activités identifiées sur la matrice primaire-secondaire du LNH

ID	NAF	CITP
1	Recherche et développement	Professions intermédiaires des sciences physiques et techniques
2	Cokéfaction, raffinage, industries nucléaires	Professions intermédiaires des sciences physiques et techniques
3	Travail des métaux	Artisans et ouvriers des métiers de l'extraction et du bâtiment
4	Métallurgie	Professions intermédiaires des sciences et techniques
5	Administration publique	Employés de bureau
6	Industrie chimique	Conducteurs d'installations et de matériels fixes et assimilés
7	Services personnels	Employés non qualifiés des services et de la vente
8	Santé et action sociale	Professions intermédiaires des sciences de la vie et de la santé
9	Industrie du papier et du carton	Professions intermédiaires des sciences physiques et techniques
10	Education	Professions intermédiaires des sciences physiques et techniques
11	Fabrication d'autres matériels de transport	Artisans et ouvriers des métiers de la métallurgie, de la construction mécanique et assimilés
12	Métallurgie	Artisans et ouvriers des métiers de la métallurgie, de la construction mécanique et assimilés
13	Travail des métaux	Artisans et ouvriers des métiers de la métallurgie, de la construction mécanique et assimilés
14	Culture et production animale, chasse et services annexes	Agriculteurs et ouvriers qualifiés de l'agriculture commerciale
15	Commerce et réparation automobile	Artisans et ouvriers des métiers de la métallurgie, de la construction mécanique et assimilés
16	Industrie chimique	Artisans et ouvriers des métiers de l'extraction et du bâtiment
17	Santé et action sociale	Artisans et ouvriers de la mécanique de précision, des métiers d'art, de l'imprimerie et assimilés
18	Non renseigné	Conducteurs d'installations et de matériels fixes et assimilés
19	Agriculture, chasse, services annexes	Agriculteurs et ouvriers de l'agriculture et de la pêche de subsistance
20	Non renseigné	Professions intermédiaires des sciences de la vie et de la santé
21	Industrie de l'habillement et des fourrures	Autres artisans et ouvriers des métiers de type artisanal
22	Construction	Artisans et ouvriers des métiers de l'extraction et du bâtiment
23	Commerce de détail et réparation d'articles domestiques	Modèles, vendeurs et démonstrateurs
24	Commerce et réparation automobile	Artisans et ouvriers des métiers de l'extraction et du bâtiment
25	Cokéfaction, raffinage, industries nucléaires	Artisans et ouvriers des métiers de l'extraction et du bâtiment
26	Industrie automobile	Professions intermédiaires des sciences physiques et techniques
27	Commerce de gros et intermédiaires du commerce	Conducteurs de véhicules et d'engins lourds de levage et de manœuvre
28	Transports terrestres	Conducteurs de véhicules et d'engins lourds de levage et de manœuvre
29	Production et distribution d'électricité, de gaz et de chaleur	Artisans et ouvriers des métiers de la métallurgie, de la construction mécanique et assimilés
30	Agriculture, chasse, services annexes	Artisans et ouvriers des métiers de la métallurgie, de la construction mécanique et assimilés
31	Autres industries extractives	Manœuvres des mines, du bâtiment et des travaux publics, des industries manufacturières et des transports
32	Services personnels	Personnel des services directs aux particuliers et des services de protection et de sécurité
33	Industrie textile	Conducteurs de machines et ouvriers de l'assemblage
34	Administration publique	Agriculteurs et ouvriers qualifiés de l'agriculture et de la pêche destinées aux marchés
35	Industrie chimique	Artisans et ouvriers des métiers de la métallurgie, de la construction mécanique et assimilés
36	Services fournis principalement aux entreprises	Professions intermédiaires des sciences physiques et techniques
37	Extraction de houille, de lignite et de tourbe	Professions intermédiaires des sciences physiques et techniques
38	Industrie du caoutchouc et des plastiques	Artisans et ouvriers des métiers de la métallurgie, de la construction mécanique et assimilés
39	Non renseigné	Professions intermédiaires des sciences physiques et techniques
40	Education	Directeurs de société
41	Industries alimentaires	Conducteurs de machines et ouvriers de l'assemblage
42	Agriculture, chasse, services annexes	Directeurs de société
43	Industrie textile	Autres artisans et ouvriers des métiers de type artisanal
44	Activités récréatives, culturelles et sportives	Spécialistes des sciences physiques, mathématiques et techniques
45	Fabrication d'autres produits minéraux non	Conducteurs de machines et ouvriers de l'assemblage

