



HAL
open science

Modèles de mélange en analyse de survie en présence de données groupées : application à la tremblante du mouton

Fabien Corbière

► **To cite this version:**

Fabien Corbière. Modèles de mélange en analyse de survie en présence de données groupées : application à la tremblante du mouton. Sciences du Vivant [q-bio]. Université Victor Segalen - Bordeaux II, 2007. Français. NNT: . tel-00195495

HAL Id: tel-00195495

<https://theses.hal.science/tel-00195495>

Submitted on 11 Dec 2007

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

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

Université Victor Segalen Bordeaux 2

Année 2007

Thèse n° 1465

THESE

pour le

DOCTORAT DE L'UNIVERSITÉ BORDEAUX 2

Mention : Sciences Biologiques et Médicales

Option : Epidémiologie et Intervention en Santé Publique

Présentée et soutenue publiquement

Le 4 décembre 2007

Par

Fabien CORBIÈRE

Né le 02 mars 1975 à Avranches

**Modèles de mélange en analyse de survie en
présence de données groupées :
application à la tremblante du mouton**

Membres du jury

Monsieur Ahmadou ALIOUM, Professeur, Université de Bordeaux	Président
Monsieur Vincent DUCROCQ, Directeur de recherche, INRA Jouy-en-Josas	Rapporteur
Monsieur Christian DUCROT, Directeur de recherche, INRA Theix	Rapporteur
Monsieur Francis BARILLET, Ingénieur de Recherche, INRA Toulouse	Membre
Monsieur Pierre JOLY, Maître de Conférences, Université de Bordeaux	Directeur de thèse

Remerciements

À Monsieur Ahmadou Alioum

Je vous remercie d'avoir accepté de présider ce jury. Trouvez ici l'expression de ma gratitude.

À Monsieur Vincent Ducrocq

Vous m'avez fait un grand honneur en acceptant de juger ce travail. Votre grande expérience en modélisation des données de survie m'apporta beaucoup.

Je vous remercie sincèrement d'avoir accepté d'être le rapporteur de ma thèse.

À Monsieur Christian Ducrot

Je suis très honoré que vous ayez accepté de juger ce travail. Votre connaissance de l'épidémiologie animale et de la tremblante en particulier m'apporteront un éclairage nouveau sur mon travail.

Je vous remercie vivement d'avoir accepté d'être le rapporteur de ma thèse.

À Monsieur Francis Barillet

Je ne pourrai jamais assez te remercier pour tout ce que tu m'as apporté. Tu as été des plus motivants, voyant toujours l'intérêt épidémiologique des méthodes que nous développons. Ton grand enthousiasme et ton soutien m'ont beaucoup aidé à réaliser ce travail. Je t'en suis profondément reconnaissant. Je suis persuadé que nous continuerons à collaborer étroitement.

À Monsieur Pierre Joly

Tu m'as pris sous ton aile il y a maintenant près de 4 ans, et m'as accordé toute ta confiance. Ce fut un réel bonheur de travailler avec toi. Pendant les moments de doute, ta bonne humeur et ta gentillesse ont su me remotiver. Merci infiniment pour tout ça et le reste.

Un immense merci à tous ceux qui ont contribué à la mise en place et à la réalisation du suivi des élevages dans les Pyrénées Atlantiques

Je ne remercierais jamais assez Francis Fidèle pour le travail énorme qu'il produit chaque jour dans les Pyrénées Atlantiques avec les éleveurs, pour sa bonne humeur et sa disponibilité. *Milesker anitz!*, et à bientôt chez les basques, pour quelques nouvelles séries de biopsies. Ps : à quand le pique-nique ?

À Claude Soulas, Jean-Marc Arranz, Xavier Aguerre et toutes les personnes du CDEO pour leur soutien et leur aide dans cette belle aventure.

Merci aux éleveurs qui ont accepté les contraintes liées à l'étude, pour leur accueil toujours chaleureux et pardon pour les réunions parfois tardives.

Aux stagiaires qui sont passés à la SAGA avant moi, et qui ont contribué chaque fois à l'amélioration des fichiers de données.

Merci enfin aux brebis... pour leur bon fromage.

Merci à tous ceux de l'École Vétô de Toulouse, mes collègues et amis...

À Caroline et Olivier tout d'abord, pour les week-end et les nuits passées devant la table à inclusion et le microscope, pour les séances de tuerie dans la joie et la bonne humeur... pour les soirées autour d'un bon pâté et d'une bonne bouteille et les remarques pertinentes de la Vérole, du Bâtard, du Ché, d'Adrien et des autres... Le livre d'or s'en souvient encore...

À Pierrette et Sév, pour le travail admirable qu'elles produisent tous les jours au labo... Sans vous, tout cela n'existerait pas.

À ceux qui sont passés et ne sont pas restés : merci à Armèle, Kader et bien sûr au plus grand biopsieur, Arnaud dit "Le Louche". Vous avez contribué grandement à faire que ce travail existe.

Merci enfin à toute l'équipe de bovine de m'avoir permis de m'évader pendant 4 ans pour faire ce petit rapport, et d'avoir assumé ma charge de travail : Hervé, les 2 Gilles, et les petits, Denis, Adrien et à tous les autres ... Promis je me remets au travail... après les vacances.

Un merci tout particulier à François Schelcher. En plus d'être l'un des instigateurs de ce projet, vous m'avez mis sur la voie de ma réalisation professionnelle. Je vous dois beaucoup.

Un immense merci à tous ceux que j'ai eu le plaisir de croiser à l'ISPED

À Estelle, Sandy, Gaelle pour les moments de franche rigolade et les soirées déjantées... Promis, je reviendrai vous voir souvent car vous allez me manquer.

À ceux du bureau 45, Cécile, Jérémie, Julien... et aux nouveaux : Julia, Cécile, Etienne. Merci pour votre aide quand je ne comprenais rien au $\text{\LaTeX} 2_{\epsilon}$, Fortran, Marquardt et autres algorithmes nébuleux... Je ne sais pas si j'ai progressé mais je vous dois beaucoup.

Un merci tout particulier à Danaëlle, qui m'a choyé, nourri, encouragé pour le travail et pour le reste. C'est quoi déjà une matrice définie positive ?

À tous les enseignants de l'ISPED qui m'ont appris un autre métier que celui de soigner des vaches et des moutons. A Pierre, Daniel, Hélène, Virginie, Alioum... Vos conseils et votre disponibilité m'ont toujours été d'un grand secours. Je retourne à mes animaux, mais je garderai toujours un oeil sur la biostatistique. En espérant pouvoir continuer à travailler avec vous.

Un grand merci enfin à toute l'équipe de l'ISPED. C'est une grande famille qui me manquera.

À mon papa,

À ma maman,

À ma soeurette et sa petite (et grandissante) famille,

À mon frerot, par monts et par vaux.

À ma demoiselle coquelicot...pour la vie

Table des matières

1	Introduction	15
1.1	Diversité biologique des agents des EST	18
1.2	Un agent transmissible non conventionnel : le Prion	20
1.3	Physiopathologie des ESST	23
1.4	Diagnostic des EST	25
1.5	Contrôle génétique de la susceptibilité	28
1.6	Épidémiologie de la tremblante du mouton	33
1.6.1	Distribution géographique	33
1.6.2	Prévalence de la maladie et incidence	34
1.6.3	Sources d’infectiosité	37
1.6.4	Vecteurs de transmission	39
1.6.5	Transmission inter-individuelle	40
1.6.6	Transmission entre élevages et persistance de l’infection	42
1.7	Difficultés des études épidémiologiques sur la tremblante	44
1.7.1	Définition de l’événement étudié	44
1.7.2	Durée d’incubation et espérance de vie	46
1.7.3	Difficultés liées au diagnostic	47
1.7.4	Schémas d’études et biais induis	47
1.8	Données groupées	52
1.9	Objectifs et plan	53

<i>TABLE DES MATIÈRES</i>	12
2 Modèles de survie avec fraction non à risque	55
2.1 Notations et définitions	56
2.1.1 Analyse des données de survie	56
2.1.2 Fonctions associées aux distributions de survie	57
2.1.3 Censure et troncature	59
2.1.4 Implication dans l'analyse de données de survie en présence d'une fraction non à risque	61
2.2 Estimation de la fraction non à risque : approches non paramétriques	63
2.2.1 Estimateur non paramétrique de la fraction à risque	63
2.2.2 Temps de suivi	64
2.2.3 Tester la présence d'une fraction non à risque	65
2.3 Prise en compte d'une fraction non à risque : différents modèles	66
2.3.1 Approche par modèles de mélange	66
2.3.2 Approche par modèle à risque cumulé borné ou modèle de promotion	67
2.3.3 Relations entre les modèles	69
2.4 Incorporation de variables explicatives	69
2.4.1 Modélisation de la fraction non à risque	69
2.4.2 Modélisation des temps de survie	70
2.4.3 Interprétation des paramètres de régression	71
2.5 Modèles de régression paramétriques et semi-paramétriques	73
2.5.1 Modèles paramétriques	73
2.5.2 Modèles semi-paramétriques	76
2.6 Estimation des modèles de survie avec fraction non à risque	78
2.6.1 Écriture de la vraisemblance	78
2.6.2 Estimation des paramètres	79
2.6.3 Estimation de la variance des paramètres	81
2.6.4 Identifiabilité	82
2.7 Modèles de survie avec fraction non à risque et effets aléatoires	84

2.7.1	Modèles de survie avec fraction non à risque et fragilité individuelle	85
2.7.2	Modèle de survie avec fraction non à risque et données groupées	87
2.8	Utilisation pratique des modèles de survie avec fraction non à risque .	90
2.8.1	Temps de suivi	91
2.8.2	Tester la présence d'une fraction non à risque	91
3	Programmation d'une Macro SAS	93
4	Approche par vraisemblance pénalisée	105
4.1	Article méthodologique	107
4.2	Extension au modèle de promotion	140
5	Modèle de mélange avec fraction non à risque et effets aléatoires	145
5.1	Les modèles	147
5.1.1	Modèle à deux effets aléatoires indépendants	147
5.1.2	Modèle à deux effets aléatoires non indépendants	148
5.1.3	Modèle à un effet aléatoire commun	149
5.2	Estimation des paramètres	150
5.2.1	Écriture de la vraisemblance	150
5.3	Estimation	151
5.4	Etude par simulations	152
5.4.1	Générations des données	152
5.4.2	Résultats	153
6	Application à la tremblante du mouton : une première analyse	157
6.1	Contexte	158
6.1.1	Les différents programmes de recherche	158
6.1.2	Contraintes liées à la Réglementation	158
6.2	Une première analyse	161
6.3	Analyse par vraisemblance pénalisée	173

6.3.1	Effets des variables explicatives	174
6.3.2	Fractions à risque estimées	177
6.3.3	Fonctions de survie et de risque	177
7	Application à la tremblante du mouton : une analyse plus large	181
7.1	Matériels et méthodes	182
7.1.1	Schéma d'étude	182
7.1.2	Population d'étude	182
7.1.3	Nature et origines des informations recueillies	184
7.1.4	Recueil et circuit des informations	190
7.1.5	Analyse statistique	190
7.2	Résultats : analyse descriptive des données	194
7.2.1	Échantillon d'étude	194
7.2.2	Temps de suivi	198
7.2.3	Génotype au gène PrP	200
7.3	Résultats de l'analyse multifactorielle	205
7.3.1	Description des modèles évalués	205
7.3.2	Résultats de l'analyse 1	208
7.3.3	Résultats de l'analyse 2	213
7.3.4	Estimation des fractions infectées	219
7.4	Discussion	222
8	Discussion & perspectives	229
9	Bibliographie	239
10	Annexes	263
10.1	Liste des publications et communications scientifiques	264
10.2	Autres publications liées à la tremblante	265
10.3	Résultats de l'analyse multifactorielle	291

Chapitre 1

Introduction

Les Encéphalopathies Subaiguës Spongiformes Transmissibles (ESST) sont des maladies neurodégénératives progressives rencontrées chez l'homme et différentes espèces de mammifères. Chez l'homme, différentes entités cliniques ont été décrites : le Kuru, le syndrome de Gerstmann-Sträussler-Scheinker (GSS), l'insomnie fatale familiale (IFF), la maladie de Creutzfeldt-Jakob (MCJ), et récemment le nouveau variant de la MCJ (nv-MCJ). Chez les autres mammifères, plusieurs entités cliniques ont été décrites et rattachées au groupe des ESST : tremblante des petits ruminants domestiques (mouton, chèvre) et sauvages en captivité (mouflon), maladie du dépérissement chronique (CWD) des ruminants sauvages captifs ou en liberté (antilopes, élans, cervidés), encéphalopathie du vison d'élevage, encéphalopathie spongiforme bovine (ESB). Le développement de l'épidémie d'ESB depuis le début des années 1980 semble être à l'origine de l'apparition de l'Encéphalopathie Spongiforme Féline (ESF) chez les chats domestiques et les grands fauves en captivité, et du nouveau variant de la MCJ (nv-MCJ) chez l'homme (Bruce et al., 2002).

Les ESST se caractérisent par une phase d'incubation (asymptomatique) longue par rapport à l'espérance de vie moyenne de l'espèce dont est issu l'individu atteint. Cette période d'incubation se chiffre en dizaines de mois chez les petits et grands ruminants et en années ou dizaines d'années chez l'homme. La phase symptomatique dure généralement plusieurs semaines à plusieurs mois et se traduit par une aggravation lente de symptômes nerveux. Ceux-ci sont éminemment variables selon les espèces, associant à la fois des troubles moteurs (incoordination motrice, instabilité posturale, tremblements, myoclonies), de la sensibilité (démangeaisons incoercibles), des facultés cognitives (démence, accès d'humeur). Dans tous les cas, l'issue de la maladie est fatale.

Aucune lésion macroscopique spécifique, directement imputable à la maladie, n'est décelable. Le diagnostic repose sur la mise en évidence, par histologie conventionnelle, de lésions spécifiques, limitées au Système Nerveux Central (SNC), de vacuolisation neuronale périkaryonique et neuropilaire. L'apparition de ces lésions dans l'encéphale ne précède que de peu les manifestations cliniques de la maladie.

Ce n'est que dans les années 30 que le caractère transmissible par inoculation de ces maladies fut démontré, par Jean Cuillé et Paul Louis Chelle, qui reproduisirent une tremblante clinique chez des ovins 24 mois, après injection d'un broyat de cerveaux issus d'ovins atteints (Cuillé et Chelle, 1936). De même, l'inoculation de système nerveux central issu d'individus atteints, à des individus de la même espèce ou d'espèce différente (rongeurs de laboratoire), permet généralement de reproduire la maladie, ou du moins une maladie neuro-dégénérative fatale présentant des caractères lésionnels typiques d'EST (Chandler, 1961 ; Fraser, 1976).

La connaissance de la physiopathologie des ESST a largement évolué depuis les années 1980, grâce, notamment, à l'utilisation des modèles expérimentaux sur rongeurs de laboratoire et aux avancées de la biologie cellulaire et moléculaire. En revanche, en raison des longues durées d'incubation qui caractérisent ces affections et des difficultés à établir un diagnostic précoce, l'épidémiologie de certaines ESST demeure mal connue. De ce fait, la modélisation mathématique a été utilisée que ce soit, par exemple, pour mieux comprendre la dynamique de la tremblante dans les troupeaux atteints (Woolhouse et al., 1998 ; Stringer et al., 1998 ; Matthews et al., 1999 ; Hagenaars et al., 2000 ; Gravenor et al., 2000), pour estimer l'importance de l'épidémie d'ESB en France (Donnelly, 2000 ; Supervie et Costagliola, 2004 ; Supervie et Costagliola, 2007) ou au Royaume-Uni (Donnelly et al., 1997), ou pour évaluer le risque d'ESB chez les ovins (Ferguson et al., 2002 ; Kao et al., 2003). Une des principales limites de ces approches mathématiques est le nombre important de paramètres définissant chaque modèle, et dont la plupart sont inconnus et arbitrairement fixés (Hagenaars et al., 2003). Dans le cas de la tremblante classique, le risque de contamination et la durée d'incubation de la maladie sont des paramètres clés des modèles dynamiques, mais leur estimation repose sur des données limitées obtenues dans des troupeaux expérimentaux (Matthews et al., 2001, Hagenaars et al., 2001). L'estimation, à partir d'études épidémiologiques réalisées en conditions de terrain, de ces paramètres et l'identification des facteurs de risque individuels les influençant apparaissent donc importantes, mais rencontrent un certain nombre de

difficultés méthodologiques.

Nous présentons dans cette introduction quelques éléments de biologie et de physiopathologie des ESST utiles à la compréhension du problème posé. Nous nous intéressons ensuite plus spécifiquement à l'épidémiologie de la tremblante classique du mouton, avant d'aborder les difficultés de l'interprétation des résultats des études épidémiologiques sur la tremblante.

1.1 Diversité biologique des agents des EST

Notion de souches

Le passage en série de deux isolats de tremblante ovine à des chèvres a conduit à l'apparition de deux "lignées" stables développant des signes cliniques très différents ; l'une dite "scratchy" (démangeaison) et l'autre dite "drowsy" (sommolence) (Pattison et Millson, 1961).

A partir de ces constatations s'est développé le concept de diversité biologique de l'agent étiologique causal des EST, aboutissant à la notion de souche. De nombreux autres éléments viennent étayer ce principe tels que : (i) la variété des symptômes observés chez des souris (Carp et al., 1984) ou des moutons (Foster et Dickinson, 1988) inoculés avec différents isolats d'EST, (ii) la possibilité ou l'impossibilité de transmettre certains isolats lors d'inoculations hétérospécifiques (Kimberlin et Walker, 1978a), (iii) l'extrême variabilité de thermostabilité de l'agent issu d'isolats différents (Dickinson et Taylor, 1978). En l'absence de toute connaissance précise sur la nature biologique de l'agent infectieux, les travaux menés dans les années 1960 - 1970 par l'équipe de Dickinson au NPU (Neuropathogenesis Unit) (Dickinson, 1976), ont fait de la transmission à des lignées de souris l'outil de référence pour le typage (caractérisation des souches) et pour l'étude de la bio-diversité de ces agents infectieux. Lorsqu'un isolat d'EST est inoculé plusieurs fois sur la même lignée de souris, on constate en cas de transmission effective une réduction progressive de la durée d'incubation de la maladie et une augmentation rapide de l'effica-

ité de la transmission (nombre de souris atteintes / nombre de souris inoculées). Ces caractères se stabilisent généralement après 3 ou 4 passages en série (Fraser et Dickinson, 1973 ; Bruce, 1979). Ainsi, après plusieurs passages en série d'un isolat, sur 5 lignées de souris standards, la durée d'incubation et le profil lésionnel sont les deux paramètres quantifiables qui permettent théoriquement de caractériser une souche (Bruce et Fraser, 1991).

La dénomination de tremblante ovine recouvre ainsi plusieurs souches d'agent infectieux qui diffèrent par leur durée d'incubation, leur profil lésionnel et leur efficacité de transmission. La mise en évidence, en 2003 en Norvège, d'un isolat (Nor98) de tremblante associé à un profil lésionnel atypique et touchant des ovins particulièrement âgés ou généralement peu sensibles voire résistants à la tremblante (Benestad et al., 2003) a conduit à la distinction majeure entre les souches de tremblante dites "classiques" et "atypiques". Ces dernières, décrites désormais dans plusieurs pays d'Europe (Orge et al., 2004 ; Buschmann et al., 2004, Konold et al., 2006, Seuberlich et al., 2007) pourraient former une seule et même entité (Arsac et al., 2007), bien que des profils lésionnels différents soient rapportés (Nentwig et al., 2007). Des différences épidémiologiques majeures existent aussi entre la tremblante dite "classique" et la tremblante "atypique", la tremblante atypique étant observée de manière sporadique au sein des troupeaux (1 à 3 cas au maximum décrits) sur des animaux relativement âgés (Luhken et al., 2007).

Notion de barrière d'espèce

Lors de l'inoculation d'une souche issue d'une espèce animale à une autre, différents phénomènes sont susceptibles de se produire. Il est tout d'abord possible qu'aucune transmission ne soit observée sur l'espèce cible ou que la maladie ne soit transmise en premier passage qu'à un faible pourcentage d'animaux inoculés. Certaines espèces comme le lapin semblent ainsi totalement réfractaires aux EST. D'autres, comme le porc – inoculé par voie intracérébrale avec de l'ESB issue de bovin – ou la souris – avec la souche 263K de hamster – ne semblent que partiellement et peu réceptifs

(Kimberlin et Walker, 1978b, Ryder et al., 2001).

Ensuite, au fur et à mesure des passages sur la nouvelle espèce hôte, la durée d'incubation va progressivement diminuer et l'efficacité de la transmission augmenter, jusqu'à devenir stable (Fraser, 1976).

Certaines souches passées à une autre espèce ont des propriétés modifiées et montrent un phénotype radicalement différent de celui de la souche originelle. Ainsi, le passage en série de la souche murine 22C sur le hamster, puis sa ré-inoculation à la souris, amène à une variation du phénotype et à l'émergence d'une nouvelle souche chez la souris (Kimberlin et al., 1989). Toutefois, ce phénomène n'est pas systématique. Certaines souches, comme la souche 139A (stable sur la souris), sont modifiées lors d'un passage en série sur le hamster, tandis qu'elles demeurent stables lors d'un passage en série sur le rat. (Kimberlin et al., 1987). De même, des souches telles que celle de l'ESB (Bruce et al., 2002) ou la souche murine ME7 (Kimberlin et Walker, 1978a) possèdent un phénotype extrêmement stable après passages interspécifiques. L'ensemble des phénomènes observés lors du passage trans-spécifique des souches est attribué à l'existence d'une barrière d'espèce. Cependant cette notion de barrière d'espèce dépend aussi bien des facteurs liés à la souche elle-même, qu'à certaines composantes de l'hôte.

1.2 Un agent transmissible non conventionnel : le Prion

La transmissibilité de la maladie par inoculation prouve le caractère infectieux des EST. Cependant, l'agent transmissible s'illustre par son caractère atypique qui lui a valu la qualification d'Agent Transmissible Non Conventionnel (ATNC), par opposition aux agents conventionnels que sont virus, bactéries ou agents parasitaires.

Absence de réponse immune humorale

L'une des caractéristiques principales de l'infection par l'agent de la tremblante est l'absence de réponse immunitaire humorale (Porter et al., 1973 ; Kascsak et al., 1987). Une telle propriété pose évidemment le problème de l'impossibilité de la détection de la maladie par des techniques sérologiques usuelles, mais aussi la question d'une éventuelle "immuno-tolérance" de l'hôte vis-à-vis de l'agent pathogène.

Propriétés physico-chimiques de l'ATNC

L'ATNC se montre particulièrement résistant aux différents procédés utilisés pour inactiver les agents conventionnels : radiations, chaleur sèche, chaleur humide, procédés chimiques ou enzymatiques (Sutton et al., 2006 ; Taylor, 2004). Cette exceptionnelle résistance aux procédés physico-chimiques de décontamination varie en fonction des isolats d'ATNC, mais également de l'état physique du matériel biologique soumis à la décontamination (lyophilisation préalable, taille de l'échantillon) (Kimberlin et al., 1983).

La persistance de l'ATNC dans le milieu extérieur semble aussi exceptionnelle : la réduction du titre infectieux d'un homogénat de cerveau de hamster, expérimentalement infecté, n'était que de 2 à 3 log après 3 ans passés enfoui dans la terre (Brown et Gajdusek, 1991). Des élans sains peuvent se contaminer à partir du sol où ont été entreposés deux ans auparavant des cadavres d'élans naturellement atteints par la CWD (Miller et al., 2006).

Nature de l'agent infectieux

L'observation au microscope électronique de la fraction synaptosomale-mitochondriale de préparations de cerveaux d'ovins atteints de tremblante, de patients atteints de MCJ ou de kuru et d'élans atteints de CWD, a permis l'identification d'agrégats fibrillaires d'une vingtaine de nanomètres de diamètre et de 100 à 200 nanomètres de long (Merz et al., 1981). Ces agrégats, baptisés Scrapie Associated Fibrils (Merz

et al., 1983), sont retrouvés dans les tissus des individus atteints d'ESST naturelle ou expérimentale et sont absents des préparations issues de tissus d'animaux sains (Merz et al., 1984). Leur inoculation sous forme purifiée a permis de reproduire la maladie (Diringer, 1984; Prusiner et al., 1983). Le constituant majeur des SAF est un polypeptide codé par un gène cellulaire (Oesch et al., 1985; Basler et al., 1986). La protéine codée par ce gène est à la fois exprimée chez les individus sains et atteints. Toutefois, chez les individus sains, cette protéine est sensible à la protéinase K, alors que chez les individus atteints elle y est partiellement résistante (Oesch et al., 1985).

Par la suite, une nomenclature particulière destinée à rendre compte des propriétés de cette protéine a fait son apparition (Meyer et al., 1986) : la protéine est dénommée PrP pour '**pro**tease **resistant** **pro**tein'; le gène codant pour la protéine PrP a été baptisé PRP; la protéine retrouvée chez les individus sains est appelée la Protéine Prion Cellulaire (PrPc); la protéine retrouvée chez les individus atteints est décrite sous le nom de Protéine Prion Scrapie ou résistante (PrPsc ou PrPres).

La PrPsc a la même séquence primaire que l'isoforme normale et diffère seulement par sa structure tertiaire, riche en feuillets β . Les particularités physico-chimiques de la PrPsc ont été attribuées à sa structure tertiaire particulière lui conférant une grande stabilité (Muramoto et al., 1996). Cette structure pathologique interférerait également avec le catabolisme naturel de la protéine et aboutirait à une accumulation progressive de celle-ci (Prusiner, 1991; Harris, 1999). En outre, la PrPsc serait synthétisée à partir de l'isoforme normale PrPc et en dériverait par un processus post-traductionnel (Meyer et al., 1986; Prusiner, 1991).

Très précocement a été formulée l'hypothèse qui selon laquelle les ATNC seraient une protéine infectieuse, dépourvue d'acides nucléiques, et douée de propriétés de répliation (Griffith, 1967). En 1982, suite à la découverte du polypeptide constituant

les SAF, Stanley Prusiner a repris et développé cette hypothèse, en baptisant cet agent putatif prion pour “Proteineaceous infectious particle” (Prusiner, 1982). La découverte du gène PRP (Oesch et al., 1985) et de la dualité PrP^c/PrP^{sc} sont venues compléter cette hypothèse. Bien que l’hypothèse Prion (PrP^{sc} = agent infectieux) soit désormais largement admise, l’hypothèse virale d’abord envisagée (Diringer, 1984 ; Dickinson et Outram, 1988 ; Diringer, 1991) n’est pas abandonnée (Manuelidis, 2007, Manuelidis et al., 2007).

1.3 Physiopathologie des ESST

Il est désormais bien établi que la voie de contamination principale est digestive pour la tremblante classique (Pattison et al., 1974 ; Hadlow et al., 1982), et la CWD (Williams et Young, 1980). La consommation de tissus infectieux semble également être la source des cas d’ESB chez les fauves en captivité (Wilesmith et al., 1991), des cas de Kuru (Gajdusek, 1977) et de nv-MCJ chez l’homme (Bruce et al., 1997 ; Hill et al., 1997). La voie de pénétration transcutanée par scarification et la voie conjonctivale sont également possibles (Carp, 1982 ; Taylor et al., 1996), mais leur implication lors de contamination naturelle reste hypothétique.

De nombreux éléments du schéma de dissémination de l’agent dans l’organisme sont à présent établis (Beekes et McBride, 2007). Ce schéma de dissémination est fortement dépendant du modèle animal (modèle rongeur ou hôte naturel), de l’isolat (tremblante classique, atypique, ESB, nv-MCJ, ou CWD), de la dose de l’inoculum, de la voie d’inoculation utilisée (voie orale, intrapéritonéale, sous-cutanée ou intracérébrale), ainsi que de certaines caractéristiques de l’hôte (génotype PRP en particulier chez les ovins). Cependant, plusieurs phases caractéristiques semblent communes aux différentes entités pathologiques et espèces cibles. Les observations réalisées sur des ovins naturellement infectés (van Keulen et al., 1996 ; van Keulen et al., 1999 ; van Keulen et al., 2002 ; Andreoletti et al., 2000 ; Heggebo et al., 2002 ; Heggebo et al., 2003) ou des patients atteints de nv-MCJ (Bruce et al., 2001), ainsi

que les résultats obtenus sur divers modèles expérimentaux rongeurs (Kimberlin et al., 1989; Maignien et al., 1999; Beekes et McBride, 2000), bovins (Wells et al., 1994; Terry et al., 2003), ovins (Jeffrey et al., 2001; Bellworthy et al., 2005; Ersdal et al., 2005) ou chez les cervidés sauvages (Sigurdson et al., 1999), permettent de distinguer :

- une phase de lympho-invasion, qui ne semble pas indispensable à l'infection mais la favorise. Cette phase se caractérise par une contamination précoce des structures lymphoïdes du tube digestif (GALT, **G**ut **A**ssociated **L**ymphoid **T**issue), puis des noeuds lymphatiques associés, et conduit progressivement à l'accumulation de PrPsc dans toutes les formations lymphoïdes secondaires,
- une phase de neuro-invasion au cours de laquelle de la PrPsc va s'accumuler d'abord dans les neurones du système nerveux autonome périphérique, puis dans ceux du système nerveux central.
- une phase de dissémination centrifuge, à partir du système nerveux central, vers des structures périphériques tels que le tissu musculaire.

La dissémination au système sanguin, retrouvée dans les modèles expérimentaux murins (Diringer, 1984; Brown et al., 1999) ou ovins (Houston et al., 2000; Hunter et al., 2002), semble pouvoir avoir lieu précocement dans la phase préclinique de la maladie.

Ce schéma connaît des exceptions notables. Un certain nombre d'études rapporte l'absence de détection de PrPres dans les organes lymphoïdes d'ovins ARR/VRQ, ARQ/ARQ, ou ARQ/VRQ cliniquement atteints ou en phase d'incubation dont le système nerveux central fournissait un résultat clairement positif (van Keulen et al., 1996; Jeffrey et al., 2002; Langeveld et al., 2006; Ligios et al., 2006; Monleon et al., 2005; O'Rourke et al., 2000). Un phénomène semblable pourrait exister chez les animaux atteints de tremblante atypique (ou Nor98), quelque soit leur génotype, mais les données disponibles sont trop peu nombreuses pour conclure (Benestad et al., 2003; Simmons et al., 2007). Chez les bovins atteints d'ESB clinique, aucune accumulation importante de PrPsc ou d'infectiosité n'est retrouvée dans les organes

lymphoïdes (Somerville et al., 1997 ; Hill et al., 1999). Chez les ovins, en revanche, lors d'inoculation orale par l'ESB, une accumulation précoce de PrPsc est observée dans les lymphocentres (Jeffrey et al., 2001). Il semble donc que l'existence d'une répllication lymphoïde primaire dépende à la fois de la souche de prion, mais aussi du terrain génétique de l'hôte.

Dans tous les cas, la phase de neuro-invasion centrale est tardive. Chez des ovins Romanov de génotype VRQ/VRQ issus d'élevages fortement infectés, les premiers dépôts de PrPres dans l'obex (noyaux du nerf vague) ne sont observés qu'à partir de l'âge de 9 ou 10 mois alors que de la PrPsc est mise en évidence de manière quasi systématique dès 3 mois dans le GALT ou dans les amygdales (Andreoletti et al., 2000 ; van Keulen et al., 2000).

1.4 Diagnostic des EST

Les outils diagnostiques des EST relèvent d'un double objectif : (i) permettre la détection pré-clinique la plus précoce possible des individus atteints, et (ii) confirmer une suspicion clinique. Une détection pré-clinique est, dans l'espèce humaine, indispensable au développement de thérapeutiques (encore à définir) préventives de l'atteinte neurodégénérative. Pour les espèces animales, la mise en oeuvre de tests pré-cliniques, en l'absence d'autres moyens de lutte, comme l'outil génétique chez les ovins, semble nécessaire à une prophylaxie efficace et acceptable socialement et économiquement.

La suspicion clinique d'EST est souvent tardive, les symptômes de la maladie étant peu évocateurs en début d'évolution. La suspicion est par ailleurs difficile à confirmer du vivant de l'individu. La palette des examens complémentaires, même en médecine humaine, reste limitée et d'une fiabilité relative. Le diagnostic est le plus souvent réalisé de façon post-mortem.

Les outils diagnostiques utilisés dans les EST humaines et animales, reposent sur :

- l'identification des lésions vacuolaires pathognomoniques associées à une gliose dans l'encéphale par histologie conventionnelle. Toutefois des artéfacts sur des prélèvements de tissu mal ou tardivement fixés ne sont pas rares et rendent le diagnostic par cette technique sinon impossible du moins très difficile. Par ailleurs les lésions vacuolaires peuvent être absentes ou peu intenses. Enfin l'histopathologie, comme outil de détection en phase d'incubation (dépistage pré-clinique), s'avère peu efficace du fait de l'apparition tardive des lésions de spongiose.
- la mise en évidence des Scrapie Associated Fibrils (SAF) (Merz et al., 1984). Cette méthode bien que lourde (mise en oeuvre et équipements requis) est considérée comme extrêmement spécifique et permet le diagnostic à partir de tissus autolysés (Scott et al., 1992; Stack et al., 1993), mais souffre d'une sensibilité réduite (Wells et al., 1994). Si cette technique paraît peu adaptée au diagnostic de première intention ou au criblage à grande échelle, elle n'en demeure pas moins l'un des outils possibles pour la confirmation d'un résultat obtenu par une autre méthode.
- la détection immuno-chimique de l'isoforme anormale de la protéine prion (PrP^{sc}) par techniques immunohistochimiques (Bell et al., 1997), ELISA ou Western-Blot. Plusieurs tests basés sur les techniques ELISA et Western-Blot ont été validés par la Commission Européenne pour le diagnostic de l'ESB et de la tremblante des petits ruminants. L'intérêt de ces techniques réside dans une grande facilité de mise en oeuvre, une grande rapidité d'exécution et la possibilité de traiter de grandes séries d'échantillons. Leur sensibilité est pour certaines au moins aussi bonne que celle de l'histologie, de l'immunohistochimie et de la détection des SAF (Grassi et al., 2001; Grassi, 2003). Par ailleurs et contrairement à l'histologie, ces méthodes peuvent être utilisées sur des échantillons autolysés. Toutefois, certains isolats de tremblante des petits ruminants (isolats Nor98) sont difficiles à détecter par les méthodes Western-Blot.

- la mise en évidence d’infectiosité par transmission aux rongeurs de laboratoire. Cette méthode a le grand avantage de révéler la présence d’infectiosité, indépendamment de la présence de PrPsc, mais reste longue, lourde et coûteuse et ne peut servir de méthode de première intention pour le diagnostic. Par ailleurs, la barrière d’espèce peut, avec certaines souches, empêcher la transmission primaire. Toutefois la méthode garde tout son intérêt pour le typage de souches et permet notamment de différencier les souches de tremblante de l’ESB.

Le dépistage préclinique ou dans les phases cliniques précoces est un objectif majeur en santé animale et humaine. La recherche de PrPsc dans les organes lymphoïdes secondaires d’un accès relativement aisé est une alternative intéressante. Le prélèvement est ensuite soumis à une détection immunohistochimique de la PrPsc. La méthode a été initialement développée chez les ovins infectés par la tremblante. Il est possible de détecter de la PrPsc dans les amygdales d’ovin en incubation de la maladie plus d’un an avant les signes cliniques (Schreuder et al., 1996 ; van Keulen et al., 1996 ; Schreuder et al., 1998). Cependant l’apparition de PrPsc dans les amygdales varie notamment en fonction du génotype de l’animal (Schreuder et al., 1998). Pour certains génotypes, une biopsie d’amygdale négative n’implique donc pas forcément que l’animal soit indemne. Plus récemment des alternatives à la biopsie d’amygdale ont été proposées chez les ovins et les ruminants sauvages, par prélèvement du tissu lymphoïde de la membrane nictitante (“3^{ieme} paupière”) (O’Rourke et al., 2000) ou de la muqueuse rectale (Gonzalez et al., 2005 ; Gonzalez et al., 2006 ; Espenes et al., 2006) . Ces approches se sont révélées légèrement moins sensibles que la biopsie d’amygdale, en raison d’un nombre souvent réduit de follicules lymphoïdes rendant l’interprétation délicate. Chez l’homme, une biopsie d’amygdale ou le prélèvement de l’appendice permet de conforter une suspicion de nv-MCJ (Hill et al., 1997 ; Hilton et al., 1998). Toutefois les biopsies tissulaires ne peuvent être utilisées dans tous les cas. En effet chez les bovins atteints d’ESB ou chez les individus atteints de MCJ sporadique l’accumulation de PrPsc dans les organes lymphoïdes secondaires n’est

pas ou exceptionnellement détectable (Wells et al., 1998).

D'autres techniques à visée diagnostique, moins invasives mais de spécificité réduite, sont utilisées chez l'homme. Elles sont basées (i) sur la détection de marqueurs biochimiques ou moléculaires, (ii) sur des modifications fonctionnelles tissulaires associées à la maladie, ou (iii) des modifications lésionnelles mises en évidence par des techniques d'imagerie médicale.

Des efforts très soutenus sont consentis actuellement à la mise au point et à la validation de tests qui seraient à la fois précoces, spécifiques, sensibles et peu invasifs, en recourant notamment aux fluides corporels. La mise en évidence de PrPsc dans l'urine, évaluée sur 100 patients atteints de CJD familial, sporadique et du nv-CJD s'est révélée d'une faible sensibilité (Head et al., 2005). Différentes compagnies privées et équipes de recherche travaillent actuellement à l'élaboration de tests diagnostiques sanguins applicables aux maladies à prion. Toutefois, les résultats actuels, même si certaines approches semblent prometteuses, ne permettent pas d'envisager à court terme l'élaboration d'un outil diagnostique préclinique robuste.

1.5 Contrôle génétique de la susceptibilité

Modulation de la sensibilité par le génotype PRP

Chez les ovins le risque d'infection (susceptibilité) par la tremblante classique est étroitement associé au polymorphisme de trois codons (136, 154, 171) du gène PRP (chromosome 13), codant pour la PrPc, sans que les mécanismes cellulaires et moléculaires soient complètement élucidés (Laplanche et al., 1993; Hunter et al., 1992; Hunter et al., 1994). Des 12 (2x2x3) combinaisons possibles sur les codons 136 (T/A/V), 154 (R/H) et 171 (Q/R/H/K) seulement six combinaisons ont été observées. Les études expérimentales et observationnelles portant sur l'association entre polymorphisme du gène PrP et susceptibilité à la tremblante classique ont conduit à distinguer :

(a) des animaux résistants, de génotype $A_{136}R_{154}R_{171}/A_{136}R_{154}R_{171}$ (plus simplement noté ARR/ARR), qui, placés dans un milieu fortement contaminant ou inoculés par voie orale, ne déclenchent pas la maladie. Toutefois le caractère absolu de cette résistance a récemment été remise en question. Deux cas d'infection naturelle par la tremblante classique, associés à des signes cliniques ont été rapportés (Groschup et al., 2007). Une première suspicion avait déjà été faite en 1995 sur un Suffolk japonais (Ikeda et al., 1995), sans qu'une erreur de typage génétique n'ait pu être définitivement écartée. La contamination par voie intra cérébrale avec de l'ESB, est possible (Houston et al., 2003). La présence de PrP^{sc} a été récemment mise en évidence dans la rate d'un animal ARR/ARR contaminé par voie orale avec de l'ESB (Andreoletti et al., 2006, voir annexes), montrant la possibilité, dans des conditions expérimentales particulières (forte dose d'inoculum), d'un portage asymptomatique chez les ovins de ce génotype.

(b) les animaux hétérozygotes ARR, qui, dans les conditions naturelles d'exposition à l'agent, ne développent les signes cliniques de la maladie que dans une (très) faible proportion. Dans un élevage Romanov fortement contaminé moins de 2% des animaux ARR/ARQ sont ainsi atteints (Elsen et al., 1999; Diaz et al., 2005). Ces données sont confirmées par plusieurs études cas témoins réalisées en Islande (Thorgeirsdottir et al., 1999), Norvège (Tranulis et al., 1999), Irlande (O'Doherty et al., 2002), Grèce (Billinis et al., 2004) en Espagne (Acin et al., 2004) ou en France (Barillet et al., 2002).

(c) des animaux sensibles (génotype ARQ/ARQ) et très sensibles (génotypes VRQ/VRQ, ARQ/VRQ) qui, placés en milieu contaminant, développent la maladie dans une forte proportion. Sur plus de 1500 moutons de race Romanov, issus d'un troupeau français naturellement infecté (troupeau de Langlade), les animaux homozygotes VRQ semblent les plus sensibles à la tremblante (66% d'animaux atteints) (Elsen et al., 1999; Diaz et al., 2005). Cette sensibilité est élevée mais moindre pour les

animaux hétérozygotes VRQ/ARQ (52% de cas) et homozygotes ARQ/ARQ (38%). La plupart des études menées en Europe confirment ces observations (Baylis et al., 2002 ; Baylis et al., 2004 ; Thorgeirsdottir et al., 1999 ; Billinis et al., 2004).

En Grande-Bretagne une étude cas témoins comparant les fréquences génotypiques de plus de 1543 cas cliniques confirmés sur une période de 45 mois et celles de près de 15000 animaux cliniquement sains issus de 44 élevages a permis d'établir le risque de chacun des 15 génotypes décrits en terme de nombre de cas par année et par million d'ovins. Ces résultats, présentés dans le tableau 1.1, offrent une confirmation à grande échelle des différentes études menées jusqu'alors (Baylis et al., 2004).

TAB. 1.1 – Susceptibilité à la tremblante classique des différents génotypes ovins [d'après Baylis et al., 2004]

Génotype	Cas par an par million	IC 95%
ARR/ARR	0.0	0.0-0.3
ARR/AHQ	0.3	0.0-1.6
ARR/ARQ	0.4	0.2-0.8
ARR/ARH	0.0	0.0-2.9
AHQ/AHQ	5.0	2.3-10.9
ARQ/AHQ	8.7	6.2-12.1
AHQ/ARH	0.0	0.0-20.6
ARH/ARH	2.0	0.5-7.3
ARQ/ARH	5.2	2.2-12.1
ARQ/ARQ	36.9	33.1-41.0
ARR/VRQ	6.3	4.5-8.6
AHQ/VRQ	0.7	0.0-3.7
ARQ/VRQ	225.4	214.7-236.0
VRQ/VRQ	544.5	490.5-602.9

La sensibilité relative des animaux d'un génotype donné peut varier en fonction de l'environnement et de la souche de tremblante (Benestad et al., 2003 ; Foster et Dickinson, 1988 ; Goldmann et al., 1994 ; Hunter, 1997). En effet si dans un troupeau

particulier de race Suffolk le génotype ARQ/ARQ a pu être associé à la majorité des cas de tremblante (Hunter et al., 1997), dans un autre troupeau de structure génétique PrP comparable, mais en race Texel, ce même génotype ARQ/ARQ ne correspond qu'à une minorité de cas (Dawson et al., 1998). Lors d'inoculation par voie intracérébrale d'animaux ARQ/ARQ avec la souche CH1641 tous les animaux vont déclarer la maladie, alors que des ovins du même génotype inoculés avec la souche SSBP1 ne seront pas atteints cliniquement (Goldmann et al., 1994). La souche Nor98, isolée en Norvège et responsable de la tremblante dite "atypique", semble avoir un tropisme pour l'allèle AHQ et l'allèle AF₁₄₁RQ et a été décrite plusieurs fois chez des ovins de génotypes ARR/ARR (Benestad et al., 2003; Buschmann et al., 2004; Moum et al., 2005; Arsac et al., 2007; Moreno et al., 2007; Luhken et al., 2007).

Durée d'incubation

Les animaux cliniquement atteints sont le plus souvent âgés de 2 à 5 ans. Dans les conditions naturelles, cependant, une grande variabilité de la durée de survie des animaux infectés est observée. La maladie semble rare après 8 ans et avant 18 mois, bien que quelques cas de tremblante sur des animaux âgés de 12, voire 10 ou 7 mois aient été rapportés (Baylis et al., 2002; Hunter et al., 1992; Hunter, 1997; Redman et al., 2002; Wineland et al., 1998; Hoinville, 1996). Les données européennes (EFSA, 2007), relatives à l'âge de détection des animaux par les programmes de surveillance active et passive, concernent les animaux suspects cliniques et ceux détectés en phase pré-clinique. La médiane de l'âge à la détection est de 3.5 ans (intervalle interquartiles 2.5-4.5). Des observations similaires sont réalisées à partir des données anglaises du Scrapie Notification Database sur 1543 cas cliniques confirmés sur une période de 45 mois (août 1998-avril 2002) : 80% des animaux meurent entre 2 et 4 ans et seulement 3% déclarent les symptômes à plus de 7 ans (Baylis et al., 2004).

Les études pour lesquelles une information sur le génotype des animaux malades est disponible sont peu nombreuses, et portent sur des effectifs souvent réduits. Une

des études la mieux documentée concerne 276 cas issus d'un troupeau expérimental de race Romanov fortement infecté (Elsen et al., 1999). L'âge moyen à la mort des animaux malades était de 2.4 ans, et les animaux de génotype VRQ/VRQ étaient en moyenne atteints respectivement 91 et 141 jours avant les ARQ/VRQ et les ARQ/ARQ. D'autres travaux menés dans d'autres races et sur des effectifs réduits ne mettent pas en évidence de différences entre génotypes sensibles (Tranulis, 2002) ou indiquent une tendance inverse (Hunter et al., 1996). Certains auteurs font état d'une diminution de l'âge à la mort (et donc peut-être d'une réduction de la durée d'incubation) avec le temps dans des élevages expérimentaux fortement infectés, sans que les génotypes des animaux ne soient renseignés (Foster et Dickinson, 1989, Redman et al., 2002).

Utilisation de l'outil génétique dans la lutte contre les EST chez les ovins

Les programmes de lutte contre la tremblante ont été longtemps fondés sur une stratégie d'abattage total des cheptels atteints, avec comme résultat une fréquente ré-infection des animaux nouvellement introduits. Cette situation est probablement liée à la forte rémanence de l'agent infectieux dans le milieu d'élevage et à sa résistance aux procédés de décontamination. Aux USA, les campagnes d'éradication menées dès 1952 n'ont pas été, d'un point de vue global, couronnées de succès. Seule l'Islande semble en voie de réussir par l'application d'une telle stratégie. La politique islandaise en la matière apparaît toutefois comme extrêmement contraignante et coûteuse. Compte tenu des connaissances actuelles, l'utilisation de l'allèle ARR comme moyen de lutte contre la tremblante classique est apparue comme une stratégie efficace, fondée sur l'absence de portage sain et la résistance extrêmement forte des ovins ARR/ARR aux souches de tremblante classique (Dawson et al., 1998). Des programmes de lutte fondés en grande partie sur la sélection d'animaux résistants ont d'abord débuté en Grande-Bretagne, en France et aux Pays-Bas, avant d'être étendus à l'ensemble des pays membres de l'Union Européenne en 2001. En France, un programme national de sélection des ovins pour la résistance à la trem-

blante a été initié en 2002, sous l'égide du Ministère de l'Agriculture, de l'Alimentation, de la Pêche et des Affaires Rurales pour une durée initiale de 5 ans, prolongée depuis pour une deuxième phase de 3 ans. Une présentation détaillée de ce programme et des résultats préliminaires sont présentés en annexe, dans l'article de Barillet et al. (2002).

1.6 Épidémiologie de la tremblante du mouton

Dans cette partie, nous évoquerons des éléments d'épidémiologie de la tremblante classique du mouton. La tremblante atypique ne sera pas traitée.

La plupart des études épidémiologiques menées jusqu'alors ont eu pour objectifs (i) d'estimer la prévalence de la maladie à l'échelle nationale ou européenne, (ii) de déterminer les facteurs de risque de l'acquisition et du développement de la maladie à l'échelle de l'élevage, (iii) de déterminer les facteurs de risque génétiques et environnementaux liés à la maladie à l'échelle individuelle.

Depuis le 14 juin 1996, date à laquelle la tremblante des petits ruminants est devenue une maladie à déclaration obligatoire, 1099 foyers de tremblante ovine et caprine ont été confirmés en France (source DGAL, données au 6 septembre 2007), à partir des réseaux de surveillance passive (animaux suspects cliniques), actif (sondage à l'abattoir ou à l'équarrissage), ou du Contrôle Sanitaire Officiel (pour les élevages en sélection).

1.6.1 Distribution géographique

Décrite pour la première fois en Angleterre en 1732, la tremblante des petits ruminants est reconnue aujourd'hui dans de nombreux pays d'Europe et du monde (Brown et Bradley, 1998, Detwiler et Baylis, 2003). Dans de nombreux pays (Etats-Unis, Australie, Canada, Colombie, Kenya, Nouvelle-Zélande, Norvège, Afrique du Sud, Inde) l'importation d'ovins en provenance de Grande-Bretagne est fortement suspectée d'être à l'origine de l'introduction de la maladie entre 1938 et 1977. La

distribution mondiale de la tremblante est cependant difficile à établir avec précision, compte tenu de l'absence de réseau d'épidémiologie active ou passive dans la majorité des pays.

En Europe, le réseau de surveillance active initié en 2002 a permis de mettre en évidence la maladie dans la plupart des pays, exceptés le Danemark, l'Estonie, la Lituanie, la Lettonie, la Hongrie, Malte, l'Autriche, la Bulgarie et le Luxembourg. L'absence de cas positif dans ces pays ne semble pas être due à un défaut de détection (EFSA, 2007).

En France, la tremblante a été mise en évidence dans la plupart des régions où l'élevage ovin est présent (figure 1.1). Malgré un sous-échantillonnage dans certains départements, l'origine des différences observées entre régions semble liée à une forte hétérogénéité géographique des types de production (allaitant, laitier) et au fort attachement des différents bassins à certaines races (Manech tête Rousse en Pyrénées Atlantique et Lacaune dans le Rayon Roquefort). La tremblante classique est particulièrement présente dans le département des Pyrénées Atlantiques, qui concentre près de 40 % des foyers de tremblante classique détectés au 1^{er} septembre 2007 (source AFSSA, données non publiées).

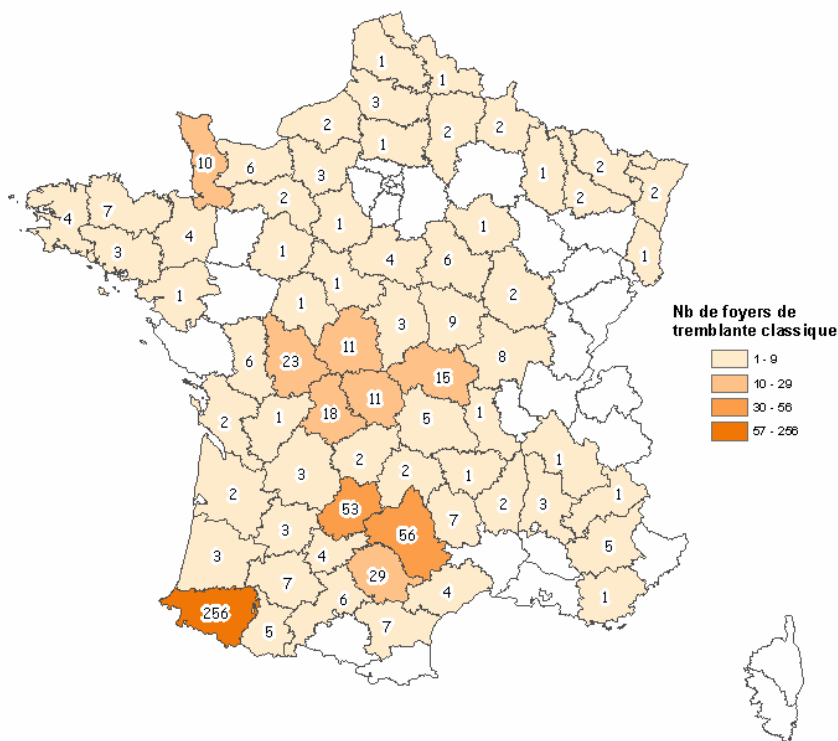
1.6.2 Prévalence de la maladie et incidence

La prévalence de la tremblante classique peut s'exprimer à deux échelles : celle de l'élevage (proportion d'élevages infectés) ou individuelle (proportion d'animaux infectés).

Assez peu d'études ont été menées pour estimer, à l'échelle d'un pays, la proportion d'élevages infectés. En raison (i) de la probable faible prévalence de la maladie dans la plupart des troupeaux, (ii) du faible nombre d'animaux testés par troupeaux au travers des programmes de surveillance active (à titre d'exemple, en France entre 2002 et 2006, 47% des élevages n'ont eu qu'un seul animal testé, et seulement 25% ont eu plus de 6 animaux testés, source AFSSA, données non publiées), (iii) du

Foyers de tremblante classique détectés en France
entre le 14 juin 1996 et le 1^{er} septembre 2007
par les réseaux d'épidémiologie-surveillance
(surveillance passive, active et contrôle sanitaire officiel)

Source : AFSSA 2007 – Données non publiées



Note : Les foyers de tremblante non typée détectés par la surveillance passive ont été inclus. Les foyers de tremblante détectés par la surveillance active qui n'ont pas pu faire l'objet de typage de souche n'ont pas été inclus (échantillons non analysables).

FIG. 1.1 – Répartition des foyers incidents de tremblante classique ovine en France.

manque de sensibilité de l'obex en phase préclinique, la capacité de ces programmes de surveillance à détecter un élevage infecté apparaît très faible. Une étude par simu-

lations Monte-Carlo (Hopp et al., 2003), menée à partir des données de surveillance norvégiennes a estimé que, dans les meilleures conditions d'échantillonnage et de sensibilité diagnostique, moins de 9% des troupeaux atteints pourraient être mis en évidence.

Sur la base des enquêtes postales réalisées en Grande Bretagne en 1998 (Hoinville et al., 2000) et 2002 (Sivam et al., 2003 ; Sivam et al., 2006), entre 11.7% et 14.9% des éleveurs britanniques estiment avoir déjà eu au moins un cas clinique dans leur troupeau et entre 1% et 2.7% pensent en avoir vu au cours des 12 derniers mois. Dans les îles Shetland la proportion des élevages qui ont connu au moins une fois la tremblante est évaluée à 28.5% [IC 95% 24.5 - 32.8] (Gubbins et al., 2006).

Aux Pays-Bas, selon ce même type d'enquêtes menées en 1990 (Schreuder et al., 1993), 6.1 % [IC 95% 3.8 - 8.4] des éleveurs pensent avoir eu au moins un animal malade dans leur élevage. Ces études, réalisées à partir d'auto-questionnaires, sont fortement tributaires du taux de réponse, parfois faible, et de la capacité des éleveurs à reconnaître les signes cliniques de la maladie : dans l'enquête menée au Pays-Bas seuls 40% des éleveurs répondants pensent savoir identifier un animal cliniquement atteint. Malgré ces défauts, ces résultats laissent supposer que, dans ces pays, un nombre considérable d'élevages a pu être infecté.

L'incidence clinique annuelle intra-troupeau est, elle aussi, très variable : sur la base des mêmes questionnaires, sa valeur médiane est estimée entre 0.4% et 1.3% mais des incidences de 20% à 30% ont été rapportées (Schreuder et al., 1993 ; Morgan et al., 1990, Hoinville, 1996 ; Hoinville et al., 2000).

Les données fournies par les programmes de surveillance active à l'échelle nationale et européenne fournissent un ordre de grandeur de la prévalence individuelle à l'abattoir (animaux cliniquement sains de plus de 18 mois) et à l'équarrissage (population à risque).

Le programme de sondage mené, depuis 2002, dans l'Europe des 15 puis des 25 et en Grande-Bretagne et Roumanie a permis de tester (tests rapides sur l'obex) plus de 2 663 000 petits ruminants à l'abattoir et à l'équarrissage (données 2002-

2006, actualisées au 1^{er} juillet 2007, dont 77.7 % d'ovins) (EFSA, 2007). Sur cette période, la prévalence de la tremblante (classique et atypique confondue) observée à l'échelle individuelle variait, pour les 15 pays ayant détecté des cas, de 1.3 pour 10 000 (Finlande) à plus de 746.2 pour 10 000 (Chypre) à l'abattoir. Conformément aux hypothèses initiales, la prévalence individuelle à l'équarrissage (animaux à risque) est plus élevée : de 4.3 (Suède) à près de 1603 (Chypre) cas pour 10 000. La prévalence annuelle moyenne observée chez les ovins à l'échelle européenne sur la période 2002-2006 varie entre 5.9 et 9.4 cas pour 10 000 à l'abattoir et entre 12.1 et 27.8 cas pour 10 000 à l'équarrissage. La forte hétérogénéité, tant spatiale que temporelle, dans la mise en oeuvre des programmes de surveillance active entre les différents États membres, rend cependant difficile la comparaison des résultats obtenus entre chaque pays (Del Rio Vilas et al., 2007).

En France, plus spécifiquement, 1 035 218 prélèvements d'obex ont été réalisés dans le cadre de la surveillance active entre 2002 et 2006 (dont 64.0 % d'ovins et 58.5% des prélèvements réalisés à l'abattoir). Plus de 63% de l'ensemble des prélèvements ont été effectués pendant l'année 2006, suite à une modification du plan de sondage initial vers l'exhaustivité à partir du mois d'avril 2006. La prévalence brute observée de la tremblante ovine classique à l'abattoir, estimée à 4.8 [IC 95% : 2.7 - 7.8] cas pour 10000 en 2002, semble avoir connu une baisse en 2006 (0.8 [IC 95% : 0.0 - 2.0] cas pour 10 000). A l'image de beaucoup de pays européens, la prévalence brute observée à l'équarrissage est plus élevée et relativement constante : 12.7 [IC 95% : 8.5 - 18.9] cas pour 10 000 en 2002 et 7.0 [IC 95% : 5.4 - 8.2] cas pour 10 000 en 2006.

1.6.3 Sources d'infectiosité

La recherche d'infectiosité ou de PrPres a été réalisée dans la plupart des sécrétions et excréments ou organes pouvant être à l'origine de contaminations inter-individuelles. Les résultats de ces études, pour la plupart conduites avant 1990, doivent être in-

interprétés avec précaution en tenant compte du seuil de détection des techniques utilisées et de la pertinence de l'échantillon animal (espèce, effectifs, génotype PRP, tremblante naturelle ou expérimentale, phase clinique ou préclinique).

Des résultats expérimentaux suggèrent la présence d'une très faible infectiosité dans l'urine de certaines espèces (bovin, hamster, homme)(Shaked et al., 2001). Aucune donnée comparable n'est disponible chez les petits ruminants. Dans le cas de la Maladie du Dépérissement chronique des Cervidés sauvages (CWD), la présence d'infectiosité dans la salive de cerf a été récemment démontrée par inoculation orale (Mathiason et al., 2006).

La coexistence d'une infection ou inflammation lympho-proliférative pourrait favoriser l'excrétion de PrP pathogène : chez la souris inoculée par de la tremblante, l'induction d'une néphrite lymphocytaire expérimentale a permis l'excrétion de PrPres infectieuse dans l'urine (Seeger et al., 2005). De la PrPres a aussi été mise en évidence dans le tissu mammaire d'une brebis atteinte d'une mammite et du Visna-Maedi, suggérant la possibilité d'une excrétion dans le lait (Ligios et al., 2005).

Par ailleurs, une infectiosité a été mise en évidence dans certains organes ou tissus, comme les glandes salivaires, la muqueuse nasale et les intestins (Pattison et Millson, 1962 ; Hadlow et al., 1980 ; Hadlow et al., 1982 ; Vascellari et al., 2007).

Dans le sang, la présence d'infectiosité a été démontrée sur des modèles murins (Diringer, 1984 ; Brown et Bradley, 1998) et plus récemment chez des brebis atteintes et en incubation de tremblante naturelle et d'ESB (Houston et al., 2000 ; Hunter et al., 2002), et chez le cerf atteint de CWD (Mathiason et al., 2006). Le rôle de la transfusion sanguine (sang non déleucocyté) dans la transmission du vCJD est fortement suspecté chez 4 patients (Lefrere, 2007). Chez les ovins, une transmission effective de la maladie est possible par l'utilisation de sang complet prélevé sur des donneurs à mi-incubation. Même si les volumes transfusés sont élevés, les durées d'incubation chez les animaux récepteurs sont proches de celles obtenues par une injection intraveineuse d'une grande quantité d'infectiosité (0.2 g de cerveau issu

de bovin ESB), ce qui suggère une infectiosité circulante importante (Hunter et al., 2002).

L'infectiosité du placenta est bien démontrée, sans être constante. Cette infectiosité est dépendante à la fois du génotype au gène PRP de la mère et de celui des foetus (Tuo et al., 2002, Andreoletti et al., 2002, Lacroux et al., 2007, voir annexes). Le placenta des foetus de génotype sensible (VRQ/VRQ) issus d'une brebis VRQ/VRQ en incubation accumule de la PrPres. A l'inverse, l'insémination d'une brebis VRQ/VRQ en incubation de tremblante avec de la semence de bélier ARR/ARR résulte en un agneau hétérozygote ARR et un placenta ne contenant pas de PrPres. De même, le placenta issu d'une brebis ARR/VRQ en incubation ne semble pas pouvoir accumuler de PrPres, quelque soit le génotype, même sensible, du produit. La présence de PrPres placentaire est possible dès la première mise-bas (Andreoletti et al., 2002) et a été décelée jusqu'à 470 jours avant l'apparition des premiers signes cliniques sur la brebis (Race et al., 2002). Les niveaux de PrPres observés dans les placentas sont d'environ 50 à 100 fois moins élevés que dans les obex des animaux atteints en phase clinique et cinq à huit fois plus élevés que dans les organes lymphoïdes secondaires (Andreoletti et al., 2002).

1.6.4 Vecteurs de transmission

Dans les conditions naturelles, différents vecteurs de transmission de la tremblante classique ont été suggérés. Les acariens des fourrages ont été suspectés (Wisniewski et al., 1996), sans que cela puisse être confirmé. Dans des conditions expérimentales, des larves de mouches de l'espèce *Sarcophaga carnaria*, alimentées sur des encéphales de hamsters infectés par la tremblante, semblent susceptibles de recontaminer des hamsters après ingestion (Post et al., 1999), mais la pertinence de cette étude pour mimer des conditions naturelles est très faible.

Un vaccin dirigé contre le louping ill, préparé à partir de suspensions de cerveaux de moutons infectés par la tremblante, a été incriminé dans la transmission de la ma-

ladié en Grande Bretagne (Greig, 1950). Plus récemment, le rôle d'un vaccin dirigé contre *Mycoplasma agalactiae* a été mis en cause dans l'augmentation de l'incidence de la tremblante en Italie en 1997 -1998 (Caramelli et al., 2001).

1.6.5 Transmission inter-individuelle

Lors d'exposition naturelle, il semble que la contamination est lieu autour de la naissance. De la PrPres est mise en évidence dans les plaques de Peyer iléales ou les amygdales dès l'âge de 2 mois chez des agneaux VRQ/VRQ nés dans des élevages fortement infectés (Heggebo et al., 2000, Andreoletti et al., 2000 ; van Keulen et al., 2002). La contamination à l'âge adulte semble aussi possible (Hourrigan et al., 1979 ; Ryder et al., 2004) mais son importance, comparée à la contamination néonatale est mal documentée. Dans les conditions naturelles d'infection, la transmission horizontale, ou latérale, est bien établie (Brotherston et al., 1968 ; Dickinson et al., 1974 ; Hourrigan et al., 1979 ; Ryder et al., 2004 ; Foster et al., 2006). L'importance relative des modalités directes et/ou indirectes de la transmission horizontale est mal connue. Les essais ont pour la plupart été réalisés avant la mise en évidence de l'importance du génotype PrP, et souffrent de ce fait de nombreux biais, mais permettent d'émettre plusieurs hypothèses.

La persistance à long terme dans l'environnement de l'infectiosité de tissus infectés (Brown et Gajdusek, 1991 ; Miller et al., 2006), et les connaissances sur la résistance aux agents physiques et chimiques des prions, suggèrent la possibilité d'une transmission indirecte à partir de l'environnement contaminé.

La transmission de la mère à l'agneau recouvre plusieurs modalités biologiques (post-natale et ante-natale), difficiles à distinguer. Selon les études (Hoinville, 1996), le risque relatif de développer la tremblante pour les agneaux issus de brebis infectées par rapport à des agneaux du même troupeau mais issus de brebis cliniquement saines était de 1.85 [IC 95% 1.37-2.51] à 4.80 [IC 95% 3.12-7.40]. L'absence d'information sur les génotypes des brebis, des béliers et des agneaux – ces études ont été

conduites dans les années 60 et 70 – ne permet cependant pas de faire la part des choses entre l'intervention d'une contamination directe mère-agneaux et une sensibilité génétique particulière des agneaux issus des mères infectées. La part de risque attribuable au statut infectieux des parents était relativement faible dans ces études (entre 21% et 32%) tendant à montrer que la majeure partie des contaminations seraient due à une transmission horizontale entre animaux sans lien de parenté.

Dans ces mêmes études, le risque de transmission post-natale a été évalué par le retrait des agneaux de l'environnement contaminé, à des âges différents. Le risque de tremblante clinique semble s'accroître avec la durée de l'exposition dans un environnement contaminé. Les mêmes réserves quant aux facteurs de confusion potentiels non maîtrisés réduisent cependant la portée des résultats obtenus. L'allaitement artificiel, associé à un retrait des agneaux de la mère, conduit à réduire l'incidence de tremblante par rapport à l'allaitement maternel, sans qu'il soit possible d'en préciser la raison (Elsen et al., 1999). La réduction de la durée de contact entre les agneaux et les mères infectés pourrait expliquer ces observations.

L'existence d'une transmission verticale au sens strict (transplacentaire, antenatale), n'est pas clairement documentée. L'infectiosité du sperme n'a pas été démontrée et celle des ovaires et de l'utérus ont été rapportées par une seule équipe (Hourrigan, 1990). Toutefois la transmission *in utero* n'a pas été démontrée formellement. Les résultats des essais de transfert embryonnaire sont discordants et certains éléments manquent pour pouvoir conclure avec certitude. Le transfert d'embryons prélevés sur des brebis infectées naturellement ou expérimentalement, à des brebis issues d'élevages indemnes, n'a conduit, après une période d'observation de 60 mois, à aucun signe clinique de maladie sur les agneaux ou les brebis receveuses, ni à la mise en évidence de PrPres dans l'encéphale (Foote et al., 1993 ; Wang et al., 2001). Toutefois le génotype PRP des receveuses et des agneaux n'a pas été précisé. Une équipe britannique conclut à une transmission possible de la tremblante par l'embryon ovin sans toutefois pouvoir exclure une contamination horizontale des agneaux (Foster et al., 1992 ; Foster et al., 1996).

Le rôle du placenta dans la transmission est souvent mis en avant. Lors de la mise-bas, les annexes foetales souillent abondamment le périnée et la mamelle, ainsi que l'environnement. L'agneau peut alors se contaminer lors de la tétée ou par léchage. Le placenta peut être ingéré par une brebis voisine et ainsi la contaminer. L'infectiosité ne serait pas modifiée par le passage digestif et se retrouverait alors dans les fèces (Dickinson, 1976). La contamination de l'environnement (litière, pâturage) est donc possible à partir de ces matières fécales infectées ou par persistance du placenta lui-même, dans la zone d'agnelage.

1.6.6 Transmission entre élevages et persistance de l'infection

Les enquêtes postales menées en Grande Bretagne (Hoinville et al., 2000 ; Sivam et al., 2006 ; McLean et al., 1999, McIntyre et al., 2006) ou les études cas-témoins réalisées en Norvège (Hopp et al., 2001), Irlande (Healy et al., 2004) ou en France (Philippe et al., 2005) ont permis de mettre en évidence un certain nombre de facteurs de risque d'acquisition et de la persistance de la maladie à l'échelle de l'élevage. L'achat d'animaux (reproducteurs ou agneaux), en provenance d'élevages atteints ou de statut sanitaire inconnu et le partage de pâturages apparaissent comme les principaux facteurs de risque d'introduction de la tremblante au sein d'un troupeau. La maladie est plus fréquemment observée dans des troupeaux de grande taille. Parmi les facteurs susceptibles de favoriser la persistance de la maladie dans l'exploitation, le nombre de bâtiments, le nombre de moutons en bergerie l'hiver, le fait de ne pas séparer les mâles des femelles au cours de la période d'agnelage, le fait que les agnelages aient lieu en parc collectif plutôt qu'en cases individuelles ou qu'ils aient toujours lieu au même endroit, l'incorporation des placentas dans le fumier et la dissémination de ce dernier dans les pâturages, se sont révélés à risque. L'observation de la maladie clinique chez les animaux nés et élevés sur l'exploitation (cas secondaires) est plus fréquente lorsque la plupart des agnelles de renouvellement

sont issues de l'exploitation que quand elles sont achetées (McLean et al., 1999). L'achat d'aliments lactés s'est révélé à risque dans une étude cas-témoins menée en France (Philippe et al., 2005).

Une série d'observations réalisées en Islande suggère un rôle de l'environnement dans le maintien de l'infectiosité dans les troupeaux (Georgsson et al., 2006). Un programme d'éradication de la tremblante a été mis en oeuvre dans ce pays depuis 1978. Le pays a été divisé en zones de quarantaine, ce qui a permis de constituer des zones indemnes de tremblante, à partir desquelles les exploitations atteintes de tremblante ont pu être repeuplées. Le programme d'éradication, bâti avec les éleveurs, comprend des mesures draconiennes : le troupeau est euthanasié et les cadavres sont enfouis à 1.5 m de profondeur, avec plantation d'arbres sur les zones d'enfouissement ; le foin récolté sur l'exploitation pendant deux ans est éliminé ; des mesures strictes de désinfection, de dératisation et de lutte contre les acariens sont mises en oeuvre ; les bergeries impossibles à nettoyer correctement sont détruites et brûlées ; le sol autour des bâtiments est éliminé, et les alentours recouverts de terre nouvelle, de gravier ou d'asphalte. Après deux années de vide sanitaire, le repeuplement est autorisé avec des ovins issus de zones indemnes de tremblante. De 1978 à 1997, ce plan d'éradication de la tremblante a été appliqué dans 800 exploitations (145 000 ovins euthanasiés), 500 d'entre elles ayant reconstitué un troupeau ovin depuis. Parmi elles, la tremblante est réapparue dans 25 exploitations, soit dans 5% des cas. La réapparition de la maladie est intervenue le plus souvent deux à trois ans après le repeuplement du troupeau. La principale hypothèse pour expliquer la réapparition de la tremblante est le maintien de l'agent pathogène dans les exploitations, en dépit des mesures mises en oeuvre.

1.7 Difficultés des études épidémiologiques sur la tremblante

1.7.1 Définition de l'événement étudié

Une difficulté majeure dans l'étude de la tremblante est liée à la définition de l'événement étudié. En raison de la longue durée d'incubation de la tremblante, une dissociation temporelle importante existe entre le phénomène de contamination par l'agent infectieux et l'apparition des signes cliniques. On peut alors définir un modèle multi-état présenté dans la figure 1.2 où chaque état correspond à un état de santé. L'épidémiologiste peut être à la fois intéressé par les facteurs qui influencent la contamination des animaux (transition de l'état 1, "animal sain" à l'état 2, "animal contaminé", et par ceux qui modulent la durée d'incubation de la maladie chez les animaux infectés (transition de l'état 2, à l'état 3, "animal malade"). L'absence de diagnostic précoce de l'infection ne permet pas de connaître la proportion des animaux qui se contaminent ni la date à laquelle les animaux infectés se sont contaminés. Les possibilités d'étudier le délai entre l'exposition à l'agent infectieux et la contamination (transition 1-2) sont donc limitées. De même les états 3 et 4 ("mort") sont généralement confondus chez les animaux malades, car même si l'expression clinique de la maladie peut durer, dans les conditions naturelles, plusieurs semaines, le début de la phase symptomatique peut être frustré et difficilement détectable. On est alors généralement amené à étudier le "délai à la mort par tremblante" au lieu de la durée d'incubation, *stricto-sensu*. Chez les animaux nés dans un milieu contaminé, ce délai est équivalent à l'âge, tandis que pour les animaux exposés plus tardivement, le temps de présence dans le milieu contaminé peut être utilisé. Par simplicité, le terme de durée d'incubation sera souvent (abusivement) employé par la suite pour décrire ce délai.

L'étude des facteurs de risque de contamination peut s'affranchir de la notion

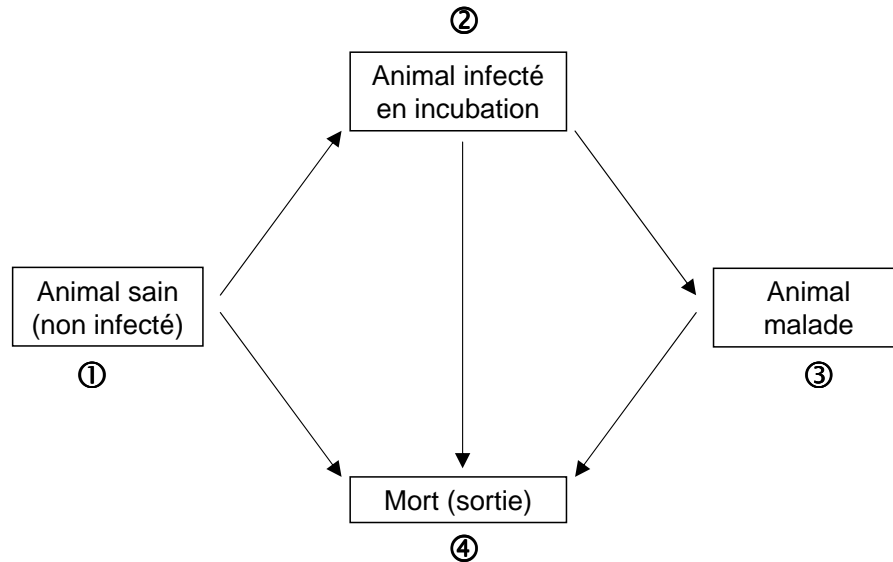


FIG. 1.2 – Modèle multi-états adapté à la tremblante.

de temps. Dans ce cas les animaux infectés (en incubation ou en phase clinique) seront comparés aux animaux non infectés et l'événement d'intérêt sera le fait d'être contaminé. Ce sont principalement ces facteurs qui intéressent les biologistes car ils ont une implication majeure en santé animale et humaine. Ces mesures d'association pourront se faire dans le cadre d'études transversales, d'études cas-témoins ou encore par une analyse en cas témoins dans une enquête de cohorte.

L'étude du "délai à la mort" s'intéresse intrinsèquement à la notion de temps et suppose une période de suivi. Pour cette raison elle n'est généralement possible que dans le cadre de suivis longitudinaux (enquête de cohorte). La population d'étude est alors restreinte aux seuls animaux contaminés et l'événement d'intérêt est la mort par tremblante.

La définition de l'événement étudié est donc d'une importance capitale dans l'étude de la tremblante, car elle conditionne à la fois le schéma d'étude et la population cible. Malheureusement, les caractéristiques même de la maladie (longue durée d'incubation), ou les difficultés liés à l'absence de diagnostic précoce, rendent souvent ces aspects difficilement dissociables et contrôlables (Hoinville96, Detwiler03).

1.7.2 Durée d'incubation et espérance de vie

Les longues durées d'incubation qui caractérisent les ESST nécessitent des durées d'observation étendues. En conditions expérimentales (animaux inoculés) ou dans certains élevages expérimentaux dédiés à l'étude de la tremblante (troupeau de Langlade, NPU), les animaux peuvent généralement être suivis pendant plusieurs années. A condition que les durées d'incubation ne dépassent pas l'espérance de vie "biologique" (Dickinson et al., 1975) et que les maladies intercurrentes ne conduisent pas à une mortalité précoce et importante, les durées d'observation permettent l'expression des signes cliniques de tremblante chez la majeure partie des animaux infectés.

En revanche, les études épidémiologiques menées dans les troupeaux commerciaux sont soumises à des contraintes zootechniques qui peuvent limiter la durée du suivi. Les durées d'observation sont variables selon les types d'élevages (élevages laitiers *vs* élevages allaitants) et au sein même d'un élevage les animaux ont des durées de vie économique très hétérogènes. Dans ces conditions, une partie seulement des animaux infectés vivent suffisamment longtemps pour pouvoir exprimer la maladie. Un des facteurs déterminants pour la pertinence de telles études est donc la durée de vie des animaux suivis, qui conduit à une sous-estimation d'autant plus importante du nombre d'animaux infectés qu'elle est réduite.

Une étude menée dans 15 troupeaux infectés en Grande Bretagne (Tongue et al., 2004) indique une grande variabilité de l'âge médian des animaux, entre 1 et 4 ans

selon les troupeaux. Moins de 16% des ovins présents dans ces élevages avaient 5 ans et plus. Pour une durée de survie médiane de 3 ans, les modélisations réalisées par Matthews et al. (2001) et Hagenaaers et al. (2006) concluent que 70% à 80% des animaux infectés seraient réformés avant l'apparition des signes cliniques.

1.7.3 Difficultés liées au diagnostic

Les connaissances de physiopathologie des ESST indiquent que, dans le processus de dissémination de la maladie dans l'organisme, l'atteinte du système nerveux central, et plus spécifiquement de l'obex, est relativement tardive. De très nombreuses études ont basé leurs résultats sur la mise en évidence de lésions de vacuolisation par histologie conventionnelle ou sur la recherche de PrPsc dans l'obex, par ELISA ou Western-Blot (Hunter et al., 1996; Acin et al., 2004, O'Doherty et al., 2002, Billinis et al., 2004). Les animaux en phase précoce de l'infection sont alors ignorés car non détectés. Une sous-estimation du nombre d'animaux infectés est donc probable (Hoinville, 1996; Hagenaaers et al., 2006). L'estimation de la prévalence intra-troupeau et des rapports de cotes (odds ratios) pourrait donc être biaisée. L'impact de ce biais de classification des animaux en phase préclinique est cependant difficile à apprécier mais pourrait s'avérer important. Ainsi une étude récente rapporte que l'analyse par immunohistochimie sur un grand nombre de structures nerveuses et lymphoïdes périphériques a permis de détecter 13 cas supplémentaires ($13/53 = 24.5\%$) par rapport aux résultats fournis par les tests rapides (ELISA TeSeE Biorad) réalisés sur l'obex (Reckzeh et al., 2007).

1.7.4 Schémas d'études et biais induits

Etudes transversales

La plupart des résultats disponibles sur les facteurs de risque individuels, principalement génétiques, liés à la tremblante ovine classique reposent sur des études transversales (Vascellari et al., 2005; Thorgeirsdottir et al., 1999 Tranulis et al.,

1999; O’Doherty et al., 2002; Billinis et al., 2004; Acin et al., 2004). Les enquêtes transversales portent sur les animaux de la population d’étude présents au moment de l’enquête et mesurent classiquement l’exposition et/ou l’état de santé à ce même moment. L’inclusion des animaux dans l’étude est conditionnée à leur présence au moment de l’enquête. Ces enquêtes ont la particularité de mesurer la prévalence de la maladie, c’est à dire le pourcentage d’animaux malades dans une population donnée, et non l’incidence. Le recueil des données pour ce type d’études a souvent lieu lors de l’abattage total ou partiel de troupeaux infectés et des prélèvements biologiques (généralement l’encéphale, et dans les études plus récentes, certains organes lymphoïdes) sont réalisés pour établir le diagnostic de tremblante. Cette approche, si elle a l’avantage d’être relativement aisée à mettre en place et de fournir un résultat dans un laps de temps restreint, présente des inconvénients pouvant être à l’origine de biais dans le cadre de l’étude d’une maladie à longue durée d’incubation telle que la tremblante.

Les difficultés liées au diagnostic, évoquées précédemment, conduisent à un **biais de classification** des animaux infectés pré-cliniques. Ce phénomène est d’autant plus dommageable que seul l’obex est prélevé. De plus, si la durée d’incubation (et donc la positivation de l’obex ou des organes lymphoïdes) est liée à certaines caractéristiques individuelles, par exemple génétiques, l’estimation de l’influence des facteurs de risque étudiés est alors biaisée, sans que le sens et l’importance de ce biais soient appréciables.

Un phénomène de **troncature à gauche**¹ est aussi rencontré dans ce type d’étude, pouvant conduire à un biais de sélection. En effet les animaux morts de tremblante avant la mise en place de l’étude ne sont pas observés. Seuls pourront être observés les animaux non-infectés et les animaux infectés (en phase préclinique avancée ou en phase clinique) toujours vivants lors de la réalisation des prélèvements. L’échantillon d’étude, constitué à partir des animaux les plus “résistants”, peut donc

¹une définition statistique stricte des termes de troncature et de censure est donnée au paragraphe 2.1.3

ne pas être représentatif des animaux infectés.

Enfin, l'antériorité de l'exposition aux facteurs de risque par rapport à la maladie peut être difficile à établir.

Ce type d'étude présente aussi l'inconvénient de ne pas permettre d'étudier l'influence des facteurs de risque sur la durée d'incubation de la maladie. Les observations des animaux en phase préclinique sont censurées¹ à droite, tandis que la troncature à gauche, évoquée précédemment, a sélectionné un échantillon d'animaux particuliers. Tout au plus peut-on décrire les données observées, sans pouvoir juger de leur pertinence et de leur représentativité.

Etudes cas témoins

Un autre schéma d'étude fréquemment employé est celui des études cas-témoins. Le principe de ce type d'enquête consiste à comparer la fréquence d'exposition antérieure à un facteur de risque dans un groupe de "cas", atteints de la maladie étudiée, et dans un groupe de "témoins" indemnes de la maladie. La mesure d'association utilisée est l'odds ratio. Dans la plupart de ces études, les "cas" sont représentés par des animaux en phase clinique de la maladie où chez lesquels la présence de PrPres a pu être mise en évidence dans différents organes. Les "témoins" sont des animaux sains, contemporains des cas, issus soit des élevages infectés, soit d'autres élevages, considérés comme non infectés. Ces enquêtes se révèlent souvent plus pertinentes pour la mesure d'association entre facteurs de risque et maladie que les études transversales, car elles sont moins sujettes aux biais de sélection. Elles sont de plus bien adaptées à l'étude de maladies rares pour lesquelles le délai depuis le début l'exposition est relativement long, et sont plus aisément et rapidement mises en place.

Leurs inconvénients majeurs sont liés aux biais possibles dus en particulier à la constitution du groupe des animaux témoins et au recueil rétrospectif des informations. Si ces derniers sont issus des élevages infectés étudiés, alors il peut être difficile de s'affranchir du biais de classification dû aux difficultés de diagnostic. L'apparie-

ment sur l'âge, ou sa prise en compte comme variable d'ajustement dans les modèles de régression, est parfois utilisé pour limiter ce biais, mais peut être rendu difficile par des effectifs limités. Les témoins peuvent aussi être choisis dans des élevages considérés comme sains. Dans ce cas le biais rencontré concerne l'exposition à la maladie. En effet, par définition, les animaux issus des élevages sains ne sont pas exposés à l'agent infectieux, et la mesure de l'association entre certains facteurs génétiques ou environnementaux est alors faussée.

Un autre inconvénient est que ce type d'enquêtes ne s'intéresse pas, par essence, à la fréquence de la maladie au sein d'un groupe donné. Les mesures de prévalence ou d'incidence ne sont généralement pas possibles. De même que précédemment, les enquêtes cas-témoins sont rarement adaptées à l'étude des durées d'incubation et des facteurs associés.

Etudes de cohorte

Les enquêtes de cohorte consistent à comparer la mortalité observée dans un ou plusieurs groupes d'animaux initialement indemnes de tremblante et définis en fonction de leurs expositions à un facteur de risque soupçonné. Elles nécessitent une période de suivi entre l'inclusion et l'observation de la maladie. Le terme de cohorte est utilisé pour désigner le ou les groupes d'animaux suivi(s) au cours du temps. Il peut s'agir d'une ou de plusieurs générations (animaux nés la même année), ou encore de tous les animaux présents à un temps donné dans un ou plusieurs élevages infectés. L'analyse des enquêtes de cohorte consiste essentiellement à comparer le taux d'incidence observé chez les sujets exposés à celui d'un groupe de référence, le rapport des deux taux étant le risque relatif associé à l'exposition.

Bien que généralement considérées comme les enquêtes les plus puissantes et les moins sujettes aux biais de sélection et de classification, les enquêtes de cohorte sont très peu employées dans le cadre de la tremblante classique. En effet, leur méthodologie est relativement lourde. La faible incidence de la tremblante au sein des troupeaux infectés exige de suivre des groupes d'animaux de grande taille. La

complétude du suivi sur plusieurs années nécessite des moyens importants. De plus la durée de suivi nécessaire à l'expression des signes cliniques chez les animaux infectés n'est parfois pas compatible avec les impératifs zootechniques. A ce jour très peu de travaux publiés ont fait appel à ce schéma d'étude. Les plus remarquables concernent le troupeau expérimental de Langlade appartenant à l'INRA, où un suivi longitudinal a été mis en place depuis 1993 (Elsen et al., 1999, Diaz et al., 2005).