	métalliques	
46	Commerce de détail, à l'exception des automobiles et des motocycles	Commerçants et vendeurs
47	Agriculture, chasse, services annexes	Professions intermédiaires des sciences physiques et techniques
48	Transports par eau	Professions intermédiaires des sciences physiques et techniques
49	Travail des métaux	Autres artisans et ouvriers des métiers de type artisanal
50	Administration publique et défense ; sécurité sociale obligatoire	Métiers qualifiés du bâtiment et assimilés, sauf électriciens
51	Transports terrestres	Employés non qualifiés des services et de la vente
52	Recherche et développement	Spécialistes des sciences de la vie et de la santé
53	Recherche et développement	Employés non qualifiés des services et de la vente
54	Transports terrestres	Dirigeants et gérants
55	Travail des métaux	Professions intermédiaires des sciences physiques et techniques
56	Administration publique	Artisans et ouvriers des métiers de l'extraction et du bâtiment
57	Commerce et réparation d'automobiles et de motocycles	Métiers qualifiés de la métallurgie, de la construction mécanique et assimilés
58	Administration publique	Manœuvres des mines, du bâtiment et des travaux publics, des industries manufacturières et des transports
59	Edition, imprimerie, reproduction	Conducteurs de machines et ouvriers de l'assemblage
60	Non renseigné	Agriculteurs et ouvriers qualifiés de l'agriculture et de la pêche destinées aux marchés
61	Santé et action sociale	Employés de réception, caissiers, guichetiers et assimilés
62	Fabrication de machines et appareils électriques	Manœuvres des mines, du bâtiment et des travaux publics, des industries manufacturières et des transports
63	Education	Spécialistes de l'enseignement
64	Fabrication de produits en caoutchouc et en plastique	Conducteurs de machines et d'installations fixes
65	Transports aériens	Directeurs de société
66	Activités extra-territoriales	Sans objet
67	Non renseigné	Employés de bureau
68	Fabrication de machines et appareils électriques	Conducteurs de machines et ouvriers de l'assemblage
69	Fabrication de machines de bureau et de matériel informatique	Employés de bureau
70	Industrie du cuir et de la chaussure	Autres artisans et ouvriers des métiers de type artisanal
71	Commerce et réparation automobile	Conducteurs de machines et ouvriers de l'assemblage
72	Non renseigné	Conducteurs de machines et ouvriers de l'assemblage
73	Non renseigné	Artisans et ouvriers des métiers de la métallurgie, de la construction mécanique et assimilés
74	Assurance	Agriculteurs et ouvriers qualifiés de l'agriculture et de la pêche destinées aux marchés
75	Construction	Conducteurs de véhicules et d'engins lourds de levage et de manœuvre
76	Santé et action sociale	Employés non qualifiés des services et de la vente
77	Transports aériens	Manœuvres des mines, du bâtiment et des travaux publics, des industries manufacturières et des transports
78	Industrie chimique	Spécialistes des sciences physiques, mathématiques et techniques
79	Assainissement, voirie et gestion des déchets	Spécialistes des sciences physiques, mathématiques et techniques
80	Travail des métaux	Conducteurs de machines et ouvriers de l'assemblage
81	Services personnels	Professions intermédiaires des sciences physiques et techniques
82	Edition, imprimerie, reproduction	Artisans et ouvriers de la mécanique de précision, des métiers d'art, de l'imprimerie et assimilés
83	Transports terrestres	Artisans et ouvriers des métiers de la métallurgie, de la construction mécanique et assimilés
84	Imprimerie et reproduction d'enregistrements	Métiers qualifiés de l'artisanat et de l'imprimerie
85	Fabrication de machines et appareils électriques	Professions intermédiaires des sciences physiques et techniques
86	Non renseigné	Artisans et ouvriers des métiers de l'extraction et du bâtiment
87	Agriculture, chasse, services annexes	Agriculteurs et ouvriers qualifiés de l'agriculture et de la pêche destinées aux marchés
88	Administration publique	Artisans et ouvriers des métiers de la métallurgie, de la construction mécanique et assimilés
89	Non renseigné	Non renseigné
90	Industrie chimique	Conducteurs de machines et ouvriers de l'assemblage
91	Travail des métaux	Dirigeants et gérants
92	Fabrication de machines et équipements	Professions intermédiaires des sciences physiques et techniques

93	Non renseigné	Manœuvres des mines, du bâtiment et des travaux publics, des industries manufacturières et des transports
94	Industrie du papier et du carton	Artisans et ouvriers de la mécanique de précision, des métiers d'art, de l'imprimerie et assimilés
95	Production et distribution d'électricité, de gaz et de chaleur	Artisans et ouvriers des métiers de l'extraction et du bâtiment
96	Administration publique et défense ; sécurité sociale obligatoire	Métiers qualifiés de la métallurgie, de la construction mécanique et assimilés
97	Assainissement, voirie et gestion des déchets	Manœuvres des mines, du bâtiment et des travaux publics, des industries manufacturières et des transports
98	Santé et action sociale	Spécialistes des sciences de la vie et de la santé
99	Production et distribution d'électricité, de gaz et de chaleur	Professions intermédiaires des sciences physiques et techniques
100	Commerce et réparation d'automobiles et de motocycles	Professions intermédiaires des sciences et techniques
101	Administration publique et défense ; sécurité sociale obligatoire	Professions intermédiaires des sciences et techniques
102	Réparation et installation de machines et d'équipements	Métiers qualifiés du bâtiment et assimilés, sauf électriciens
103	Non renseigné	Spécialistes des sciences physiques, mathématiques et techniques
104	Autres industries extractives	Conducteurs d'installations et de matériels fixes et assimilés
105	Travail du bois et fabrication d'articles en bois	Artisans et ouvriers des métiers de l'extraction et du bâtiment
106	Métallurgie	Manœuvres des mines, du bâtiment et des travaux publics, des industries manufacturières et des transports
107	Transports aériens	Artisans et ouvriers de la mécanique de précision, des métiers d'art, de l'imprimerie et assimilés
108	Industrie chimique	Employés de bureau
109	Fabrication d'équipements de radio, télévision et communication	Artisans et ouvriers des métiers de la métallurgie, de la construction mécanique et assimilés
110	Industrie automobile	Conducteurs de machines et ouvriers de l'assemblage
111	Fabrication d'autres matériels de transport	Professions intermédiaires des sciences physiques et techniques
112	Santé et action sociale	Professions intermédiaires des sciences physiques et techniques
113	Industrie du caoutchouc et des plastiques	Conducteurs de machines et ouvriers de l'assemblage
114	Services fournis principalement aux entreprises	Artisans et ouvriers des métiers de l'extraction et du bâtiment
115	Administration publique	Personnel des services directs aux particuliers et des services de protection et de sécurité
116	Agriculture, chasse, services annexes	Manœuvres de l'agriculture, de la pêche et assimilés
117	Non renseigné	Artisans et ouvriers de la mécanique de précision, des métiers d'art, de l'imprimerie et assimilés
118	Fabrication de machines et équipements	Artisans et ouvriers des métiers de la métallurgie, de la construction mécanique et assimilés
119	Services fournis principalement aux entreprises	Employés non qualifiés des services et de la vente
120	Transports terrestres	Employés de bureau
121	Fabrication de machines et appareils électriques	Artisans et ouvriers des métiers de la métallurgie, de la construction mécanique et assimilés
122	Commerce de détail et réparation d'articles domestiques	Directeurs de société
123	Location sans opérateur	Conducteurs de machines et ouvriers de l'assemblage
124	Extraction de minerais d'uranium	Professions intermédiaires des sciences physiques et techniques
125	Commerce et réparation automobile	Manœuvres des mines, du bâtiment et des travaux publics, des industries manufacturières et des transports
126	Fabrication de machines et équipements	Conducteurs de machines et ouvriers de l'assemblage
127	Industrie automobile	Artisans et ouvriers de la mécanique de précision, des métiers d'art, de l'imprimerie et assimilés
128	Industrie automobile	Artisans et ouvriers des métiers de la métallurgie, de la construction mécanique et assimilés
129	Fabrication d'autres matériels de transport	Artisans et ouvriers des métiers de l'extraction et du bâtiment
130	Industrie du papier et du carton	Conducteurs de machines et ouvriers de l'assemblage
131	Industrie du cuir et de la chaussure	Conducteurs de machines et d'installations fixes
132	Transports terrestres et transport par conduites	Métiers qualifiés de la métallurgie, de la construction mécanique et assimilés
133	Fabrication de produits en caoutchouc et en plastique	Eboueurs et autres travailleurs non qualifiés
134	Administration publique et défense ; sécurité sociale obligatoire	Eboueurs et autres travailleurs non qualifiés
135	Culture et production animale, chasse et services annexes	Agriculteurs, pêcheurs, chasseurs et cueilleurs de subsistance

136	Industrie du cuir et de la chaussure	Métiers qualifiés de l'artisanat et de l'imprimerie
137	Activités pour la santé humaine	Professions intermédiaires de la santé
138	Industrie chimique	Conducteurs de machines et d'installations fixes
139	Activités des ménages en tant qu'employeurs de personnel domestique	Agriculteurs et ouvriers qualifiés de l'agriculture commerciale
140	Fabrication de produits métalliques, à l'exception des machines et des équipements	Aides de ménage
141	Industrie chimique	Métiers qualifiés de l'artisanat et de l'imprimerie
142	Travaux de construction spécialisés	Métiers qualifiés du bâtiment et assimilés, sauf électriciens
143	Industrie chimique	Spécialistes des sciences techniques
144	Activités d'architecture et d'ingénierie ; activités de contrôle et analyses techniques	Spécialistes des sciences techniques
145	Industrie pharmaceutique	Professions intermédiaires de la santé
146	Fabrication de produits métalliques, à l'exception des machines et des équipements	Métiers qualifiés du bâtiment et assimilés, sauf électriciens
147	Travail du bois et fabrication d'articles en bois et en liège, à l'exception des meubles ; fabrication d'articles en vannerie et sparterie	Métiers de l'alimentation, du travail sur bois, de l'habillement et autres métiers qualifiés de l'industrie et de l'artisanat
148	Fabrication de produits métalliques, à l'exception des machines et des équipements	Professions intermédiaires des sciences et techniques
149	Imprimerie et reproduction d'enregistrements	Manœuvres des mines, du bâtiment et des travaux publics, des industries manufacturières et des transports

Annexe 4 : Liste de tous les codes et libellés des motifs recensées pour les trois exemples d'application du RNV3P

Code	Label	Asthme	LNH	Vessie
21120601	1,3-butadiene	-	Oui	-
21140202	dichlorométhane	-	Oui	-
21170204	trichloroéthylène	-	Oui	Oui
21170205	tétrachloroéthylène	Oui	Oui	Oui
21230000	hydrocarbure alicyclique insaturé - terpène	Oui	-	-
21300000	hydrocarbure aromatique et dérivé	-	-	Oui
21310100	benzène	-	Oui	Oui
21310200	toluène	Oui	-	Oui
21310600	styrène	Oui	-	Oui
21320000	hydrocarbure aromatique polycyclique	-	Oui	Oui
21460000	thiazole, isothiazole et dérivé	Oui	-	-
24170000	hydroquinone	Oui	-	-
27100000	aldéhyde aliphatique	Oui	-	-
27110000	aldéhyde aliphatique sature	Oui	-	-
27110100	formaldéhyde	Oui	Oui	-
28110100	acétone	Oui	-	Oui
28420000	benzisothiazoline-3-one	Oui	-	-
31000000	microbiologique	Oui	-	-
31200000	champignon, moisissure	Oui	-	-
31211100	levure	Oui	-	-
31240300	alternaria	Oui	-	-
31240700	aspergillus	Oui	-	-
32110110	acarien ou mite ou gale	Oui	-	-
32110300	crustacé	Oui	-	-
32110500	insecte	Oui	-	-
32220000	mammifère et cétacé	Oui	-	-
32220400	bœuf ou veau ou vache	Oui	-	-
32220800	chat	Oui	-	-
32230000	oiseau	Oui	-	-
33300000	arbre et plante	Oui	-	-
33310700	caféier	Oui	-	-
33320000	fleur et plante verte	Oui	-	-
33320500	armoise	Oui	-	-
33321600	tabac	Oui	-	Oui
33330000	graminée céréale	Oui	-	-
33330200	blé	Oui	-	-
33330500	mais	Oui	-	-
33350000	plante alimentaire, légume ou fruit	Oui	-	-
33350100	ail	Oui	-	-
41320220	houille, charbon	-	-	Oui
41320610	silice	Oui	-	Oui
41320612	silice cristalline	Oui	-	Oui
42110000	amiante (fibre)	Oui	Oui	Oui
51210000	ambiance thermique chaude	Oui	-	-
51220000	ambiance thermique froide	Oui	-	-
51320000	courant d'air	Oui	-	-
51410000	ambiance humide	Oui	-	-
56300000	rayonnement ionisant	-	Oui	-
61200000	port de charge	Oui	-	-
61300000	effort physique	Oui	-	-
13CL1000	chlore	Oui	-	-
13CL4100	acide chlorhydrique	Oui	-	Oui
13CL6100	hypochlorite	Oui	-	-
13F04100	acide fluorhydrique	Oui	-	-