En présence d'un diagnostic de la maladie fiable et précoce, les enquêtes de cohorte sont adaptées à l'exploration des facteurs de risque associés à la fois à la contamination et à ceux pouvant influencer la durée d'incubation. L'analyse des données peut être relativement aisée. Une analyse en cas témoins entre les animaux infectés et les animaux non infectés permet de répondre à la question du risque de contamination. Puis la sélection de la sous population des animaux détectés positifs permet l'étude des facteurs influençant la durée d'incubation.

Cependant, compte tenu de l'absence de diagnostic de certitude précoce, de la longue durée d'incubation de la tremblante et de l'existence de suivis individuels hétérogènes, l'analyse des données de ces études fait généralement appel aux modèles classiques d'analyse de données de survie. L'événement d'intérêt est alors la mort par tremblante clinique et la variable d'intérêt est l'âge à la mort par tremblante. Les observations des animaux cliniquement sains à la fin de leur période de suivi sont considérés comme censurés¹ à droite. Ce type d'analyses permet l'étude des facteurs de risque de mourir de tremblante clinique à un temps donné.

Cette approche statistique, bien que classique, présente cependant plusieurs limites conduisant à des biais. En effet le groupe des animaux cliniquement sains est composé des animaux non infectés, qui ne présenteront jamais les signes cliniques de la maladie, et les animaux infectés en phase d'incubation, qui, si le temps de suivi le permettait, mourraient un jour de la maladie. Malheureusement, une approche par un modèle de survie classique ne fait pas le distinguo entre ces deux groupes d'animaux, qui participent pour la même information dans le modèle.

L'analyse des données de survie fait l'hypothèse importante, mais parfois ignorée,

que tous les sujets suivis sont à risque de connaître un jour l'événement d'intérêt et le connaîtront un jour, si le temps de suivi est suffisamment long (potentiellement infini). Dans le cadre de l'étude de la tremblante, les résultats des études transversales montrent qu'une proportion non négligeable, voire majoritaire, des animaux ne s'infectent pas, même dans un milieu fortement contaminé (Hoinville, 1996 ; Jeffrey et al., 2002). Si l'événement d'intérêt est défini par la mort par tremblante, les animaux non infectés ne peuvent pas être considérés comme à risque de connaître cet événement. Tout au plus peut-on les considérer comme étant à risque de s'infecter. Si l'on est intéressé par l'estimation de la fonction de risque des animaux infectés, l'incorporation des animaux non infectés dans l'effectif d'étude conduit à sous-estimer le risque réel. Inversement, la survie des animaux infectés sera surestimée. Une analyse réalisée sur les seuls animaux observés infectés (cas cliniques et animaux testés positifs) conduit à ne sélectionner que ceux dont on a pu montrer qu'ils sont positifs. Par un phénomène de troncature¹ à droite, on est amené à la sélection d'un sous-ensemble des animaux réellement infectés, dont la représentativité est difficilement vérifiable. Ce phénomène est accentué si le suivi est basé sur la seule surveillance clinique.

L'utilisation des modèles de survie classiques s'avère théoriquement impropre à l'analyse des facteurs de risque de contamination, puisque seuls les animaux cliniquement atteints sont bien identifiés et considérés comme infectés, les animaux infectés en incubation étant confondus avec les animaux non-infectés.

1.8 Données groupées

Si la pertinence des enquêtes épidémiologiques dépend principalement du schéma d'étude adopté pour répondre à la question posée, leur puissance est fortement liée à l'effectif des animaux inclus. Les résultats obtenus à partir de grands échantillons auront généralement plus de poids que ceux tirés d'étude de taille limitée. Compte tenu de l'incidence souvent faible de la tremblante au sein des troupeaux infectés,

les études privilégient le suivi de grands troupeaux (Hunter et al., 1996 ; Elsen et al., 1999 ; Redman et al., 2002 ; Diaz et al., 2005) ou l'inclusion d'animaux appartenant à plusieurs élevages (Acin et al., 2004 ; O'Doherty et al., 2002 ; Billinis et al., 2004). Cette dernière approche pose le problème de la corrélation potentielle existant entre les animaux d'un même élevage. En effet, l'élevage est une entité relativement close où les conditions de vie (alimentation, reproduction, logement...) sont partagées par tous les animaux ou par certains groupes d'animaux (animaux nés la même année). Ces conditions de vie sont généralement propres à chaque élevage. L'information relative à ces différents facteurs est souvent limitée, et il peut subsister une corrélation des observations liée à un ensemble de facteurs non mesurés. La quantité d'infectiosité présente dans l'environnement, par exemple, n'est pas mesurable en pratique mais est commune à tous les animaux vivant dans le même élevage ou étant nés la même année dans cet élevage. Une première solution consiste à inclure l'élevage en tant que variable explicative, mais on se trouve face à un problème d'estimation d'un grand nombre de paramètres si le nombre d'élevages est élevé. Une autre solution est de tenir compte d'une corrélation résiduelle par l'inclusion dans le modèle d'un (ou de plusieurs) effets aléatoires.

1.9 Objectifs et plan

Nous avons vu que les études épidémiologiques de la tremblante sont rendues difficiles par l'existence de biais difficilement maîtrisables et que, même dans le cas de schémas d'étude *a priori* bien adaptés, la nature des données rend difficile leur exploitation statistique. L'exploration de nouvelles approches statistiques, adaptées aux données disponibles est donc nécessaire.

Idealement, les modèles utilisés doivent permettre une estimation séparée des risques de contamination et des durées d'incubation en tenant compte :

- du fait que même exposée à un milieu fortement infectant, une partie des animaux ne se contaminera pas ;

- de l’incertitude quant au statut sanitaire des animaux cliniquement sains, liée à l’absence de diagnostic précoce et fiable de l’infection ;
- de la présence de données incomplètes, liée en particulier à l’hétérogénéité des périodes de suivi des animaux étudiés à l’origine de données censurées à droite ;
- de l’existence de données groupées, dont l’unité de regroupement peut être définie par l’élevage auquel les animaux appartiennent, ou le groupe d’animaux nés la même année au sein de chaque élevage (millésime ou cohorte de naissance).

Le chapitre suivant s’attache à présenter un type de modèles d’analyse de survie, dont le nom anglais “cure models” ou “survival models with long term survivors”, peut être traduit par “modèles de survie prenant en compte une fraction non à risque”, ou plus simplement “modèles de survie avec fraction non à risque”. Nous y exposerons les différentes méthodes existantes et montrerons en quoi ce type d’approche est adapté à l’analyse des données épidémiologiques issues d’études de cohorte sur la tremblante. Dans le chapitre 3, nous présenterons un programme d’estimation pour ces modèles. Une méthode d’estimation des paramètres de régression et de la fonction de risque par vraisemblance pénalisée sera présentée dans le chapitre 4. Le chapitre 5 sera consacré à la prise en compte de données groupées, et des simulations seront présentées. Les chapitre 6 et 7 concerneront l’application de ces méthodes à l’étude des facteurs de risque de contamination par la tremblante classique du mouton et des facteurs influençant les durées d’incubation. Les données utilisées sont issues d’un suivi longitudinal rétrospectif et prospectif réalisé dans plusieurs troupeaux infectés des Pyrénées-Atlantiques (France). Une discussion des méthodes et des résultats, tant statistiques que biologiques, conclura ce rapport.

Chapitre 2

Modèles de survie avec fraction non à risque

2.1 Notations et définitions

De nombreux ouvrages sont consacrés à l'analyse des données de survie (Kalbfleisch et Prentice, 1980 ; Cox, 1984 ; Klein et Moeschberger, 1997). Dans ce chapitre nous rappellerons quelques définitions utiles à la suite de l'exposé.

2.1.1 Analyse des données de survie

L'analyse des données de survie est l'étude de la survenue, au cours du temps, d'un événement précis pour un ou plusieurs groupes d'individus donnés. Cet événement, souvent appelé décès, peut aussi bien être le décès d'un individu que la survenue d'une maladie, la réponse à un traitement ou la panne d'une machine. Chaque observation est définie par :

- **Une date d'origine**. Cela peut être la date de naissance du sujet, si l'on étudie l'âge du sujet lorsque survient l'événement ou la date de mise en contact avec un agent infectieux, si l'on l'étudie la durée d'incubation d'une maladie infectieuse. Chaque individu a une date d'origine différente sur le calendrier, mais la mesure qui nous intéresse est le délai depuis cette date. La date d'origine définit pour chaque individu le temps 0.
- Pour permettre la comparaison des durées de survie entre les individus, une définition précise de **l'événement d'intérêt** est nécessaire. S'il s'agit du décès provoqué par une maladie, il faut s'assurer que chaque décès est effectivement dû à la maladie étudiée, et non à des causes intercurrentes.
- La **durée de survie** est alors définie comme le délai entre la date d'origine et la survenue de l'événement d'intérêt.

Les durées de survie correspondent à des variables aléatoires positives, de distribution le plus souvent dissymétrique, rendant difficile leur description par les lois de distribution usuelles.

Les individus ou groupes d'individus sont susceptibles de différer pour un ou plusieurs facteurs. Ces facteurs, dénommés *variables explicatives* ou *covariables* peuvent

expliquer une différence importante de la durée de survie des sujets étudiés. Leurs effets sont analysés par des modèles de régression. Il peut s'agir de facteurs individuels (sexe, âge, paramètres biologiques relatifs à une maladie, paramètres génétiques..), ou liés à un essai thérapeutique (appartenance au groupe de traitement ou au groupe placebo, dosage médicamenteux...). L'analyse des données de survie s'attache alors à la description des temps de survie et à voir dans quelle mesure ils dépendent des ces variables explicatives.

2.1.2 Fonctions associées aux distributions de survie

Soit T la variable aléatoire positive correspondant à la durée de survie. La loi de probabilité de T peut être caractérisée par plusieurs fonctions liées entre elles.

La fonction de densité de probabilité, notée $f(t)$:

$$f(t) = \lim_{\Delta t \rightarrow 0^+} \frac{Pr(t \leq T < t + \Delta t)}{\Delta t}$$

$f(t)\Delta t$ est donc la probabilité de connaître l'événement d'intérêt entre t et $t + \Delta t$.

La fonction de répartition, notée $F(t)$:

$$F(t) = Pr(T \leq t) = \int_0^t f(u)du$$

$F(t)$ définit la probabilité de connaître l'événement d'intérêt entre 0 et t . Cette fonction est monotone et l'on a

$$F(0) = 0 \text{ et } \lim_{t \rightarrow \infty} F(t) = 1$$

La fonction de survie, notée $S(t)$:

$$S(t) = P(T > t) = 1 - F(t)$$

Cette fonction représente la probabilité de connaître l'événement d'intérêt au-delà du temps t . C'est une fonction monotone décroissante telle que

$$S(0) = 1 \text{ et } \lim_{t \rightarrow \infty} S(t) = 0$$

La fonction de risque, notée $\lambda(t)$:

$$\begin{aligned} \lambda(t) &= \lim_{\Delta t \rightarrow 0^+} \frac{\text{Pr}(t \leq T < t + \Delta t | T \geq t)}{\Delta t} \\ &= \frac{f(t)}{S(t)} \\ &= -\frac{d \log(S(t))}{dt} \end{aligned}$$

$\lambda(t)\Delta t$ représente, quand Δt est petit, la probabilité “approchée” pour un individu de connaître l’événement d’intérêt au temps t , conditionnellement au fait qu’il est encore à risque de le connaître (encore “vivant”) juste avant t . Cette fonction est aussi appelée *risque instantané de décès*.

La fonction de risque cumulé, notée $\Lambda(t)$:

$$\Lambda(t) = \int_0^t \lambda(u) du$$

Par manipulation des définitions précédentes, on retrouve facilement les relations suivantes :

$$\begin{aligned} f(t) &= -\frac{dS(t)}{dt} \\ S(t) &= \exp\left(-\int_0^t \lambda(u) du\right) \\ S(t) &= \exp(-\Lambda(t)) \\ f(t) &= \lambda(t) \exp\left(-\int_0^t \lambda(u) du\right) = \lambda(t)S(t) \end{aligned}$$

La distribution de la durée de survie T peut être décrite par l’une quelconque de ces fonctions. Toutefois l’une des plus intéressantes est la fonction de risque $\lambda(t)$ car elle est une description probabiliste du futur immédiat du sujet “encore à risque” et reflète des différences entre les modèles souvent moins visibles au travers des fonctions de répartition ou de survie. En épidémiologie, elle peut, dans certains cas, s’interpréter en termes d’incidence.

2.1.3 Censure et troncature

Une caractéristique des données de survie est qu'elles comportent le plus souvent des observations incomplètes en raison, en particulier, des phénomènes de censure et de troncature.

Censure à droite

Lorsque à la fin de la durée d'observation, un sujet n'a pas connu l'événement d'intérêt, son délai d'apparition de l'événement sera dit censuré à droite et sa durée d'observation constituera le délai de censure C . Si le délai de censure C est une variable aléatoire indépendante de la durée de survie T , alors la censure est considérée comme aléatoire. Une écriture courante pour représenter les données de survie censurées à droite est d'associer à chaque individu un couple de variables aléatoires (Y, δ) relié à la variable T par le modèle :

$$Y = \min(T, C)$$

$$\text{et } \delta = \begin{cases} 0 & \text{si } T > C \\ 1 & \text{si } T \leq C \end{cases}$$

où δ est un indicateur de censure à droite égal à 1 si la variable de survie T est observée (on parlera alors d'observation non censurée) et égal à 0 si la variable de censure C est observée (on parlera alors d'observation censurée à droite).

Censure à gauche

Une observation est dite censurée à gauche si l'on sait seulement que l'événement s'est produit avant une certaine date, sans qu'il soit possible d'en connaître la date exacte. On considère le couple de variables aléatoires (Z, δ) relié à la variable T par le modèle

$$Y = \max(T, C)$$

$$\text{et } \delta = \begin{cases} 0 & \text{si } T < C \\ 1 & \text{si } T \geq C \end{cases}$$

La censure à gauche est un cas symétrique de la censure à droite. On définit δ un indicateur de censure à gauche égal à 1 si la variable de survie T est observée et égal à 0 si la variable de censure C est observée (on parlera alors de d'observation censurée à gauche).

Censure par intervalles

On qualifera une observation de censurée par intervalles si au lieu d'observer la variable durée de survie T , on observe deux valeurs \mathcal{L} et \mathcal{R} , avec $\mathcal{L} < \mathcal{R}$, telles que la seule information dont on dispose sur T est que $\mathcal{L} < T < \mathcal{R}$. Encore une fois la distribution de T est supposée indépendante de \mathcal{L} et \mathcal{R} .

Troncature à gauche

Une observation est dite tronquée à gauche si elle n'est observable que conditionnellement à un autre événement. Si le temps de survie T est inférieur au temps d'entrée dans la cohorte \mathcal{L} , le sujet n'appartient pas à l'échantillon d'étude, c'est à dire que la durée de survie T n'est observable que conditionnellement au fait que $T > \mathcal{L}$. La variable \mathcal{L} , appelée variable de troncature à gauche est supposée indépendante de T .

Troncature à droite et troncature par intervalle

La définition de la troncature à droite est similaire à celle de la troncature à gauche : T est tronquée à droite si elle n'est observable qu'à la condition $T < \mathcal{R}$, la variable de troncature \mathcal{R} étant supposée indépendante de T . Le problème de troncature à droite se rencontre par exemple lorsque l'on étudie des registres. Les sujets figurent dans ces registres qu'à partir du moment où ils ont connus l'événement d'intérêt. Dans le cas de la tremblante du mouton, s'intéresser uniquement aux animaux ayant présenté des signes cliniques de la maladie conduit, nous l'avons vu, à l'exclusion de ceux qui sont en phase d'incubation.

Si l'observation de la variable durée de survie T est conditionnée par le fait que $\mathcal{L} < T < \mathcal{R}$, on dit que T est tronquée par intervalle. Les variables \mathcal{L} et \mathcal{R} sont des variables de troncature, supposées indépendantes de T . Il apparaît donc que la troncature à droite est un cas particulier de la troncature par intervalle avec $\mathcal{L} = 0$ et que la troncature à gauche est une troncature par intervalle avec $\mathcal{R} = +\infty$.

2.1.4 Implication dans l'analyse de données de survie en présence d'une fraction non à risque

Définissons l'événement d'intérêt comme étant le fait de mourir au stade clinique de tremblante. Dans ce qui précède nous avons vu que la fonction de survie de T est monotone décroissante telle que $\lim_{t \rightarrow \infty} S(t) = 0$. Cette propriété correspond à l'hypothèse, émise en introduction, que tous les animaux suivis sont à risque de connaître l'événement d'intérêt et le connaîtront un jour si le temps de suivi est suffisamment long. Il est évident que seuls les animaux infectés peuvent connaître cet événement (Jeffrey et al., 2002). Lorsque des animaux, même exposés pendant une longue période à l'agent infectieux, demeurent non infectés, on observe alors $\lim_{t \rightarrow \infty} S(t) > 0$ et la fonction de survie de la population est dite impropre. Ceci peut être observé sur la représentation graphique de l'estimateur de Kaplan-Meier de la fonction de survie (figure 2.1). Après un certain temps, la fonction de survie estimée reste constante, en plateau, à une valeur non nulle, en raison de l'absence de nouvel événement (mort par tremblante).

Pour que la condition $\lim_{t \rightarrow \infty} S(t) = 0$ soit vérifiée, cette fonction de survie doit s'appliquer aux seuls individus à risque de connaître l'événement étudié – les animaux infectés dans le cas de la tremblante. Nous avons donc besoin de deux définitions de la fonction de survie. Dans la suite de l'exposé, la fonction de survie attachée aux individus “à risque” ou **fonction de survie conditionnelle**, sera notée $S(t|U = 1)$, où U est une variable binaire telle que $U = 1$ si l'individu est à risque et $U = 0$ sinon. Elle correspond à la fonction de survie de T conditionnelle-

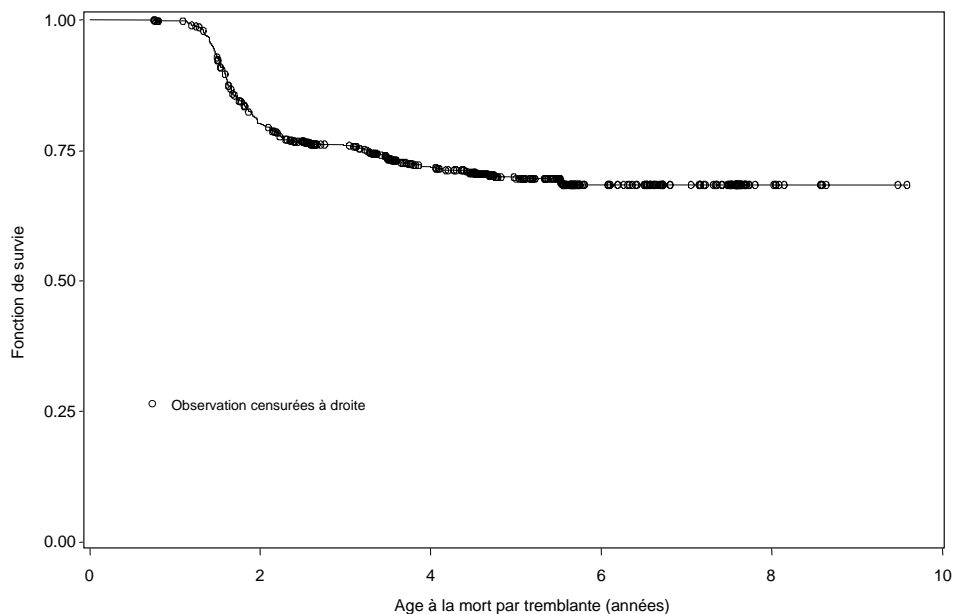


FIG. 2.1 – Exemple de fonction de survie marginale en plateau.

ment au fait d’être à risque. La fonction de survie de la population totale, incluant les individus à risque et non à risque, dite **fonction de survie marginale**, sera notée $S_{pop}(t)$. On définit de même les fonctions conditionnelles et marginales de risque, de densité de probabilité, de risque cumulé, et de répartition.

D’après les définitions précédentes, dans le cas d’un modèle de survie avec fraction non à risque, nous avons donc,

$$\lim_{t \rightarrow \infty} S_{pop}(t) > 0 \text{ et } \lim_{t \rightarrow \infty} S(t|U = 1) = 0$$

Une difficulté majeure dans l’approche qui nous intéresse est le fait que le statut des individus n’est pas complètement connu. De manière certaine, nous savons que les sujets qui connaissent l’événement d’intérêt appartiennent au groupe des individus à risque. Les sujets qui ne connaissent pas l’événement peuvent être soit des individus non à risque (il peut s’agir d’individus “guéris”, ou “non infectés” selon le type d’événement et de processus biologique étudié) ou des individus à risque (“non

guéris”, “infectés”) dont le temps de survie est censuré à droite avant l’événement. La proportion “réelle” d’individus à risque, que nous noterons p , n’est donc *a priori* pas connue. L’estimation de p , de $S(t|U = 1)$ et de l’influence de variables explicatives sur ces quantités ont fait l’objet de nombreux travaux, que nous présentons dans la suite de ce rapport.

2.2 Estimation de la fraction non à risque : approches non paramétriques

2.2.1 Estimateur non paramétrique de la fraction à risque

L’estimation non paramétrique de p a fait l’objet de développements théoriques détaillés, que l’on pourra retrouver dans Maller et Zhou (1996). Supposons que la fonction de distribution des temps de survie, $F_{pop}(t)$ soit continue et notons $t_1 < t_2 < \dots < t_n$ les temps de survie, potentiellement censurés à droite. Maller et Zhou (1995) montrent que, si et seulement si le temps de suivi est suffisamment long, alors

$$\hat{p}_n = \hat{F}_n(t_n), \quad (2.1)$$

est un estimateur consistant et asymptotiquement sans biais de la proportion d’individus à risque, p , où $\hat{F}_n(t)$ est l’estimateur de Kaplan-Meier de la fonction $F_{pop}(t)$ défini par

$$\hat{F}_n(t_j) = 1 - \prod_{j:t_j \leq t} \left(1 - \frac{d_j}{N_j}\right)$$

avec d_j , le nombre d’événement au temps t_j et N_j le nombre d’individus “à risque” (i.e. présent dans l’échantillon et n’ayant pas encore connu l’événement) juste avant t_j . De même, on peut montrer (Maller et Zhou, 1996) que

$$\begin{aligned} \hat{F}_n(t|U = 1) &= \hat{F}_n(t)/\hat{p}_n, & t \geq 0, \hat{p}_n < 1 \\ \hat{S}_n(t|U = 1) &= (\hat{S}_n(t) + \hat{p}_n - 1)/\hat{p}_n \end{aligned} \quad (2.2)$$

sont des estimateurs consistants et asymptotiquement sans biais, respectivement, de la fonction de répartition des temps de survie $F(t|U = 1)$ et de la fonction de survie $S(t|U = 1)$ des individus à risque.

Les propriétés évoquées précédemment sont établies asymptotiquement. Asymptotiquement, $\hat{F}_n(t)$ est un estimateur sans biais de $F_{pop}(t)$. Dans le cas d'une population finie, $\hat{F}_n(t)$ est un estimateur biaisé de $F_{pop}(t)$ et on peut montrer (Maller et Zhou, 1996) que le biais est donné par

$$B_n(t) = -E \left\{ \frac{I(t_n \leq t)(1 - \hat{F}_n(t_n))(F_{pop}(t) - F_{pop}(t_n))}{1 - F_{pop}(t_n)} \right\} \quad (2.3)$$

Il en découle que $\hat{F}_n(t) \leq F_{pop}(t)$ en tout t et par conséquent $\hat{p}_n \leq p$. Dans les conditions d'utilisation, sur des échantillons de taille souvent réduite, l'estimateur \hat{p} sous-estime donc la fraction à risque p .

2.2.2 Temps de suivi

Un point important dans cette approche est de définir ce que signifie la condition “si et seulement si le temps de suivi est suffisamment long”, condition nécessaire et suffisante à la vérification asymptotique de (2.1) et (2.2). Soit $C_{pop}(t)$ la fonction de distribution des temps de censure dans la population d'étude et, pour toute fonction de répartition $K(t)$, définissons τ_K tel que $\tau_K = \inf\{t \geq 0; K(t) = 1\}$. Sous l'hypothèse d'un phénomène de censure indépendant de l'événement d'intérêt, si $0 < p \leq 1$, la condition de “suivi suffisamment long” équivaut à

$$\tau_{F(t|U=1)} \leq \tau_C. \quad (2.4)$$

En d'autres termes, cette condition équivaut à ce que le dernier temps d'événement (possiblement censuré) des individus à risque $\tau_{F(t|U=1)}$ soit inférieur ou égal au dernier temps de censure de la population d'étude. Si cette condition n'est pas vérifiée, alors un certain nombre d'événements peuvent théoriquement exister au delà de la période de suivi maximum. Dans ce cas, il n'est pas possible de savoir quelle pro-

portion des données censurées à la fin de la période de suivi appartient au groupe à risque et l'estimation de p par $\hat{F}_n(t_n)$ est biaisée.

Maller et Zhou (1996) proposent de tester non paramétriquement l'hypothèse $H_0 : \tau_{F(t|U=1)} > \tau_C$. Le rejet de H_0 au seuil α conduit à accepter l'hypothèse $H_1 : \tau_{F(t|U=1)} \leq \tau_C$ qui garantit un temps de suivi suffisamment long. A l'inverse, ne pas rejeter H_0 conduit à conclure que le temps de suivi n'est pas suffisamment long pour écarter la présence d'une fraction non à risque ou que si une fraction non à risque existe, alors $\hat{p} = \hat{F}_n(t_n)$ est un estimateur biaisé de p . Ce test est basé sur la distance $t_n^* - t_n$ entre le dernier temps de survie non censuré, t_n^* , et le dernier temps de censure t_n . Les auteurs montrent que sous des conditions de régularité appropriées la statistique $\alpha_n = (1 - N_n/n)^n$, où N_n est le nombre d'observations non censurées dans l'intervalle $[2t_n^* - t_n, t_n^*]$, est une valeur approchée de la p-value du test de l'hypothèse H_0 . Il s'avère cependant que ce test est trop conservateur. Ils proposent alors d'utiliser les percentiles de la distribution de α (ou de manière équivalente de $q_n = N_n/n$) obtenus par simulations. Malheureusement, cette distribution dépend de $F_{pop}(t)$ et de $C_{pop}(t)$, qui ne sont pas connues, et ne peut qu'être approchée.

2.2.3 Tester la présence d'une fraction non à risque

Tester l'hypothèse de l'existence d'une fraction non à risque équivaut à tester l'hypothèse inverse $H_0 : F_{pop}(\tau_C) = 1$, où τ_C est le dernier temps de censure. Si H_0 est acceptée, alors on est conduit à rejeter l'existence d'une fraction non à risque. Sous H_0 , il est possible, avec une probabilité non nulle pour n fixé, que la dernière observation soit censurée, de sorte que $\hat{p}_n = \hat{F}_n(t_n) < 1$.

De façon pratique l'hypothèse H_0 n'est pas rejetée si $\hat{p}_n \geq c_\alpha$, où c_α est le α^{ieme} percentile de la distribution de \hat{p}_n calculé sous H_0 . Généralement on choisit $\alpha = 0.05$. De même que précédemment, H_0 ne peut être testé formellement, car la distribution de \hat{p}_n dépend de $F_{pop}(t)$ et $C_{pop}(t)$ elles mêmes inconnues, et on a là aussi recours à des résultats de simulations.

Plusieurs tests non paramétriques, dérivés du test du log-rank en analyse de survie “classique”, ont été proposés pour comparer les fractions non à risque estimées entre deux groupes, sous l’hypothèse d’homogénéité des distributions des temps de survie des individus à risque (Gray et Tsiatis, 1989) ou sans faire cette hypothèse (Laska et Meisner, 1992; Sposto et al., 1992). A l’inverse, Lee (1995) propose un test permettant de comparer les fonctions de survie des individus à risque de deux groupes, sous l’hypothèse que les fractions non à risque sont identiques.

2.3 Prise en compte d’une fraction non à risque : différents modèles

2.3.1 Approche par modèles de mélange

L’approche la plus intuitive, introduite par Boag (1949) puis par Berkson et Gage (1952), est basée sur l’idée qu’une population caractérisée par une fonction de survie impropre peut être envisagée comme un mélange de deux sous populations, l’une étant à risque de connaître l’événement d’intérêt, l’autre n’étant pas à risque. Soit U la variable aléatoire binaire, non observée, valant 1 si l’on appartient au groupe à risque et 0 sinon. La fonction de survie marginale impropre peut être représentée par

$$S_{pop}(t) = E \{ [S(t|U = 1)]^U \} = (1 - p) + pS(t|U = 1) \quad (2.5)$$

où U prend les valeurs 1 avec la probabilité p et 0 avec la probabilité $1 - p$, avec

$$1 - p = Pr(T = \infty) = Pr(U = 0) \quad (2.6)$$

et $S(t|U = 1)$ est la fonction de survie conditionnellement au fait d’appartenir à la sous-population à risque ($U = 1$), c’est à dire

$$S(t|U = 1) = Pr(T > t | T < \infty) \quad (2.7)$$

Les fonctions de densité de probabilité et de risque marginales sont alors données par

$$f_{pop}(t) = pf(t|U = 1) \quad (2.8)$$

$$\lambda_{pop}(t) = \frac{f_{pop}(t)}{S_{pop}(t)} = \frac{pf(t|U = 1)}{pS(t|U = 1) + 1 - p} \quad (2.9)$$

Notons que dans (2.5), p , qui représente la fraction à risque, est un paramètre inconnu que nous cherchons à estimer. Si $p = 1$, c'est à dire si tous les individus sont à risque, on retrouve le modèle de survie classique à partir des égalités (2.5) et (2.8).

2.3.2 Approche par modèle à risque cumulé borné ou modèle de promotion

Une autre approche consiste à considérer que la fonction de risque cumulé marginale $\Lambda_{pop}(t) = \int_0^t \lambda_{pop}(u)du$ admet une limite finie positive, notée θ , quand t tend vers ∞ . Dans ce cas, la fonction de survie marginale est donnée par

$$S_{pop}(t) = \exp\{-\theta F(t)\}, \quad \theta > 0, t \geq 0 \quad (2.10)$$

où $F(t) = \Lambda_{pop}/\theta$ est la fonction de répartition d'une variable aléatoire non négative, croissante et continue, telle que $F(0) = 0$ et $F(\infty) = 1$. Dans ce modèle, l'existence d'une limite θ à $\Lambda_{pop}(t)$ lui vaut le nom de modèle à risque cumulé borné (BCH = bounded cumulative hazard model), proposé par Tsodikov (1998). La fraction de la population considérée comme non à risque est donnée par

$$1 - p = \lim_{t \rightarrow \infty} S_{pop}(t) = \exp(-\theta) \quad (2.11)$$

De manière intéressante, cette approche peut aussi avoir une interprétation mécanistique. Proposée par Yakovlev (Yakovlev et al., 1993; Yakovlev et al., 1994) et développée dans un cadre Bayésien par Chen et al. (1999), elle découle d'une interprétation intuitive du phénomène de récurrence d'un cancer. Soit l'individu i ,

($i = 1, \dots, n$), et N_i le nombre de cellules cancéreuses persistant après un traitement chirurgical ou médicamenteux chez l'individu i , distribué selon un loi de Poisson $N_i \sim \mathcal{P}(\theta)$. Les individus non à risque sont ceux pour lesquels $N_i = 0$. Pour $N_i = k$, $k > 0$, soit X_{ij} , ($j = 1, \dots, k$) la variable aléatoire correspondant au temps nécessaire pour que la j^{ieme} cellule cancéreuse produise une métastase ou une résurgence locale visible. X_{ij} est aussi appelé temps de promotion ou temps de progression de la cellule j . Considérons X_{ij} , ($j = 1, \dots, k$) i.i.d. suivant une loi de répartition commune $F(X) = 1 - S(X)$ ne dépendant pas de N_i . Les N_i et X_{ij} ne sont pas observés mais on peut définir le temps de résurgence ou de métastase chez l'individu i par $T_i = \min(X_{ij}, j = 1, \dots, k)$.

La fonction de survie marginale est donnée, pour l'individu i par

$$\begin{aligned}
 S_{pop}(t_i) &= Pr(N_i = 0) + \sum_{k=1}^{\infty} Pr(T_i > t | N_i = k) Pr(N_i = k) \\
 &= Pr(N_i = 0) + \sum_{k=1}^{\infty} Pr(X_{i1} > t, \dots, X_{ik} > t | N_i = k) Pr(N_i = k) \\
 &= \exp(-\theta) + \sum_{k=1}^{\infty} S(t)^k \frac{\theta^k}{k!} \exp(\theta) \\
 &= \exp\{-\theta + \theta S(t)\} \\
 &= \exp\{-\theta F(t)\}
 \end{aligned} \tag{2.12}$$

où l'on retrouve l'expression (2.10). L'idée de "temps de promotion" des cellules cancéreuses vaut son nom à cette approche.

Les autres fonctions d'intérêt sont données par

$$f_{pop}(t) = \theta f(t) \exp\{-\theta F(t)\} \tag{2.13}$$

$$\lambda_{pop}(t) = \theta f(t) \tag{2.14}$$

où $f(t) = \frac{dF(t)}{dt}$ est la fonction de densité de probabilité associée à T .

2.3.3 Relations entre les modèles

Les modèles de mélange et de promotion précédemment décrits, s'ils ont une origine et une interprétation différentes, sont étroitement liés. En effet on a les relations

$$1 - p = e^{-\theta}, \quad S(t|U = 1) = Pr(T \geq t|N \geq 1) = \frac{e^{-\theta F(t)} - e^{-\theta}}{1 - e^{-\theta}} \quad (2.15)$$

$$\theta = -\ln(1 - p), \quad F(t) = \frac{\ln\{1 - p + pS(t|U = 1)\}}{\ln(1 - p)} \quad (2.16)$$

Notons que dans la première relation $S(t|U = 1)$ est fonction de θ . De même, dans la seconde, $F(t)$ est fonction de p .

Les relations suivantes découlent naturellement de (2.15)

$$f(t|U = 1) = \theta f(t) \frac{e^{-\theta F(t)}}{1 - e^{-\theta}} \quad (2.17)$$

$$\lambda(t|U = 1) = \frac{f(t|U = 1)}{S(t|U = 1)} = \theta f(t) \frac{e^{-\theta F(t)}}{e^{-\theta F(t)} - e^{-\theta}} \quad (2.18)$$

Ainsi, bien que les modèles de mélange et de promotion sont différemment spécifiés, ces relations sont utiles pour comparer les estimateurs des fractions à risque et des fonctions marginales et conditionnelles obtenus par les deux modèles.

2.4 Incorporation de variables explicatives

L'inclusion de variables explicatives permet d'explorer l'effet de facteurs de risque supposés sur la fraction à risque et sur le délai de survenue de l'événement d'intérêt chez les individus considérés à risque.

2.4.1 Modélisation de la fraction non à risque

Modèles de mélange

Dans les modèles de mélange (2.5) la paramétrisation de p est souvent réalisée par un modèle logistique tel que

$$\ln \left\{ \frac{p(\mathbf{z})}{1-p(\mathbf{z})} \right\} = \beta_0 + \beta_1 z_1 + \cdots + \beta_q z_q = \boldsymbol{\beta}' \mathbf{z} \quad (2.19)$$

où \mathbf{z} est le vecteur des variables explicatives et $\boldsymbol{\beta}$ est le vecteur des paramètres de régression associés à \mathbf{z} , contenant un intercept β_0 (Farewell, 1982; Taylor, 1995; Peng et Dear, 2000, Sy et Taylor, 2000). La fraction à risque dans le groupe de référence est donc donnée par $p_0 = \exp(\beta_0)/\{1 + \exp(\beta_0)\}$. D'autres fonctions de lien ont été proposées (Peng, 2003; Lam et al., 2005) telles que la fonction probit

$$\Phi^{-1}(p(\mathbf{z})) = \boldsymbol{\beta}' \mathbf{z} \quad (2.20)$$

où Φ est la fonction de distribution de la loi normale, ou encore la fonction complémentaire log-log

$$\log \{ -\log(1-p(\mathbf{z})) \} = \boldsymbol{\beta}' \mathbf{z}. \quad (2.21)$$

Modèles de promotion

Dans les modèles de promotion, la modélisation de la fraction à risque se fait au travers de θ . La paramétrisation la plus communément employée est $\theta(\mathbf{z}) = \exp(\boldsymbol{\beta}' \mathbf{z})$, où $\boldsymbol{\beta}$ contient un intercept Tsodikov (1998).

2.4.2 Modélisation des temps de survie

Outre la modélisation de la fraction à risque, la modélisation du délai de survenue de l'événement d'intérêt peut faire intervenir des variables explicatives. Dans le modèle de mélange, cette paramétrisation est attachée à la fonction de survie conditionnelle $S(t|U=1)$. Dans le cas des modèles à risque borné, c'est la fonction de survie des temps de promotion $S(t)$ qui est concernée (Broet et al., 2001; Tsodikov, 2002). Le modèle à risque proportionnels (PH) est le plus communément utilisé (Kuk et Chen, 1992; Peng et Dear, 2000; Sy et Taylor, 2000), mais des approches par des modèles de survie accélérée (AFT) (Yamaguchi, 1992; Li et Taylor, 2002;

Lambert et al., 2004 ; Zhang et Peng, 2007) ou des modèles additifs (Yin et Ibrahim, 2005) ont aussi été proposées.

La fonction de survie marginale s'exprime alors par

$$S_{pop}(t|\mathbf{z}, \mathbf{x}) = p(\mathbf{z})S_0(t|U = 1)^{\exp(\boldsymbol{\gamma}'\mathbf{x})} + 1 - p(\mathbf{z}) \quad (2.22)$$

dans le cas du modèle de mélange avec fraction non à risque, et

$$S_{pop}(t|\mathbf{z}, \mathbf{x}) = \exp \left\{ -\theta(\mathbf{z})[1 - S_0(t)^{\exp(\boldsymbol{\gamma}'\mathbf{x})}] \right\} \quad (2.23)$$

pour le modèle de promotion, avec \mathbf{x} le vecteur de variables explicatives, identique ou non à \mathbf{z} , et $\boldsymbol{\gamma}$ le vecteur des paramètres de régression associés. Dans les deux modèles, $S_0(\cdot)$ désignent les fonctions (conditionnelle) de survie de base.

Notons que le modèle de mélange (2.22), n'est pas un modèle à risque marginaux proportionnels, même si les risques conditionnels le sont. En revanche, par construction du modèle de promotion (voir égalité 2.14), si deux groupes ont la même fonction de répartition $F(t)$ ne dépendant pas de variables explicatives, alors les risques marginaux sont proportionnels. Ce n'est plus le cas si $F(t)$ est fonction de variables explicatives.

Si un modèle de survie accélérée est utilisé, la fonction de survie marginale est donnée par

$$S_{pop}(t|\mathbf{z}, \mathbf{x}) = p(\mathbf{z})S(\log(t) - \boldsymbol{\gamma}'\mathbf{x}|U = 1) + 1 - p(\mathbf{z}) \quad (2.24)$$

$$S_{pop}(t|\mathbf{z}, \mathbf{x}) = \exp \left\{ -\theta(\mathbf{z})(1 - S(\log(t) - \boldsymbol{\gamma}'\mathbf{x})) \right\} \quad (2.25)$$

où (2.24) s'applique au modèle de mélange et (2.25) au modèle de promotion.

2.4.3 Interprétation des paramètres de régression

Chacun des modèles est associé à deux vecteurs de paramètres $\boldsymbol{\gamma}$ et $\boldsymbol{\beta}$. Le vecteur $\boldsymbol{\beta}$ caractérise l'influence des variables explicatives \mathbf{z} sur la fraction à risque, tandis que le vecteur $\boldsymbol{\gamma}$ traduit l'influence des variables explicatives \mathbf{x} , potentiellement différentes de \mathbf{z} , sur le délai de survenue de l'événement chez les individus à

risque. Dans la littérature consacrée aux modèles de mélange, certains auteurs utilisent le terme “incidence” pour la modélisation de la fraction à risque et “latence” pour la modélisation de la survie conditionnelle (Taylor, 1995; Sy et Taylor, 2000; Peng et Dear, 2000; Zhang et Peng, 2007). En revanche, l’existence d’une fraction non à risque est qualifiée “d’effet à long terme” (“long term effect”) dans le cadre des modèles de promotion tandis que l’effet des variables explicatives sur la survie conditionnelle de l’événement est qualifiée “d’effet à court terme” (“short term effect”) (Broet et al., 2001; Tsodikov, 2002).

Incidence ou effet à long terme

Dans le modèle de mélange, si la fraction non à risque est modélisée par un modèle logistique, l’interprétation des coefficients de régression $\beta' = (\beta_0, \beta_1, \dots, \beta_l)$ est simple. Pour une variable binaire $z^{(k)}$, comme dans le modèle logistique classique, $\exp(\beta_k)$, $k = 1, \dots, l$, correspond à l’odds ratio (ou rapport de cotes) pour la variable $z^{(k)}$,

$$OR_{z^{(k)}} = \exp(\beta_k) = \frac{Pr(U = 1 | z^{(k)} = 1) / [1 - Pr(U = 1 | z^{(k)} = 1)]}{Pr(U = 1 | z^{(k)} = 0) / [1 - Pr(U = 1 | z^{(k)} = 0)]}, \quad k = 1, \dots, l.$$

Le coefficient β_0 (ou intercept) permet d’estimer p_0 , la fraction à risque dans le groupe de référence, pour lequel $z^{(k)} = 0, \forall k, k = 1, \dots, l$. En effet, $\hat{p}_0 = \exp(\hat{\beta}_0) / [1 + \exp(\hat{\beta}_0)]$.