16NA5100	hydroxyde de sodium	Oui	-	-
17CA5100	hydroxyde de calcium	Oui	-	-
18CO1000	cobalt	Oui	-	-
18CR1000	chrome	Oui	-	-
18CR6100	chromate	Oui	-	-
18FE3000	oxyde de fer	Oui	-	-
18NI1000	nickel	Oui	-	-
18W01000	tungstène	Oui	-	-
19PB1000	plomb	-	-	Oui
1AH03100	peroxyde d'hydrogène	Oui	-	-
1AH06104	hydrure de sulfure	Oui	-	-
1AN04100	acide nitrique	Oui	-	-
1AN05100	ammoniac (NH3)	Oui	-	Oui
1AN06202	trichlorure d’azote	Oui	-	-
1AN06600	compose d'ammonium quaternaire	Oui	-	-
1AN066ZZ	autre compose d'ammonium quaternaire	Oui	-	-
1AN07100	ammoniaque (NH4OH)	Oui	-	-
1AP04100	acide phosphorique	Oui	-	-
1AS04100	acide sulfurique	Oui	-	-
1AS06100	sulfite	Oui	-	-
1AS06400	persulfate	Oui	-	-
2132ZZ00	autre hydrocarbure aromatique polycyclique	-	-	Oui
214B0000	dioxine et dérivé	-	Oui	-
27110C00	glutaraldéhyde	Oui	-	-
2A100000	acide carboxylique aliphatique et dérivé	Oui	-	-
2A110200	acide acétique	Oui	-	-
2A510000	acide peracétique	Oui	-	-
2AZ00000	autre acide et peracide carboxylique	Oui	-	-
2C600000	anhydride phtalique	Oui	-	-
2CZ00000	autre anhydride d'acide carboxylique	Oui	-	-
2D160000	acrylate	Oui	-	-
2D170000	méthacrylate	Oui	-	-
2D170100	méthacrylate de méthyle	Oui	-	-
2D330000	phtalate	Oui	-	-
2D620000	sulfate et sulfite	Oui	-	-
2D630200	dérivé organophosphoré	-	Oui	-
2H220000	sel de sodium de la N-chloro-p-toluènesulfonamide	Oui	-	-
2J110000	amine aliphatique	Oui	-	-
2J11ZZ00	autre amine aliphatique	Oui	-	-
2J150100	éthanolamine	Oui	-	-
2J160000	dérivé nitrosé d'amine aliphatique	-	-	Oui
2J200000	amine alicyclique et dérivé	Oui	-	-
2J300000	amine aromatique et dérivé	Oui	-	Oui
2J310000	amine aromatique	-	-	Oui
2J310A00	4,4'-diaminobiphenyle	-	-	Oui
2J310G00	phénylènediamine	Oui	-	-
2J330600	4,4'-méthylene-bis-orthochloraniline	-	-	Oui
2M300000	cyanate et isocyanate	Oui	-	-
2M340000	toluène-2,4-diisocyanate	Oui	-	-
2M360000	diphénylméthanediisocyanate	Oui	-	-
2M380000	hexaméthylenediisocyanate	Oui	-	-
2M3Z0000	autre cyanate et isocyanate	Oui	-	-
3211011Z	autres acariens, mites et gales	Oui	-	-
32220D00	chien	Oui	-	-
32220T00	hamster	Oui	-	-
32221A00	rat	Oui	-	-
32221F00	souris	Oui	-	-
33310R00	hêtre	Oui	-	-
33311G00	sapin	Oui	-	-

3331XX01	bois exotique sans précision	Oui	-	-
33330A00	sarrasin	Oui	-	-
33330B00	seigle	Oui	-	-
33330C00	soja	Oui	-	-
33350L00	oignon	Oui	-	-
3VY20000	Rongeur	Oui	-	-
A1110100	poussière d'alliage léger (aluminium...)	Oui	-	-
A1220100	poussière de métaux durs carbures métalliques	Oui	-	Oui
A122ZZ00	autre poussière métallique	Oui	-	Oui
A1340600	poussière de ciment, aluminosilicate de calcium	Oui	-	-
A1360000	plancher maison	Oui	-	-
A13C0110	laine de roche	Oui	-	-
A13C0120	laine verre	Oui	-	-
A13Z0100	poussière de chantier	Oui	-	-
A1510000	fumée et gaz de soudage	Oui	Oui	Oui
A1510100	fumée et gaz de soudage d'aluminium	Oui	-	-
A1510200	fumée et gaz de soudage d'inox	Oui	-	-
A15Z0000	autre produit de soudage	-	Oui	-
A1610000	peinture, vernis, laque, mastic	Oui	Oui	Oui
A1610210	peinture, vernis, laque, mastic polyuréthanne phase solvant	Oui	-	-
A1610B10	peinture, vernis, laque, mastic sans précision en poudre	Oui	-	-
A1610B20	peinture, vernis, laque, mastic sans précision en phase solvant	Oui	-	Oui
A1610B30	peinture, vernis, laque, mastic sans précision en phase aqueuse	Oui	-	Oui
A1620100	teinture pour bois	-	-	Oui
A1630000	adhésif, colle	Oui	-	Oui
A1630200	adhésif, colle cyanoacrylate	Oui	-	-
A1630700	adhésif, colle polyuréthane	Oui	-	-
A1630800	adhésif, colle urée formol	Oui	-	-
A1640000	résine	Oui	-	-
A1640300	résine époxydique	Oui	-	Oui
A1660000	solvant, diluant	Oui	Oui	Oui
A1660400	solvant de nettoyage	Oui	-	-
A1670000	matière plastique, caoutchouc	Oui	Oui	-
A1670113	thermoplastique polyacrylique ou polymétacrylique	Oui	-	-
A1670116	thermoplastique polyester	Oui	-	-
A1670122	plastique thermodurcissable urée-formaldéhyde	Oui	-	-
A167012B	plastique thermodurcissable polyuréthane	Oui	-	-
A1670140	produit de décomposition thermique de matière plastique	Oui	-	-
A1670141	produit de décomposition thermique du PVC	Oui	-	-
A1670210	caoutchouc, latex naturel	Oui	-	-
A1670220	produit de vulcanisation du caoutchouc	Oui	-	-
A1670230	produit dégagé lors de l'utilisation normale du caoutchouc	Oui	-	-
A1690000	matière colorante	Oui	-	Oui
A16B0200	encre	Oui	-	Oui
A16C0000	produit pour l'alimentation	Oui	-	-
A16D0000	huile, graisse et fluide d’usinage industriel	Oui	-	Oui
A16D0100	huile de coupe, fluide d'usinage	Oui	-	Oui
A16D0110	huile de coupe pleine, fluide d'usinage pleine	-	-	Oui
A16D01ZZ	autre huile de coupage	Oui	-	-
A16D0200	huile et graisse lubrifiante	Oui	-	Oui
A16D0210	huile et graisse lubrifiante minérale	-	-	Oui
A16H0120	essence	-	-	Oui
A16H0500	gaz d’échappement	Oui	-	Oui
A16H0510	diesel gaz d'échappement	Oui	-	Oui
A16J0000	produit noir	-	-	Oui
A16J0100	bitume	-	-	Oui
A16J0400	goudron	-	-	Oui
A16J0500	suie et autre produit dégagé lors de l'utilisation de produit noir	-	Oui	Oui
A16J0520	suie de combustion du charbon	-	-	Oui

A16K0000	produit d'entretien	Oui	-	-
A16K0200	décapant	Oui	-	-
A16K0300	dégraissant	Oui	-	-
A16K0410	détergent désinfectant pour locaux (sols, murs, surfaces) domestique et professionnel	Oui	-	-
A16K0420	détergent désinfectant pour matériel médical, instruments	Oui	-	-
A16K0470	nettoyant désinfectant alimentaire	Oui	-	-
A16K0900	lessive détergent	Oui	-	-
A16L0100	produit capillaire	Oui	-	Oui
A16L0110	décolorant, éclaircissant capillaire	Oui	-	-
A16L0120	oxydant et décolorant (cosmétique)	Oui	-	-
A16L0150	teinture (cosmétique)	Oui	-	-
A16L0310	parfum et substance aromatique	Oui	-	-
A16L0350	antiseptique, bactéricide pour la peau (sauf SHA solutions hydro-alcooliques)	Oui	-	-
A16L0400	produit cosmétique divers	Oui	-	-
A16M0100	compose chimique carcinogène	Oui	-	-
A16N0000	produit photographique	-	-	Oui
A16P0000	produit pesticide (biocide, insecticide, rodenticide, fongicide, engrais)	Oui	Oui	Oui
A16P0300	insecticide	Oui	Oui	-
A16P0350	insecticide organophosphoré	Oui	-	-
A16P0500	herbicide	-	Oui	-
A16P0600	fongicide	Oui	-	-
A16S0000	produit de décomposition thermique d'autre produit divers	Oui	-	-
A16W0110	poussière de papier carton	Oui	-	-
A16W0810	fumée d'explosif	Oui	-	Oui
A16W0920	fumée de tabac	-	-	Oui
A1800000	enzyme	Oui	-	-
A1810000	enzyme protéolytique	Oui	-	-
A1820100	alpha amylase	Oui	-	-
A1880000	hydrolase	Oui	-	-
A1900000	produit pharmaceutique	Oui	-	-
A2110000	poussière organique	Oui	-	-
A211Y000	déchet de poussière organique	Oui	-	-
A2210000	poussière inorganique	Oui	-	Oui
A2220100	carbone (fibre)	Oui	-	-
A31B0000	œuf	Oui	-	-
A31D0000	plume, duvet	Oui	-	-
A31F0000	fourrure, feutre	Oui	-	-
A31N0000	protéine animale	Oui	-	-
A31P0200	fromage	Oui	-	-
A31R0000	poussière d'origine animale	Oui	-	-
A3310000	aromate épice	Oui	-	-
A3370000	pollens	Oui	-	-
A33D0000	huile essentielle	Oui	-	-
A33K0500	colophane	Oui	-	-
A33K0K00	latex d'hévéa	Oui	-	-
A33K0M00	protéine de latex	Oui	-	-
A33M0100	fibre végétale textile	Oui	-	-
A33M0130	coton (fibre)	Oui	-	-
A33MY000	déchet de fibre végétale	Oui	-	-
A33N0000	poussière d'origine végétale	Oui	-	-
A33N0200	poussière de farine	Oui	-	-
A33N0300	poussière de bois	Oui	-	Oui
B11M0000	traitement phytosanitaire	-	Oui	-
D1120000	équipement de protection	Oui	-	-

Annexe 5 : Liste des 145 motifs identifiés pour les aides-soignants par la Surveillance Observationnelle en 2013. Les libellés des motifs ont été fournis par les acteurs du réseau IDEWE.

ID	Ordre I	Ordre II	Nœuds	Poids	Statut*
1	**malaise et fatigue**	**-**	**119**	**239**	**A**
2	**rhinopharyngites aigues, rhumes de cerveau**	**-**	**92**	**164**	**A**
3	**grippe, avec manifestation, autre**	**-**	**63**	**120**	**A**
4	**lumbago**	**-**	**67**	**116**	**A**
5	**symptôme relatif à la région dorsale, autre**	**-**	**59**	**86**	**A**
6	**influenza**	**-**	**39**	**74**	**A**
7	**angines, amygdalites aigues**	**-**	**36**	**69**	**A**
8	**rachialgie, sans autre précision**	**-**	**37**	**62**	**A**
9	**troubles dépressifs, non classés ailleurs**	**-**	**43**	**57**	**A**
10	**gastro-entérite et colite non infectieuse, autre et sans autre précision**	**-**	**33**	**52**	**A**
11	**colite, entérite et gastro-entérite infectieuses**	**-**	**25**	**50**	**A**
12	**affection de la région dorsale, autre, sans autre précision**	**-**	**31**	**46**	**A**
13	**infection aigue de voie respiratoire supérieure, sans autre précision**	**-**	**32**	**45**	**A**
14	**pharyngites aigues**	**-**	**35**	**43**	**B**
15	**hernie discale, sans autre précision, sans myélopathie**	**-**	**28**	**41**	**B**
16	**épicondylite, sans autre précision**	**-**	**32**	**41**	**A**
17	**douleur articulaire, d'épaule**	**-**	**22**	**39**	**A**
18	**asthme non spécifié, sans autre précision**	**-**	**29**	**37**	**A**
19	**bronchite aigue**	**-**	**24**	**33**	**B**
20	**bronchites, sans autre précision**	**-**	**27**	**32**	**B**
21	**symptôme concernant les systèmes nerveux et musculo-squelettique, autre**	**-**	**21**	**32**	**A**
22	**cervicalgie**	**-**	**24**	**31**	**B**
23	**réaction aigue au stress, sans autre précision**	**-**	**24**	**28**	**A**
24	**dyspepsie et trouble fonctionnel de l'estomac, autre**	**-**	**21**	**27**	**A**
25	**trouble bipolaire, sans autre précision**	**-**	**19**	**26**	**B**
26	**diarrhée**	**-**	**20**	**26**	**B**
27	**sinusite chronique, sans autre précision**	**-**	**18**	**24**	**A**
28	**capsulite adhésive de l'épaule**	**-**	**20**	**24**	**A**
29	**otite moyenne, sans autre précision**	**-**	**14**	**23**	**B**
30	**bronchite chronique obstructive, avec exacerbation (aigue)**	**-**	**19**	**23**	**A**
31	**réaction aigue au stress, autre**	**-**	**16**	**22**	**B**
32	**migraine**	**-**	**18**	**22**	**B**
33	**migraine, sans autre précision**	**-**	**18**	**22**	**B**
34	**colite, entérite et gastro-entérite présumées infectieuses**	**-**	**16**	**21**	**A**
35	**varice de membre inférieur, asymptomatique**	**-**	**14**	**21**	**A**
36	**gastrite et gastroduodénite, sans autre précision, sans mention d'hémorragie**	**-**	**13**	**21**	**A**
37	**libération du canal carpien**	**-**	**15**	**20**	**B**
38	**malaise et fatigue, sans autre précision**	**-**	**16**	**19**	**D**
39	**arthroscopie, genou**	**-**	**12**	**19**	**C**
40	**infection virale sans autre précision**	**-**	**12**	**18**	**C**
41	**hernie thoracique noyau pulpeux ou lombaire sans myélopathie**	**-**	**13**	**18**	**D**
42	**traitement chirurgical d'une hernie discale**	**-**	**11**	**16**	**C**
43	Trouble de l'adaptation avec humeur dépressive	-	14	14	D
44	myalgie et myosite, sans autre précision	-	10	14	D
45	entorse de région cervicale	-	5	14	D
46	nausée isolée	-	13	13	D
47	extraction dentaire chirurgicale	-	12	12	D
48	syndrome du canal carpien	-	11	12	D
49	sinusite aigue	-	12	12	D
50	troubles de disques intervertébraux	-	7	12	D