L’interprétation des paramètres de régression du modèle de promotion n’est pas directe. Si $F(t)$ ne dépend pas de variables explicatives, alors d’après (2.14), on a, pour deux sujets ayant les caractéristiques z_1 et z_2 pour la variable $z^{(k)}$,

$$\frac{h_{pop}(t, z_1)}{h_{pop}(t, z_2)} = e^{(z_1 - z_2)\beta_k}$$

et $e^{(z_1 - z_2)\beta_i}$ peut être interprété comme le risque relatif marginal (au niveau de la population) de connaître l’événement au temps t . Cette relation n’est plus valide si $F(t)$ dépend de variables explicatives. De manière générale, si $\beta_i > 0$ la fraction à risque $p_1 = Pr(U = 1 | z^{(i)} = 1)$ sera plus grande que $p_2 = Pr(U = 1 | z^{(i)} = 0)$.

Latence ou effet à court terme

Dans le modèle de mélange, l'interprétation des paramètres de régression des variables explicatives γ ne présente généralement pas de difficultés. En revanche, cette interprétation est à nouveau difficile dans le cas du modèle de promotion. En effet, en ré-écrivant le modèle de promotion sous la forme d'un modèle de mélange en utilisant (2.18), le rapport des fonctions de risque conditionnel n'est plus constant et dépend du temps. Selon Broet et al (2001) et Tsodikov (2002), $F(t, \mathbf{x})$ peut s'interpréter comme l'effet dépendant du temps des variables explicatives \mathbf{x} sur la distribution des temps de survie des individus à risque.

2.5 Modèles de régression paramétriques et semi-paramétriques

Dans le cadre du modèle de mélange (2.5) nous nous intéressons à la distribution de T chez les individus à risque et dont la fonction de survie est donnée par $S(t|U = 1)$. Dans le cadre du modèle de promotion (2.10), la fonction de répartition des temps de "promotion", $F(t) = 1 - S(t) = \Lambda_{pop}(t)/\theta$ est la fonction d'intérêt.

Plusieurs approches ont été proposées. Elles peuvent être divisées en deux grands groupes : les approches paramétriques et les approches semi-paramétriques, qui définissent les modèles de survie avec fraction non à risque du même nom.

2.5.1 Modèles paramétriques

Lors d'une approche paramétrique, on fait l'hypothèse que la distribution des données de survie est une fonction connue qui est entièrement déterminée par un vecteur de paramètres ϕ de dimension finie. Ces approches permettent l'estimation de fonctions lisses et présentent la plus grande efficacité si les hypothèses de distribution sont vérifiées. Leur principal défaut est de faire des hypothèses fortes sur la distribution des temps de survie, qui peuvent s'avérer difficilement vérifiables en

pratique.

Dans le cadre des modèles de survie avec fraction non à risque de nombreuses fonctions ont été proposées, principalement pour les modèles de mélange. Citons l'utilisation de la distribution exponentielle (Jones et al., 1981; Goldman, 1984; Ghitany et al., 1995), de la distribution de Weibull (Farewell, 1982), ou lognormale (Boag, 1949; Gamel et al., 1990). Le vecteur de paramètres ϕ est de dimension 1 (loi exponentielle) ou 2 pour la fonction de risque de base. Ces fonctions, classiquement utilisées en analyse des données de survie ont la particularité d'être monotones décroissantes (si le paramètre de forme $\rho < 1$) et permettent de s'assurer que $\lim_{t \rightarrow +\infty} S(t|U = 1) = 0$. Nous décrivons brièvement ces fonctions simples dans le cadre du modèle de mélange. Ces fonctions s'appliquent aussi, pour le modèle de promotion, à la fonction $F(t) = 1 - S(t) = \Lambda_{pop}(t)/\theta$.

Loi exponentielle

La fonction conditionnelle de risque, est donnée par

$$\lambda(t|U = 1, \rho) = \rho$$

où $\phi = \rho$ est le paramètre de la loi exponentielle, définissant un risque constant pour tout t . Les fonction de densité de probabilité et de répartition sont données par

$$\begin{aligned} f(t|U = 1, \rho) &= \rho \exp(-\rho t) \\ F(t|U = 1, \rho) &= 1 - \exp(-\rho t) \end{aligned}$$

Une extension du modèle exponentiel est le modèle exponentiel par morceaux (Chen et al., 1999; Ibrahim et al., 2001). L'axe des temps est divisé en n intervalles et le risque est considéré comme constant au sein de chacun de ces intervalles, de sorte que $\phi = (\rho_1, \dots, \rho_n)$.

Loi de Weibull

C'est une généralisation de la loi exponentielle, dépendant de deux paramètres $\phi = (\rho, \kappa)$.

$$\begin{aligned}\lambda(t|U = 1, \rho, \kappa) &= \rho^\kappa \kappa t^{\kappa-1} \\ f(t|U = 1, \rho, \kappa) &= \rho^\kappa \kappa t^{\kappa-1} \exp\{-(\rho t)^\kappa\} \\ F(t|U = 1, \rho, \kappa) &= 1 - \exp\{-(\rho t)^\kappa\}\end{aligned}$$

Une des généralisations est la loi de Weibull translatée, dépendante de 3 paramètres $\phi = (\rho, \kappa, c)$, dont la densité de probabilité est :

$$f(t|U = 1, \rho, \kappa, c) = \rho^\kappa \kappa (t - c)^{\kappa-1} \exp\{-(\rho(t - c))^\kappa\} \quad (2.26)$$

Elle est particulièrement intéressante lorsque le risque est nul pendant une période assez longue, ce qui peut être observé lorsque t représente l'âge.

Loi de Gompertz

La distribution de Gompertz est un cas particulier (Cantor et Shuster, 1992 ; Gieser et al., 1998). Dans la reparamétrisation proposée par Cantor et Shuster (1992), la fonction de survie marginale prend la forme

$$S_{pop}(t) = \exp\left\{\frac{\alpha}{\beta}(1 - e^{\beta t})\right\}, \quad t \geq 0 \quad (2.27)$$

où $\alpha > 0$ et β peut être positif ou négatif. Si $\beta > 0$ alors $\lim_{t \rightarrow +\infty} S_{pop}(t) = 0$. A l'inverse pour $\beta < 0$ on observe $\lim_{t \rightarrow +\infty} S_{pop}(t) = \exp(\alpha/\beta) > 0$. Cette propriété suggère la possibilité d'utiliser la distribution de Gompertz dans le cas où la présence d'une fraction non à risque est suspectée, la valeur de β permettant d'en tester l'existence.

Autres approches paramétriques

Les lois Gamma Généralisée et F Généralisée sont des généralisations de la plupart des distributions communes. Leur utilisation a été proposée dans le cadre du

modèle de mélange (Yamaguchi, 1992 ; Peng et al., 1998) mais doit cependant faire face à des problèmes d'estimation des paramètres.

2.5.2 Modèles semi-paramétriques

Les lois paramétriques usuelles n'offrent pas suffisamment de diversité et de flexibilité pour s'adapter à toutes les situations. C'est pourquoi des approches non-paramétriques, basées sur la vraisemblance partielle de Cox (Cox, 1972) ou sur la vraisemblance marginale ont été proposées pour le modèle de mélange et le modèle de promotion.

Une approche non paramétrique a été proposée par Kuk et Chen (1992) puis par Taylor (1995) pour les modèles de mélange. Un estimateur non paramétrique des fonctions conditionnelles de base de risque cumulée $\Lambda_0(t|U = 1)$ ou de survie $S_0(t|U = 1)$ peut être obtenu par des approches comparables à celles utilisées pour le modèle de Cox classique (Breslow, 1972 ; Kalbfleisch et Prentice, 1980). Notons $\tau_1 < \dots < \tau_k$ les différents temps d'événement ordonnés, d_j le nombre d'événements au temps τ_j , D_j le sous échantillon des individus connaissant l'événement au temps τ_j et R_j le sous échantillon des individus à risque juste avant τ_j .

Un estimateur non paramétrique de la fonction conditionnelle de risque cumulé, basé sur la vraisemblance partielle est donné par :

$$\hat{\Lambda}_0(t|U = 1) = \sum_{j \in D_j} \frac{d_j}{\sum_{i \in R_j} g_i \exp(\hat{\gamma}' \mathbf{x}_i)} \quad (2.28)$$

où g_i est l'estimation de la probabilité d'être à risque pour l'individu i , détaillée dans le paragraphe (2.6.2)(Peng et Dear, 2000 ; Sy et Taylor, 2000). Cet estimateur est une modification de l'estimateur de Breslow (1972) pour le modèle de Cox classique, avec l'ajout d'une pondération du dénominateur par g_i . L'estimateur de $S_0(t|U = 1)$ est alors fourni par $\hat{S}(t|U = 1) = \exp\{-\hat{\Lambda}_0(t|U = 1) \exp(\hat{\gamma}' \mathbf{x})\}$. Notons qu'en l'absence de variables explicatives (ou pour $\gamma = 0$) et si tous les individus sont à risque ($g_i \equiv 1$), cet estimateur se réduit à l'estimateur de Nelson-Aalen.

Une autre approche est basée sur la vraisemblance marginale et utilise une généralisation de l'estimateur non paramétrique du maximum de vraisemblance de Kaplan-Meier. De manière analogue à Kalbfleisch et Prentice (1980), on peut montrer (Sy et Taylor, 2000) qu'un estimateur de la fonction conditionnelle de survie de base est fourni par

$$\hat{S}_0(t|U = 1) = \hat{S}_0(\tau_{j-1} + 0|U = 1) = \prod_{j:\tau_j < t} \hat{\alpha}_j \quad (2.29)$$

où $\hat{\alpha}_j$, $j = 1, \dots, k$ sont des paramètres non négatifs définis uniquement aux temps d'événement, avec $\alpha_0 = 1$, solutions de

$$\sum_{i \in D_j} \frac{\exp(\gamma' z_i)}{1 - \alpha_j^{\exp(\gamma' z_i)}} = \sum_{i \in R_j} g_i \exp(\gamma' z_i), \quad j = 1, \dots, k. \quad (2.30)$$

En l'absence d'ex aequo en τ_j , (2.30) a une solution analytique et la vraisemblance profilée $\tilde{L}(\hat{\gamma}, \hat{\alpha})$ peut être maximisée par rapport à α . Dans le cas contraire, les estimateurs du maximum de vraisemblance $\hat{\gamma}$ et $\hat{\alpha}$ sont obtenus conjointement par maximisation de la vraisemblance par un processus itératif de type Newton-Raphson.

De nouveau l'expression (2.30) est identique à celle obtenue dans le modèle de Cox classique, pondérée par g_i . Une approche semblable est adoptée par Tsodikov (1998) dans le cadre des modèles de promotion pour l'estimation de $F(t)$. L'estimateur non-paramétrique de la fonction de survie (ou de risque cumulé) est donc une fonction en escalier, continue à droite et constante entre deux événements.

Dans des conditions de censure raisonnables, et pour des raisons d'identifiabilité, il importe que $\hat{S}_0(t|U = 1)$ approche 0 pour $t > t_{\max}$, où t_{\max} est le dernier temps de suivi. Cependant dans certains cas, les estimateurs basés sur (2.28) ou (2.29) peuvent conduire à des valeurs très supérieures à 0 en t_{\max} . Une solution, proposée par Taylor (1995) et reprise par Sy et Taylor (2000) et Peng et Dear (2000) est d'imposer $S_0(t|U = 1) = 0$ pour tout $t \geq \tau_k$, où τ_k est le dernier temps d'événement observé. De manière identique Tsodikov (1998) impose $F(\tau_k) = 1$ dans le modèle de promotion. Cette contrainte, communément appelée "zero tail constraint" assure

une convergence plus rapide et l'identifiabilité du modèle (Sy et Taylor, 2000; Li et al., 2001; Tsodikov et al., 2003). Elle revient cependant à considérer que tous les individus censurés à droite après le dernier événement τ_k sont non à risque. Cette hypothèse, qui peut paraître forte, semble cependant acceptable quand le temps de suivi de la population est long par rapport au temps de survenue des événements (Taylor, 1995). Pour limiter cette contrainte, Peng et Carriere, 2002 proposent de compléter $S_0(t|U = 1)$ pour $t > \tau_k$ par un modèle paramétrique de type exponentiel ou de Weibull. Les paramètres de ces modèles sont estimés par la méthode du maximum de vraisemblance avec une contrainte de continuité en τ_k . Sur la base de simulations, ces approches (“*tail completion methods*”) fournissent des estimateurs moins biaisés de la fraction non à risque et de la fonction de survie conditionnelle dans les cas de censure précoce (Peng et Carriere, 2002). Une approche similaire peut être adoptée pour les modèles de promotion (Tsodikov, 2001).

2.6 Estimation des modèles de survie avec fraction non à risque

Les paramètres inconnus du modèle, que l'on cherche à estimer, sont $\Theta = (\beta, \gamma, \phi)$, où β est le vecteur des paramètres de régression traduisant l'effet des variables explicatives \mathbf{z} sur la fraction à risque, γ le vecteur des paramètres de régression traduisant l'effet des variables explicatives \mathbf{x} sur la distribution des temps de survie des individus à risque et ϕ le vecteur des paramètres définissant la fonction de distribution de base des temps de survie (ou de promotion).

2.6.1 Écriture de la vraisemblance

L'estimation de Θ peut être obtenue par la méthode du maximum de vraisemblance. La principale hypothèse est l'indépendance entre les temps de survie et les temps de censure. Quel que soit le modèle considéré, modèle de mélange ou de pro-

motion, l'écriture de la vraisemblance procède de la même manière. Supposons que les données soient sous la forme $(t_i, \delta_i, \mathbf{x}_i, \mathbf{z}_i), i = 1, \dots, n$, où δ_i est l'indicateur de censure tel que $\delta_i = 1$ si t_i est un temps d'événement et $\delta_i = 0$ si t_i est un temps de censure et n le nombre de sujets. La contribution à la vraisemblance complète de l'individu i est $L_i(\Theta) = f_{pop}(t_i)$ si $\delta_i = 1$ et $L_i(\Theta) = S_{pop}(t_i)$ si $\delta_i = 0$. Sous l'hypothèse d'indépendance des observations, la vraisemblance complète est donnée par

$$L(\Theta) = \prod_{i=1}^n L_i(\Theta_i)$$

Dans le cas du modèle de mélange avec fraction non à risque, la vraisemblance complète s'écrit donc

$$L(\Theta) = \prod_{i=1}^n \left\{ p_i(\mathbf{z}_i) f(t_i | U=1, \mathbf{x}_i) \right\}^{\delta_i} \times \left\{ (1 - p_i(\mathbf{z}_i)) + p_i(\mathbf{z}_i) S(t_i | U=1, \mathbf{x}_i) \right\}^{1-\delta_i} \quad (2.31)$$

En l'absence de fraction non à risque, i.e. $p(\mathbf{z}_i) = 1, \forall \mathbf{z}_i$, la vraisemblance (2.31) revient à la vraisemblance du modèle de survie classique.

Dans le cas du modèle de promotion, la vraisemblance prend la forme

$$L(\Theta) = \prod_{i=1}^n \left\{ \theta_i(\mathbf{z}_i) f(t_i, \mathbf{x}_i) \right\}^{\delta_i} \times \exp\{-\theta_i(\mathbf{z}_i) F(t_i, \mathbf{x}_i)\} \quad (2.32)$$

2.6.2 Estimation des paramètres

Le vecteur ϕ est de faible dimension dans les modèles paramétriques et la maximisation directe de $L(\Theta)$ peut se faire par un algorithme itératif de type Newton-Raphson ou Marquardt (Marquardt, 1963). En revanche, dans le cas des modèles semi-paramétriques ϕ est souvent de grande dimension (un paramètre par temps d'événement + 1) et rend difficile l'utilisation des algorithmes précédents, car ils requièrent le calcul des dérivées premières et secondes de la vraisemblance par rapport à chaque paramètre du modèle.

Approche semi-paramétrique : algorithme EM

L'utilisation de l'algorithme "Expectation-Maximization" a été proposée par Taylor (1995) puis Sy et Taylor (2000) et Peng et Dear (2000) pour l'estimation des paramètres Θ du modèle de mélange à risques conditionnels proportionnels.

Soit U_i la variable aléatoire telle que $U_i = 1$ si le sujet i est à risque et $U_i = 0$ sinon, $i = 1, \dots, n$. En supposant que la variable aléatoire U_i est observée, la vraisemblance du modèle de mélange avec fraction non à risque est la somme de deux parties indépendantes $l(\cdot) = l_I(\boldsymbol{\beta}; \mathbf{u}) + l_S(\boldsymbol{\gamma}, \Lambda_0; \mathbf{u})$ telles que :

$$l_I(\boldsymbol{\beta}; \mathbf{u}) = \log \prod_{i=1}^n p_i(\mathbf{z}_i)^{u_i} (1 - p_i(\mathbf{z}_i))^{1-u_i} \quad (2.33)$$

$$l_S(\boldsymbol{\gamma}, \Lambda_0; \mathbf{u}) = \log \prod_{i=1}^n \lambda(t|U=1, \mathbf{x}_i)^{\delta_i u_i} S(t|U=1, \mathbf{x}_i)^{u_i} \quad (2.34)$$

Cette décomposition est mise à profit pour estimer séparément (2.33) et (2.34). En effet $L_I(\cdot)$ correspond à la logvraisemblance d'un modèle de régression pour données binaires (modèle logistique par exemple) ayant $\mathbf{u} = (u_1, \dots, u_n)$ pour variable réponse et $l_S(\cdot)$ est la logvraisemblance d'un modèle de survie classique sur l'ensemble des sujets avec $u_i = 1$. L'algorithme EM permet d'estimer la probabilité $Pr(u_i = 1)$, sachant les variables explicatives et les temps de survie.

L'algorithme Maximisation-Expectation procède de la manière suivante :

Étape 1 : Attribution de valeurs initiales $\boldsymbol{\beta}^{(0)}$, $\boldsymbol{\gamma}^{(0)}$ et $S_0^{(0)}(t|U=1)$. L'obtention de $\boldsymbol{\beta}^{(0)}$ peut être réalisée en ajustant un modèle de régression pour données binaires aux données observées en fixant $u_i^{(0)} = \delta_i$. L'ajustement d'un modèle de survie aux données non censurées ($\delta_i = 1$) fournit $\boldsymbol{\gamma}^{(0)}$. L'estimateur $\hat{S}_0^{(0)}(t|U=1)$ est alors l'estimateur de Nelson-Aalen ou de Kaplan-Meier en l'absence de censure et fournit l'estimateur $\boldsymbol{\phi}^{(0)}$.

Étape 2 : Etape E (Espérance, Expectation) : à l'itération r cette étape consiste à

calculer l'espérance de \mathbf{u} sachant la valeurs des autres paramètres courants. Cette espérance est donnée par :

$$\begin{aligned} g_i^{(r)} &= E\{u_i | \boldsymbol{\beta}^{(r)}, \boldsymbol{\gamma}^{(r)}, S_0^{(r)}(t|U=1)\} \\ &= \delta_i + (1 - \delta_i) \frac{p_i^{(r)}(\mathbf{z}_i) S^{(r)}(t|U=1, \mathbf{x}_i)}{1 - p_i^{(r)}(\mathbf{z}_i) + p_i^{(r)}(\mathbf{z}_i) S^{(r)}(t|U=1, \mathbf{x}_i)} \end{aligned} \quad (2.35)$$

Étape 3 : Etape M (Maximisation) : cette étape consiste en la maximisation séparée de L_I et L_S en remplaçant u_i par son espérance $g_i^{(r)}$.

L'algorithme itère entre les étape 2 et 3 jusqu'à convergence des estimateurs et des vraisemblances.

Des variantes de l'algorithme EM sont proposées par Tsodikov et al. (2003) pour l'estimation des paramètres dans un modèle de promotion semi-paramétrique.

2.6.3 Estimation de la variance des paramètres

La matrice asymptotique de variance-covariance des estimateurs des paramètres $\boldsymbol{\beta}, \boldsymbol{\gamma}$ peut être estimée en inversant la matrice d'information de Fisher des dérivés secondes de la logvraisemblance de (T, λ) . L'utilisation de l'algorithme de Newton-Raphson fournit directement la matrice d'information. Dans le cas des modèles semiparamétriques estimés par l'algorithme EM, Sy and Taylor (2000) utilisent les approches de Louis (1982) pour estimer la matrice d'information observée basée sur la logvraisemblance complète. Cette matrice peut cependant être de grande dimension du fait des paramètres définis en chaque temps d'événement pour la fonction de survie et des problèmes numériques peuvent survenir lors de l'inversion de la matrice d'information (Tsodikov et al., 2003). Une approximation basée sur la vraisemblance profilée est proposée par les mêmes auteurs (Sy et Taylor, 2001) et fournit des résultats satisfaisants. D'autre part, dans le cadre du modèle de mélange logistique-Cox, ils montrent que dans des conditions de censure modérée et lorsque la fraction non à risque n'est pas trop faible, les variances des estimateurs $\hat{\boldsymbol{\beta}}$ et $\hat{\boldsymbol{\gamma}}$ peuvent être raisonnablement approchées en inversant séparément les matrices d'informa-

tion basées respectivement sur les logvraisemblances (2.33) et (2.34) en remplaçant les u_i par \hat{g}_i , les estimateurs du maximum de vraisemblance des g_i obtenus à convergence de l'algorithme EM par (2.35). Les covariances entre les paramètres $\hat{\beta}$ et $\hat{\gamma}$ ne sont cependant pas disponibles.

De manière plus simple, une estimation de la variance des paramètres par des méthodes de bootstrap (Efron et Tibshirani, 1993 ; Carpenter et Bithell, 2000) ou par imputation multiple est souvent utilisée (Peng et Dear, 2000 ; Peng, 2003 ; Lam et al., 2005 ; Zhang et Peng, 2007). Ces approches sont cependant coûteuses en temps de calcul.

2.6.4 Identifiabilité

De par leur construction, les modèles de mélange en analyse de survie doivent souvent faire face au problème d'identifiabilité des paramètres (Bohning et Seidel, 2003 ; Choi et Zhou, 2002). Ceci se traduit par l'existence d'une vraisemblance n'ayant pas une forme quadratique mais présentant plusieurs modes. Dans ce cas les algorithmes de maximisation itératifs tel que l'algorithme de Newton-Raphson peuvent s'avérer infructueux dans la recherche des estimateurs du maximum de vraisemblance (Cantor et Shuster, 1992). Plusieurs solutions peuvent être trouvées correspondant à des paramètres de régression totalement différents. Dans le cas du modèle de mélange avec fraction non à risque qui nous intéresse, une situation possible est celle où des paramètres correspondant à une importante fraction à risque ($p \simeq 1$) et à une fonction de survie conditionnelle impropre ($S(t|U = 0) \gg 0$ pour $t \geq \tau_k$, où τ_k est le dernier temps d'événement observé) aboutissent à une valeur de vraisemblance très proche à celle obtenue par des paramètres correspondant à une faible fraction à risque ($p \ll 1$) et à une fonction de survie conditionnelle approchant 0 pour $t \geq \tau_k$. Le modèle de promotion, peut aussi, dans une moindre mesure, poser des problèmes d'identifiabilité.

Li et al. (2001) ont étudié l'identifiabilité de plusieurs modèles de mélange avec

fraction non à risque et des modèles de promotion. Dans le cas du modèle de mélange avec fraction non à risque où la distribution des temps de survie est non spécifiée et ne dépend pas de variables explicatives, c'est à dire le modèle donnée par

$$S_{pop}(t, \mathbf{z}) = p(\mathbf{z})S(t|U = 1) + 1 - p(\mathbf{z}), \quad (2.36)$$

ils montrent que :

- le modèle (2.36) n'est pas identifiable si $p(\mathbf{z})$ n'est pas paramétriquement spécifié, sauf si $p(\mathbf{z}) = 1$, pour tout \mathbf{z} ou si $S(T^+|U = 1) = 0$, où T^+ est grand (potentiellement infini).
- le modèle (2.36) n'est pas identifiable si $p(\mathbf{z})$ est une constante p . Ce modèle correspond au modèle non-paramétrique décrit dans le paragraphe (2.2) dans le cas d'une seule population. Laska et Meisner (1992) montrent que l'estimateur du maximum de vraisemblance de $S_{pop}(t)$ est alors l'estimateur de Kaplan-Meier $\hat{S}_n(t)$, et qu'il existe plusieurs solutions possibles pour p et $S(t|U = 1)$ si le dernier temps de survie est un temps de censure.
- le modèle (2.36) est identifiable si $p(\mathbf{z})$ est paramétriquement spécifié (par une fonction logistique par exemple). C'est le modèle proposé par Taylor (1995).

Les auteurs montrent aussi l'identifiabilité des modèles où la fonction conditionnelle de survie prend une forme paramétrique (modèle de mélange paramétriques) et du modèle semi-paramétrique de mélange à risques proportionnels de Kuk et Chen (1992), Sy et Taylor (2000) et Peng et Dear (2000), où la fonction conditionnelle de survie dépend de variables explicatives \mathbf{x} selon un modèle à risque proportionnels.

L'analyse du modèle de promotion montre que le $S_{pop}(t, \mathbf{z}) = \exp[\theta(\mathbf{z})F(t)]$ n'est pas identifiable si :

- $\theta(\mathbf{z})$ et $F(t)$ sont laissés complètement non spécifiés.
- $F(t)$ est complètement non spécifié et $\theta(\mathbf{z}) = \exp(\gamma_0 + \boldsymbol{\gamma}'\mathbf{x})$ où $\gamma_0 =$ constante $\neq 0$.

En revanche si $F(t)$ est non spécifié et $\theta(\mathbf{z}) = \exp(\boldsymbol{\gamma}'\mathbf{x})$ ou si $F(t)$ prend une forme paramétrique, alors le modèle devient identifiable. Imposer $F(T^+|U = 1) = 1$ rend

aussi le modèle identifiable.

Les problèmes d'identifiabilité des paramètres peuvent être limités en choisissant des valeurs de départ obtenues à partir d'un modèle plus simple, comme cela est fait dans l'étape d'initialisation de l'algorithme EM. Ceci ne garantit cependant pas d'atteindre le maximum de vraisemblance global si la vraisemblance présente des modes (maxima locaux) importants.

2.7 Modèles de survie avec fraction non à risque et effets aléatoires

La prise en compte d'une fraction non à risque dans l'analyse de données de survie permet de tenir compte d'une certaine hétérogénéité de la population, vis à vis du risque de subir l'événement étudié. Une hétérogénéité supplémentaire peut provenir de l'omission de variables pertinentes, individuelles ou communes à un groupes de sujets, dans le modèle, soit parce qu'elles ne sont pas suspectées d'avoir une influence sur l'événement étudié soit parce qu'elles ne sont pas mesurées. Il peut s'agir de facteurs génétiques, environnementaux, géographiques ou alimentaires par exemple. Lorsque cette hétérogénéité non prise en compte provient d'un ensemble de facteurs individuels non mesurés, les observations sont bien indépendantes. Un modèle de survie avec fraction non à risque classique peut donc être utilisé pour analyser les données. Cependant, pour tenir compte de l'hétérogénéité des observations, l'introduction d'une variable de fragilité individuelle sous la forme d'un effet aléatoire sera préférable.

Lorsque l'on considère des données groupées, l'hétérogénéité entre groupes peut être expliquée par un ensemble de facteurs propres à chaque groupe. Cependant une corrélation des observations dans chaque groupe peut subsister, en raison de facteurs non observés. La prise en compte de cette corrélation résiduelle peut se faire par l'utilisation d'un modèle à fragilités partagées, en introduisant dans le modèle un

ou plusieurs effets aléatoires communs à tous les membres d'un même groupe.

2.7.1 Modèles de survie avec fraction non à risque et fragilité individuelle

Dans un modèle à fragilité individuelle simple (Vaupel et al., 1979), la fonction de risque pour l'individu i est donnée par

$$\lambda_i(t|Y_i, \mathbf{x}_i) = Y_i \lambda(t, \mathbf{x}_i) \quad (2.37)$$

où $\lambda_i(t|Y_i, \mathbf{x}_i)$ est la fonction de risque conditionnelle à l'effet aléatoire Y_i . Les variables aléatoires Y_i , non observées, sont propres à chaque individu et sont supposés identiquement et indépendamment distribués selon une fonction $F(Y)$. La variable de fragilité individuelle permet de tenir compte d'une hétérogénéité des risques : ainsi, des petites valeurs de Y_i sont associées à un risque faible de subir l'événement. À l'inverse, des grandes valeurs de Y_i caractérisent un risque élevé. Diverses fonctions de distribution $F(Y)$ ont été envisagées. Parmi les plus communes citons la loi Gamma, Gamma inverse, lognormale, ou positive stable (Hougaard, 1984; Hougaard, 1986). Cependant pour toutes ces lois, $Pr(Y = 0) \approx 0$ et aucune de ces distributions ne permet donc de tenir compte d'une fraction non à risque, définie par un risque nul (Price et Manatunga, 2001).

Le modèle de mélange peut être vu comme un cas particulier d'un modèle à fragilité individuelle (Sy et Taylor, 2000; Tsodikov et al., 2003). Notons Y la variable aléatoire, non observée, définissant l'appartenance d'un individu au groupe "à risque" ($Y = 1$) et "non à risque" ($Y = 0$). Conditionnellement à Y le risque pour l'individu i peut s'écrire $\lambda(t_i|Y_i, \mathbf{x}_i) = Y_i \lambda(t_i|Y_i = 1, \mathbf{x}_i)$, où les variables de fragilité Y_i , $i = 1, \dots, n$, suivent une loi de Bernoulli $\mathcal{B}(p)$. La variable Y est partiellement observée, puisque $Y = 1$ pour les sujets qui connaissent l'événement d'intérêt.

Cependant, dans le modèle de mélange avec fraction non à risque, conditionnellement à $Y = 1$, le risque est considéré comme identique pour tous les individus

de la strate définie par les variables explicatives \mathbf{x} et la prise en compte d'une hétérogénéité supplémentaire, non mesurée, parmi les individus à risque n'est donc pas possible. Cela le devient quand la variable de fragilité Y suit une distribution continue ayant une densité de probabilité non nulle en 0. Longini et Halloran (1996) proposent une distribution de Y telle que

$$\begin{aligned} Pr(Y = 0) &= p, & Pr(Y > 0) &= 1 - p \\ Y|Y > 0 \equiv X \sim f(), & E(X) &= 1, & \text{var}(X) = 1/\theta \end{aligned}$$

où $f(X) > 0$ pour tout X . Ces auteurs utilisent la fonction Gamma pour la distribution $f(X)$. La fonction Gamma inverse est aussi proposée par Price et Manatunga (2001).

Aalen (1988) propose une distribution composite de Poisson pour Y , faite d'un mélange de lois Gammas tel que :

$$Y = \begin{cases} X_1 + X_2 + \dots + X_N & \text{si } N > 0 \\ 0 & \text{si } N = 0 \end{cases}$$

avec $N \sim P(-\xi/\theta)$, $\theta < 0$ et X_1, X_2, \dots, X_N i.d.d. de fonction de densité $\text{Gamma}(\xi, -\theta)$. Cette fonction de distribution de Y a une probabilité non nulle en 0 et permet donc de tenir compte d'une fraction non à risque.

La fonction de survie marginale (ou observée) s'obtient en intégrant sur Y la fonction de survie conditionnelle à Y . Une formule générale de la fonction de survie marginale peut être obtenue par la transformation de Laplace. Dans le cas du modèle de mélange avec fragilité gamma telle que $X \sim \text{Gamma}(\theta, \theta)$ proposé par Longini et Halloran (1996) la fonction de survie marginale est donnée par

$$S_{pop}(t) = (1 - p) + p \left[1 + \frac{\Lambda(t|Y > 0)}{\theta} \right]^{-\theta} \quad \theta > 0 \quad (2.38)$$

Quand $\theta \rightarrow \infty$, X est dégénérée en 1 et le modèle (2.38) revient au modèle de mélange simple.

Pour le modèle de Aalen (1988) la fonction de survie marginale est

$$S_{pop}(t) = \exp \left\{ -\frac{\xi}{\theta} \left[\left(1 + \frac{\Lambda(t|Y > 0)}{\xi} \right)^\theta - 1 \right] \right\}, \quad \theta < 0, \xi > 0$$

Price et Manatunga (2001) ont comparé, sur la base du critère d'Akaike (Akaike, 1973), les modèles de mélange simples, modèles de mélange à fragilité Gamma et à fragilité Gaussienne inverse, ainsi que l'approche proposée par Aalen (1988). A partir de données sur des patients traités contre la leucémie, et dans un modèle sans variables explicatives, ils concluent à la supériorité des modèles avec fragilité. La prise en compte d'une hétérogénéité supplémentaire parmi les individus à risque, par l'intermédiaire d'une variable de fragilité ayant une probabilité non nulle en 0 peut donc, dans certains cas, s'avérer intéressante.

2.7.2 Modèle de survie avec fraction non à risque et données groupées

Dans le paragraphe précédent, la variable de fragilité Y , attachée à l'individu, permet de traiter des problèmes d'hétérogénéité dus à des variables explicatives individuelles non observées. Une autre utilisation de ces modèles à effets aléatoires est de modéliser la dépendance statistique (Clayton, 1978). Cette dépendance peut être présente dans l'étude de données groupées, répétées ou récurrentes. Nous nous intéresserons plus particulièrement au cas des données groupées, qui se pose dans notre application du fait du regroupement des animaux par élevages.

Les approches sont dérivées du modèle à fragilité partagée proposée par Clayton (1978) et sont appliquées au modèle de survie de mélange (Yau et Ng, 2001) et au modèle de promotion (Yin, 2005). La variable de fragilité Y est cette fois spécifique à chaque groupe i , ($i=1, \dots, G$) et partagée par tous les individus j , ($j = 1, \dots, n_i$) d'un même groupe.

Modèle de promotion à effet(s) aléatoire(s)

En reprenant l'interprétation mécanistique du modèle de promotion (2.12), on peut montrer (Yin, 2005) que la fonction de survie de la population, conditionnelle-

ment à Y , pour le sujet j du groupe i est donnée par

$$S_{pop}(t_{ij}|\mathbf{z}_{ij}, Y_i) = \exp[\theta(\mathbf{z}_{ij}) + \theta(\mathbf{z}_{ij}) \exp(-Y_i \Lambda(t_{ij}))] \quad (2.39)$$

La fonction de risque correspond à (2.39) est

$$\lambda_{pop}(t_{ij}|\mathbf{z}_{ij}, Y_i) = \theta(\mathbf{z}_{ij}) Y_i \lambda(t_{ij}) \exp(-Y_i \Lambda(t_{ij})) \quad (2.40)$$

où $\lambda(t) = d\Lambda(t)/d(t)$. En procédant de manière analogue au modèle à risque proportionnels et fragilité partagée, un autre modèle de promotion avec fragilité partagée peut être obtenu (Yin05). A partir de la fonction de risque de la population conditionnellement à Y_i , ce modèle devient

$$\begin{aligned} \lambda_{pop}(t_{ij}|\mathbf{z}_{ij}, Y_i) &= Y_i \lambda_{pop}(t_{ij}|\mathbf{z}_{ij}) \\ &= \theta(\mathbf{z}_{ij}) Y_i f(t_{ij}) \end{aligned} \quad (2.41)$$

Dans l'un ou l'autre des modèles (2.39) ou (2.40), les variables de fragilité Y_i , $i = 1, \dots, G$, sont distribuées selon une loi Gamma(η, η) tel que $E(Y_i) = 1$ et $\text{Var}(Y_i) = 1/\eta$. La dépendance de $\lambda(t)$ et $f(t)$ de variables explicatives dans (2.40) et (2.41) n'a pas été traitée. Notons de plus que si ces modèles tiennent compte d'une corrélation entre les temps d'événements des individus d'un même groupe à travers Y , cette corrélation n'est pas prise en compte en ce qui concerne la fraction non à risque $1 - p = \exp(-\theta(\mathbf{z}))$.

Modèle de mélange à effet(s) aléatoire(s)

L'utilisation d'un modèle de mélange permet, de manière intuitive, de prendre en compte facilement une corrélation intra-groupe à la fois en ce qui concerne la fraction à risque p et la distribution des temps de survie des individus à risque par l'introduction de deux effets aléatoires (Yau et Ng, 2001). En utilisant le modèle logistique, la fraction à risque p , conditionnellement à l'effet aléatoire V , s'exprime par

$$p(\mathbf{z}_{ij}|V_i) = \frac{\exp(\boldsymbol{\beta}' \mathbf{z}_{ij} + V_i)}{1 + \exp(\boldsymbol{\beta}' \mathbf{z}_{ij} + V_i)} \quad i = 1, \dots, n_i; j = 1, \dots, G$$

où $V_i, (i = 1, \dots, G)$ sont considérées i.i.d et de fonction de densité $\mathcal{N}(0, \sigma_V)$. Des valeurs élevées de V_i indiquent que les individus du groupe i auront une probabilité plus importante d'être à risque. De même, l'introduction d'un effet aléatoire W dans un modèle à risques proportionnels permet de tenir compte de la corrélation intra-groupe entre les temps d'événement :

$$\lambda(t_{ij}|U = 1, \mathbf{x}_{ij}, W_i) = \lambda_0(t_{ij}|U = 1) \exp(\boldsymbol{\gamma}' \mathbf{x}_{ij} + W_i) \quad i = 1, \dots, n_i; j = 1, \dots, G$$

où $W_i, (i = 1, \dots, G)$ sont considérées i.i.d et de fonction de densité $\mathcal{N}(0, \sigma_W)$. Dans le modèle proposé par Yau et Ng (2001) les effets aléatoires V et W sont considérés comme indépendants. Des valeurs élevées de W_i indiquent que les individus à risque du groupe i présentent un risque accru de connaître l'événement d'intérêt au temps t . Notons qu'en prenant $Y_i = \exp(W_i)$, on retrouve la notation communément associée aux modèles à fragilité partagée telle que

$$\lambda(t_{ij}|U = 1, \mathbf{x}_{ij}, Y_i) = Y_i \lambda_0(t_{ij}|U = 1) \exp(\boldsymbol{\gamma}' \mathbf{x}_{ij}).$$

Dans ce modèle, les valeurs de $\hat{\sigma}_V$ et $\hat{\sigma}_W$ permettent de tester l'existence d'une corrélation intra-groupe. Si σ_V et σ_W sont significativement différents de 0, alors il existe un effet groupe à la fois sur la probabilité d'être à risque et sur la survenue de l'événement d'intérêt chez les individus à risque. Si seul σ_V est significativement différent de 0, l'effet groupe a une influence sur la fraction à risque, mais pas sur la fonction de survie des individus à risque. Le raisonnement inverse s'applique à σ_W .

Yau et Ng (2001) adaptent les méthodes du BLUP (Best Linear Unbiased Predictor) et du maximum de la vraisemblance restreinte (REML) (McGilchrist, 1993 ; McGilchrist et Yau, 1995 ; Yau et McGilchrist, 1998) pour l'estimation des paramètres de régression et de σ_V et σ_W . La fonction de survie conditionnelle est spécifiée paramétriquement (Weibull) et les variances des estimateurs des paramètres de régression sont obtenues par bootstrap.

Approche Marginale

L'utilisation d'effets aléatoires nécessite de spécifier la structure de corrélation des effets aléatoires. Une approche marginale a été récemment proposée par Peng et al (2007) dans le cadre des modèles de mélange avec fraction non à risque. Cette approche est robuste à la mauvaise spécification de la structure de corrélation des données. Une extension de l'algorithme EM (Peng et Dear, 2000 ; Sy et Taylor, 2000) est utilisée, qui permet d'estimer la fonction de survie conditionnelle semi-paramétriquement. Un estimateur corrigé et "robuste" (dit estimateur "sandwich") de la matrice de variance-covariance des estimateurs des paramètres de régression permet d'ajuster la matrice de variance-covariance usuelles en tenant compte des possibles associations intra-groupes des événements.

2.8 Utilisation pratique des modèles de survie avec fraction non à risque

L'utilisation d'un modèle de survie prenant en compte la présence d'une fraction non à risque doit se faire avec précaution (Sposto et al., 1992 ; Farewell, 1986 ; Laska et Meisner, 1992 ; Yamaguchi, 1992 ; Maller et Zhou, 1996). En premier lieu il convient de vérifier les conditions d'utilisation de ce type de modèle. Les limitations majeures sont liées à la présence effective d'une fraction non à risque et à la durée d'observation des individus suivis. Lorsque la période d'observation est relativement réduite, l'utilisation d'un modèle de survie avec fraction non à risque n'est pas indiquée. L'hypothèse de l'existence d'une fraction non à risque est souvent fondée sur des aspects biologiques, cliniques ou épidémiologiques mais il convient de vérifier que le choix de ce type de modèle procure une adéquation aux données supérieure aux modèles de survie classiques.

2.8.1 Temps de suivi

Si le temps de suivi n'est pas suffisamment long, de sorte qu'une faible proportion seulement des individus à risque aura connu l'événement d'intérêt, l'estimateur paramétrique $\hat{F}_n(t_n)$ aura tendance à sous-estimer la fraction à risque p si le dernier temps de suivi t_n est un temps de censure. A l'inverse, si t_n est un temps d'événement, alors $\hat{p} = \hat{F}_n(t_n) = 1$ et on considérera qu'il n'y a pas de fraction non à risque (Laska et Meisner, 1992, Maller et Zhou, 1995). Les modèles paramétriques ou semi-paramétriques conduisent également à des résultats biaisés. A partir d'une étude par simulations, Yu et al. (2004) montrent que les estimateurs de la fraction non à risque et de la fonction de survie des individus à risque sont relativement peu biaisés dès lors que le temps de suivi est supérieur à la médiane des temps de décès des individus à risque. Ces résultats sont cependant obtenus en supposant que l'ensemble des individus sont suivis pendant un temps fixe. Ces conclusions sont sans doute trop optimistes en présence d'un phénomène de censure. Les simulations réalisées par différents auteurs (Sy et Taylor, 2000 ; Peng et Dear, 2000 ; Lam et al., 2005), tendent à montrer que les biais sont d'autant plus importants que la censure est précoce et importante.

2.8.2 Tester la présence d'une fraction non à risque

En premier lieu, l'existence d'une fraction non à risque doit être testée, ce qui revient à tester l'hypothèse $H_0 : p = 1$. L'approche non-paramétrique (chapitre 2.2.3), basée sur la distribution de $\hat{p}_n = \hat{F}_n(t_n)$ sous l'hypothèse nulle d'absence de fraction non à risque nécessite l'utilisation de simulations de la distribution de \hat{p} sous l'hypothèse H_0 , ce qui en limite l'utilisation pratique. Le rapport de vraisemblance permet de tester l'hypothèse nulle. La statistique de test est $LRS = -2[\ln(L(\hat{\theta}_{H_0})) - \ln(L(\hat{\theta}, \hat{p}))]$, où $L(\hat{\theta}_{H_0})$ est la vraisemblance du modèle sous l'hypothèse nulle (absence de fraction non à risque) c'est à dire le modèle de survie de survie classique. Le paramètre d'intérêt p se situe cependant sur la limite

de l'espace des paramètres, et les résultats asymptotiques classiques ne peuvent pas être utilisés. Dans le modèle de mélange exponentiel sans variables explicatives (p =constante et $S(t|U = 1) = \exp(\rho t)$), Maller et Zhou (1996) montrent que, sous H_0 , LRS suit asymptotiquement un mélange de distribution χ_1^2 et χ_0^2 avec des poids égaux à 0.5. Ces résultats sont aussi applicables lorsque $S(t|U = 1)$ est modélisée par une loi de Weibull, Gamma ou lognormale (Peng et al., 2001) et dépend de variables explicatives.