51	**opération du système musculo-squelettique, autre**	**-**	**9**	**12**	**C**
52	rhinopharyngites aigues, rhumes de cerveau	malaise et fatigue	5	12	D
53	hernie discale cervicale, sans myélopathie	-	10	11	D
54	symptôme concernant le système digestif, autre	-	8	11	D
55	grippe, avec manifestation, autre	malaise et fatigue	6	11	D
56	ulcère de l'estomac	-	10	10	D
57	maladie des insertions ligamentaires et tendineuses, sans autre précision	-	7	10	D
58	hernie discale lombaire, sans myélopathie	-	7	9	D
59	**lumbago**	**symptôme relatif à la région dorsale, autre**	**9**	**9**	**D**
60	cholécystectomie	-	8	8	D
61	pyrexie d'origine indéterminée	-	7	8	D
62	influenza	malaise et fatigue	3	8	D
63	diarrhée présumée infectieuse	-	7	7	D
64	angine de poitrine, autre et sans autre précision	-	7	7	D
65	hypotension, sans autre précision	-	7	7	D
66	pneumonies à microorganisme, sans autre précision	-	7	7	D
67	sciatique	-	6	7	D
68	épine calcanéenne	-	6	7	D
69	rhinopharyngites aigues, rhumes de cerveau	grippe, avec manifestation, autre	4	7	D
70	troubles dépressifs, non classés ailleurs	rachialgie, sans autre précision	5	7	D
71	rhinopharyngites aigues, rhumes de cerveau	rachialgie, sans autre précision	7	7	D
72	histoplasmose, sans autre précision, avec pneumonie	-	6	6	D
73	eczéma, dermatite de contact, sans autre précision	-	4	6	D
74	affection articulaire, autre, d'épaule	-	6	6	D
75	réparation de la coiffe des rotateurs	-	6	6	D
76	V22, label inconnu	-	6	6	D
77	grippe, avec manifestation, autre	symptôme relatif à la région dorsale, autre	4	6	D
78	angines, amygdalites aigues	malaise et fatigue	4	6	D
79	Tumeur maligne du sein chez la femme, sans précision	-	5	5	D
80	excision de lésion d'amygdale et de végétations adénoïdes	-	5	5	D
81	trouble dysthymique	-	5	5	D
82	céphalée de stress	-	5	5	D
83	ligature et décapage des varices veineuses site non précisé	-	5	5	D
84	sinusite chronique, maxillaire	-	5	5	D
85	maladie de dent et de parodonte, sans autre précision	-	5	5	D
86	néphrite et néphropathie avec lésion rénale pathologique, sans autre précision	-	5	5	D
87	cystite, sans autre précision	-	5	5	D
88	vomissements abondants dans la grossesse	-	5	5	D
89	grossesse et troubles mentaux, épisode de soins non précisé	-	5	5	D
90	excision ou destruction de l'utérus et de structure soutenante, autre	-	5	5	D
91	arthropathie, autre, localisation non précisée	-	5	5	D
92	bursite, autre	-	5	5	D
93	interventions obstétricales, autre	-	5	5	D
94	995.3, label inconnu	-	4	5	D
95	angines, amygdalites aigues	grippe, avec manifestation, autre	3	5	D
96	rhinopharyngites aigues, rhumes de cerveau	symptôme relatif à la région dorsale, autre	4	5	D
97	lumbago	épicondylite, sans autre précision	5	5	D
98	infection aigue de voie respiratoire supérieure, sans autre précision	malaise et fatigue	4	5	D
99	douleur articulaire, d'épaule	malaise et fatigue	5	5	D
100	rachialgie, sans autre précision	malaise et fatigue	3	5	D
101	dorsopathie	-	4	4	D
102	syndrome cervico-brachial diffus	-	4	4	D
103	étourdissement et éblouissement	-	4	4	D
104	nausée avec vomissement	-	4	4	D
105	arthroscopie, autre site spécifié	-	4	4	D