De tels tests d'hypothèse n'ont pas été développés dans le cadre des modèles semi-paramétriques, ni lorsque des variables explicatives sont ajoutées. Si p est modélisé par un modèle logistique, la valeur de l'estimateur de l'intercept β_0 et son intervalle de confiance peuvent être utilisés. Même si cela ne constitue pas un test formel, des grandes valeurs de $\hat{\beta}_0$, conduisant à des valeurs de \hat{p} proches de 1, sont en défaveur de la présence d'une fraction non à risque.

Chapitre 3

Programmation d'une Macro SAS

Dans ce chapitre nous présentons un programme d'estimation pour les modèles de survie de mélange avec fraction non à risque. Les logiciels de statistiques usuels (Splus, R, SAS, SPSS..) n'incluent pas, dans leurs formes initiales, de fonctions permettant l'analyse de telle données. A ce jour, une extension du programme Splus (en fait écrit en langage C) est proposée par Peng et al. (1998) pour les modèles de mélange paramétriques de la famille F généralisée (package GFCURE). Le package "SEMICURE" de Splus (Peng, 2003), permet d'estimer des modèles semi-paramétriques, mais ses possibilités sont relativement réduites. Le programme "CANSURV" (Yu et al., 2005) est destiné à l'analyse de données de survie relative, pour des données issues de registres. Des programmes BUGS illustrant les approches bayésiennes proposés par Ibrahim et al. (2001) sont disponibles sur le site internet de l'ouvrage, mais ne présentent pas de caractère général. Ce travail était aussi motivé par la nécessité de réaliser des études de simulation pour vérifier les conditions d'application des modèles de survie avec fraction non à risque à l'analyse des données sur la tremblante du mouton.

Ce travail a fait l'objet d'une publication dans *Computer Methods and Programs in Biomedicine*. La macro SAS est mise à disposition sur le site internet de l'ISPED (<http://www.isped.u-bordeaux2.fr/recherche/biostats/FR-biostats-accueil.htm> programmes). Depuis, nous avons programmé les modèles de promotion paramétriques ainsi que les modèles paramétriques et semi-paramétriques (constants par morceaux) avec effets aléatoires de distribution normale.

Résumé de l'article

Nous présentons une macro SAS permettant d'estimer des modèles de mélange paramétriques et/ou semiparamétriques pour données de survie avec fraction non à risque. L'estimation des modèles paramétriques fait appel à l'algorithme de Newton-Raphson de la procédure NLMIXED. Un algorithme EM, basé sur les travaux de Sy et Taylor (2000) et Peng et Dear (2000) est proposé pour l'estimation des modèles se-

miparamétriques. Les variances sont estimées par différentes méthodes de bootstrap pour les modèles semiparamétriques et par inversion de la matrice d'information pour les modèles paramétriques. L'utilisation du programme proposé est illustrée par l'analyse d'un jeu de données issu d'un essai clinique sur le cancer.



ELSEVIER

journal homepage: www.intl.elsevierhealth.com/journals/cmpb



A SAS macro for parametric and semiparametric mixture cure models

Fabien Corbière*, Pierre Joly

EMI E0338 Biostatistique, Institut de Santé Publique et Développement, Université Bordeaux 2, 146 rue Léo Saignat, 33076 Bordeaux Cedex, France

ARTICLE INFO

Article history:

Received 7 March 2006
 Received in revised form
 31 October 2006
 Accepted 31 October 2006

Keywords:

Survival analysis
 Cure models
 EM algorithm
 Parametric and semiparametric models
 SAS macro

ABSTRACT

Cure models have been developed to analyze failure time data with a cured fraction. For such data, standard survival models are usually not appropriate because they do not account for the possibility of cure. Mixture cure models assume that the studied population is a mixture of susceptible individuals, who may experience the event of interest, and non-susceptible individuals that will never experience it. The aim of this paper is to propose a SAS macro to estimate parametric and semiparametric mixture cure models with covariates. The cure fraction can be modelled by various binary regression models. Parametric and semiparametric models can be used to model the survival of uncured individuals. The maximization of the likelihood function is performed using SAS PROC NL MIXED for parametric models and through an EM algorithm for the Cox's proportional hazards mixture cure model. Indications and limitations of the proposed macro are discussed and an example in the field of cancer clinical trials is shown.

© 2006 Elsevier Ireland Ltd. All rights reserved.

1. Introduction

Models for survival analysis typically assume that everybody in the study population is susceptible to the event of interest and will eventually experience this event if the follow-up is sufficiently long. In recent years, there has been an increasing interest in modelling survival data with long term survivors. Such data may arise from clinical trials, in which, even after an extended follow-up, no further events of interest are observed. Some people in the population may be considered as cured or non-susceptible (cured). Failing to account for such cured subjects would lead to incorrect inferences. Moreover, researchers may be interested in estimating the cured fraction.

Mixture cure models assume that the studied population is a mixture of susceptible (uncured) individuals, that

may experience the event of interest, and non-susceptible (cured) individuals, that will never experience it [1]. This approach allows to estimate simultaneously whether the event of interest will occur, which is called incidence, and when it will occur, given that it can occur, which is called latency.

Let U be the indicator denoting an individual is susceptible ($U = 1$) or non-susceptible ($U = 0$) to the event of interest and T is a non-negative random variable denoting the failure time of interest, defined only when $U = 1$. The mixture cure model is given by

$$S(t|\mathbf{x}, \mathbf{z}) = \pi(\mathbf{z})S(t|U = 1, \mathbf{x}) + 1 - \pi(\mathbf{z}) \quad (1)$$

where $S(t|\mathbf{x}, \mathbf{z})$ is the unconditional survival function of T for the entire population, $S(t|U = 1, \mathbf{x}) = P(T > t|U = 1, \mathbf{x})$ the

* Corresponding author. Tel.: +33 5 57 57 45 79; fax: +33 5 56 24 00 81.

E-mail address: fabien.corbier@isped.u-bordeaux2.fr (F. Corbière).

0169-2607/\$ - see front matter © 2006 Elsevier Ireland Ltd. All rights reserved.

doi:10.1016/j.cmpb.2006.10.008

survival function for susceptible individuals given a covariate vector $\mathbf{x} = (x_1, \dots, x_p)'$, and $\pi(\mathbf{z}) = P(U = 1|\mathbf{z})$ is the probability of being susceptible given a covariate vector $\mathbf{z} = (z_1, \dots, z_q)'$, which may include the same covariates as \mathbf{x} . The survival function of cured patients can be set to one for all finite values of t because they will never experience the event of interest. Note that $S(t|\mathbf{x}, \mathbf{z}) \rightarrow 1 - \pi(\mathbf{z})$ as $t \rightarrow \infty$. When $\pi(\mathbf{z}_i) = 1$ for all \mathbf{z}_i , i.e. when there is no cured fraction, the mixture cure model reduces to the standard survival model. Various parametric and semiparametric specifications of $S(t|U = 1)$ have been proposed, leading to parametric and semiparametric mixture cure models [2].

In the next section, a brief description of parametric and semiparametric mixture cure models is presented as well as computational methodology. The macro and its requirements are described in Section 3 and a simulation study is shown in Section 4. In Section 5 an illustrative example is provided.

2. Computational methods and theory

2.1. Parametric and semiparametric mixture cure models

The effect of \mathbf{z} on the probability of $\pi(\mathbf{z})$ can be modelled by the use of binary regression models, with logit link

$$\text{logit}(\pi(\mathbf{z})) = \beta_0 + \beta_1 z_1 + \dots + \beta_q z_q = \boldsymbol{\beta}'\mathbf{z}$$

where β_0 is the intercept and $\boldsymbol{\beta}$ is the vector of regression parameters associated to \mathbf{z} . Other regression models include the probit link

$$\Phi^{-1}(\pi(\mathbf{z})) = \boldsymbol{\beta}'\mathbf{z}$$

where Φ is the distribution function of a standard normal distribution, and the complementary log-log link

$$\log(-\log(1 - \pi(\mathbf{z}))) = \boldsymbol{\beta}'\mathbf{z}$$

The conditional latency distribution $S(t|U = 1)$ can take the form of parametric or semiparametric distributions. Among the parametric models, exponential (EXP), Weibull (WB), log-normal (LN) and loglogistic (LG) are commonly used to model survival data. After reparameterization [3], these distributions can be expressed as

$$S(t|U = 1) = \begin{cases} \exp[-\exp(\log t - \mu)], & \text{exponential;} \\ \exp\left[-\exp\left(\frac{\log t - \mu}{\sigma}\right)\right], & \text{Weibull;} \\ 1 - \Phi\left(\frac{\log t - \mu}{\sigma}\right), & \text{lognormal;} \\ \left[1 + \exp\left(\frac{\log t - \mu}{\sigma}\right)\right]^{-1}, & \text{loglogistic;} \end{cases} \quad (2)$$

Covariates can be included by parameterizing μ such as $\mu = \boldsymbol{\gamma}'\mathbf{x}$, where $\boldsymbol{\gamma}$ represents the vector of unknown regression parameters. These models are also known as parametric accelerated failure time (AFT) mixture cure models [4]. Since \mathbf{x} acts multiplicatively on the scale parameter μ , it

accelerates or decelerates the failure time of susceptible individuals.

In proportional hazards (PH) models, the conditional distribution of T is modelled by

$$S(t|U = 1, \mathbf{x}) = S_0(t|U = 1)^{\exp(\boldsymbol{\gamma}'\mathbf{x})} = \exp\left(-\exp(\boldsymbol{\gamma}'\mathbf{x}) \int_0^t \lambda_0(v|U = 1) dv\right) \quad (3)$$

where $S_0(t|U = 1)$ and $\lambda_0(t|U = 1)$ are the baseline conditional survival and hazard functions, respectively. The conditional cumulative hazard function is $\Lambda(t|U = 1) = \Lambda_0(t|U = 1) \exp(\boldsymbol{\gamma}'\mathbf{x})$, where $\Lambda_0(t|U = 1) = \int_0^t \lambda_0(v|U = 1) dv$. If $S_0(t|U = 1)$ is left arbitrary, the model is defined as the Cox's proportional hazards mixture cure model [5]. Note that the Weibull (and exponential) models are both AFT and PH models.

Through the vectors of regression parameters $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$, parametric and semiparametric mixture cure models are able to separate the covariate effects on the incidence and the latency.

2.2. Likelihood

Suppose the data are of the form $(t_i, \delta_i, \mathbf{x}_i, \mathbf{z}_i)$, $i = 1, \dots, n$, where δ_i is the censoring indicator with $\delta_i = 1$ if t_i is uncensored and $\delta_i = 0$ otherwise. The likelihood contribution for individual i is $\pi_i(\mathbf{z}_i)f(t_i|U = 1, \mathbf{x}_i)$ for $\delta_i = 1$ and $(1 - \pi_i(\mathbf{z}_i)) + \pi_i(\mathbf{z}_i)S(t_i|U = 1, \mathbf{x}_i)$ for $\delta_i = 0$, where $f(\cdot) = S(\cdot)\lambda(\cdot)$ is the conditional probability density function of T . The observed full likelihood is then given by

$$L(\boldsymbol{\gamma}, \boldsymbol{\beta}) = \prod_{i=1}^n \{\pi_i(\mathbf{z}_i)f(t_i|U = 1, \mathbf{x}_i)\}^{\delta_i} \times \{(1 - \pi_i(\mathbf{z}_i)) + \pi_i(\mathbf{z}_i)S(t_i|U = 1, \mathbf{x}_i)\}^{1-\delta_i} \quad (4)$$

When no cured fraction is assumed, i.e. $\pi(\mathbf{z}_i) = 1$ for all \mathbf{z}_i , the likelihood function (4) reduces to the likelihood of the standard survival model.

2.3. Estimation procedures

This section presents a brief description of the procedures used to estimate the parameters by maximizing the likelihood (4). A more detailed presentation can be found in Peng and Dear [6] and Sy and Taylor [7].

2.3.1. Maximization of the likelihood function

To estimate $\boldsymbol{\gamma}$, we must specify the failure time distribution of uncured subjects. For parametric mixture cure models, $f(\cdot|U = 1)$ and $S(\cdot|U = 1)$ can be defined by a few unknown parameters in (4). Therefore, maximum likelihood estimates are obtained via usual optimization methods as the Newton-Raphson method (PROC NLMIXED [8]). Asymptotic standard errors are obtained by inverting the Fisher's information matrix of second order derivatives of $\log(L)$.

Unlike in the standard Cox's proportional hazard, where little information is lost by eliminating $S_0(t)$, one cannot eliminate $S_0(t|U = 1)$ in the Cox' PH mixture cure model without losing information about $\boldsymbol{\beta}$. The EM algorithm provides a simple and efficient way to estimate separately $\boldsymbol{\beta}$, $\boldsymbol{\gamma}$ and $S_0(t|U = 1, \mathbf{x}_i)$.

From the introduction, U is the random variable denoting an individual is susceptible ($U = 1$) or non-susceptible ($U = 0$). It follows that, if $\delta_i = 1$ then $u_i = 1$, and if $\delta_i = 0$ then u_i is not observed, where u_i is the value taken by the random variable U_i . Given the u_i 's, the complete-data full log-likelihood (4) is the sum of two independent components, l_1 , which depends only on β , and l_S , which depends only on γ and Λ_0 where

$$l_1(\beta; \mathbf{u}) = \log \prod_{i=1}^n \pi(\mathbf{z}_i)^{u_i} (1 - \pi(\mathbf{z}_i))^{1-u_i} \quad (5)$$

and

$$l_S(\gamma, \Lambda_0; \mathbf{u}) = \log \prod_{i=1}^n \lambda(t|U = 1, \mathbf{x}_i)^{\delta_i u_i} S(t|U = 1, \mathbf{x}_i)^{u_i} \quad (6)$$

with \mathbf{u} the vector of u_i values. The EM algorithm starts with initial values $\beta^{(0)}$, $\gamma^{(0)}$ and $S_0^{(0)}(t|U = 1)$. The E step in the (r)th iteration calculates the expectation of the complete log-likelihood function with respect to \mathbf{u} , conditional on the observed data and $\beta^{(r)}$, $\gamma^{(r)}$ and $S_0^{(r)}(t|U = 1)$, the estimates of β , γ and $S_0(t|U = 1)$ at the r th iteration. This is given by the following conditional expectation

$$\begin{aligned} y_i^{(r)} &= E\{u_i | \beta^{(r)}, \gamma^{(r)}, S_0^{(r)}(t|U = 1)\} \\ &= \delta_i + (1 - \delta_i) \frac{\pi^{(r)}(\mathbf{z}_i) S^{(r)}(t|U = 1, \mathbf{x}_i)}{1 - \pi^{(r)}(\mathbf{z}_i) + \pi^{(r)}(\mathbf{z}_i) S^{(r)}(t|U = 1, \mathbf{x}_i)} \end{aligned} \quad (7)$$

which is the r th estimator of the probability of the i th individual being susceptible. Given $y_i^{(r)}$ the M step in the ($r + 1$)th iteration maximizes the expected complete log-likelihood function with respect to β and γ to obtain $\beta^{(r+1)}$, $\gamma^{(r+1)}$ and $S_0^{(r+1)}(t|U = 1)$. The algorithm iterates until convergence on estimates of β , γ and $S_0(t|U = 1)$.

Peng and Dear [6] and Sy and Taylor [7] proposed a Cox's partial-likelihood-type method to estimate γ semiparametrically without specifying $\lambda_0(t|U = 1)$. For this method (6) is approximated by

$$\sum_{j=1}^k \left[\gamma' s_{(j)} - d_j \log \left(\sum_{i \in R_j} u_i \exp(\gamma' \mathbf{x}_i) \right) \right] \quad (8)$$

where k is the number of distinct uncensored failures times, d_j the number of uncensored observations at t_j , $s_{(j)}$ the sum of covariate vectors associated with the uncensored observations at t_j and R_j is the risk set at t_j .

Note that (5) is the log-likelihood function of a binary regression model, and that (6) is similarly the log-likelihood function of the standard Cox's PH model, with the addition of the offset variable $\log(u_i)$. Therefore, the M step of the EM algorithm is equivalent to the separated maximization of l_1 and l_S with standard regression models for binary variables (PROC LOGISTIC) and failure time data (PROC PHREG), respectively.

2.3.2. Estimation of the conditional baseline survival function in semiparametric models

The estimation of $S_0(t|U = 1)$ is of concern here, because it is needed in (7). Two non-parametric methods are discussed in Peng and Dear [6] and Sy and Taylor [7]. The first one is based on a profile likelihood estimate of $\Lambda_0(t|U = 1)$, similar to the Breslow's likelihood for the standard Cox's PH model [9]. The

second one is derived from the product limit estimator (PLE) after Kalbfleisch and Prentice [10].

In order to obtain a good estimation for γ and β it is important for $\hat{S}_0(t|U = 1)$ to approach 0 as $t \rightarrow \infty$. However, the estimates from the Breslow or the PLE methods do not approach zero as $t \rightarrow \infty$ when there are censored survival times after t_k , where t_k is the last observed failure time. Setting $\hat{S}_0(t|U = 1) = 0$ for all $t \geq t_k$ allows for a proper distribution function for susceptible individuals and avoid identifiability problems [11,12]. However, this zero-tail constraint implies that individuals with survival times greater than t_k are all considered as non-susceptible or cured, which may appear to be a strong assumption. Peng [13] proposed that $\hat{S}_0(t|U = 1)$ decreases from $\hat{S}_0(t_k + 0|U = 1)$ to zero smoothly for all $t > t_k$ and considered the exponential and Weibull distribution functions to complete the tail of the conditional baseline survival function. For the exponential distribution function $\hat{S}_0(t|U = 1) = \exp(-\hat{\zeta}t)$ for $t > t_k$, where $\hat{\zeta}$ satisfies $\exp(-\hat{\zeta}t) = \hat{S}_0(t_k + 0|U = 1)$. For the Weibull distribution function, $\hat{S}_0(t|U = 1) = \exp(-(\hat{\zeta}t)^{\hat{\rho}})$ for $t > t_k$, where $\hat{\zeta}$ and $\hat{\rho}$ are the maximum likelihood estimates based on all observations.

2.3.3. Variance estimation in semiparametric models

The standard errors of estimated parameters are not directly available, because of the EM algorithm. The variance of $\hat{\gamma}$ is particularly difficult to compute, because it involves $S_0(t|U = 1)$ through u_i . Multiple imputation and bootstrap methods have been proposed to estimate the observed information matrix of $\hat{\gamma}$ [6,14]. However, simulations studies (not shown here) indicated that the variance estimated by inverting the Fisher's information matrix of second derivatives when convergence in parameter estimates and likelihood is attained may perform quite well. Non-parametric bootstrap methods [15,16] are implemented in the proposed macro. Resampling from the original dataset is performed through PROC MULTTEST with the BOOTSTRAP option [17]. Computation of bootstrap confidence intervals follows the same conventions of %JACKBOOT macro (<http://support.sas.com/ctx/samples/index.jsp?sid=479>).

3. Program description

3.1. Required parameters

The SAS macro, called PSPMCM, has the basic function of fitting parametric or semiparametric mixture cure models for individual data. The parameters and their description are summarized in Table 1.

The dataset (DATA) is assumed to be entered with one record per individual, with the failure/censoring time (TIME), the censoring indicator (CENSCOD), and covariate vector (VAR). For categorical variables and second or higher order interactions, dummy variables have to be created in a previous data step. In the VAR statement, the name of covariates are separated by blanks. Each variable is followed, into brackets, by the option I, S or IS, which indicates whether it is included as a covariate in the incidence part only (I), the survival part only (S), or both of them (IS). In addition, when plots of survival functions are requested (see below), the values at which the

Table 1 – Parameters required by the SAS macro PSPMCM

Parameter	Description
DATA	SAS data set name to fit the mixture cure model
ID	Subject identification
CENSCOD	Censoring status (1 = event, 0 = censored)
TIME	Failure/censoring time
VAR	E.g.: GENDER(I S,0) AGE(S,10). List of covariates with indicator for incidence (I) or survival (S) and value for survival functions plots
LINK	Link function for incidence regression model
MODEL	Conditional survival function specification
ALPHA	Significance level for confidence interval
BASELINE	Y for output of the conditional baseline survival function
SPLOT	Y for plots of the conditional and marginal survival functions
PLOTFIT	Y for goodness of fit measures
SUOMET	Breslow's or PL method to compute of $\hat{S}_0(t U = 1)$
TAIL	Tail completion method (ZERO, ETAIL, WTAIL, NONE)
MAXITER	Maximum number of iterations for the EM algorithm
CONVCRT	Convergence criterium
FAST	Y for output at convergence, N otherwise
BOOTSTRAP	Y for bootstrap methods computation, N otherwise
NSAMPLE	Number of replicates for bootstrap computation
STRATA	Name of the variable for stratified bootstrap resampling
BOOTMET	Type of bootstrap confidence intervals
GESTIMATE	Y for bar charts and Q-Q plots of parameter estimates

survival estimates are plotted are specified after a comma, as shown in the example given in Table 1. Other required statements to specify the model are:

INCPART: to model the incidence part of the model, the user can choose between binary regression models with the logit (LOGIT), probit (PROBIT) or complementary log log (CLOGLOG) links. By default the logistic regression model (LOGIT) is assumed.

SURVPART: specifies the form of the conditional baseline survival function. Parametric models including exponential (EXP), weibull (WEIB), loglogistic (LLOG) and lognormal (LOGN), and the semiparametric Cox's proportional hazards model (COX) are available.

ALPHA: sets the significance level used for the confidence limits for the hazard ratios and odd's ratios. The value must be between 0 and 1. The default value is 0.05, which results in the calculation of a 95% confidence interval.

BASELINE: when set to Y, indicates that the conditional baseline survival function estimate $\hat{S}_0(t|U = 1)$, and parameter estimates $\hat{\beta}$, $\hat{\gamma}$ and $\hat{\mu}$, $\hat{\sigma}$ (for parametric models) are written to the BASELINE dataset. If bootstrap resampling is requested for the Cox mixture cure model (see below), the BASELINE.T dataset will moreover contain the estimates for all bootstrap replicates. The default value is N (no output).

SPLOT: when set to Y, the estimated conditional survival curve $\hat{S}_0(t|U = 1, \mathbf{x}_i, \mathbf{z}_i)$ and marginal survival curve $\hat{S}_0(t|\mathbf{x}_i, \mathbf{z}_i)$ for individuals with covariate vectors $\mathbf{x}_i, \mathbf{z}_i$ are plotted. The values of \mathbf{x}_i and \mathbf{z}_i at which the survival function estimates are computed are specified for each covariate after a comma in the VAR statement. The default value is N.

PLOTFIT: when set to Y, the macro computes for each stratum defined by the covariate vectors \mathbf{x}_i and \mathbf{z}_i the observed

(empirical) marginal survival curve $S^{(obs)}(t|\mathbf{x}_i, \mathbf{z}_i)$ which is the Kaplan–Meier estimate for the whole stratum (including right censored subjects). The plots of $(S^{(obs)}(t|\mathbf{x}_i, \mathbf{z}_i), \hat{S}(t|\mathbf{x}_i, \mathbf{z}_i))$ versus t is a visual tool to examine the goodness of the model prediction. The correlation coefficient between $S^{(obs)}(t|\mathbf{x}_i, \mathbf{z}_i)$ and $\hat{S}(t|\mathbf{x}_i, \mathbf{z}_i)$ is also computed for each stratum and provides an appropriate measure of the goodness of fit [18]. P–P plots of the KM estimates versus the fitted values are also plotted. The default value is N.

When the Cox PH mixture cure model is requested, additional options are available:

SUOMET: indicates whether the Breslow-type method (CH) or the product limit estimator (PL) is utilized to estimate the conditional baseline survival function. The default value is PL.

TAIL: indicates whether a constraint or a tail completion method is used to estimate $\hat{S}_0(t|U = 1)$. The option TAIL = ZERO specifies that the zero tail constraint is utilized (i.e. $\hat{S}_0(t|U = 1) = 0$ for $t > t_k$). ETAIL or WTAIL specify that the exponential and Weibull tail completion methods are used, respectively. NONE indicates that no tail constraint is used, but identifiability and convergence problems may rise with this option. The default value is ZERO.

MAXITER: is the maximum number of iterations to perform. If convergence is not attained the displayed output and all output data sets created by the procedure contain results that are based on the last maximum likelihood iteration. The default value is MAXITER = 200.

CONVCRT: sets the convergence criterion. The default value is 10^{-5} . The iterations are considered to have converged when the maximum relative change in the parameters

Table 2 – Simulation results on biases and MSE of regression parameters

Parameter	Method	True values of $(\beta_0, \beta_1, \gamma_1)$			
		(0,0,0)		(1.3863, -1, -1)	
		Biases	MSE	Biases	MSE
$\hat{\beta}_0$	Weibull mixture	0.0034	0.0194	0.0175	0.0345
	Cox's PH mixture	-0.0019	0.0193	0.0157	0.0337
$\hat{\beta}_1$	Weibull mixture	-0.0216	0.0419	0.0296	0.0628
	Cox' PH mixture	-0.0227	0.0412	0.0301	0.0610
$\hat{\gamma}_1$	Weibull mixture	-0.0070	0.0057	-0.0075	0.0211
	Cox's PH mixture	-0.0107	0.0065	-0.0160	0.0226

and likelihood estimates between iteration steps is less than the value specified.

FAST: when set to Y, parameter estimates and their standard errors (computed by inverting the matrix of second derivatives when convergence is attained) are written to the FAST_INC and FAST_SURV datasets, respectively. Although the standard errors may be underestimated [19], they may be of concern, since they do not require extensive bootstrap computation. The default value is Y.

The following options are available when bootstrap confidence intervals are requested for the Cox's mixture cure model:

BOOTSTRAP: when set to Y, indicates that non-parametric resampling with replacement from the original data set is performed. The default value is N.

NSAMPLE: is the number of bootstrap replicates that are produced by PROC MULLTEST.

STRATA: identifies a single variable to use as a stratification variable in PROC MULLTEST. Stratified resampling may better mimic the observed data.

BOOTMET: specifies the type of bootstrap confidence intervals to compute. These include the percentile (PTCL), hybrid method (HYB), normalized bias corrected (BOOTN), bias corrected (BC) and accelerated bias corrected (BCA) confidence intervals. Jackknife after bootstrap (JACK) is also available. The option BOOTMET = ALL indicates that all methods are requested.

GESTIMATE: when set to Y, Q-Q plots and bar charts of the distribution of parameter estimates over the bootstrap replicates are produced. This allows the user to check graphically the asymptotic normal distribution of parameter estimates and thus to check the validity of the different bootstrap confidence intervals. If the distribution is overdispersed relative to the normal distribution, the validity of the percentile based confidence interval is questionable.

4. Simulation study

In this study, data are generated from a logistic-Weibull mixture cure model, where $\pi(z) = \exp(\beta_0) / [1 + \exp(\beta_0 + \beta_1 z)]$. $S(t|U = 1, z) = \exp[-(\lambda t)^\rho \exp(\gamma_1 z)]$. The covariate z is fixed by design and is binary. We set $\lambda = 0.5$ and $\rho = 1$. Censoring times are generated according to the uniform distribution

U[0, 15]. The results given below are based on $n = 200$ with 500 replications.

Table 2 presents the estimated biases and MSE from the logistic-Weibull and logistic-Cox PH mixture cure models of three regression parameters β_0, β_1 and γ_1 based on simulated data for two different configurations. The first one corresponds to $\pi(z) = 0.5$ that is 50% of the population is cured. In this setting, the covariate z has no effect neither on incidence ($\beta_1 = 0$) nor on latency ($\gamma_1 = 0$). The second configuration corresponds to $\pi(z = 0) = 0.8$ and $\pi(z = 1) = 0.6$, meaning that 20% of the population is cured in one group and 40% in the other. It can be seen that point estimates have little bias. We also simulated a case with a continuous covariate, and similarly there are not substantial departures from the simulated values.

5. Application

The macro is applied to the melanoma data from the Eastern Cooperative Oncology Group (ECOG) phase III clinical trial e1684 [20]. The dataset can be downloaded at <http://merlot.stat.uconn.edu/~mhchen/survbook>. Our purpose here is not to perform a detailed analysis, but rather to illustrate the use of the PPSMCM macro. An extensive analysis of this dataset is provided in [21,22]. Briefly, the aim of e1684 clinical trial was to evaluate high dose interferon alpha-2b (IFN) regimen against placebo as postoperative adjuvant therapy. After deleting missing informations a total of $n = 284$ are used in the analysis. Only three covariates are included here, both in the incidence and latency parts. These covariates are treatment (0, control group; 1 IFN group), gender (0 for male, 1 for female) and age, which is continuous and centered to the mean. The response variable is taken to be the relapse-free survival (RFS) in years. The Weibull and Cox PH mixture cure models are applied to the data. The cured fraction is modelled by a logistic regression model. For the Cox PH mixture model the zero tail constraint option is specified to compute the conditional baseline survival function estimate. The bootstrap resampling is stratified on the treatment variable, which is the variable of interest. The number of bootstrap replicates is set to 3000 and all the bootstrap confidence intervals computation methods are requested as well as the plot of the distribution of parameter estimates. The measures of goodness of fit are also requested.

For the Cox mixture cure model, the macro is invoked by the following statement:


```
%pspmcm( DATA= e1684, CENSCOD= fcensor, TIME= RFS,
VAR= treatment(I S,1) gender(I S,.) age (I S,.),
INCPART=logit, SURVPART= Cox, TAIL= zero , SUOMET= pl,
MAXITER= 200, CONVCRIT= 1e-5, ALPHA= 0.05, FAST= Y,
BOOTSTRAP= Y, NSAMPLE= 3000, STRATA= treatment,
BOOTMET= ALL, GESTIMATE= Y,
PLOTFIT= Y)
run;
```

The Kaplan–Meier (KM) plot of the survival function estimate for each treatment group is shown in Fig. 1, together with the estimates from the Weibull and Cox PH mixture cure models. The KM survival function estimate levels off at the right tail and exhibits a long and stable plateau, which ensures the applicability of the mixture cure model approach. The three plots are nearly identical, which allows to think that the model adequacy is good. The correlation coefficients between the KM estimates and the fitted values indicate a good fit for both treatment groups for the Cox mixture cure model ($r = 0.9971$ for both groups). From the Weibull mixture cure model, the fit is better for the treatment group ($r = 0.9984$) than for the control group ($r = 0.9860$).

The parameter estimates from the Weibull and Cox PH mixture cure model are quite similar (Table 3). Moreover, the standard errors from the Cox PH mixture model estimated by using the inverse of the Hessian matrix at convergence of parameter estimates and likelihood are quite similar to those obtained from the Weibull mixture model.

The results indicate that the treatment effect is significant in the incidence (logistic) part, despite a low significance level ($p = 0.040$ for the Weibull model and $p = 0.034$ for the Cox PH

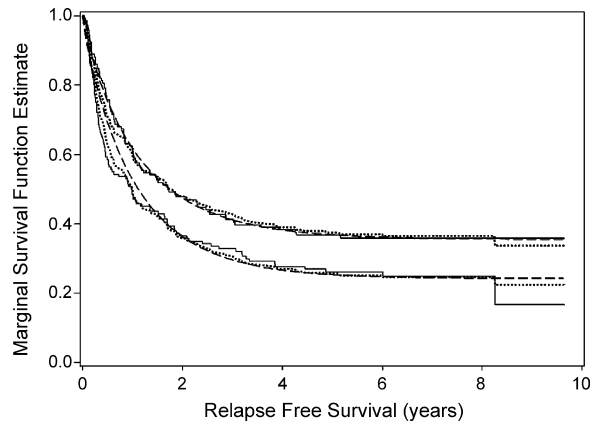


Fig. 1 – Marginal survival function curves for ECOG e1684 data. Upper lines, treatment group; lower lines, control group; solid lines, Kaplan–Meier estimates; dashed lines, Weibull mixture cure model estimates; dotted lines, Cox mixture cure model estimates.

Table 3 – Estimates (standard errors) for the ECOG e1684 data set from the Weibull and Cox’s PH mixture cure models

	Weibull mixture model	Cox PH mixture model	
	Estimate (S.E.) ^a	Estimate (S.E.) ^b	BCA 95% CI ^c
Logistic model			
Intercept	1.197(0.239)	1.298(0.236)	0.788; 2.312
Treatment: control	-0.565(0.272)*	-0.574(0.271)*	-1.328; 0.057
Gender: male	-0.061(0.275)	-0.082(0.274)	-0.812; 0.613
Age (per year)	0.014(0.010)	0.018(0.010)	-0.007; 0.062
Survival model			
Treatment: control	-0.104(0.159)	-0.149(0.144)	-0.500; 0.192
Gender: male	0.131(0.152)	0.106(0.147)	-0.268; 0.471
Age (per year)	-0.007(0.006)	-0.007(0.005)	-0.021; 0.006
Intercept (Weibull)	-0.285(0.316)	-	-
Shape (Weibull)	1.088(0.063)	-	-

^a Standard errors obtained using the inverse of the Hessian matrix.

^b Standard errors obtained using the inverse of the Hessian matrix at convergence of parameters for the estimates and likelihood.

^c Bias corrected accelerated bootstrap confidence interval from 3000 replicates.

* p-Value < 0.05.

Table 4 – Estimates (standard errors) for the ECOG e1684 data set from the standard Weibull and Cox PH models

	Standard Weibull model	Standard Cox PH model
	Estimate (S.E.)*	Estimate (S.E.)*
Treatment: control	-0.375(0.144)*	-0.360(0.144)*
Gender: male	0.001(0.148)	-0.018(0.147)
Age (per year)	-0.007(0.005)	0.005(0.005)
Intercept (Weibull)	-0.614(0.115)	-
Shape (Weibull)	0.586(0.035)	-

* p-Value < 0.05.

model), but not in the latency part. This means that the treatment increases the cured fraction, but do not delay the relapse in uncured patients. No other covariate has significant effect, neither on incidence, nor latency.

Invoking the model with the treatment as the only covariate leads to cured fraction estimations of 22.4% (95% confidence interval: 19.1–26.1) for the control group and 33.5% (95% CI: 23.7–44.7) for the IFN group from the Cox PH mixture model. The estimated cured fractions from the Weibull mixture model are 24.2% (95% CI: 20.7–28.1) for the control group and 35.4% (95% CI: 25.5–46.8) for the IFN group.

Applying the standard Weibull and Cox PH survival models (Table 4), would lead to conclude that the treatment had a significant effect on the survival. These models, however, do not account for the possibility of cure and may lead to misinterpretation of covariate effects.

Fig. 2 shows the Q-Q plots for the treatment covariate from the logistic-Cox PH mixture cure model over the 3000 bootstrap replicates. The distribution from the logistic part is overdispersed relative to the normal distribution. Thus, the bias corrected and accelerated bootstrap confidence intervals for parameter estimates is preferred and reported in Table 3. These confidence intervals are slightly larger than those obtained using the variance of parameter estimates. The effect of treatment on incidence is no longer significant (approximated p-value: 0.078).

6. Discussion

We propose a simple SAS macro to estimate parametric and semiparametric mixture cure models for individual data. The proposed SAS macro makes use of standard regression procedures available in the SAS/STAT package since SAS version 7 (PROC NLMIXED is not available in older versions). A limitation of our program is that time dependant variables are not handled, because in this case PROC PHREG does not provide the baseline survival function.

Parametric mixture cure models are simpler to implement than the semiparametric Cox mixture cure model since they do not require the EM algorithm. However, fully-parametric mixture models may not be flexible enough when analyzing biological data, since they involve strong assumptions about the survival distribution function of uncured subjects. A generalized F mixture model has been proposed [27] which makes less distributional assumptions, but computational difficulties may arise.

Alternatively, data can be analyzed utilizing statistical models that account for heterogeneity among individuals. These models, also known as frailty models, differ from cure models in that they assume all individuals eventually experience the event of interest with varying risk that are greater than zero [23,24]. The proportion of individuals considered to

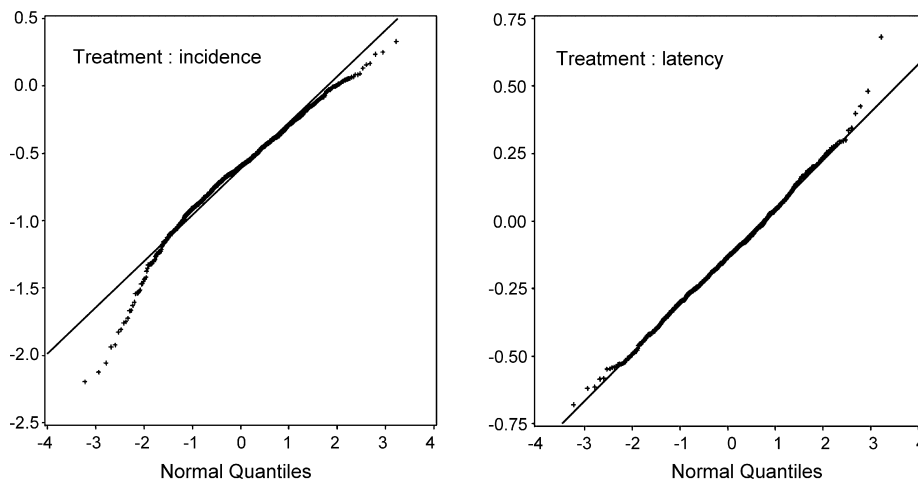


Fig. 2 – Q-Q plots of parameters estimates for the treatment covariate from the Cox mixture cure model based on 3000 bootstrap replicates. Logistic part (left) and survival part (right).

be cured in the former models are generally considered as having a low risk of experiencing the event in the latter models. The mixture cure model is a special case of a multiplicative frailty model, in which the hazard for an individual, conditional on U , can be written as $\lambda(t|U, \mathbf{x}) = U\lambda(t|U = 1, \mathbf{x})$. As a frailty variable, U is not entirely observable since an individual becomes labelled as $U = 1$ if an event is observed. Usually frailties are assumed to follow a distribution as the gamma, inverse Gaussian or positive stable distribution. These frailty distributions do not allow individuals to have zero risks, hence the standard frailty models do not account for a cured proportion. As an extension of the parametric family, Aalen [25] considered a compound Poisson distribution, which allows a positive probability for the risk to be zero. However, these models only account for heterogeneity among individuals in the latency part but not in the incidence one. Parametric mixture cure models with random effects have recently been proposed [26], but the choice of the frailties' distribution and of their variance matrix may be an important issue.

As stressed by many others [10,28] there are potential problems in applying the mixture cure model in cases where it may not be adequately justified. The use of such models should be restricted to problems in which there are strong biological evidences of the presence of a cured fraction. Another element of caution is that of sufficient follow-up. The levelling off of the Kaplan-Meier curve of the marginal survival function to non-zero proportions, the presence of a long and stable plateau together with a heavy censoring at the tail may also provide graphical evidences of the presence of non-susceptible subjects to ensure the applicability of the mixture cure models [29].

7. Availability

The SAS macro PSPMCM is available to the public at no charge at <http://www.isped.u-bordeaux2.fr/recherche/biostats/FR-biostats-accueil.htm#programmes>. A simulated dataset is also provided to illustrate the program.

REFERENCES

- [1] V.T. Farewell, The use of mixture models for the analysis of survival data with long-term survivors, *Biometrics* 38 (1982) 1041-1046.
- [2] Y. Peng, K.C. Carriere, An empirical comparison of parametric and semiparametric cure models, *Biomed. J.* 8 (2002) 1002-1014.
- [3] J.W. Gamel, E.A. Weller, M.N. Wesley, E.J. Feuer, Parametric cure models of relative and cause-specific survival for grouped survival times, *Comput. Methods Programs Biomed.* 61 (2000) 99-110.
- [4] D.R. Cox, D. Oakes, *Analysis of Survival Data*, Chapman and Hall, London, New York, 1984, pp. 62-111.
- [5] D.R. Cox, Regression models and life-tables (with discussion), *J. R. Stat. Soc. Ser. B* 34 (1972) 187-220.
- [6] Y. Peng, K.B. Dear, A nonparametric mixture model for cure rate estimation, *Biometrics* 56 (2000) 237-243.
- [7] J.P. Sy, J.M. Taylor, Estimation in a Cox proportional hazards cure model, *Biometrics* 56 (2000) 227-236.
- [8] SAS Institute Inc., *The NL MIXED Procedure, SAS/STAT User's guide, Version 8*, Cary, NC, 2000, pp. 2419-2504.
- [9] N.E. Breslow, Contribution to the discussion of D.R. Cox, *J. R. Stat. Soc. Ser. B* 34 (1972) 216-217.
- [10] J.D. Kalbfleisch, R.L. Prentice, *The Statistical Analysis of Failure Time Data*, John Wiley, New York, 1980, pp. 84-85.
- [11] J.M. Taylor, Semi-parametric estimation in failure time mixture models, *Biometrics* 51 (1995) 899-907.
- [12] C.S. Li, J.M.G. Taylor, J.P. Sy, Identifiability of cure models, *Stat. Probab. Lett.* 54 (2001) 389-395.
- [13] Y. Peng, Estimating baseline distribution in proportional hazards cure models, *Comput. Stat. Data Anal.* 42 (2003) 187-201.
- [14] K.F. Lam, D.Y. Fong, O.Y. Tang, Estimating the proportion of cured patients in a censored sample, *Stat. Med.* 24 (2005) 1865-1879.
- [15] A.C. Davison, D.V. Hinkley, *Bootstrap Methods and their Application*, Cambridge University Press, Cambridge, 1997.
- [16] J. Carpenter, J. Bithell, Bootstrap confidence intervals: when, which, what? A practical guide for medical statisticians, *Stat. Med.* 19 (2000) 1141-1164.
- [17] S.R. Cole, Simple bootstrap statistical inference using the SAS system, *Comput. Methods Programs Biomed.* 60 (1999) 79-82.
- [18] R.A. Maller, S. Zhou, *Survival Analysis with Long-term Survivors*, Wiley, Chichester, UK, 1996.
- [19] H.B. Fang, G. Li, J. Sun, Maximum likelihood estimation in a semiparametric logistic/proportional-hazards mixture model, *Scand. J. Stat.* 32 (2005) 59-75.
- [20] J.M. Kirkwood, M.H. Strawderman, M.S. Ernststoff, T.J. Smith, E.C. Borden, R.H. Blum, Interferon alfa-2b adjuvant therapy of high-risk resected cutaneous melanoma: the Eastern Cooperative Oncology Group Trial EST 1684, *J. Clin. Oncol.* 14 (1996) 7-17.
- [21] J.G. Ibrahim, M.H. Chen, D. Sinha, Bayesian semiparametric models for survival data with a cure fraction, *Biometrics* 57 (2001) 383-388.
- [22] J.G. Ibrahim, M.H. Chen, D. Sinha, *Cure Rate Models, Bayesian Survival Analysis*, Springer, New York, 2001 (Chapter 5).
- [23] O.O. Aalen, Heterogeneity in survival analysis, *Stat. Med.* 7 (1988) 1121-1137.
- [24] P. Hougaard, P. Myglegard, K. Borch-Johnsen, Heterogeneity models of disease susceptibility, with application to diabetic nephropathy, *Biometrics* 50 (1994) 1178-1188.
- [25] O.O. Aalen, Modelling heterogeneity in survival analysis by the compound Poisson distribution, *Ann. Appl. Probab.* 2 (1988) 951-972.
- [26] K.K.W. Yau, A.S.K. Ng, Long-term survivor mixture model with random effects: application to a multi-centre clinical trial of carcinoma, *Stat. Med.* 20 (2001) 1591-1607.
- [27] Y. Peng, K.B. Dear, J.W. Denham, A generalized f mixture model for cure rate estimation, *Stat. Med.* 17 (1998) 813-830.
- [28] V.T. Farewell, D.A. Sprott, The use of a mixture model in the analysis of count data, *Biometrics* 44 (1988) 1191-1194.
- [29] B. Yu, R.C. Tiwari, K.A. Cronin, E.J. Feuer, Cure fraction estimation from the mixture cure models for grouped survival data, *Stat. Med.* 23 (2004) 1733-1747.

.

Chapitre 4

Approche par vraisemblance pénalisée

Dans ce chapitre nous présentons une méthode d'estimation par vraisemblance pénalisée pour les modèles de survie avec fraction non à risque. Les modèles paramétriques sont d'une estimation relativement aisée mais ont l'inconvénient de faire de fortes hypothèses sur la distribution des temps de survie qui peuvent s'avérer difficilement vérifiables. A l'inverse les modèles semiparamétriques (Kuk et Chen, 1992 ; Taylor, 1995 ; Sy et Taylor, 2000 ; Peng et Dear, 2000 ; Tsodikov, 1998) présentent une souplesse importante mais ne fournissent pas d'estimateurs lisses des fonctions de risque et de survie. D'autre part, dans les approches par maximisation de la vraisemblance, les variances des estimateurs des paramètres de régression ne sont pas directement disponibles en raison de l'utilisation de l'algorithme EM. Enfin, l'instabilité des estimateurs non-paramétriques de la fonction de survie quand t est proche de la fin des temps d'observation, conduit à contraindre la fonction de survie à être nulle au delà du dernier temps d'événement. Les limites des modèles paramétriques et semi-paramétriques ont motivé l'approche par vraisemblance pénalisée. Dans le cadre des modèles de mélange avec fraction non à risque cette approche permet :

- de fournir un estimateur lisse des fonctions de risque et de survie (conditionnelles ou marginales)
- de ne pas imposer de contrainte sur la fonction de survie conditionnelle qui reste définie au delà du dernier temps d'événement.
- d'obtenir des estimateurs des variances des paramètres de régression par inversion de la matrice d'information de Fisher basée sur la vraisemblance complète du modèle.

Ce travail fait l'objet d'un article soumis sous forme révisée à *Statistics in Medicine* dont nous présentons la deuxième version. Une extension de l'approche par vraisemblance pénalisée au modèle de promotion fait l'objet d'une étude par simulations. Quelques remarques concluent ce chapitre.

4.1 Article méthodologique

Résumé de l'article

Une classe de modèles de survie a été développée pour l'analyse de données de survie avec une fraction non à risque, ou "guérie". Dans ce contexte, les modèles de survie classiques ne sont généralement pas appropriés car ils ne tiennent pas compte de la possibilité d'être non à risque. Les modèles de survie de mélange considèrent que la population d'étude est composée d'individus à risque de connaître l'événement d'intérêt et d'individus non à risque. Dans cette approche une des difficultés est l'estimation de la fonction de survie de base des individus à risque. L'estimation des variances des estimateurs des paramètres de régression peut aussi poser problème.

Dans cet article, nous proposons une approche par vraisemblance pénalisée fournissant un estimateur lisse et souple de la fonction de risque des individus non guéris. Cette approche permet aussi d'obtenir les variances des paramètres directement par inversion de la matrice d'information de Fisher basée sur la vraisemblance des données observées. Les performances de cette approche sont évaluées au travers de simulations. Les propriétés et les limites de la méthode proposée sont discutées et une application dans le domaine du cancer est présentée.

A penalized likelihood approach for mixture cure models

Fabien Corbiere^{1,2}, Daniel Commenges^{1,2}, Jeremy M.G Taylor³ and Pierre Joly^{1,2}

¹ INSERM U875 Biostatistique, Bordeaux, F-33076, France

² Université Victor Ségalen Bordeaux 2, Bordeaux, F-33076, France.

³ Department of Biostatistics, 1420 Washington Heights, University of Michigan, Ann Arbor, MI 48109, USA.

Abstract : Cure models have been developed to analyze failure time data with a cured fraction. For such data, standard survival models are usually not appropriate because they do not account for the possibility of cure. Mixture cure models assume that the studied population is a mixture of susceptible individuals, who may experience the event of interest, and non-susceptible individuals that will never experience it. Important issues in mixture cure models are estimation of the baseline survival function for susceptibles and estimation of the variance of the regression parameters. The aim of this paper is to propose a penalized likelihood approach which allows for flexible modeling of the hazard function for susceptible individuals using M-splines. This approach also permits direct computation of the variance of parameters using the inverse of the Hessian matrix. Properties and limitations of the proposed

method are discussed and an illustration from a cancer study is presented.

keywords : survival analysis, mixture cure models, penalized likelihood, splines.

Corresponding author : Fabien CORBIERE, U875, ISPED, Université de Bordeaux 2, 146 rue Léo Saignat, 33076 Bordeaux cedex, France. Tel: (33) 5 57 57 45 79; Fax: (33) 5 56 24 00 81; e-mail : fabien.corbiere@isped.u-bordeaux2.fr

1 Introduction

Models for survival analysis typically assume that everybody in the study population is susceptible to the event of interest and will eventually experience this event if the follow-up is sufficiently long. However, in some diseases there is strong biological evidence that, when considering endpoints other than natural death, the event would never occur for some fraction of the subjects. The assumption that some individuals will never experience the event may also be based solely on empirical considerations, such as the presence of a large number of long-term survivors. The presence of 'non-susceptible' individuals (or 'immune' or 'cured' regarding the event of interest) is suggested by a Kaplan-Meier plot of the marginal survival distribution function, which shows a long and stable plateau with heavy censoring at the right extreme. In recent years, there has been an increasing interest in modeling survival data with long term survivors. Failing to account for such cured subjects would lead to overestimation of the survival of susceptible subjects. Mixture cure models assume that the studied population is a mixture of susceptible (uncured) individuals, that may experience the event of interest, and non-susceptible (cured) individuals, that will never experience it. This approach allows us to estimate simultaneously whether the event of interest will occur, which is called incidence in the related literature, and when it will occur, given that it can occur, which is called latency. An important issue arising from mixture cure models is estimation of the baseline survival function for uncured individuals. Fully parametric distributions were first proposed, including exponential [1-3], lognormal [4-5], Weibull [6], Gompertz [7-8] and Burr XII [9]. However, fully parametric mixture cure models may not be flexible enough when analyzing biological data, since they involve strong assumptions

about the survival function of uncured patients. The generalized F distribution was proposed by Peng and al [10], which makes less distributional assumptions, but computational difficulties may arise. Another approach is based on a non-parametric estimate of the survival function of uncured individuals, and this defines the area of semiparametric mixture cure models. However, unlike the standard Cox proportional hazards model (PH) [11], the baseline survival distribution function in semiparametric proportional hazards mixture cure models cannot be eliminated as a nuisance parameter in the likelihood function and an estimation method more complicated than maximising the partial likelihood is needed. A semiparametric PH specification for the time to recurrence of the disease was proposed by Kuk and Chen [12], but a Monte Carlo approximation of the likelihood was required, which complicates parameter estimation. Assuming no covariates effect the latency, Taylor [13] used a weighted Kaplan-Meier estimator to estimate the failure time distribution and proposed an EM-algorithm. Peng and Dear [14] further extended this work by allowing the failure time distribution to depend on explanatory variables, but Monte Carlo techniques were still required to estimate the variance of the estimates. In the same way, Sy and Taylor [15,16] use the EM-algorithm to estimate the parameters, but proposed to estimate the asymptotic variance of the estimator by the inverse of the observed information matrix, based on the observed marginal likelihood function using the discrete proportional hazards model. Methods based on the EM-algorithm adopted the zero-tail constraint proposed by Taylor [13] to avoid identifiability problems. This constraint states that the conditional survival function is set to zero for censored times greater than the largest uncensored event time, which ensures a proper estimate. However, it implies that individuals with

survival times greater than the last event are all considered as non susceptible or cured, which may appear to be a strong assumption, especially in the case of poor follow-up beyond the period when events occur or in the case of heavy censoring. Peng [17] further proposed different tail-completion methods to allow the conditional survival estimate to decrease slowly to zero after the largest event time. Based on simulations studies, he also showed that the zero tail constraint could lead to an overestimated cured fraction.

Another drawback of semiparametric approaches is that a smooth estimate of the hazard function, which often has a meaningful interpretation, is not available. An alternative approach is to define the estimator of the conditional hazard function non-parametrically as the function which maximizes a penalized likelihood. The solution is then approximated using splines [18]. Such an approach has been proposed in standard survival models and multi-state models under the proportional hazards assumptions for right censored and left truncated data [19-20]. In this paper we shall apply the penalized likelihood approach to mixture cure models. This method allows us to obtain flexible and smooth estimates of the conditional and marginal hazard and survival functions and does not require a tail constraint. The asymptotic variances of the estimators are obtained using the inverse of the information matrix. The model is presented in the next section. The penalized likelihood approach and the estimation procedure is described in section 3. In section 4 a simulation study is shown. An illustrative example is provided in section 5.

2 The Model

2.1 Mixture cure models

Let U be the unobserved indicator denoting whether an individual is susceptible ($U = 1$) or non susceptible ($U = 0$) to the event of interest and T is a nonnegative random variable denoting the failure time of interest. Taking into account the possibility that a fraction of the population can be considered as immune or cured, the marginal, population-based survival and probability density functions are given by

$$S_{pop}(t|\mathbf{x}, \mathbf{z}) = \pi(\mathbf{z})S(t|U=1, \mathbf{x}) + 1 - \pi(\mathbf{z}) \quad (1)$$

$$f_{pop}(t|\mathbf{x}, \mathbf{z}) = \pi(\mathbf{z})f(t|U=1, \mathbf{x}) \quad (2)$$

where $S_{pop}(t|\mathbf{x}, \mathbf{z})$ and $f_{pop}(t|\mathbf{x}, \mathbf{z})$ are the unconditional (marginal) survival and probability density functions of T for the entire population respectively. $S(t|U=1, \mathbf{x}) = P(T > t|U=1, \mathbf{x})$ is the survival function for susceptible individuals given a covariate vector $\mathbf{x} = (x_1, \dots, x_p)'$, and $\pi(\mathbf{z}) = P(U=1|\mathbf{z})$ is the probability of being susceptible given a covariate vector $\mathbf{z} = (z_1, \dots, z_q)'$, which may include the same covariates as \mathbf{x} . Mixture models assume that $S(t|U=1, \mathbf{x})$ is proper (i.e. total mass equals 1), so that $\lim_{t \rightarrow \infty} S(t|U=1, \mathbf{x}) = 0$ and $\lim_{t \rightarrow \infty} S_{pop}(t|\mathbf{x}, \mathbf{z}) = 1 - \pi(\mathbf{z})$. The zero-tail constraint proposed by Taylor [13] assumes that there is a t^* such that for all $t \geq t^*$, $S(t|U=1, \mathbf{x}) = 0$. Usually one takes $t^* = t_\kappa$, where t_κ is the last observed failure time. When $\pi(\mathbf{z}) = 1$ for all \mathbf{z} , i.e. when there is no cured fraction, the mixture cure model reduces to the standard survival model.

The proportion of cured patients can be modeled by the usual binary regression models. Farewell [6], Kuk and Chen [12], Peng and Dear [14] and Sy and Taylor [15]

all used a logistic regression model with logit link, $\text{logit}(\pi(\mathbf{z})) = \boldsymbol{\beta}'\mathbf{z}$, where $\boldsymbol{\beta}$ is the vector of regression parameters associated with \mathbf{z} , including an intercept. Other regression models include the probit link and the complementary log-log link [21].

In parametric mixture cure models, $S(t|U=1)$ is specified by a few parameters, whereas in semiparametric mixture cure models it is left arbitrary. In proportional hazards (PH) models, the conditional distribution of T is modeled by

$$\begin{aligned} S(t|U=1, \mathbf{x}) &= S_0(t|U=1)^{\exp(\boldsymbol{\gamma}'\mathbf{x})} \\ &= \exp\left(-\exp(\boldsymbol{\gamma}'\mathbf{x}) \int_0^t \lambda_0(v|U=1)dv\right) \end{aligned} \quad (3)$$

where $S_0(t|U=1)$ and $\lambda_0(t|U=1)$ are the baseline conditional survival and hazard functions respectively. The marginal, population based, hazard function is then given by

$$\begin{aligned} \lambda_{pop}(t|x, z) &= \frac{f_{pop}(t|\mathbf{x}, \mathbf{z})}{S_{pop}(t|\mathbf{x}, \mathbf{z})} \\ &= \frac{\pi(\mathbf{z})\lambda_0(t|U=1)\exp(\boldsymbol{\gamma}'\mathbf{x})S_0(t|U=1)^{\exp(\boldsymbol{\gamma}'\mathbf{x})}}{\pi(\mathbf{z})S_0(t|U=1)^{\exp(\boldsymbol{\gamma}'\mathbf{x})} + 1 - \pi(\mathbf{z})} \end{aligned} \quad (4)$$

From (4) it can be seen that at the population level, the proportional hazard assumption no longer holds. If $S_0(t|U=1)$ is left arbitrary, the model is the semi-parametric Cox proportional hazards mixture cure model as described by Kuk and Chen [12], Peng and Dear [14] and Sy and Taylor [15].

2.2 Likelihood function

Suppose the data are of the form $(t_i, \delta_i, \mathbf{x}_i, \mathbf{z}_i), i = 1, \dots, n$, where δ_i is the censoring indicator with $\delta_i = 1$ if t_i is uncensored and $\delta_i = 0$ otherwise. It follows that,

if $\delta_i = 1$ then $u_i = 1$, and if $\delta_i = 0$ then u_i is not observed, where u_i is the value taken by the random variable U_i . Under the assumption of type I censoring, the likelihood contribution for individual i is $\pi_i(\mathbf{z}_i)f(t_i|u_i = 1, \mathbf{x}_i)$ for $\delta_i = 1$ and $1 - \pi_i(\mathbf{z}_i) + \pi_i(\mathbf{z}_i)S(t_i|u_i = 1, \mathbf{x}_i)$ for $\delta_i = 0$. The observed full likelihood is given by

$$L(\boldsymbol{\gamma}, \boldsymbol{\beta}) = \prod_{i=1}^n \left\{ \pi_i(\mathbf{z}_i) \lambda_0(t_i|u_i=1) e^{\boldsymbol{\gamma}' \mathbf{x}_i} e^{-\Lambda_0(t_i|u_i=1) \exp(\boldsymbol{\gamma}' \mathbf{x}_i)} \right\}^{\delta_i} \times \left\{ (1 - \pi_i(\mathbf{z}_i)) + \pi_i(\mathbf{z}_i) e^{-\Lambda_0(t_i|u_i=1) \exp(\boldsymbol{\gamma}' \mathbf{x}_i)} \right\}^{(1-\delta_i)} \quad (5)$$

where $\Lambda_0(t|U=1) = \int_0^t \lambda_0(v|U=1) dv$ is the conditional cumulative hazard function.

When no cured fraction is assumed, i.e. $\pi(\mathbf{z}_i) = 1$ for all \mathbf{z}_i , the likelihood function (5) reduces to the likelihood of the standard survival model.

3 The penalized likelihood approach

3.1 Penalized likelihood function

Most often the conditional hazard function can be expected to be smooth. A possible means to introduce such *a priori* knowledge is to penalize the likelihood by a term which takes large value for rough functions. We choose the roughness penalty function to be the L2 norm of the second derivative of the conditional hazard function.

$$pl(\boldsymbol{\gamma}, \boldsymbol{\beta}) = l(\boldsymbol{\gamma}, \boldsymbol{\beta}) - \kappa \int \lambda_0''(v|U=1)^2 dv \quad (6)$$

where l is the full log-likelihood given in (5) and $\kappa > 0$ a positive smoothing parameter which controls the trade-off between the fit to the data and the smoothness

of the function. Maximization of (6) in the desired class of function defines the maximum penalized likelihood estimators (MPnLE) $\hat{\lambda}_0(\cdot|U = 1)$, $\hat{\beta}$ and $\hat{\gamma}$.

3.2 Approximation by splines

The estimator $\hat{\lambda}_0(\cdot|U = 1)$ cannot be computed explicitly but can be approximated by a linear combination of m cubic M-splines which are normalized B-splines. An M-spline of order k is defined as

$$M_j(x|k) = \begin{cases} \frac{k[(x-t_j)M_j(x|k-1)+(t_{j+k}-x)M_{j+1}(x|k-1)]}{(k-1)(t_{j+k}-t_j)}, & t_j \leq x < t_{j+k}, \\ 0, & \text{elsewhere,} \end{cases}$$

with

$$M_j(x|1) = \begin{cases} \frac{1}{(t_{j+1}-t_j)} & \text{if } t_j \leq x < t_{j+1}, \\ 0 & \text{elsewhere,} \end{cases}$$

where t_1, \dots, t_m is a sequence of increasing knots. Each $M_j(x|k)$ is zero outside the interval $[t_j, t_{j+k}]$ and hence is nonzero over k intervals, and over each interval there are k nonzero M-splines. We use splines of order 4 (also called cubic splines). To each spline, we associate an I-spline, i.e.,

$$I_j(x|k) = \int_0^x M_j(u|k)du.$$

Each M_j is a piecewise polynomial of degree $k - 1$ and each associated I_j is a piecewise polynomial of degree k defined as (for $t_j \leq x < t_{j+1}$)

$$I_h(x|k) = \begin{cases} 0 & \text{if } h > j, \\ \sum_{l=h}^j (t_{l+k+1} - t_l) \frac{M_l(x|k+1)}{k+1} & \text{if } j - k + 1 \leq h < j, \\ 1 & \text{if } h < j - k + 1. \end{cases}$$

Note that M-splines are nonnegative and I-splines are monotonically increasing. The monotonicity constraint for a function represented on a basis of I-splines can thus be fulfilled by constraining the coefficients to be positive. The estimator $\hat{\Lambda}_0(\cdot|U=1)$ can be approximated by a linear combination of m I-splines $\tilde{\Lambda}_0(\cdot|U=1) = \sum_{j=1}^m g(\boldsymbol{\theta}_j)I_j(\cdot)$ where $g(\boldsymbol{\theta}_j) \geq 0 \forall j$. In practice we use $g(\boldsymbol{\theta}_j) = \boldsymbol{\theta}_j^2$. By differentiating, we get the conditional baseline hazard function with the same vector of coefficients $\boldsymbol{\theta} = (\theta_1, \dots, \theta_m)^T$, such that $\tilde{\lambda}_0(\cdot|U=1) = \sum_{j=1}^m g(\boldsymbol{\theta}_j)M_j(\cdot)$. The resulting function is a smooth positive function, defined on the positive real line and has continuous second derivatives. A spline function is completely defined by a sequence of increasing knots (t_1, \dots, t_l) and the coefficients $\boldsymbol{\theta}$. The first knot is set at 0 and the last knot is set at the last (possibly censored) follow-up time point, so that the hazard function is defined on the whole follow-up period, including censored times greater than the last observed event time. Other knots can be either put equidistantly between the first and last one, or can be defined taking into consideration the observed times. Therefore, we have $m = l + 2$ parameters to estimate the conditional hazard function. Theoretically, the more knots, the better the approximation. Increasing the number of knots does not deteriorate the MPnLE: this is because the degree of smoothing in the penalized likelihood method is tuned by the smoothing parameter κ and not by the number of splines. On the other hand, once a sufficient number of knots is established, there is no advantage in adding more. Moreover, the more knots, the longer the running time, especially if the method involves a search for the smoothing parameter. Furthermore, some numerical problems (slow convergence or failure to invert the Hessian matrix) can arise, particularly for a large number of knots. Generally we found that a good choice for the number of knots is

between 7 and 15. Also in our experience, we found that setting many knots after the last observed failure time could lead to numerical instabilities and convergence problems.

3.3 Estimation procedure

We implemented the method with a Fortran program. The approximation $\tilde{\lambda}_0(t|U = 1)$ of $\lambda_0(t|U = 1)$ is the function belonging to the space generated by the basis of splines which maximizes $pl(\lambda)$. The estimate $\hat{\theta}$ of θ for a fixed κ is obtained by maximizing the log-likelihood using the robust Marquardt [22] algorithm, which is a combination of a Newton-Raphson algorithm and a steepest descent algorithm. This algorithm has the advantage of being more stable than the Newton-Raphson algorithm while preserving its fast convergence property near the maximum. Convergence is attained when the difference between two consecutive log-likelihoods is small, the coefficients are stable and the gradient is small enough. This maximization method performs well in most cases. To avoid the identifiability problem, that may sometimes arise due to the absence of constraint on the conditional survival function, initial MPnLE values for θ and γ may first be obtained by fitting a standard survival model to the uncensored subjects.

The variance of regression parameters are computed using the inverse of the matrix of the second derivatives of the penalized likelihood.

3.4 Selection of the smoothing parameter κ

Because we model the conditional hazard function and we do not constrain the conditional survival function at the right tail, parameters estimates may be sensitive

to the choice of κ .

In our estimation method, an empirical estimate of κ can be provided or the smoothing parameter can be chosen by maximizing an approximate cross-validation score [23],

$$LCV_a = l(\cdot) - Tr(H_{pl}^{-1} H_l)$$

where l is the likelihood given in (5) with the corresponding Hessian matrix H_l and H_{pl} is the Hessian matrix based on the penalized likelihood [17]. However, in some cases, the automatic selection of the smoothing parameter may give unsatisfactory results, for example in the presence of very late failure times when the risk set is very small, the MPnLE may suggest the absence of a cured fraction (i.e. $\pi(\mathbf{z}) \approx 1$ and $\hat{S}(t_{\max}|U = 1) > 0$), where t_{\max} is the last follow-up times. In those cases a heuristic choice of κ , can be used to choose between several propositions, to yield an estimate of $\hat{S}(t_{\max}|U = 1)$ close to zero. However, the existence of very late failure times may also preclude the application of mixture cure models because it may indicate the absence of a cured fraction or insufficient follow-up.

4 Simulation study

In this section we present the results from two simulation studies to illustrate the performance of the proposed estimation method.

The first simulation study aims at evaluating the performance of the model in a regression context. Data were generated from the logistic-Weibull mixture cure model. Specifically, the cured fraction is modeled through $1 - \pi(z) = [1 + \exp(\beta_0 +$

$\beta_1 z)^{-1}$ and the latency part is given by $S(t|U = 1) = \exp(-(\xi t)^\rho \exp(\gamma z))$. The covariate z was fixed by design and binary, with half of the individuals having $z = 1$. Censoring time C were derived from an exponential distribution with censoring rate λ_c . Each individual was followed for at most $\tau_{max} = 10$, so that the data were (t, δ, z) where $t = \min(C, 10)$ when $U_i = 0$ and $t = \min(T, C, 10)$ when $U_i = 1$. Various values of β_0 , β_1 , γ and λ_c were considered.

Values for β_0 were either -1.3863 , 0 or 1.3863 , so that the cured fraction was 80%, 50% or 20% in the baseline group ($z=0$), respectively. Values for β_1 were either 0 or -1 , meaning that the covariate z has no effect on the incidence part or that it increases the cured fraction in individuals with $z = 1$, respectively. In the same way, values for γ were either 0 or 0.5 , meaning that the covariate z has no effect on the latency part or that it accelerates the occurrence of failure in susceptible individuals with $z = 1$. We set $\xi = 0.9$ and $\rho = 1.5$. The censoring rate λ_c was either 0.1 or 0.4 representing mild and heavy censoring respectively. The penalized likelihood approach was applied to each sample with seven equidistantly distributed knots defining the hazard function. The smoothing parameter was chosen automatically for each sample using the cross-validation method explained previously.

For each configuration, 550 samples were generated, with sample size 200. For each sample, the value of $\hat{S}_0(t_{max}|U = 1)$ was recorded, where t_{max} is the last follow-up time. The few cases (less than 5 per cent) for which $\hat{S}_0(t_{max}|U = 1) \gg 0$ using the automatic search of the smoothing parameter, or for which convergence or numerical problems occurred, were eliminated, because it was too fastidious, in a simulation context, to look for an heuristic value of κ which would perform well.

Table 1 and 2 present the estimated biases for regression parameters β_0 , β_1 and γ

based on 500 samples. To evaluate the adequacy of estimated variances, we compared the empirical variance, i.e. the variance of estimates, with the mean of estimated variances. The coverage rates of the normal approximation 95% confidence interval, based on the point estimates and the estimated variances are also shown for each regression parameter. Results are only presented here for $\beta_1 = -1$ and $\gamma_1 = 0.5$ but conclusions were almost identical for other simulated values.

It can be seen from tables 1 and 2 that point estimates have little biases in most cases. The biases on $\hat{\beta}_1$ and $\hat{\gamma}$ are generally higher in case of heavy censoring, meaning that a reasonable proportion of non-susceptibles surviving for long enough is needed to ensure a good estimate of the incidence and latency effect of covariates. In addition, the empirical variance and the mean of estimated variances agree pretty well and coverage rates are good in most cases with mild censoring, suggesting that the variance estimators perform quite well. In case of heavy censoring and small proportions of susceptible individuals ($\beta_0 = -1.3863$), the estimate $\hat{\gamma}$ is not stable, due to an insufficient number of observed events, which is reflected by high observed and estimated variances. However, the variances estimators still perform well and coverage rates are still reasonable.

In the second simulation study, we focused on the estimation of the baseline conditional hazard function. To illustrate the flexibility of the conditional hazard function estimator based on M-splines, we generated a mixture of two Weibull hazard functions, such that $S(t|U = 1) = pS_1(t|U = 1) + (1 - p)S_2(t|U = 1)$ where $S_1(t|U = 1) = \exp(-(\xi_1 t)^{\rho_1} \exp(\gamma Z))$ and $S_2(t|U = 1) = \exp(-(\xi_2 t)^{\rho_2} \exp(\gamma Z))$. We set $\xi_1 = 1.0$, $\rho_1 = 1.5$, $\xi_2 = 0.1$, $\rho_2 = 5.0$, $\gamma = 0.5$ and $p = 0.4$. The resulting baseline conditional hazard function is shown in figure 1. The censoring times were

generated using the uniform distribution $U[0; 20]$. As previously, a logistic model was used to model the cured fraction such that $1 - \pi(z) = [1 + \exp(\beta_0 + \beta_1 z)]^{-1}$, with $\beta_0 = 0$ and $\beta_1 = -0.5$.

Under this setting 500 random samples were generated with sample size 1000, 400 and 200. The penalized-likelihood approach was applied with an automatic search for the smoothing parameter with fourteen equidistantly distributed knots. The performances of the estimates of the conditional hazard and survival functions were evaluated in term of bias at different time points, i.e. 0.6, 3.5, 6.5 and 10. The variance of $\hat{\lambda}_0(t|U = 1)$ was computed using the delta method, and we compared the empirical variance, i.e. the variance of estimates, with the mean of estimated variances. The results are shown in table 3. With sample size 1000, bias are small at all time points and suggest good properties of the estimators. Biases increase with smaller sample sizes. Empirical variances and the means of estimated variances agree in most cases and coverage rate are satisfactory. The largest variances are observed at time point 10, when the number in the set at risk is reduced. The good fit between the simulated conditional baseline hazard function and some randomly selected estimated curves is illustrated graphically on figure 1.

5 Application

To illustrate the penalized likelihood approach, we consider the data from 672 patients with squamous cell carcinoma of the tonsil [24]. The data have already been explored using a semiparametric PH mixture model [15] and a semiparametric accelerated mixture cure model [25]. In this example, local recurrence is considered as the event and failure time is time from initial treatment to local recurrence. After

an approximate 15 years follow-up period, 206 events were observed. The Kaplan-Meier plot of the marginal survival function levels off with a large and stable plateau at the right tail, ensuring the appropriateness of the mixture cure model approach (not shown here). Six covariates are considered here, including T-stage which has four levels (T1; T2; T3; T4), neck node status which has two levels (negative (N0) and positive (N+)), sex, and the continuous variables age, total dose and overall treatment duration.

In our application of the penalized likelihood approach, we used 14 knots, placed every 0.6 years between 0 and 7.2 years and at the last follow-up time (14.46 years). The results were not sensitive to the location of knots. However, because of the absence of observed event after 8.15 years, adding more knots beyond the last event time did not improve the likelihood significantly, and led in some cases to convergence breakdown. The smoothing parameter was chosen automatically based on the cross-validation criterion. Standard errors for the parameter estimates were computed using the inverse of the second derivative of the penalized likelihood.

Table 4 gives the results from proposed penalized likelihood estimation method and from the Cox PH mixture cure model of Sy and Taylor [14]. Parameters estimates are very similar between the two estimation methods and variance estimates are in close agreement. T stage acts significantly on the incidence part, patients with high T stage being cured less. The effect of T stage is also significant on latency ($p=0.003$), with patients with T stage = 2 having delayed recurrences. Node is found to be marginally significant in both the incidence ($p= 0.098$) and latency ($p=0.052$) parts, with higher occurrence rate and earlier recurrence times for those with positive nodes. The total dose and treatment duration are highly significant for

incidence but not for latency. Higher doses reduce the recurrence rate while patients with long treatment duration are at high risk to fail. Sex is not significant on either incidence or latency. While it is not significant for the incidence part, the effect of age is significant on latency, local recurrence being delayed in older patients.

Figure 2 and 3 illustrate the effect of age and T stage on incidence and latency. The estimated conditional and marginal hazard functions and the corresponding survival functions based on the penalized likelihood approach were computed, under the proportional hazard assumption, for a male, node positive, patient total dose 64 Gray and overall treatment duration 45 days. The number at-risk in the susceptible group, based on the model estimates, was 352.6 at $t=0$, 202.38 at six months, 26.22 at two years, 6.2 at 5 years and only 2.24 after 8 years of follow-up. The effect of age (figure 1) is computed for fixed T stage = 2. The opposite effect of age on incidence and latency is illustrated by survival and hazard curves that cross. The effect of T stage is computed for fixed age = 60 years. Similarly the marginal hazard curves cross after 1 year, illustrating the different effects of T stage on incidence and latency.

Because of a very long follow-up period beyond the last failure time, which ensures the appropriateness of the zero-tail constraint, the estimated conditional and marginal survival curves from the Cox PH mixture cure model (not shown here) are similar (although not smooth) to those obtained with our approach.

Interestingly, the conditional hazard curves for both age and T stage show a bimodal shape between 0 and 4 years, with hazard peaking first around 0.5 year and then around 2.5 years. This was not modified by changing the location of knots. The hazard function is not plotted beyond 4 years because of the very small number

in the risk set beyond this date but goes to infinity after the last event time (not shown). Note that this bimodal shape could not be identified using a Weibull mixture cure mixture model, and illustrates the flexibility of the approximation of the hazard function using splines. The reason for the two modes in the hazard is unclear. It may be partly due to scheduling of the patient's visits to the doctor's clinic, at which time recurrences may be first detected. Also there may be other biological parameters than those explored here, which may play a role on the distribution of failure times among uncured patients.

6 Discussion

In this paper we propose a new estimation method applied to mixture cure models. Our work was mainly motivated by two major issues regarding estimation procedures in this kind of model: the estimation of the conditional survival/hazard function, and the closely related estimation of the variance of regression parameters. The proposed approximation of the conditional hazard function using splines yields a smooth and much more flexible estimate than standard fully parametric distributions. Moreover, estimation could be performed by direct maximization of the full likelihood function, whereas the E-M algorithm is mostly used in semiparametric mixture cure models [14-15,25-26]. Although this estimation procedure works well without specifying any tail constraint, we did not completely rule out the identifiability issues commonly met with mixture cure models. The choice of the smoothing parameter κ based on the approximate cross-validation score may lead to an improper conditional survival estimate in the presence of late failure time. This issue corresponds, however, to cases where the application of mixture cure models may not be justified due to insufficient follow-up or because the presence of a cured fraction is questionable. However, in our numerical experience, a heuristic choice of κ always yielded a satisfactory solution.

Alternatively, data can be analyzed utilizing statistical models that account for heterogeneity among individuals. These models, also known as frailty models, differ from cure models in that they assume all individuals eventually experience the event of interest with varying risks that are greater than zero [27-28]. The proportion of individuals considered to be cured in the former models are generally considered as having a low risk of experiencing the event in the latter models. As noted by Sy

and Taylor [15], the mixture cure model is a special case of a multiplicative frailty model, in which the hazard for an individual, conditional on U , can be written as $\lambda(t|U, \mathbf{x}) = U\lambda(t|U = 1, \mathbf{x})$. As a frailty variable, U is not entirely observable since an individual becomes labelled as $U = 1$ if an event is observed. Usually frailties are assumed to follow a distribution such as the gamma, inverse Gaussian or positive stable distribution. These frailty distributions do not allow individuals to have zero risks (i.e. $Pr(U = 0) = 0$), and the standard frailty models do not account for a cured proportion. As an extension of the parametric family, Aalen [29] considered a compound Poisson distribution, which allows a positive probability for the risk to be zero. A comparison of parametric mixture cure models, gamma frailty mixture models and compound Poisson models was performed by Price and Manutunga [30] in a one-sample, no-covariate analysis of a leukemia data set. Their results suggest that gamma frailty mixture models and compound Poisson models could perform better than parametric mixture cure models. However, these models only account for heterogeneity among individuals in the latency part, not in the incidence one. Parametric mixture cure models with random effects have recently been proposed [31], but the choice of the frailty distribution and of their covariance matrix may be an important issue. Further work is needed to address these issues using a penalized likelihood approach.

References

1. Jones DR, Powles RL, Machin D, Sylvester RJ. On estimating the proportion of cured patients in clinical studies. *Biometrie-Praximetric*, 1981; **21**:1 – 11.
2. Goldman AI. Survivorship analysis when cure is a possibility: A Monte Carlo study. *Statistics in Medicine*, 1984; **3**: 153 – 163.
3. Ghitany ME, Maller RA, Zhou S. Exponential mixture models with long-term survivors and covariates. *Journal of Multivariate Analysis*, 1994; **49**:218 – 241.
4. Boag JW. Maximum likelihood estimates of the proportion of patients cured by cancer therapy. *Journal of the Royal Statistical Society, Series B*, 1949; **11**:15 – 44.
5. Gamel JW and Vogel RL. Comparison of parametric and non-parametric survival methods using simulated clinical data. *Statistics in Medicine*, 1997; **16**:1629 – 1643.
6. Farewell VT. The use of mixture models for the analysis of survival data with long-term survivors. *Biometrics* 1982; **38**: 1041 – 1046.
7. Gordon NH. Application of the theory of finite mixtures for the estimation of "cure" rates of treated cancer patients. *Statistics in Medicine*, 1990; **9**: 397 – 407
8. Cantor AB and Shuster JJ. Parametric versus non-parametric methods for estimating cure rates based on censored survival data. *Statistics in Medicine*, 1992; **11**:931 – 937.

9. Shao Q and Zhou X. A new parametric model for survival data with long-terms survivors. *Statistics in Medicine*, 2004; **23**:3525 – 3543.
10. Peng Y, Dear KB, Denham JW. A generalized F mixture model for cure rate estimation. *Statistics in Medicine*, 1998; **17**:813 – 830.
11. Cox DR. Regression models and life-tables (with discussion). *Journal of the Royal Statistical Society, Series B* 1972; **34**: 187 – 220.
12. Kuk AYC and Chen C. A mixture model combining logistic regression with proportional hazards regression. *Biometrika*, 1992; **79**:531 – 541.
13. Taylor JMG. Semi-parametric estimation in failure time mixture models. *Biometrics*, 1995; **51**: 899 – 907.
14. Peng Y and Dear KB. A nonparametric mixture model for cure rate estimation. *Biometrics*, 2000; **56**: 237 – 243.
15. Sy JP and Taylor JMG. Estimation in a Cox proportional hazards cure model. *Biometrics*, 2000; **56**: 227 – 236.
16. Sy, J.P. and J.M.G. Taylor. Standard errors for the Cox proportional hazards cure model. *Mathematical and Computer Modelling*, 2001; **33**: 1237 – 1251.
17. Peng Y. Estimating baseline distribution in proportional hazards cure models. *Computational Statistics and Data Analysis*, 2003; **42** : 187 – 201.
18. Eilers PHC and Marx BD. Flexible smoothing with B-splines and penalties. *Statistical Science*,1996; **11**; 89 – 121.

19. Joly P, Commenges D, Letenneur L. A penalized likelihood approach for arbitrarily censored and truncated data: Application to age-specific incidence of dementia. *Biometrics*, 1998; **54**:185 – 194.
20. Joly P, Commenges D, Helmer C, Letenneur L. A penalized likelihood approach for an illness-death model with interval-censored data: application to age-specific incidence of dementia. *Biostatistics*, 2002; **3**: 433 – 443.
21. Lam KF, Fong DY, Tang OY. Estimating the proportion of cured patients in a censored sample. *Statistics in Medicine*, 2005; **24**: 1865 – 1879.
22. Marquardt D. An algorithm for least-squares estimation of nonlinear parameters. *SIAM Journal of Applied Mathematics*, 1963; **11**:431 – 441.
23. Commenges D, Joly P, Gégout-Petit A, Liqueur B. Choice between Semi-parametric Estimators of Markov and Non-Markov Multi-state Models from Coarsened Observations. *Scandinavian Journal of Statistics*, 2007; **34**:33 – 54.
24. Withers HR, Peters LJ, Taylor JMG, Owen JB, Morrison WH, Schultheiss TE, Keane T, O Sullivan B, van Dyk J, Gupta N, Wang CC, Jones CU, Doppke KP, Myint S, Thompson M, Parsons JT, Mendenhall WM, Dische S, Aird EGA, Henk JM, Bidmead MAM, Svoboda V, Chan Y, Hanlon AL, Peters TL, Hanks GE. Local control of carcinoma of the tonsil by radiation therapy: an analysis of patterns of fractionation in nine institutions. *International Journal of Radiation Oncology, Biology, Physics*, 1995; **33**:549 – 562.
25. Li CS and Taylor JMG. A semi-parametric accelerated failure time cure model. *Statistics in Medicine*, 2002; **21** : 3235 – 3247.

26. Zhang J and Peng Y. A new estimation method for the semiparametric accelerated failure time mixture cure model. *Statistics in Medicine*, 2006; DOI: 10.1002.
27. Aalen OO. Heterogeneity in survival analysis. *Statistics in Medicine*, 1988; **7**:1121 – 1137.
28. Hougaard P, Myglegard P, Borch-Johnsen K. Heterogeneity models of disease susceptibility, with application to diabetic nephropathy. *Biometrics*, 1994; **50**: 1178 – 1188.
29. Aalen OO. Modelling heterogeneity in survival analysis by the compound Poisson distribution. *Annals of Applied Probability*, 1988; **2**:951 – 972.
30. Price DL and Manatunga AK. Modelling survival data with a cured fraction using frailty models. *Statistics in Medicine*, 2001; **20**: 1515 – 1527.
31. Yau KKW and Ng ASK. Long-term survivor mixture model with random effects: application to a multi-centre clinical trial of carcinoma, *Statistics in Medicine*, 2001; **20**: 1591 – 1607.

Table 1: Simulation results: mild censoring ($\lambda_c=0.1$)

Parameter	True value	Bias	Mean variance	Empirical variance	Coverage rate (%)
Low cured fraction (20% in the baseline group)					
β_0	1.3863	0.0516	0.0837	0.0795	95.2
β_1	-1	-0.0233	0.1181	0.1267	95.4
γ	0.5	-0.0292	0.0429	0.0357	97.6
Mild cured fraction (50% in the baseline group)					
β_0	0	0.0088	0.0429	0.0416	95.5
β_1	-1	-0.0049	0.0930	0.1022	95.7
γ	0.5	0.0624	0.0982	0.0706	97.6
High cured fraction (80% in the baseline group)					
β_0	-1.3863	-0.0472	0.0739	0.0735	95.2
β_1	-1	-0.0127	0.2274	0.2325	96.6
γ	0.5	0.0632	0.2486	0.2350	96.4

Table 2: Simulation results: heavy censoring ($\lambda_c=0.4$)

Parameter	True value	Bias	Mean variance	Empirical variance	Coverage rate (%)
Low cured fraction (20% in the baseline group)					
β_0	1.3863	0.0506	0.1109	0.1164	95.8
β_1	-1	0.0920	0.1671	0.1791	94.4
γ	0.5	-0.0087	0.0521	0.0470	96.8
Mild cured fraction (50% in the baseline group)					
β_0	0	0.0186	0.0538	0.0616	96.9
β_1	-1	-0.0638	0.1177	0.1286	97.1
γ	0.5	0.1209	0.1335	0.1056	97.0
High cured fraction (80% in the baseline group)					
β_0	-1.3863	-0.0616	0.0100	0.0945	94.9
β_1	-1	0.0183	0.2788	0.2787	95.6
γ	0.5	0.1226	0.3508	0.3099	97.1

Table 3: Simulation results: bias, empirical variance ($\times 100$), mean estimated variance ($\times 100$) and coverage rate for the estimation of the conditional hazard function

Sample size	Time point	True value	Bias	Mean variance ($\times 100$)	Empirical variance ($\times 100$)	Coverage rate (%)
n = 1000	0.6	0.3430	-0.0086	0.1419	0.1655	93.5
	3.5	0.0102	0.0096	0.0062	0.0051	95.8
	6.5	0.0892	-0.0080	0.0275	0.0277	95.6
	10.0	0.5000	-0.0767	0.5604	0.4595	97.5
n = 400	0.6	0.3430	-0.0592	0.2399	0.2925	92.5
	3.5	0.0102	0.0108	0.0119	0.0087	92.6
	6.5	0.0882	-0.0090	0.0590	0.0482	95.5
	10.0	0.5000	-0.0906	1.0158	0.7965	98.7
n = 200	0.6	0.3430	-0.0959	0.3642	0.4140	90.4
	3.5	0.0102	0.0152	0.0324	0.0238	97.5
	6.5	0.0882	-0.0146	0.0992	0.0883	96.2
	10.0	0.5000	-0.0830	1.7160	1.3645	98.8

Table 4: Estimates from the penalized likelihood model mixture cure model (MPnLE) and the semiparametric PH mixture cure model.