106	chirurgies sur les articulations, autre	-	4	4	D
107	traumatisme de genou, jambe, cheville et pied	-	3	4	D
108	libération du canal carpien	syndrome du canal carpien	3	4	D
109	extraction dentaire chirurgicale	rhinopharyngites aigues, rhumes de cerveau	4	4	D
110	rhinopharyngites aigues, rhumes de cerveau	pharyngites aigues	4	4	D
111	rhinopharyngites aigues, rhumes de cerveau	influenza	3	4	D
112	infection aigue de voie respiratoire supérieure, sans autre précision	grippe, avec manifestation, autre	4	4	D
113	influenza	grippe, avec manifestation, autre	4	4	D
114	rhinopharyngites aigues, rhumes de cerveau	bronchites, sans autre précision	4	4	D
115	bronchite aigue	bronchites, sans autre précision	3	4	D
116	rhinopharyngites aigues, rhumes de cerveau	asthme non spécifié, sans autre précision	4	4	D
117	colite, entérite et gastro-entérite présumées infectieuses	gastro-entérite et colite non infectieuse, autre et sans autre précision	3	4	D
118	rhinopharyngites aigues, rhumes de cerveau	lumbago	4	4	D
119	asthme non spécifié, sans autre précision	lumbago	2	4	D
120	asthme non spécifié, sans autre précision	symptôme relatif à la région dorsale, autre	4	4	D
121	troubles dépressifs, non classés ailleurs	épicondylite, sans autre précision	4	4	D
122	pharyngites aigues	malaise et fatigue	4	4	D
123	cervicalgie	malaise et fatigue	3	4	D
124	lumbago	malaise et fatigue	3	4	D
125	symptôme relatif à la région dorsale, autre	malaise et fatigue	4	4	D
126	épicondylite, sans autre précision	malaise et fatigue	2	4	D
127	sinusite chronique, sans autre précision	influenza	2	3	D
128	infection aigue de voie respiratoire supérieure, sans autre précision	lumbago	2	3	D
129	rachialgie, sans autre précision	symptôme relatif à la région dorsale, autre	1	3	D
130	lumbago	épine calcanéenne	2	3	D
131	hernie discale, sans autre précision, sans myélopathie	malaise et fatigue	2	3	D
132	lumbago	symptôme concernant les systèmes nerveux et musculo-squelettique, autre	1	3	D
133	entorse de cheville, ligament latéral externe	-	1	2	D
134	contusion de la paroi thoracique	-	1	2	D
135	rhinopharyngites aigues, rhumes de cerveau	angines, amygdalites aigues	1	2	D
136	infection aigue de voie respiratoire supérieure, sans autre précision	bronchites, sans autre précision	1	2	D
137	hernie thoracique noyau pulpeux ou lombaire sans myélopathie	hernie discale, sans autre précision, sans myélopathie	1	2	D
138	otite moyenne, sans autre précision	lumbago	1	2	D
139	infection virale sans autre précision	malaise et fatigue	1	2	D
140	asthme non spécifié, sans autre précision	malaise et fatigue	1	2	D
141	malaise et fatigue	diarrhée	1	2	D
142	lumbago	entorse de cheville, ligament latéral externe	1	2	D
143	grippe, avec manifestation, autre	contusion de la paroi thoracique	1	2	D
144	opération du système musculo-squelettique, autre	traumatisme de genou, jambe, cheville et pied	1	2	D
145	malaise et fatigue	995.3, label inconnu	1	2	D

*En gras sont identifiés les motifs dont le statut est considéré comme actif (A : persistant, B : en déclin, C : émergent) ainsi que 3 motifs latents (D) actifs en 2012.

Annexe 6 : Liste des 66 motifs identifiés pour les aides-soignants par la Surveillance Observationnelle en 2013. Les libellés des motifs ont été fournis par les acteurs du réseau IDEWE.

ID	Ordre I	Ordre II	Ordre III	Nœuds	Poids	Statut*
1	**malaise et fatigue**	**-**	**-**	**57**	**139**	**A**
2	**rhinopharyngites aigues, rhumes de cerveau**	**-**	**-**	**46**	**91**	**A**
3	**lumbago**	**-**	**-**	**40**	**82**	**A**
4	**symptôme relatif à la région dorsale, autre**	**-**	**-**	**38**	**68**	**A**
5	**influenza**	**-**	**-**	**34**	**61**	**A**
6	**grippe, avec manifestation, autre**	**-**	**-**	**24**	**40**	**A**
7	**rachialgie, sans autre précision**	**-**	**-**	**21**	**35**	**A**
8	**affection de la région dorsale, autre, sans autre précision**	**-**	**-**	**23**	**32**	**A**
9	**épicondylite, sans autre précision**	**-**	**-**	**18**	**29**	**A**
10	**douleur articulaire, d'épaule**	**-**	**-**	**16**	**26**	**A**
11	**pharyngites aigues**	**-**	**-**	**11**	**21**	**A**
12	**arthroscopie, genou**	**-**	**-**	**12**	**20**	**B**
13	**troubles dépressifs, non classés ailleurs**	**-**	**-**	**10**	**19**	**B**
14	**dyspepsie et trouble fonctionnel de l'estomac, autre**	**-**	**-**	**12**	**17**	**A**
15	**hernie discale, sans autre précision, sans myélopathie**	**-**	**-**	**11**	**16**	**A**
16	**capsulite adhésive de l'épaule**	**-**	**-**	**8**	**16**	**A**
17	**réaction aigue au stress, sans autre précision**	**-**	**-**	**13**	**15**	**A**
18	**hernie thoracique noyau pulpeux ou lombaire sans myélopathie**	**-**	**-**	**8**	**11**	**A**
19	**colite, entérite et gastro-entérite infectieuses**	**-**	**-**	**9**	**9**	**B**
20	**maladie des insertions ligamentaires et tendineuses, sans autre précision**	**-**	**-**	**5**	**9**	**A**
21	**entorse et rupture de muscle ou tendon, sans autre précision**	**-**	**-**	**9**	**9**	**B**
22	**libération du canal carpien**	**-**	**-**	**8**	**8**	**D**
23	**syndrome du canal carpien**	**-**	**-**	**5**	**8**	**C**
24	bronchites, sans autre précision	-	-	8	8	D
25	gastro-entérite et colite non infectieuse, autre et sans autre précision	-	-	8	8	D
26	lumbago	malaise et fatigue	-	4	8	D
27	**troubles de disques intervertébraux**	**-**	**-**	**5**	**7**	**D**
28	hernie discale lombaire, sans myélopathie	-	-	7	7	D
29	**lumbago**	**affection de la région dorsale, autre, sans autre précision**	**-**	**6**	**7**	**D**
30	influenza	malaise et fatigue	-	4	7	D
31	affection de la région dorsale, autre, sans autre précision	malaise et fatigue	-	6	7	D
32	infection aigue de voie respiratoire supérieure, sans autre précision	-	-	6	6	D
33	traitement chirurgical d'une hernie discale	-	-	6	6	D
34	arthropathie, autre, localisation non précisée	-	-	5	6	D
35	**douleur articulaire, de région pelvienne et cuisse**	**-**	**-**	**6**	**6**	**D**
36	lumbago	symptôme relatif à la région dorsale, autre	-	6	6	D