	MPnLE estimates			semiparametric PH mixture model		
	Estimate	SE	(P-value)	Estimate	SE	(P-value)
Incidence Model						
Intercept	-0.102	0.940		-0.181	1.030	
T-Stage			0.013 ^a			
T_2	0.866	0.362	(0.017)	0.852	0.357	(0.017)
T_3	1.648	0.345	(< 0.001)	1.655	0.345	(< 0.001)
T_4	2.182	0.429	(< 0.001)	2.198	0.430	(< 0.001)
Node	0.341	0.206	(0.098)	0.355	0.204	(0.082)
Total dose (Gray)	-0.079	0.018	(< 0.001)	-0.077	0.018	(< 0.001)
Duration (per 10 days)	0.466	0.126	(< 0.001)	0.463	0.127	(< 0.001)
Sex: male	0.085	0.218	(0.697)	0.116	0.215	(0.590)
Age (per 10 years)	0.116	0.095	(0.222)	0.134	0.097	(0.167)
Latency Model						
T-Stage			0.003 ^a			
T_2	-0.725	0.359	(0.043)	-0.625	0.365	(0.087)
T_3	-0.041	0.378	(0.913)	-0.108	0.352	(0.759)
T_4	0.473	0.403	(0.241)	0.385	0.383	(0.315)
Node	0.383	0.197	(0.052)	0.339	0.188	(0.071)
Total dose (Gray)	-0.005	0.016	(0.739)	-0.005	0.014	(0.721)
Duration (per 10 days)	-0.036	0.076	(0.639)	-0.007	0.078	(0.920)
Sex: male	0.150	0.205	(0.465)	0.065	0.198	(0.743)
Age (per 10 years)	-0.225	0.080	(0.005)	-0.303	0.097	(0.002)

^a : Wald χ^2 with 3 df.

Captions to figures

Figure 1: The simulated and some estimated conditional hazard functions in simulation 2.

Figure 2: Estimated conditional and marginal survival (left) and hazard (right) functions from the penalized likelihood approach for four different ages. For each age, the other covariates are fixed at selected values: sex = male, total dose = 64 Gray, overall treatment duration = 45 days, Node positive, T stage = 2.

Figure 3: Estimated conditional and marginal survival (left) and hazard (right) functions from the penalized likelihood approach for four different T stages. For each T stage, the other covariates are fixed at selected values: sex = male, total dose = 64 Gray, overall treatment duration = 45 days, Node positive, age=60 years.

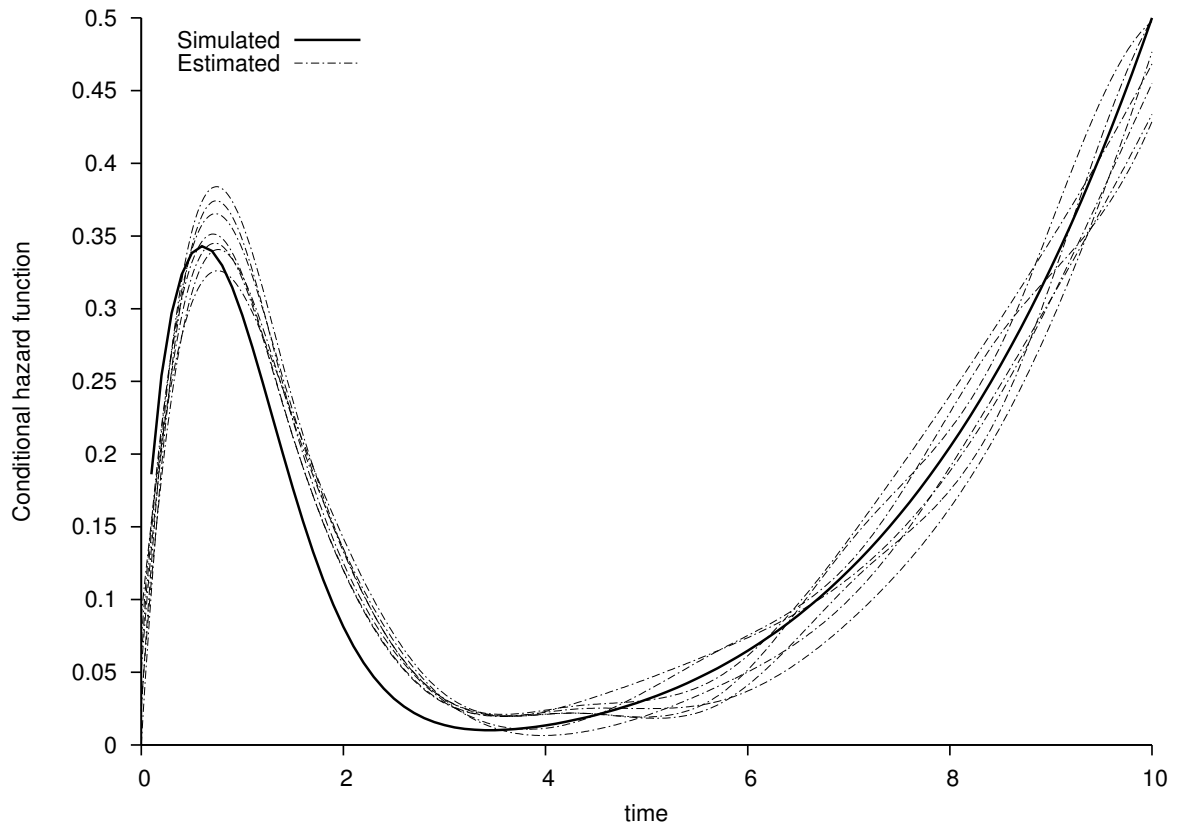


Figure 1:

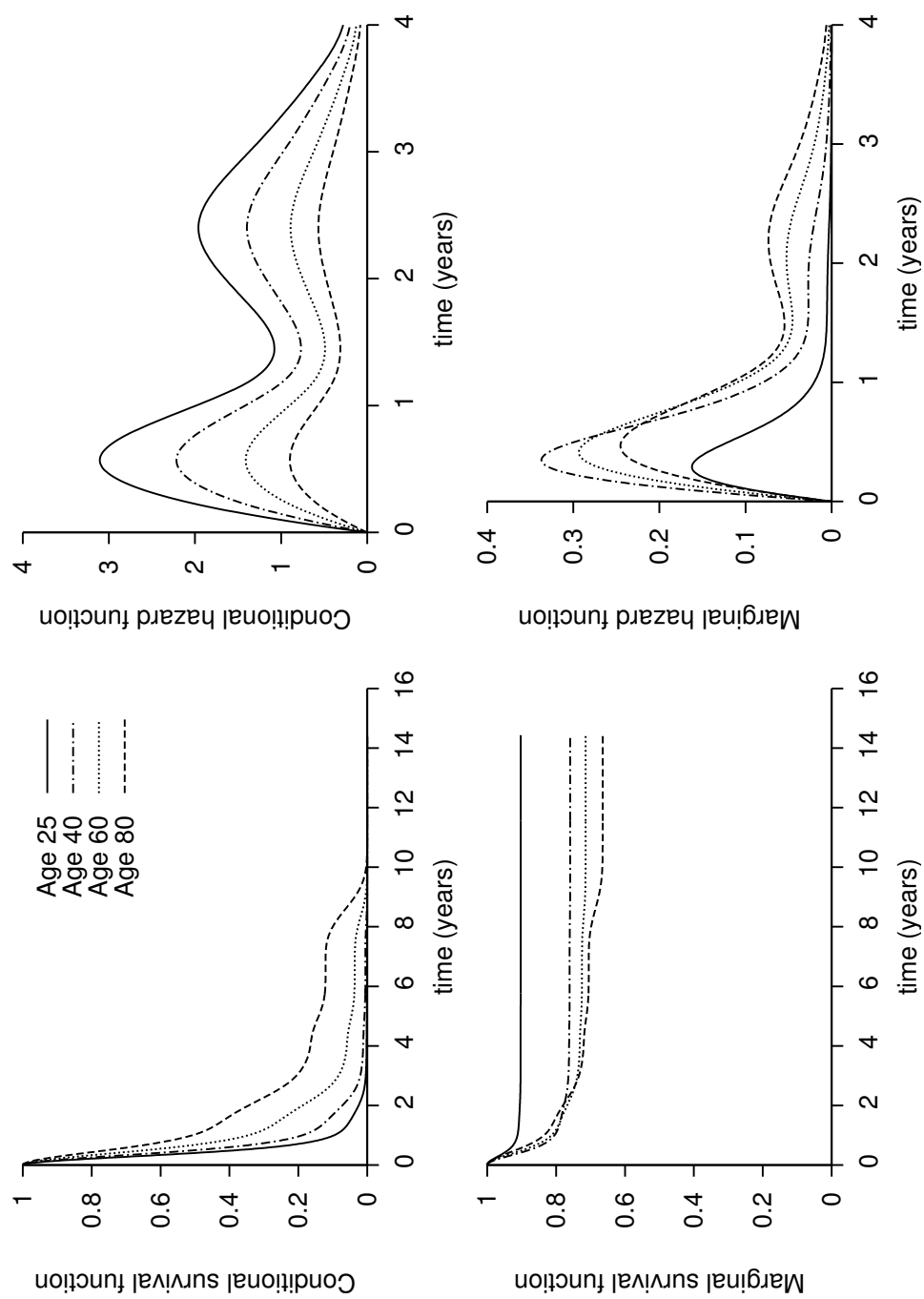


Figure 2:

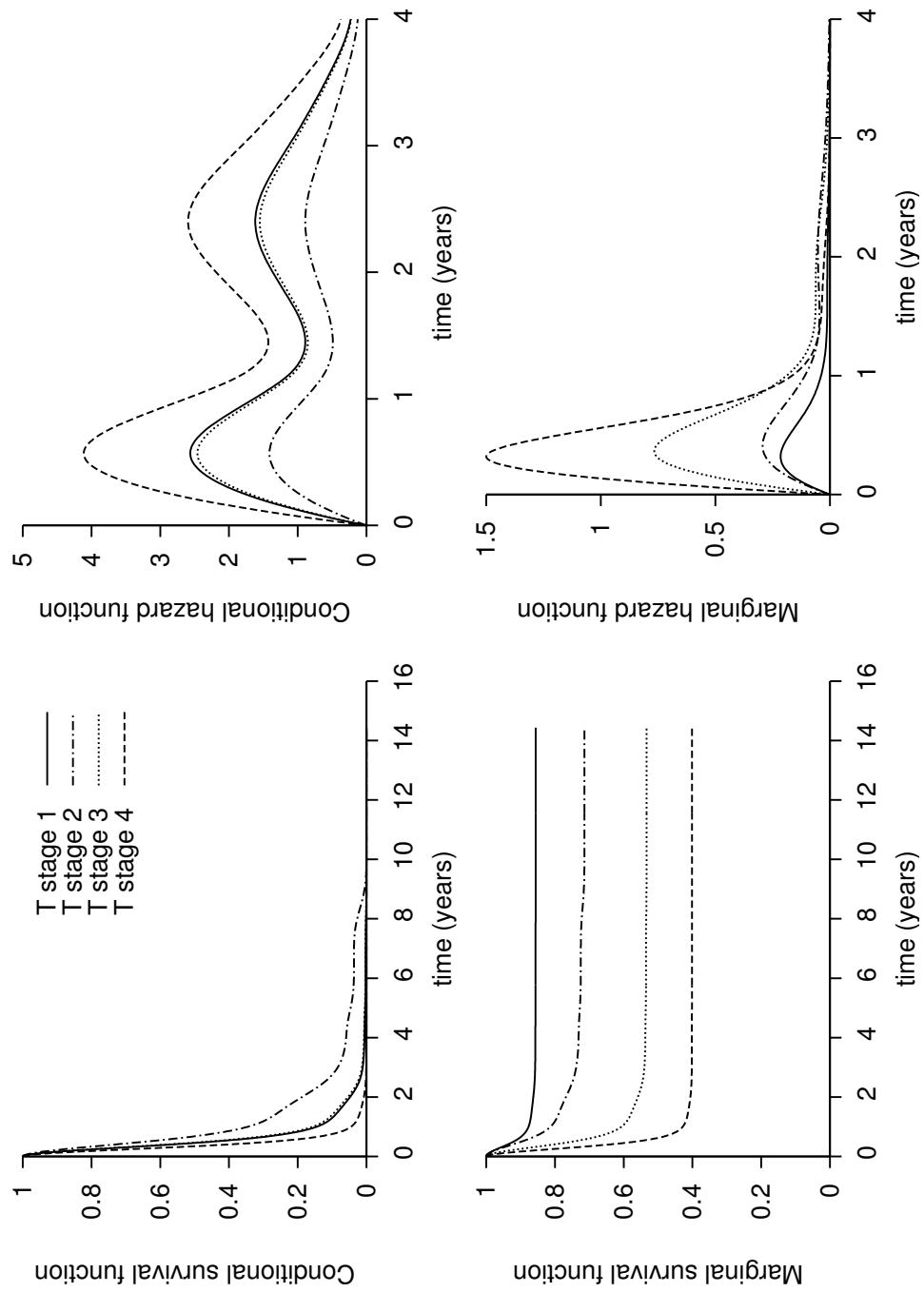


Figure 3:

4.2 Extension au modèle de promotion

Nous considérons le modèle de promotion récemment proposé par Broet et al. (2001) et Tsodikov (2002) dans lequel la fonction de survie des temps de promotion dépend de variables explicatives selon un modèle à risque proportionnels tel que

$$S_{pop}(t) = \exp \left\{ - \exp(\beta'z) [1 - S_0(t)^{\exp(\gamma'x)}] \right\} \quad (4.1)$$

Dans l'approche par vraisemblance pénalisée, la pénalisation porte sur la fonction de risque des temps de promotion $\lambda_0(t)$. La logvraisemblance pénalisée est alors donnée par

$$pl(\gamma, \beta) = l(\gamma, \beta) - \kappa \int \lambda_0''(v)^2 dv \quad (4.2)$$

La fonction de risque $\lambda_0(t)$ est, de manière comparable à ce qui est proposé pour le modèle de mélange avec fraction non à risque, approchée par une combinaison linéaire de M-splines.

Etude par simulations

Nous avons comparé les approches par modèle de mélange et par modèle de promotion à l'aide de simulations. Pour cela nous avons utilisé les données générées pour évaluer le modèle de mélange et présentées dans l'article soumis à *Statistics In Medicine*. Ces données sont générées selon un modèle de mélange (logistique-Weibull) de sorte à représenter différentes situations de censure (modérée et forte) et différentes valeurs de fractions non à risque (80%, 50% et 20%). L'interprétation des paramètres de régression du modèle (4.1) n'étant pas aussi aisée que celle du modèle de mélange, nous avons comparé les fractions à risque $\hat{p}_0 = Pr(U = 1|z = 0)$ et $\hat{p}_1 = Pr(U = 1|z = 1)$ estimées par chacun des deux modèles aux valeurs simulées pour chacune des strates de la variable binaire z définies par $z = 0$ et $z = 1$. Nous avons également calculé les valeurs de $\hat{S}_{pop}(t, z)$ et $\hat{S}(t|U = 1, z)$ pour 20 valeurs de t et comparé les valeurs estimées par les deux modèles aux valeurs simulées. Enfin, nous avons utilisé le critère de validation croisé comme moyen de comparaison des

deux modèles. Les résultats sont présentés dans le tableau 4.1 dans le cas d'un phénomène de censure modéré (25% des individus à risque ont des temps de survie censurés à droite) et le tableau 4.2 dans le cas d'une censure forte (40% des individus à risque ont des temps de survie censurés à droite).

Les biais sur les estimateurs obtenus par le modèle de mélange et le modèle de promotion sont comparables et restent acceptables dans la majorité des configurations simulées. On remarque cependant que les biais relatifs (biais absolu/valeurs simulées) sont plus élevés lorsque la censure est importante. De même lorsque les fractions à risque sont faibles (8 et 20%) les biais sont plus importants que lorsque la fraction à risque est modérée (27-50 %) en raison du faible nombre d'événements servant à l'estimation des paramètres. Dans le cas de fractions à risque élevées (60 et 80 %) un phénomène semblable, mais moins marqué, est observé. Dans ce cas, c'est le faible nombre d'individus restant en vie après le dernier événement qui peut biaiser les estimateurs.

Afin de comparer l'adéquation aux données apportée par le modèle de mélange et le modèle de promotion, nous avons utilisé le critère approché de validation croisée (Commenges et al., 2007). Pour chaque fichier simulé, nous avons calculé la différence $\Delta LCV_a = LCV_{a_m} - LCV_{a_p}$ où LCV_{a_p} est le critère approché de validation croisée calculé sous le modèle de promotion et LCV_{a_m} celui calculé sous le modèle de mélange. Les données étant simulées selon un modèle de mélange, nous nous attendions à ce que le modèle de mélange soit considéré par ce critère comme le meilleur modèle, autrement dit que $\Delta LCV_a < 0$ dans la plupart des cas. Les résultats sont présentés dans le tableau 4.3. Dans 75 % des simulations le critère de validation choisi le modèle de mélange comme étant le meilleur modèle. Dans 8.01 % des cas $\Delta LCV_a < -10^{-2}$, ce qui correspond, selon les observations de Commenges et al. (2007) à une distance "grande" à "modérée" en faveur du modèle de mélange. Dans les autres cas, la distance entre les deux approches est "faible" à "négligeable". Notons pas ailleurs que le modèle de promotion n'a dans aucun cas était considéré comme "largement" meilleur que la modèle de mélange, ce qui était attendu compte-tenu du vrai modèle

TAB. 4.1 – Comparaison du modèle de mélange et du modèle de promotion : censure modérée (25% des individus à risque sont censurés)

Paramètre	Valeurs simulées	Modèle de promotion			Modèle de mélange		
		Biais	MSE	% couv.	Biais	MSE	% couv.
<u>Fraction à risque élevée</u>							
p_0	0.8000	0.0020	0.0025	94.94	0.0010	0.0024	95.11
p_1	0.6000	0.0293	0.0038	96.15	0.0304	0.0039	96.33
$S(t U = 1, z = 0)$	0.3202*	0.0040	0.0009		0.0032	0.0008	
$S(t U = 1, z = 1)$	0.2609*	0.0046	0.0009		0.0042	0.0009	
$S_{pop}(t, z = 0)$	0.8640*	-0.0002	0.0020		0.0002	0.0019	
$S_{pop}(t, z = 1)$	0.9377*	-0.0198	0.0029		-0.0210	0.0029	
<u>Fraction à risque modérée</u>							
p_0	0.5000	0.0044	0.0032	95.89	0.0029	0.0032	96.25
p_1	0.2700	-0.0043	0.0025	96.75	-0.0038	0.0025	96.29
$S(t U = 1, z = 0)$	0.3202	0.0054	0.0018		0.0034	0.0017	
$S(t U = 1, z = 1)$	0.2609	-0.0046	0.0021		-0.0049	0.0020	
$S_{pop}(t, z = 0)$	0.8640	-0.0002	0.0023		-0.0002	0.0023	
$S_{pop}(t, z = 1)$	0.9377	0.0023	0.0018		0.0018	0.0019	
<u>Fraction à risque faible</u>							
p_0	0.2000	-0.0032	0.0034	95.57	-0.0049	0.0022	95.57
p_1	0.0800	0.0040	0.0010	96.26	0.0040	0.0010	96.26
$S(t U = 1, z = 0)$	0.3202	0.0008	0.0051		-0.0007	0.0044	
$S(t U = 1, z = 1)$	0.2609	-0.0018	0.0061		-0.0027	0.0056	
$S_{pop}(t, z = 0)$	0.8640	-0.0036	0.0015		0.0036	0.0015	
$S_{pop}(t, z = 1)$	0.9377	0.0031	0.0007		-0.0033	0.0007	

MSE = Mean Squared Error, %couv = pourcentage de couverture

* $\sum_i^{20} S(t_i|U = 1, z = .)$ et $\sum_i^{20} S_{pop}(t_i|z = .)$, sans signification

TAB. 4.2 – Comparaison du modèle de mélange et du modèle de promotion : censure forte (40% des individus à risque sont censurés)

Paramètre	Valeurs simulées	Modèle de promotion			Modèle de mélange		
		biais	MSE	% couv.	biais	MSE	% couv.
<u>Fraction à risque élevée</u>							
p_0	0.8000	0.0130	0.0039	96.72	0.0038	0.0035	98.00
p_1	0.6000	0.0505	0.0063	97.18	0.0530	0.0065	97.45
$S(t U = 1, z = 0)$	0.3202*	0.0082	0.0014		0.0032	0.0012	
$S(t U = 1, z = 1)$	0.2609*	0.0038	0.0011		0.0034	0.0011	
$S_{pop}(t, z = 0)$	0.8640*	-0.0021	0.0026		0.0003	0.0026	
$S_{pop}(t, z = 1)$	0.9377*	-0.0366	0.0045		-0.0389	0.0047	
<u>Fraction à risque modérée</u>							
p_0	0.5000	0.0092	0.0048	97.25	0.0059	0.0047	97.80
p_1	0.2700	-0.0187	0.0034	98.17	-0.0185	0.0034	98.53
$S(t U = 1, z = 0)$	0.3202	0.0067	0.0025		0.0027	0.0023	
$S(t U = 1, z = 1)$	0.2600	-0.0076	0.0030		-0.0077	0.0030	
$S_{pop}(t, z = 0)$	0.8640	-0.0025	0.0031		-0.0023	0.0031	
$S_{pop}(t, z = 1)$	0.9377	0.0208	0.0025		0.0206	0.0025	
<u>Fraction à risque faible</u>							
p_0	0.2000	-0.0061	0.0048	95.93	0.0057	0.0047	96.49
p_1	0.0800	-0.0063	0.0032	98.26	-0.0061	0.0033	98.49
$S(t U = 1, z = 0)$	0.3202	-0.0008	0.0061		-0.0022	0.0060	
$S(t U = 1, z = 1)$	0.2609	-0.0213	0.0105		-0.0214	0.0104	
$S_{pop}(t, z = 0)$	0.8640	-0.0041	0.0018		-0.0041	0.0018	
$S_{pop}(t, z = 1)$	0.9377	0.0038	0.0009		0.0037	0.0009	

MSE = Mean Squared Error, %couv = pourcentage de couverture

* $\sum_i^{20} S(t_i|U = 1, z = .)$ et $\sum_i^{20} S_{pop}(t_i|z = .)$, sans signification

TAB. 4.3 – Comparaison du modèle de mélange et du modèle de promotion selon le critère approché de validation croisé

	% des simulations	Conclusions
$\Delta LCV_a < 0$	75.00	
Valeur de ΔLCV_a		
$\leq -10^{-1}$	0.43	Distance “grande” en faveur du modèle de mélange
$] -10^{-1}; -10^{-2}]$	7.58	Distance “modérée” en faveur du modèle de mélange
$] -10^{-2}; -10^{-3}]$	39.06	Distance “faible” en faveur du modèle de mélange
$] -10^{-3}; 0]$	27.93	Distance “négligeable” en faveur du modèle de mélange
$[0; 10^{-3}[$	17.91	Distance “négligeable” en faveur du modèle de promotion
$[10^{-3}; -10^{-2}[$	7.09	Distance “faible” en faveur du modèle de promotion

réplications = 500 ; taille d'échantillon=200.

$$\Delta LCV_a = LCV_{a_{mélange}} - LCV_{a_{promotion}}$$

qui était celui de mélange. Ces résultats sont aussi compatibles avec les différences peu marquées dans les estimateurs des fractions à risque et des fonctions de survie présentés précédemment.

Chapitre 5

Modèle de mélange avec fraction non à risque et effets aléatoires

Dans notre application à la tremblante du mouton, les animaux sont regroupés par élevage et leurs observations ne peuvent plus être considérées comme indépendantes. La prise en compte de la corrélation entre les animaux d'un même élevage peut se faire par l'utilisation de modèles à effets aléatoires, présentés dans le chapitre 2.7.

Dans ce chapitre nous étendons le modèle de mélange avec fraction non à risque et fragilités partagées proposé par Yau et Ng (2001) en prenant en compte différentes structures d'effets aléatoires.

Nous considérons le modèle de mélange dans lequel un effet aléatoire est introduit dans la modélisation de la fraction à risque (incidence) et un autre dans la modélisation des temps de survie des sujets à risque (latence). La modélisation des temps de survie des individus à risque est réalisée par une fonction paramétrique de Weibull ou constante (exponentielle) par morceaux.

Nous considérons que les individus sont regroupés en G groupes indépendants. Le temps d'événement du sujet j , $j = (1, \dots, n_i)$, appartenant au groupe i , $i = (1, \dots, G)$, est noté T_{ij} et son temps de censure C_{ij} . Pour l'individu i du groupe j , on observe $Y_{ij} = \min(T_{ij}, C_{ij})$. L'indicateur de censure est noté δ_{ij} , tel que $\delta_{ij} = 1$ si Y_{ij} correspond à un temps d'événement et $\delta_{ij} = 0$ sinon. Nous supposons par ailleurs l'indépendance des temps de survie et de censure. Soit U_{ij} la variable aléatoire telle que $U_{ij} = 1$ si l'individu j du groupe i est à risque et $U_{ij} = 0$ sinon. Si $\delta_{ij} = 1$ alors $U_{ij} = 1$. En revanche pour un individu dont le temps d'observation est censuré à droite ($\delta_{ij} = 0$), U_{ij} n'est pas observée.

Dans le cadre d'un modèle de mélange avec une fraction non à risque, la fonction de survie de l'ensemble de la population (survie 'marginale') peut être écrite pour l'individu j du groupe i

$$S_{pop}(t_{ij} | \mathbf{z}_{ij}, \mathbf{x}_{ij}) = \pi(\mathbf{z}_{ij})S(t_{ij} | U_{ij} = 1, \mathbf{x}_{ij}) + 1 - \pi(\mathbf{z}_{ij}) \quad (5.1)$$

Nous considérons trois modèles en fonction de la structure des effets aléatoires :

1. un modèle à 2 effets aléatoires différents indépendants,
2. un modèle à 2 effets aléatoires différents non indépendants,

3. un modèle avec effet aléatoire commun à la partie incidence et à la partie latence.

5.1 Les modèles

5.1.1 Modèle à deux effets aléatoires indépendants

Conditionnellement aux effets aléatoires, la partie incidence, modélisée par un modèle de régression logistique est donnée par

$$\pi(\mathbf{z}_{ij}|\boldsymbol{\beta}, v_i) = \frac{\exp(\mathbf{z}_{ij}'\boldsymbol{\beta} + v_i)}{1 + \exp(\mathbf{z}_{ij}'\boldsymbol{\beta} + v_i)}$$

où v_i , $i = 1 \dots, G$, sont des variables aléatoires indépendantes de densité commune $g(v; \theta_v)$.

La partie latence, modélisée par un modèle à risques proportionnels, est, conditionnellement aux effets aléatoires, définie par

$$\lambda(t_{ij}|U=1, \boldsymbol{\gamma}, w_i) = \lambda(t_{ij}|U=1) \exp(\mathbf{x}_{ij}'\boldsymbol{\gamma} + w_i) \quad (5.2)$$

où w_i , $i = 1 \dots, G$, sont des variables aléatoires indépendantes de densité commune $g(w; \theta_w)$. Les paramètres θ_v et θ_w quantifient la variabilité des effets aléatoires. Nous considérons $g(v; \theta_v) = \mathcal{N}(0, \theta_v)$ et $g(w; \theta_w) = \mathcal{N}(0, \theta_w)$, $w \perp v$. La matrice de variance-covariance des effets aléatoires $D(\theta_v, \theta_w)$ est

$$\mathbf{D} = \begin{pmatrix} \theta_v & 0 \\ 0 & \theta_w \end{pmatrix}$$

Ce modèle caractérise une situation où la variable définissant le groupe, et à laquelle se rattache chacun des effets aléatoires indépendants, traduit l'effet de deux phénomènes non mesurés distincts, qui influencent de façon indépendante la probabilité de connaître l'événement (partie incidence), et la probabilité de survenue de l'événement d'intérêt sachant qu'il peut survenir (partie latence). Dans le cadre de

notre étude, il pourrait s'agir, à titre d'exemples, d'une caractéristique environnementale d'une part (quantité de prion présent dans l'environnement) influençant le risque d'infection, et de pratiques d'élevage modifiant l'apparition des symptômes (carences alimentaires...).

Afin d'évaluer l'existence d'un effet groupe sur la fraction non à risque, nous pouvons tester l'hypothèse $H_{00} : \theta_v = 0$, contre l'hypothèse alternative $H_{10} : \theta_v \neq 0$. La statistique de test $LRS = -2[\ln(L(0, \hat{\theta}_w)) - \ln(L(\hat{\theta}_v, \hat{\theta}_w))]$, où $L(0, \hat{\theta}_w)$ est la vraisemblance du modèle sous l'hypothèse nulle (variance de l'effet aléatoire V_i nulle) suit asymptotiquement un mélange de distribution χ_1^2 et χ_0^2 avec des poids égaux 0.5. De même tester l'hypothèse $H_{01} : \theta_w = 0$ contre l'hypothèse alternative $H_{11} : \theta_w \neq 0$, permet d'évaluer l'existence d'un effet groupe sur le risque de décès des individus à risque.

5.1.2 Modèle à deux effets aléatoires non indépendants

Il s'agit d'une généralisation du modèle précédent où u et w ne sont plus considérés comme indépendants. Dans ce cas la matrice de variance-covariance des effets aléatoires est

$$\mathbf{D} = \begin{pmatrix} \theta_v & \theta_{vw} \\ \theta_{vw} & \theta_w \end{pmatrix}$$

où θ_{vw} est la covariance entre les effets aléatoires. Ce modèle caractérise la situation où les effets aléatoires, bien que correspondant à des phénomènes sous-jacents différents, influencent de manière non-indépendante la partie incidence et la partie latence du modèle.

Pour tester l'indépendance des effets aléatoires, nous testons l'hypothèse $H_0 : \theta_{vw} = 0$ contre l'hypothèse alternative $H_1 : \theta_{vw} \neq 0$. Nous pouvons utiliser le rapport de vraisemblance entre le modèle à deux effets aléatoires indépendants et le modèle à deux effets aléatoires non indépendants. La statistique de test $LRS = -2[\ln(L(\theta_v, \theta_w, 0)) - \ln(L(\theta_v, \hat{\theta}_w, \theta_{vw}))]$, où $L(\theta_v, \theta_w, 0)$ est la vraisemblance du modèle

sous l'hypothèse nulle (effets aléatoires indépendants), suit asymptotiquement une loi de χ^2 à 1 degré de liberté. Alternativement nous pouvons utiliser le test de Wald, dont la statistique $t = \hat{\theta}_{uv} / \widehat{\text{var}}(\hat{\theta}_{uv})$ suit asymptotiquement une loi normale centrée réduite.

5.1.3 Modèle à un effet aléatoire commun

Ce modèle est une alternative au modèle à deux effets aléatoires en cas de non indépendance entre les effets aléatoires. Conditionnellement aux effets aléatoires, la partie incidence, modélisée par un modèle de régression logistique est donnée par

$$\pi(\mathbf{z}_{ij} | \boldsymbol{\beta}, v_i) = \frac{\exp(\mathbf{z}_{ij}' \boldsymbol{\beta} + v_i)}{1 + \exp(\mathbf{z}_{ij}' \boldsymbol{\beta} + v_i)}$$

où v_i , $i = 1 \dots, G$, sont des variables aléatoires indépendantes de densité commune $g(v; \theta_v)$.

La partie latence, modélisée par un modèle à risques proportionnels, est, conditionnellement aux effets aléatoires, définie par

$$\lambda(t_{ij} | U=1, \boldsymbol{\gamma}, v_i) = \lambda(t_{ij} | U=1) \exp(\mathbf{x}_{ij}' \boldsymbol{\gamma} + \beta_v v_i)$$

où β_v est un paramètre de régression associé à v_i . Nous considérons $g(v; \theta_v) = \mathcal{N}(0, \theta_v)$.

Par construction, ce modèle considère que le même phénomène est attaché à l'effet aléatoire v mais influence dans des proportions possiblement différentes la partie latence et la partie incidence. Il pourrait s'agir, par exemple, de la pression d'infection (quantité de prion disponible dans l'environnement) qui pourrait agir à la fois sur le risque d'infection et sur la durée d'incubation de la maladie (partie latence). Ces deux effets peuvent cependant ne pas avoir la même influence, ce qui est traduit par le paramètre de régression β_v . Tester l'hypothèse $H_0 : \beta_v = 0$ contre $H_1 : \beta_v \neq 0$ permet de tester la pertinence de l'introduction d'un effet aléatoire dans la modélisation de la latence.

Ce modèle, moins souple que le modèle à deux effets aléatoires non indépendants, présente cependant l'intérêt de ne nécessiter le calcul que d'une intégrale.

5.2 Estimation des paramètres

5.2.1 Écriture de la vraisemblance

Modèles à deux effets aléatoires

Sous l'hypothèse de censure indépendante et d'indépendance des observations au sein des groupes conditionnellement aux effets aléatoires v et w , la vraisemblance conditionnellement aux effets aléatoires s'écrit :

$$L(\boldsymbol{\beta}, \boldsymbol{\gamma}|v, w) = \prod_{i=1}^G L_i(\boldsymbol{\beta}, \boldsymbol{\gamma}|v_i, w_i)$$

où

$$\begin{aligned} L_i(\boldsymbol{\beta}, \boldsymbol{\gamma}|v_i, w_i) &= \prod_{j=1}^{n_i} \left\{ f(t_{ij}|\mathbf{x}_{ij}, \mathbf{z}_{ij}, v_i, w_i)^{\delta_{ij}} \times S(t_{ij}|\mathbf{x}_{ij}, \mathbf{z}_{ij}, v_i, w_i)^{1-\delta_{ij}} \right\} \\ &= \prod_{j=1}^{n_i} \left\{ \pi(\mathbf{z}_{ij}|v_i) f(t_{ij}|U=1, \mathbf{x}_{ij}, w_i) \right\}^{\delta_{ij}} \\ &\quad \times \left\{ (1 - \pi(\mathbf{z}_{ij}|v_i)) + \pi(\mathbf{z}_{ij}|v_i) S(t_{ij}|U=1, \mathbf{x}_{ij}, w_i) \right\}^{1-\delta_{ij}} \end{aligned}$$

La vraisemblance marginale s'obtient pour le groupe i en intégrant la vraisemblance conditionnelle sur chaque effet aléatoire, telle que

$$L_i(\boldsymbol{\beta}, \boldsymbol{\gamma}, \theta_v, \theta_w) = \int_v \int_w L_i(\boldsymbol{\beta}, \boldsymbol{\gamma}|v_i, w_i) g(v_i; \theta_v) g(w_i; \theta_w) dv dw, \quad w \perp v \quad (5.3)$$

lorsque les effets aléatoires sont considérés comme indépendants et

$$L_i(\boldsymbol{\beta}, \boldsymbol{\gamma}, \theta_v, \theta_w) = \int_v \int_w L_i(\boldsymbol{\beta}, \boldsymbol{\gamma}|v_i, w_i) g(v_i, w_i; \boldsymbol{\theta}) dv dw \quad (5.4)$$

lorsque les effets aléatoires sont non-indépendants.

Les groupes étant considérés comme indépendants, la vraisemblance complète est donnée par le produit des vraisemblances de chaque groupe,

$$L(\boldsymbol{\beta}, \boldsymbol{\gamma}, \theta_v, \theta_w) = \prod_{i=1}^G L_i(\boldsymbol{\beta}, \boldsymbol{\gamma}, \theta_v, \theta_w). \quad (5.5)$$

Modèle à un effet aléatoire commun

La vraisemblance, conditionnellement à l'effet aléatoire s'écrit

$$L(\boldsymbol{\beta}, \boldsymbol{\gamma}, \beta_v | v) = \prod_{i=1}^G L_i(\boldsymbol{\beta}, \boldsymbol{\gamma}, \beta_v | v_i),$$

et, en utilisant le même raisonnement que précédemment, la vraisemblance marginale pour l'ensemble des groupes est donnée par

$$L(\boldsymbol{\beta}, \boldsymbol{\gamma}, \beta_v, \theta_v) = \prod_{i=1}^G \int_v L_i(\boldsymbol{\beta}, \boldsymbol{\gamma}, \beta_v | v_i) g(v_i; \theta_v) dv$$

5.3 Estimation

Les vraisemblances ne peuvent être calculées analytiquement et nécessitent le calcul numérique de une ou deux intégrales. Dans un programme Fortran, nous avons utilisé les quadratures de Gauss-Hermite pour obtenir une approximation de ces intégrales avec un nombre de points de quadratures variant de 5 à 50. Les méthodes des quadratures adaptatives de Gauss (Pinheiro et Bates, 1995) et l'approximation de Laplace de la log-vraisemblance ont aussi été utilisées avec le logiciel SAS®(procédure NLMIXED). Nous avons pour objectif initial de développer une estimation par vraisemblance pénalisée, afin d'étendre l'approche présentée au chapitre précédent aux modèles à effets aléatoires. Des problèmes d'estimation ont cependant été rencontrés, liés à un mauvais conditionnement de la matrice Hessienne de la logvraisemblance utilisée par l'algorithme de Marquardt. A ce jour ces problèmes ne sont pas résolus et nous présentons des résultats basés sur un modèle dans lequel la distribution des temps de survie est paramétriquement spécifiée selon une loi de

Weibull. Afin d'offrir plus de souplesse, un modèle exponentiel par morceaux (Chen et al., 1999; Ibrahim et al., 2001) sera utilisé dans l'application à la tremblante du mouton.

5.4 Etude par simulations

5.4.1 Générations des données

Les données sont générées suivant un modèle logistique-weibull. La fraction à risque est modélisée par un modèle logistique tel que

$$p = \frac{\exp(\beta_0 + \beta_1 X)}{1 + \exp(\beta_0 + \beta_1 X)}$$

où X est une variable binaire (0, 1). Pour l'ensemble des simulations effectuées, β_0 est fixé à 0 et β_1 à 0.5, correspondant à une fraction à risque de 50% dans le groupe de référence ($X=0$) et à 62.24% dans le groupe $X=1$. Les temps d'événement T_{ij} des individus à risque sont générés selon un modèle de Weibull, tels que $S(T_{ij}|U_{ij} = 1, X) = \exp[-(\xi T_{ij})^\rho \exp(\gamma_1 X_{ij})]$. Nous avons fixé $\xi = 0.9$, $\rho = 1.2$ et $\gamma_1 = \ln(0.5)$. Les temps de censure C_{ij} sont générés indépendamment des T_{ij} selon une loi uniforme $U[0; 10]$. Les temps de survie observés sont alors $Y_{ij} = \min(T_{ij}, C_{ij})$. Dans ces conditions environ 15% des individus à risque ont des temps de survie censurés.

La taille de chaque groupe est de $n=20$, 50 ou 100. Le nombre de groupes G est fixé à 10 ou 50. Pour chaque jeu de simulation, $M=200$ échantillons sont générés.

Pour les modèles à deux effets aléatoires indépendants, deux combinaisons de (θ_v, θ_w) sont considérées : (0.1,0.1) et (0.5,0.5). Pour le modèle avec un effet aléatoire commun, nous avons considéré plusieurs possibilités pour (β_v, θ_v) : (0,0.2), (0,0.5), (0.5,0.2) et (0.2,0.5).

Pour chaque paramètre c du modèle, les résultats sont exprimés en termes de biais moyen, $\hat{B}(\hat{c}) = [\sum_{i=1}^M (\hat{c}_i - c_0)]/M$, où \hat{c}_i est l'estimateur du paramètre c pour l'échantillon i , et c_0 la valeur simulée. Pour chaque estimateur, nous avons aussi calculé la variance empirique, $\text{var}(\hat{c}) = [\sum_{i=1}^M (\hat{c}_i - \bar{c})^2]/(M - 1)$, où $\bar{c} = [\sum_{i=1}^M \hat{c}_i]/M$,

et la moyenne des variances asymptotiques, $[\sum_i^M \text{var}(\hat{c}_i)]/M$. Nous avons également calculé le pourcentage de couverture de l'intervalle de confiance à 95 % du paramètre c , qui correspond à la proportion de simulations telles que $c_0 \in [\hat{c}_i - 1.96\sqrt{\text{var}(\hat{c}_i)}; \hat{c}_i + 1.96\sqrt{\text{var}(\hat{c}_i)}]$.

5.4.2 Résultats

Les résultats des simulations dans le cas d'effets aléatoires indépendants sont présentés dans le tableau 5.1 dans le cas où $G = 10$ groupes sont considérés et dans le tableau 5.2 pour $G=50$ groupes. Les biais sur les paramètres de régression β_0 , β_1 et γ_1 et les variances sont d'autant plus importants que le nombre et la taille des groupes est petite et ne sont plus négligeables dans le cas où $G = 10$ et $n = 20$. L'estimation de la variance des effets aléatoires est satisfaisante dans la plupart des cas, mais les biais sont plus faibles lorsque le nombre de groupes est élevé. De même, la variance asymptotique des estimateurs de θ_v et θ_w est d'autant plus importante que le nombre de groupe est faible et que la taille des groupes est petite. Les pourcentages de couverture des paramètres de régression sont acceptables dans tous les cas pour $G=50$, mais sont médiocres pour $\hat{\theta}_v$ et $\hat{\theta}_w$ lorsque $G = 20$ et $n = 20$. Des résultats similaires, non montrés ici, sont obtenus avec le modèle à un effet aléatoire commun.