37	sinusite aigue	-	-	5	5	D
38	angines, amygdalites aigues	-	-	5	5	D
39	asthme non spécifié, sans autre précision	-	-	5	5	D
40	**hernie discale cervicale, sans myélopathie**	**-**	**-**	**5**	**5**	**D**
41	cervicalgie	-	-	3	5	D
42	maladie des insertions ligamentaires et tendineuses du poignet et du carpe	-	-	3	5	D
43	myalgie et myosite, sans autre précision	-	-	5	5	D
44	**douleur de membre, sans autre précision**	**-**	**-**	**3**	**5**	**C**
45	symptôme concernant les systèmes nerveux et musculo-squelettique, autre	-	-	5	5	D
46	rhinopharyngites aigues, rhumes de cerveau	lumbago	-	4	5	D
47	grippe, avec manifestation, autre	lumbago	-	3	5	D
48	symptôme relatif à la région dorsale, autre	malaise et fatigue	-	4	5	D
49	extraction dentaire chirurgicale	-	-	4	4	D
50	bronchite aigue	-	-	4	4	D
51	sinusite chronique, sans autre précision	-	-	4	4	D
52	pneumonies à microorganisme, sans autre précision	-	-	4	4	D
53	grippe, avec manifestation respiratoire, autre	-	-	4	4	D
54	traitement chirurgical d'une hernie inguinale bilatérale	-	-	4	4	D
55	lithiase rénale	-	-	4	4	D
56	traumatisme de genou, jambe, cheville et pied	-	-	4	4	D
57	grippe, avec manifestation, autre	malaise et fatigue	-	3	4	D
58	rhinopharyngites aigues, rhumes de cerveau	influenza	-	2	3	D
59	douleur articulaire, d'épaule	épicondylite, sans autre précision	-	2	3	D
60	symptôme relatif à la région dorsale, autre	épicondylite, sans autre précision	-	2	3	D
61	arthrodèse lombaire ou lombosacrée postérieure	-	-	1	2	D
62	influenza	lumbago	-	1	2	D
63	douleur articulaire, d'épaule	symptôme relatif à la région dorsale, autre	-	1	2	D
64	lumbago	maladie des insertions ligamentaires et tendineuses, sans autre précision	-	1	2	D
65	symptôme relatif à la région dorsale, autre	arthrodèse lombaire ou lombosacrée postérieure	-	1	2	D
66	douleur articulaire, d'épaule	symptôme relatif à la région dorsale, autre	épicondylite, sans autre précision	1	2	D

*En gras sont identifiés les motifs dont le statut est considéré comme actif (A : persistant, B : en déclin, C : émergent) ainsi que 5 motifs latents (D) actifs en 2012.

RESUME

L'impact de la population sur l'environnement et celui de l'environnement sur la santé humaine est indéniable et représente depuis les années 2000 de réels enjeux de Santé Publique. Il a été constaté en outre des augmentations de prévalence de cancer, de maladies respiratoires ou encore des troubles de la reproduction, dont les origines multifactorielles sont de plus en plus suspectées. Dans ce contexte, la surveillance est devenue un outil indispensable à la prise de décision en santé publique, et les réseaux de surveillances des évènements sanitaires se sont multipliés, donnant naissance à de nombreuses bases de données, parfois massives, dont les utilisations sont encore limitées.
L'objectif de ces travaux de thèse était de développer un nouveau concept de surveillance, la Surveillance Observationnelle (SO), permettant une utilisation optimale des bases de données observationnelles, généralisable à différentes bases et problématiques, et pouvant tenir compte des informations multiples qu'elles contiennent.
La SO est basée sur la méthode de l'exposome, ayant pour but de restructurer les données sous forme de réseau, permettant ainsi l'étude des associations entre les informations mais également l'étude de leur structure. Il a alors été développé plusieurs indicateurs permettant d'étudier à la fois les différentes associations d'informations recrutées au sein d'une base de données pour un évènement de santé, mais également d'étudier l'évolution de leur structure dans le temps par un suivi dynamique, ces indicateurs permettant de générer la signature unique d'un évènement de santé : le spectre. Un outil, appelé « Observational Surveillance Analysis » (OSA) permettant l'utilisation en routine de la méthodologie a également été développé dans la plateforme R, permettant une automatisation et une standardisation des résultats.
Des exemples d'applications ont permis d'illustrer le déroulement de l'analyse de la SO ainsi que sa portabilité et adaptabilité au contexte et à la problématique. Trois applications sont basées sur les données du Réseau National de Vigilance et de Prévention des Pathologies Professionnelles (RNV3P) : le cancer de la vessie, l'asthme et enfin le lymphome non hodgkinien. Trois autres applications sont basées sur la base de données du groupe de médecins du travail Belge IDEWE : les douleurs à la gorge, les aides-soignants et les agriculteurs.
Grâce aux différentes applications, il a été démontré la portabilité de la méthodologie de la SO à des bases de données différentes, mais également à des configurations d'analyse différentes, pathologie/expositions ou activité/expositions. De plus, l'outil « OSA » qui a été développé permet une utilisation facilitée pour l'analyse en routine des données et pourrait à terme être intégré dans un réseau de surveillance déjà existant.

Mots clefs : Surveillance Observationnelle, exposome, base de données, R, informations multiples

ABSTRACT

Impact of population on the environment, and conversely, is obvious and represents a real challenge for Public Health since 2000. It has been shown an increase in cancer prevalence, respiratory disease or even reproductive disorders, for which multifactorial origins are strongly suspected. In this context, surveillance has become an essential tool to decision making in public health, and surveillance networks of health events are multiplying, giving rise to numerous databases (sometimes considered as "big data"), still poorly used.
Objective of this thesis work was to develop a new concept of surveillance, the Observational Surveillance (OS). This allows an optimal use of observational databases, extendable to different kind of databases and problematics, taking into account of the multiplicity of information available.
OS is based on the exposome approach, to restructure data as a network, allowing the study of associations between information and also their structure. In this purpose, several indicators have been developed to study in the meantime the different recruited association for an event of interest, but also the evolution of their structure over time. These indicators allow generating the unique signature of the event: the spectrum. A tool, named "Observational Surveillance Analysis" (OSA), allowing the routine use of methodology, has been developed in the R platform, which permits automation and standardization of results.
Applications were used to illustrate the OS analysis and its portability and adaptability to different context and problematic. Three applications are based on the French National Occupational Diseases Surveillance and Prevention Network (RNV3P): bladder cancer, asthma and non-Hodgkin lymphoma. Three other applications are based on the Belgium occupational physicians group IDEWE: sore throat, caregivers and farmers.
Thanks to different applications, it has been demonstrated the portability of the OS methodology to different databases, and also, to different analysis configuration, disease/exposures or activity/diseases. Furthermore, the "OSA" tool which has been developed, allows an easier use to routine analysis and, in the end, could be integrated in an existing surveillance network.

Keywords: Observational surveillance, exposome, database, R, multiple information