Dans l'application sur la tremblante du mouton (voir chapitre 7), si le groupe est défini par l'élevage alors $G = 11$ et $n > 300$. Dans le cas où les animaux d'un même millésime de naissance au sein d'un élevage sont considérés comme l'unité de groupe, alors $G > 50$ et $n > 50$. Les résultats des simulations indiquent que, dans ces conditions, l'utilisation d'un modèle de mélange avec fraction non à risque et deux effets aléatoires est possible.

TAB. 5.1 – Biais moyen, variance empirique, variance asymptotique et pourcentage de couverture pour $G=10$ groupes (200 réalisations)

Paramètre	Estimation	Biais	Variance	Variance	%
	moyenne	moyen	empirique	asyp.	couverture
Nombre de groupes : 10, Taille de groupe : 20, $\theta_v = 0.1$, $\theta_w = 0.1$					
β_0	0.0244	0.0244	0.0580	0.0605	94.95
β_1	0.6339	0.1339	0.1261	0.1419	95.96
γ_1	-0.7666	-0.0734	0.0681	0.0674	93.27
θ_v	0.1011	0.0011	0.0229	0.0359	57.24
θ_w	0.0962	-0.0038	0.0127	0.0162	71.04
Nombre de groupes : 10, Taille de groupe : 50, $\theta_v = 0.1$, $\theta_w = 0.1$					
β_0	0.0170	0.0170	0.0305	0.0292	93.97
β_1	0.5166	0.0166	0.0389	0.0476	96.98
γ_1	-0.6728	0.0204	0.0236	0.0247	95.48
θ_v	0.0931	-0.0069	0.0078	0.0080	91.91
θ_w	0.0958	-0.0042	0.0055	0.0063	93.92
Nombre de groupes : 10, Taille de groupe : 100, $\theta_v = 0.1$, $\theta_w = 0.1$					
β_0	-0.0060	-0.0060	0.0170	0.0198	96.89
β_1	0.5105	0.0105	0.0255	0.0236	94.82
γ_1	-0.6941	-0.0010	0.0102	0.0123	97.93
θ_v	0.1006	0.0006	0.0051	0.0062	96.01
θ_w	0.0985	-0.0015	0.0037	0.0037	95.31
Nombre de groupes : 10, Taille de groupe : 20, $\theta_v = 0.5$, $\theta_w = 0.5$					
β_0	0.0125	0.0125	0.1143	0.1095	94.24
β_1	0.5783	0.0783	0.1335	0.1579	96.27
γ_1	-0.7644	-0.0713	0.0786	0.0761	93.56
θ_v	0.4835	-0.0165	0.1758	0.1636	82.37
θ_w	0.4975	-0.0025	0.1532	0.1542	86.61
Nombre de groupes : 10, Taille de groupe : 50, $\theta_v = 0.5$, $\theta_w = 0.5$					
β_0	0.0302	0.0302	0.0714	0.0703	94.28
β_1	0.5034	0.0034	0.0571	0.0561	93.85

suite...

... suite du tableau 5.1

Paramètre	Estimation	Biais	Variance	Variance	%
	moyenne	moyen	empirique	asypm.	couverture
γ_1	-0.6771	0.0160	0.0257	0.0268	96.41
θ_v	0.4651	-0.0349	0.1060	0.1134	91.03
θ_w	0.5013	0.0013	0.0894	0.0792	96.41
Nombre de groupes : 10, Taille de groupe : 100, $\theta_v = 0.5$, $\theta_w = 0.5$					
β_0	-0.0043	-0.0043	0.0623	0.0643	96.43
β_1	0.5077	0.0077	0.0301	0.0274	93.97
γ_1	-0.6994	-0.0063	0.0135	0.0128	95.48
θ_v	0.5258	0.0258	0.0900	0.0843	97.89
θ_w	0.5017	0.0017	0.0237	0.0200	96.83

TAB. 5.2 – Biais moyen, variance empirique, variance asymptotique et pourcentage de couverture pour $G=50$ groupes (200 réalisations)

Paramètre	Estimation	Biais	Variance	Variance	%
	moyenne	moyen	empirique	asypm.	couverture
Nombre de groupes : 50, Taille de groupe : 20, $\theta_v = 0.1$, $\theta_w = 0.1$					
β_0	-0.0048	-0.0048	0.0160	0.0115	91.96
β_1	0.5563	0.0563	0.0235	0.0219	93.47
γ_1	-0.7141	-0.0210	0.0098	0.0113	96.48
θ_v	0.0974	-0.0026	0.0054	0.0057	89.45
θ_w	0.0965	-0.0035	0.0026	0.0024	91.96
Nombre de groupes : 50, Taille de groupe : 50, $\theta_v = 0.1$, $\theta_w = 0.1$					
β_0	0.0063	0.0063	0.0071	0.0058	91.96
β_1	0.5150	0.0150	0.0084	0.0086	95.48
γ_1	-0.6931	0.0001	0.0042	0.0043	93.47
θ_v	0.0979	-0.0021	0.0018	0.0018	95.95
θ_w	0.0995	-0.0005	0.0011	0.0010	94.46
Nombre de groupes : 50, Taille de groupe : 100, $\theta_v = 0.1$, $\theta_w = 0.1$					
β_0	0.0015	0.0015	0.0073	0.0080	95.3265
β_1	0.4989	-0.0011	0.0080	0.0076	94.3878

suite...

... suite du tableau 5.2

Paramètre	Estimation	Biais	Variance	Variance	%
	moyenne	moyen	empirique	asypm.	couverture
γ_1	-0.7008	-0.0076	0.0039	0.0038	94.8980
θ_v	0.0985	-0.0015	0.0018	0.0017	95.1020
θ_w	0.0999	-0.0001	0.0012	0.0012	95.0000
Nombre de groupes : 50, Taille de groupe : 20, $\theta_v = 0.5$, $\theta_w = 0.5$					
β_0	-0.0095	-0.0095	0.0280	0.0211	92.50
β_1	0.5187	0.0187	0.0263	0.0246	96.00
γ_1	-0.6990	-0.0058	0.0143	0.0126	93.00
θ_v	0.5110	0.0110	0.0328	0.0337	93.00
θ_w	0.4986	-0.0014	0.0181	0.0211	92.50
Nombre de groupes : 50, Taille de groupe : 50, $\theta_v = 0.5$, $\theta_w = 0.5$					
β_0	0.0088	0.0088	0.0170	0.0146	93.91
β_1	0.5049	0.0049	0.0093	0.0096	93.91
γ_1	-0.6914	0.0017	0.0044	0.0046	95.94
θ_v	0.5073	0.0073	0.0187	0.0186	94.88
θ_w	0.5014	0.0014	0.0141	0.0120	94.76
Nombre de groupes : 50, Taille de groupe : 50, $\theta_v = 0.5$, $\theta_w = 0.5$					
β_0	0.0037	0.0037	0.0139	0.0127	94.51
β_1	0.5008	0.0008	0.0086	0.0071	94.87
γ_1	-0.6987	-0.0055	0.0013	0.0011	96.41
θ_v	0.5035	0.0035	0.0154	0.0159	95.89
θ_w	0.4934	-0.0066	0.0137	0.0133	93.07

Chapitre 6

Application à la tremblante du mouton : une première analyse

6.1 Contexte

6.1.1 Les différents programmes de recherche

Dès 1995 des programmes de recherche et développement ont été mis en place dans des élevages ovins laitiers de race Manech Tête Rousse atteints par la tremblante dans le département des Pyrénées Atlantiques (France). En 1999, le programme européen baptisé “Sheeprion” a été initié, avec pour objectif l’étude de la transmission de la tremblante et plus particulièrement l’identification des modes de contamination et des réservoirs de l’agent infectieux (notamment les parasites gastro-intestinaux des ovins, les acariens et les petits rongeurs). Ce programme incluait des partenaires français (Institut National de la recherche Agronomique, INRA, École Nationale Vétérinaire de Toulouse ENVT, Centre Départemental de l’Élevage Ovin, CDEO), espagnols (Centre de recherche de Derio) et islandais (Université de Reykjavik) (Gruner, 2003). En 1999 également, le projet dit “64” d’utilisation de l’outil génétique (utilisation de béliers ARR/ARR) comme moyen de lutte contre la tremblante, alternatif de l’abattage des troupeaux atteints, a débuté dans des élevages atteints du département des Pyrénées Atlantique. Depuis 2003, un autre programme européen, d’acronyme “Scrapie-Free-Sheep” (Barillet, 2003) inclut un plus grand nombre d’élevages infectés, avec pour objectif d’évaluer l’efficacité de la stratégie génétique, avec les mêmes partenaires français, plus des partenaires hollandais (ID Lelystad), italiens (Istituto Zootecnico E Ceseario per la Sardegna, Sassari) et norvégiens (École Nationale Vétérinaire d’Oslo).

6.1.2 Contraintes liées à la Réglementation

Depuis juin 1996, la tremblante ovine et caprine est considérée en France comme une Maladie Réputée Contagieuse (MRC) (décret n 96-258 du 14 juin 1996 complétant et modifiant la liste de maladies des animaux réputées contagieuses, Journal Officiel de la République Française du 15 juin 1996). A ce titre, elle est soumise aux règles

de police sanitaire fixées par l'arrêté ministériel du 28 mars 1997 (JORF du 17 avril 1997). Les principales dispositions de cet arrêté sont les suivantes :

Après suspicion d'un cas clinique de plus d'un an et information de la direction des services vétérinaires par le vétérinaire sanitaire, le Préfet prend un Arrêté Préfectoral de Mise sous Surveillance (APMS) qui impose :

- l'isolement de l'animal suspect et son euthanasie sur place ou à l'abattoir avec réalisation de prélèvements pour confirmation du diagnostic par histologie de l'encéphale ;
- le recensement et le contrôle de l'identification de l'ensemble des ovins et caprins présents ;
- la réalisation d'une enquête épidémiologique dans l'exploitation en liaison avec le vétérinaire sanitaire afin de repérer l'ensemble des animaux du cheptel susceptibles d'être atteints et/ou de transmettre la tremblante ;
- l'interdiction de sortie de l'exploitation des animaux, sauf à destination directe d'un abattoir ou d'un établissement d'études et de recherches.

La confirmation du cas index par un laboratoire agréé entraîne :

- l'euthanasie sans délai, à l'abattoir ou dans l'exploitation même, sur instruction du directeur des services vétérinaires, de tous les animaux de l'exploitation présentant des symptômes traduisant des troubles neurologiques ;
- l'isolement et le marquage spécifique des animaux repérés au terme de l'enquête épidémiologique et l'interdiction de sortie de l'exploitation de ces animaux, sauf à destination directe d'un abattoir ou d'un établissement d'études et de recherches.

L'arrêté du 28 mars 1997 a été complété par l'arrêté du 23 mars 1999, qui impose l'abattage des animaux marqués dans un délai de six mois.

Les arrêtés ministériels du 15 mars 2002, puis du 27 janvier 2003, abrogent les arrêtés précédents. Ils définissent de nouvelles mesures de police sanitaire qui tiennent compte des connaissances sur la susceptibilité génétique à la tremblante (génotype au gène PrP pour les codons 136, 154 et 171). Les principales modifications

sont les suivantes :

- l'âge des animaux pouvant être considérés comme suspects est abaissé à 6 mois ;
- le génotypage au gène PrP de l'animal suspect est effectué ;
- la confirmation du cas index entraîne la mise sous Arrêté Préfectoral Portant Déclaration d'Infection (APDI) de l'exploitation d'où provient l'animal et la mise sous APMS de toutes les exploitations où l'animal a été mis bas, ainsi que de l'exploitation où est né l'animal ;
- le génotypage au gène PrP est alors réalisé sur l'ensemble des ovins de l'exploitation sous APDI ;
- tous les ovins de génotypes sensibles (ARQ/ARQ, ARQ/AHQ, AHQ/AHQ) ou très sensibles (porteurs de l'allèle VRQ) sont marqués et isolés ;
- les animaux marqués doivent être euthanasiés dans un délai de 1 mois ;
- le repeuplement du cheptel ovin ne doit se faire qu'à l'aide de mâles résistants ARR/ARR et de femelles au moins hétérozygotes résistantes, à l'exclusion des femelles ARR/VRQ ;
- dans le cas d'un élevage ayant une forte proportion de femelles sensibles ou très sensibles et après engagement dans un programme d'amélioration génétique, l'éleveur peut être autorisé à conserver pendant un délai maximal de deux ans les femelles reproductrices sensibles, pour les accoupler avec des béliers ARR/ARR en vue du repeuplement ; passé ce délai, ces femelles doivent être abattues.

Les éleveurs adhérents au Projet 64 ont été dérogataires vis à vis de la police sanitaire de 1999, de sorte que les animaux désignés comme susceptibles d'être atteints et/ou de transmettre la tremblante n'ont pas dû être abattus dans le délai légal de 6 mois. Cette dérogation a permis le suivi individuel de chaque animal, de sa naissance à sa fin de carrière de production, ou à l'apparition des symptômes et son euthanasie pour cause de tremblante.

L'arrêté du 15 mars 2002 puis du 27 janvier 2003 a imposé, y compris pour les

élevages adhérents aux divers programmes de recherche, l'abattage soit immédiat, soit différé des brebis de génotypes sensibles. Compte tenu de la structure génétique majoritairement sensible de la Race Manech Tête Rouse, la plupart des élevages adhérents aux programmes de recherche ont bénéficié du délai de deux ans prévu par la loi. Mais comme les éleveurs engagés dans ces programmes de recherche utilisaient des béliers résistants avant 2002, la plupart ont choisi d'éliminer les femelles de génotypes sensibles en 2002 et 2003. Dans ces conditions, l'application de ces nouvelles mesures de police sanitaire n'a plus permis le suivi systématique de ces animaux, les plus à risque de développer la tremblante, jusqu'à l'apparition des signes cliniques de la maladie. Ces animaux ont cependant pour la plupart fait l'objet d'une recherche de PrPres dans l'encéphale ou les amygdales par les techniques définies par l'arrêté ministériel du 15 mars 2002. La mise en place de l'outil génétique dans la lutte contre la tremblante (utilisation de béliers ARR/ARR) a débuté pour les élevages adhérents au Projet 64 avant la mise en application de l'arrêté du 15 mars 2002 et a induit une profonde modification de la structure génétique des troupeaux suivis.

6.2 Une première analyse

Nous présentons dans cette partie une première application des modèles de survie avec fraction non à risque à l'analyse de facteurs de risque individuels de la tremblante du mouton. L'objectif de ce travail était (i) de montrer, à partir de simulations, quelles étaient les performances d'un modèle semi-paramétrique de survie avec fraction non à risque dans différentes conditions de durée de suivi (ii) de réaliser une première application sur données réelles, à partir d'un sous-échantillon des données recueillies dans le cadre des programmes dit "64" et "Sheeprion" afin de montrer l'intérêt d'une telle approche statistique. Ce travail a fait l'objet d'une publication dans *Journal of General Virology*.

L'analyse des données a été réalisée par un modèle de mélange semi-paramétrique,

à l'aide de la macro SAS présentée au chapitre 3.

Résumé de l'article

La longue durée d'incubation de la maladie et les biais de confusion liés aux durées d'observation des animaux rendent une analyse pertinente des épidémies de tremblante difficile à mener. Dans cette étude, 641 brebis de race Manech Tête Rousse, issues de six troupeaux naturellement atteints des Pyrénées Atlantiques, France, ont été suivies pendant six à neuf années. Durant cette période de suivi, 170 cas cliniques ont été confirmés et plus de la moitié des animaux réformés sans symptômes a fait l'objet d'un diagnostic post-mortem d'infection par la tremblante. L'analyse des données recueillies a fait appel à un modèle de survie avec fraction non à risque, permettant l'exploration de l'influence des variables explicatives sur le risque de contamination et sur la durée d'incubation des animaux considérés comme infectés. Des simulations ont été réalisées afin d'évaluer les conditions d'application de l'approche statistique considérée. Les résultats indiquent que les estimateurs des paramètres de régression et des durées d'incubation restent peu biaisés quand la censure n'est pas trop précoce. Cette condition est respectée dans les élevages étudiés. Les variables explicatives considérées dans l'analyse des données réelles sont le génotype au gène PRP et l'élevage d'origine. Les résultats montrent que, comparés aux animaux ARQ/ARQ, les animaux hétérozygotes ARR étaient moins à risque de s'infecter et qu'ils avaient des durées d'incubation plus longues. A l'inverse l'allèle VRQ était associé à un risque de contamination plus important, mais à des durées d'incubation comparables. Les durées d'incubation variaient significativement entre les élevages étudiés, mais un risque de contamination accru n'était pas nécessairement associé à des durées d'incubation réduites. Ces résultats suggèrent que d'autres facteurs, la nature de l'agent infectieux ou les pratiques d'élevages par exemple, interviennent dans la dynamique de l'épidémie dans les troupeaux infectés par la tremblante.

Advanced survival models for risk-factor analysis in scrapie

Fabien Corbière,^{1,2} Francis Barillet,³ Olivier Andréoletti,¹ Francis Fidelle,⁴ Nathalie Laphitz-Bordet,⁵ François Schelcher¹ and Pierre Joly²

Correspondence

Fabien Corbière

fabien.corbiere@isped.
u-bordeaux2.fr

¹UMR Interactions Hôtes Agents Pathogènes, Ecole Nationale Vétérinaire de Toulouse, 23 chemin des Capelles, 31076 Toulouse Cedex, France

²EMI 0338 (Biostatistique), Institut de Santé Publique et Développement, Université Victor Segalen Bordeaux 2, 146 rue Léo Saignat, 33076 Bordeaux Cedex, France

³Station d'Amélioration des Animaux, Institut National de la Recherche Agronomique, BP 27, 31326 Castanet-Tolosan cedex, France

⁴Centre Départemental d'Élevage Ovin, Quartier Ahetzia, 64130 Ordiarp, France

⁵Direction Départementale des Services Vétérinaires, Cours Lyautey, 64000 Pau, France

Because of the confounding effects of long incubation duration and flock management, accurate epidemiological studies of scrapie outbreaks are difficult to carry out. In this study, 641 Manech red-faced sheep from six scrapie-affected field flocks in Pyrénées Atlantiques, France, were monitored for clinical scrapie over a 6–9 year period. Over this period, 170 scrapie clinical cases were recorded and half of the culled animals were submitted for post-mortem transmissible spongiform encephalopathy diagnosis to assess their infectious status. Collected data were analysed using a 'mixture cure model' approach, which allowed for the discriminating effect of PrP genotype and flock origin on incidence and incubation period. Simulations were performed to evaluate the applicability of such a statistical model to the collected data. As expected, ARR heterozygote sheep were less at risk of becoming infected than ARQ/ARQ individuals and had a greater age at clinical onset. Conversely, when compared with ARQ/ARQ, the VRQ haplotype was associated with an increased infection risk, but not a shorter incubation period. Considering the flock effect, we observed that a high incidence rate was not associated with shorter incubation periods and that the incubation period could be significantly different in flocks harbouring similar infection risks. These results strongly support the conclusion that other parameters, such as the nature of the agent or flock management, could interfere with epidemiological dynamics of the infection in scrapie-affected flocks.

Received 27 February 2006

Accepted 3 October 2006

INTRODUCTION

Transmissible spongiform encephalopathies (TSE) are neurodegenerative disorders that occur primarily in sheep (scrapie), cattle (bovine spongiform encephalopathy; BSE) or humans (Creutzfeldt–Jakob disease; CJD). They all share similar characteristics, including long incubation periods, a progressive pattern of disease and a clinical course resulting in death (Fraser, 1976). Under natural conditions, it is considered that scrapie infection in sheep occurs mainly after an early oral contamination (around birth) (Andréoletti *et al.*, 2000; Heggebo *et al.*, 2000). Scrapie clinical onset generally occurs between 2 and 7 years (Detwiler & Baylis, 2003).

In prion diseases, the accumulation of an abnormal isoform (PrP^{Sc}) of a normal cellular protein (PrP^C) in tissues from infected individuals is currently considered as a disease hallmark. Most of the diagnostic tests currently available are

based on biochemical detection of the abnormal protein (McKinley *et al.*, 1983; Race *et al.*, 2001). However, post-mortem tests, as carried out in the current European surveillance programme (rapid test on the obex), are reliable only for detection of infected animals in the second half of the incubation period.

Because of long TSE incubation periods, data analysis is difficult without reference to flock demography and management. Indeed, infected individuals could be culled or die from other causes before clinical onset (intercurrent diseases, economic reasons). In this situation, no reliable information will be available about their infectious status (Begara-McGorum *et al.*, 2000; Ryder *et al.*, 2001; Thorgeirsdottir *et al.*, 2002; Billinis *et al.*, 2004).

Evaluation of genetic and environmental risk factors in scrapie has been conducted mainly using case–control

designs, in which a set of affected animals is compared with their healthy flock-mates or to a reference population (Hunter *et al.*, 1997; Thorgeirsdottir *et al.*, 1999; Tranulis *et al.*, 1999; O'Doherty *et al.*, 2002; Acin *et al.*, 2004; Baylis *et al.*, 2004; Billinis *et al.*, 2004; Tongue *et al.*, 2006). Such approaches have revealed that TSE susceptibility in sheep is controlled mainly by polymorphisms at codons 136 (T, V, A), 154 (R, H) and 171 (R, Q, H, K) of the PRNP gene (Cloucard *et al.*, 1995; Hunter *et al.*, 1996). V¹³⁶R¹⁵⁴Q¹⁷¹/VRQ, ARQ/VRQ and ARQ/ARQ animals are usually considered the most susceptible to scrapie, whereas homozygous or heterozygous AHQ and heterozygous ARR animals show only marginal susceptibility, ARR/ARR sheep being considered to be fully clinically resistant (Detwiler & Baylis, 2003).

Surveys based on long-term individual monitoring of an exposed population are less subject to sampling bias (Tongue *et al.*, 2006). Consequently, they should be considered as more relevant than case-control or cross-sectional studies for an accurate evaluation of the effect of environmental or genetic factors on infection rate and incubation period.

'Cure models' are part of the mixture models family (Bohning & Seidel, 2003). In 'mixture cure models', it is considered that the studied population is a mixture of susceptible (i.e. that may undergo the event of interest) and non-susceptible individuals (i.e. that will never undergo the event of interest) (Farewell, 1982). Unlike classical survival analysis, they allow a separate estimation of covariate effects on incidence and incubation length. Cure models also allow estimation of the proportion of healthy (or conversely infected) individuals in a population, including individuals that did not last the total length of the study (Lam *et al.*, 2005).

In this study, we propose a model based on a 'mixture cure models' approach for scrapie epidemiological analysis. Robustness and reliability boundaries of the model were assessed by simulations before analysing data collected over 6–9 years in six naturally scrapie-affected flocks in Pyrénées Atlantiques, France.

METHODS

Model design. In the model, death from scrapie is considered as the event of interest. Clinical scrapie cases are considered to be scrapie-infected (SI) and have uncensored observations over their lifetime. For apparently healthy sheep that are removed from the flock (right-censored records), it is not known whether they are scrapie-free (SF) or infected but removed prior to the onset of clinical scrapie. For each right-censored animal, the model computes the probability of being scrapie-infected, given its age and covariates information.

The model assumes that: (i) most or all deaths from scrapie occur during a determined age-period, (ii) monitoring is long enough for clinical onset to have appeared in most of the infected animals and (iii) the longer an animal lives, the lower the probability of it being scrapie

infected. Animals which live longer than the last observed scrapie clinical case are considered to have an extremely low probability (if not zero) of incubating scrapie.

If U is the indicator denoting SI animals (i.e. $U=1$ if the animal is scrapie-infected and $U=0$ if non-infected) and T is a non-negative random variable denoting the failure time of interest, defined only if $U=1$, the mixture cure model is given as follows:

$$S(t|\mathbf{x}, \mathbf{z}) = \pi(\mathbf{z}) S(t|U=1, \mathbf{x}) + [1 - \pi(\mathbf{z})]$$

where $S(t|\mathbf{x}, \mathbf{z})$ is the unconditional (marginal) survival function of T for the entire population under study (that is SF and SI groups) and $\pi(\mathbf{z}) = P(U=1|\mathbf{z})$ is the probability of being infected given a covariate vector $\mathbf{z} = (z_1, \dots, z_q)'$. $S(t|U=1, \mathbf{x}) = P(T > t|U=1, \mathbf{x})$ is the conditional survival function for SI animals given a covariate vector $\mathbf{x} = (x_1, \dots, x_m)'$ (it may include the same covariate as \mathbf{z}). The use of conditional attached to this function is to stress that the distribution of time refers not to the whole group of animals but only to the animals that are in the SI group. Note that $S(t|\mathbf{x}, \mathbf{z}) \rightarrow 1 - \pi(\mathbf{z})$ as $t \rightarrow \infty$, where $1 - \pi(\mathbf{z})$ represents the proportion of non-infected animals. When $\pi=1$, that is when no SF portion is assumed, the model reduces to the traditional survival model. Whether the inclusion of a proportion of SF animals leads to a significantly better fit to the data than a traditional survival model with no SF animals can be tested by the deviance test statistics proposed by Maller & Zhou (1996).

Various parametric and semiparametric approaches have been proposed for mixture cure models (Peng & Carriere, 2002; Lam *et al.*, 2005). For modelling the influence of exploratory variables on the incidence, a logistic regression model is usually chosen (Kuk & Chen, 1992; Peng & Dear, 2000):

$$\pi(\mathbf{z}) = p(U=1|\mathbf{z}) = \frac{\exp(\beta' \mathbf{z})}{1 + \exp(\beta' \mathbf{z})}$$

where β is the vector of regression parameters associated with \mathbf{z} and contains an intercept. The conditional survival function of infected animals is modelled through the semiparametric Cox proportional hazards (PH) model (Cox, 1972), which is given by

$$S(t|U=1, \mathbf{x}) = S_0(t|U=1) \exp(\gamma' \mathbf{x})$$

where γ is the vector of regression parameters associated with \mathbf{x} and $S_0(t|U=1)$ is the baseline conditional survival function, which is left unspecified.

Through the vectors of regression parameters β and γ , the mixture survival model is able to separate the effects of the covariates on incidence and latency. An estimate of the true proportion of SI animals, SI_{pop} given \mathbf{z} and \mathbf{x} , is provided by taking the mean of the individual probabilities $P_i(U=1|\mathbf{z}_i, \mathbf{x}_i)$:

$$SI_{pop} = \frac{1}{N} \sum_i^N P_i(U=1|\mathbf{z}_i, \mathbf{x}_i) \\ = \frac{1}{N} \sum_i^N \left[\delta_i + (1 - \delta_i) \frac{\pi_i(\mathbf{z}_i) S(t_i|U=1, \mathbf{x}_i)}{1 - \pi_i(\mathbf{z}_i) + \pi_i(\mathbf{z}_i) S(t_i|U=1, \mathbf{x}_i)} \right]$$

where δ_i is the censoring indicator with 1 if t_i is uncensored and 0 otherwise. Obviously, if $\delta_i=1$, then $P_i(U=1)=1$. When $\delta_i=0$, then $P_i(U=1|\mathbf{z}_i, \mathbf{x}_i)$ will depend on the survival length and will drop to zero as $t \rightarrow \infty$. Note that the better the model fits the data, through covariate vectors \mathbf{z}' and \mathbf{x}' , the more accurate is the estimation of the proportion of SI animals.

Simulation studies. Simulations were conducted to test (i) the ability of the model to estimate the proportion of SI animals and to discriminate covariate effects on the infection risk and incubation

duration and (ii) the effect of individual monitoring length on model outputs. We assumed that non-infected (SF) animals would never die from scrapie. Consequently, observations conducted on SF animals were right-censored. SI animals either died from clinical scrapie (uncensored records) or were eliminated from the flock before clinical onset (right-censored). Simulations were performed using (i) genetic and biological parameters (infection rates, ages at clinical death and flock demography) described in scrapie outbreaks or already used in mathematical modelling (simulation design 1) (Matthews *et al.*, 2001; Hagenaars *et al.*, 2003; Hopp *et al.*, 2003; Baylis *et al.*, 2004; Eglin *et al.*, 2005; Touzeau *et al.*, 2005) and (ii) ages at death from scrapie reflecting observations made in our dataset (simulation design 2).

The capacity of the model to separate covariate effects on incidence and incubation length (age at death) was assessed by generating two independent binary covariates, one (Z_1) affecting only the incidence and the other (Z_2) affecting only the incubation duration. The incidence is given the logistic form

$$\pi(z_1, z_2) = \frac{\exp(\beta_0 + \beta_1 Z_1 + \beta_2 Z_2)}{1 + \exp(\beta_0 + \beta_1 Z_1 + \beta_2 Z_2)}$$

where β_1 and β_2 are the effects of covariates Z_1 and Z_2 on the proportion of infected individuals, respectively. Since Z_2 should have no effect on the incidence, β_2 was set to 0. Thus, the proportion of infected animals is

$$\pi(z_1 = 0, z_2) = \frac{\exp(\beta_0)}{1 + \exp(\beta_0)}$$

for animals sharing $Z_1 = 0$ and

$$\pi(z_1 = 1, z_2) = \frac{\exp(\beta_0 + \beta_1)}{1 + \exp(\beta_0 + \beta_1)}$$

for those sharing $Z_1 = 1$. We set $\beta_0 = -0.5$ and $\beta_1 = -1$, so that the corresponding proportions of infected sheep are 37.7% (animals with $Z_1 = 0$) and 18.22% (animals with $Z_1 = 1$), respectively.

The log-normal distribution was used as the distribution function for life expectancies of infected animals, with survival function

$$S(t|U=1) = 1 - \Phi\left[\frac{(\ln t - \mu - \gamma_1 Z_1 - \gamma_2 Z_2)}{\sigma}\right]$$

where Φ is the distribution function of the standard normal law. In contrast to the incidence portion, γ_1 was set to 0, because Z_1 should have no effect on latency, whatever the value of Z_2 . In simulation design 1, the scale (μ) and shape (σ) parameters for the log-normal distribution function were set to 1.2 and 0.4, respectively. In the absence of censoring, events of interest (scrapie deaths) were allowed to occur at median age 3.3 years (interquartile range 2.5–4.2,) for individuals with $Z_2 = 0$. We set $\gamma_2 = 0.3$ so that infected individuals with $Z_2 = 1$ would die later, at median age 4.4 years (interquartile range 3.4–5.7). In simulation design 2, we set $\mu = \ln 2$, $\sigma = 0.5$, $\gamma_1 = 0$ and $\gamma_2 = 0.375$, so that median ages at clinical onset were 2.00 (interquartile range 1.5–2.8) and 3.00 years (interquartile range 2.1–4.2) (Fig. 1a).

Monitoring length was generated according to the Weibull distribution $W(\lambda, \rho)$, with shape (λ) and scale (ρ) parameters. Thirteen different scenarios were investigated with median life expectancy (meaning monitoring length) ranging from 2.5 to 9.5 years. These scenarios covered a large panel of flock management policies and demography. For each scenario and simulation design, 500 independent datasets, each consisting of 500 individuals, were generated and submitted to model analysis. The absolute biases $[B(\hat{c}) = \sum_i(\hat{c}_i - c_0)/500]$ and mean squared errors $[MSE(\hat{c}) = \sum_i(\hat{c}_i - c_0)^2/500]$, where \hat{c}_i are the estimates of c_0 , were computed for the five parameter estimates.

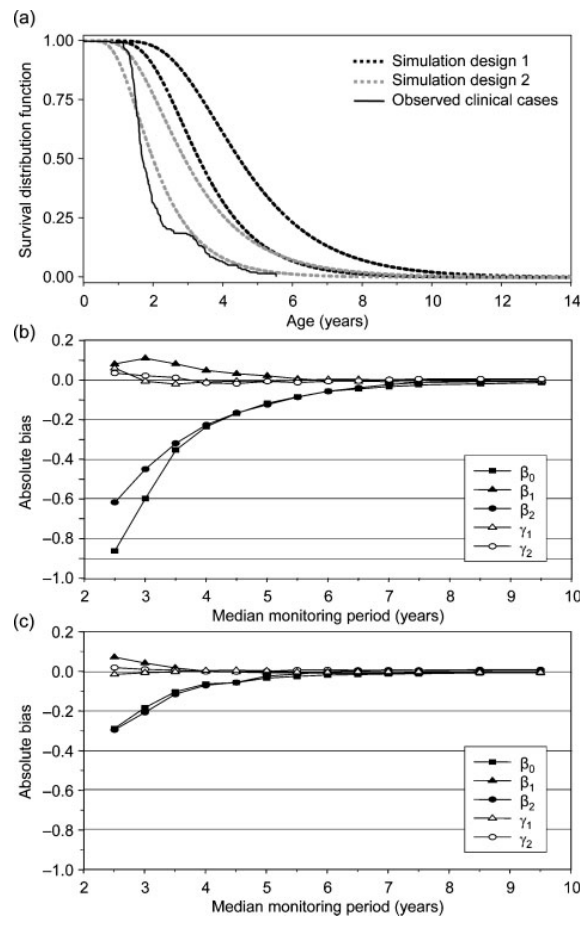


Fig. 1. (a) Simulation design for incubation length. Survival distribution functions for simulation design 1 (median age at onset 3.3 and 4.45 years) and simulation design 2 (median age at onset 2 and 3 years) and survival distribution function for observed scrapie clinical cases in the studied population are shown. (b, c) Absolute bias of estimates for each monitoring scenario (b, simulation design 1; c, simulation design 2) using the mixture cure model: ■ (β_0), proportion of infected individuals; ▲ (β_1), effect of covariate Z_1 on incidence; ● (β_2), effect of Z_2 on incidence; △ (γ_1), effect of covariate Z_1 on latency; ○ (γ_2), effect of covariate Z_2 on latency.

Sample generation and model computations were performed using SAS software (SAS-PC system, Version 8.2 for Windows; SAS Institute).

Flocks. Investigations were carried out on six naturally scrapie-affected dairy flocks, bred by private farmers, in Pyrénées Atlantiques, France. These flocks had been involved in a long-term scrapie epidemiology research project since 1994 (flock C) and 1998 (the other five flocks). Sheep were all Manech red-faced pure-breeds. Table 1 shows, for each flock, the mean flock size and the estimated year of first occurrence of scrapie. The high incidences of scrapie clinical suspicions (confirmed or not) in ewes born prior to enrolment in the research project suggest different, but nonetheless high, infection pressures (Table 1).

Table 1. Size and within-flock scrapie history for flocks A–F

Parameter	A	B	C	D	E	F
Flock size (adult ewes)*	321	331	302	463	396	380
Estimated start of scrapie outbreak	1992	1992	1993	1995	1996	1996
Number of scrapie clinical cases in ewes born before the studied birth cohorts	135	70	96	150	90	190

*Flock size at lambing of the studied birth cohorts.

At the time of enrolment, PrP polymorphism at codons 136, 154 and 171 of the PRNP gene was determined for all sheep, including breeding males and females and replacement lambs. Birth date, pedigree, date and cause of death or removal from the flock were systematically recorded for all the animals up to January 2006. No useful information could be collected for male or female lambs slaughtered at 1–3 months old.

To comply with the requirement of adequate monitoring length and the provision of high-quality data (including reliable diagnosis), only a few birth cohorts were considered for the analysis within each flock. Birth cohorts were selected in which all scrapie clinical suspicions were confirmed by histopathology and complete PrP genotype profiles were available.

The dataset submitted for analysis consisted of the 1998 birth cohort (born between October and December 1997) for flock A, the 1999 birth cohort (born between October and December 1998) for flocks B, D, E and F and the 1995, 1996 and 1997 birth cohorts in flock C (animals born in November and December 1994, 1995 and 1996, respectively) (Table 2). Only homebred animals were included in the analysis, while purchased sheep ($n=10$) were not considered. In total, our sample comprised 641 sheep, including 170 scrapie clinical cases.

Apparently healthy culled sheep were submitted for PrP^{Sc} detection on obex and palatine tonsils by ELISA (Platelia BSE detection kit; Bio-Rad) and immunohistochemistry (Andréoletti *et al.*, 2002). All the sheep included in our initial sample were not examined, either because of failure to track these animals during the elimination process or because they were still alive at the time of writing (January 2006). From these latter animals, palatine tonsils biopsies were performed each year from 2002 to 2005 (inclusive) for PrP^{Sc} immunohistochemistry detection (Andréoletti *et al.*, 2002). To account for missing information, a one-sided 95% confidence interval for the true proportion of infected animals in each PrP genotype group was calculated using the hypergeometric law.

In the mixture cure model analysis, PrP genotype and the flock were used as covariates and age at death from clinical scrapie was considered as the survival measurement. Ninety-five per cent confidence intervals for adjusted odds ratios (OR) from the logistic part and adjusted relative risks (RR) from the Cox PH part of the mixture cure model were computed using the bias corrected, accelerated bootstrap method (Davison & Hinkley, 1997).

RESULTS

Simulations

As expected, the longer an animal lived (long individual monitoring), the smaller were the mean absolute biases and mean squared errors (MSE, not shown) for the different estimates (Fig. 1b, c). Biases and MSE were acceptable for

inference when the monitoring time was longer than the median (theoretical) incubation duration. For the first set of simulated incubation times (median age at death 3.3 and 4.45 years), the estimates of the proportion of infected animals (β_0) and the effects of Z_2 on the incidence (β_2) were highly biased (absolute bias over 0.1) for median monitoring times of less than 5.5 years (Fig. 1b). For the second set of simulated incubation durations (median age at death 2 and 3.3 years), similar results were obtained for median monitoring times under 3.5 years (Fig. 1c). In our population, median age at death in scrapie-affected animals was 1.7 years (Fig. 1a; Table 3) and the median monitoring time was 4.90 years (Table 4). Under these conditions, the model outputs were expected to be relevant in the analysis of our dataset.

PrP genetic structure of the studied cohorts

Amongst the studied cohorts, the ARQ and ARR alleles were dominant, while AHQ and VRQ alleles were rare (Table 2). The genetic structure of the sample is consistent with previously reported PrP polymorphism frequencies in the Manech red-faced breed (Palhiere *et al.*, 2002).

Because of small numbers of individuals in some PrP genotypes, animals were grouped according to their level of susceptibility to classical scrapie (DEFRA, 2003). ARQ/ARQ, AHQ/AHQ and AHQ/ARQ sheep were considered in a medium-risk group (S/S; $n=343$). ARR/ARQ, ARR/AHQ and ARR/VRQ animals were included in a low-risk group (R/S; $n=212$). ARQ/VRQ, AHQ/VRQ and VRQ/VRQ animals were included in a single high-risk group (VRQ/x; $n=49$) (Table 3). Considering these PrP genotype groups, the genetic structure was not statistically different in the six selected flocks (chi-square test with 15 degrees of freedom: $\chi^2=15.95$, $P=0.38$).

Scrapie clinical cases

Clinical scrapie cases were mainly observed in ARQ/ARQ ($n=131$; 77.06%) and ARQ/VRQ ($n=28$; 16.47%) genotypes, while heterozygote ARR were poorly affected (R/S sheep: $n=9$; 5.29%) (Tables 2 and 3). High incidences in susceptible PrP genotypes ARQ/ARQ and ARQ/VRQ suggested a high infection pressure. No clinical cases were observed in ARR/ARR ($n=37$), ARR/VRQ ($n=3$), ARR/ARR ($n=3$) or AHQ/VRQ ($n=3$) animals. However, the

Table 2. PrP genotype distribution and number of clinical scrapie cases in the studied birth cohorts from flocks A–F

Group	PrP genotype	A		B			C			D	E	F	Total	Clinical cases (n)
		1998	1999	1995	1996	1997	1999	1999	1999					
S/S	ARQ/ARQ	43	42	46	43	25	45	53	34	331	131			
	AHQ/ARQ	2	–	3	3	–	3	1	–	12	1			
VRQ/x	ARQ/VRQ	8	5	3	–	6	6	10	9	47	28			
	AHQ/VRQ	–	–	1	–	–	–	–	–	1	0			
	VRQ/VRQ	–	–	–	1	–	–	–	–	1	1			
R/S	ARR/ARQ	28	24	18	15	29	33	34	25	206	9			
	ARR/AHQ	–	–	–	–	3	–	–	–	3	0			
	ARR/VRQ	1	–	–	–	–	1	1	–	3	0			
R/R	ARR/ARR	7	–	–	4	7	7	10	2	37	0			
Total		89	71	71	66	70	95	109	70	641	170			

Table 3. Clinical scrapie cases per genotype group and median age at death (clinical scrapie) in the studied birth cohorts

Percentages are given in parentheses.

Flock	R/R	R/S	S/S	VRQ/x
A	7 (7.90)	29 (32.58)	45 (50.56)	8 (8.99)
B	0 (0)	24 (33.80)	42 (59.15)	5 (7.04)
C	11 (5.31)	65 (31.40)	120 (57.97)	11 (5.31)
D	7 (7.37)	34 (35.79)	48 (50.53)	6 (6.32)
E	10 (9.17)	35 (32.11)	54 (49.54)	10 (9.17)
F	2 (2.86)	25 (35.71)	34 (48.57)	9 (12.86)
Total	37 (5.77)	212 (33.07)	343 (54.51)	49 (7.64)
Clinical cases	0	9 (4.25)	132 (38.48)	29 (59.18)
Median age at death (years) [range]	–	3.61 [3.26–4.13]	1.62 [0.77–5.54]	2.05 [1.14–4.76]

number of animals with these last three genotypes was too small to draw any conclusions.

Kaplan–Meier plots of the survival distribution functions for scrapie clinical occurrence indicate the absence of new scrapie cases after 5.54 years, whatever the genotype group considered (Fig. 2). This lack of new clinical cases fulfils two basic requirements that allow the application of the Cox PH mixture cure model, i.e. (i) most or all deaths due to scrapie occurred in a defined age period and (ii) the monitoring

length was sufficient to allow almost all infected animals to show clinical signs.

Monitoring length in clinically healthy sheep

A large number ($n=264$; 60.83%) of clinically healthy animals were eliminated by breeders for husbandry reasons at younger ages than the last observed clinical case (Table 4). However, 94.24% of these symptomless sheep lived longer than the median age at clinical onset (1.70 years), and

Table 4. Monitoring duration (years) in apparently healthy sheep from the studied birth cohorts

PrP genotype group	n	Monitoring duration (years)			
		25th centile	Median	75th centile	Maximum
R/S	203	3.58	5.59	7.09	9.09
S/S	211	3.47	4.69	5.64	8.04
VRQ/x	20	2.51	3.90	4.43	8.07
Total	434	3.50	4.90	6.36	9.09

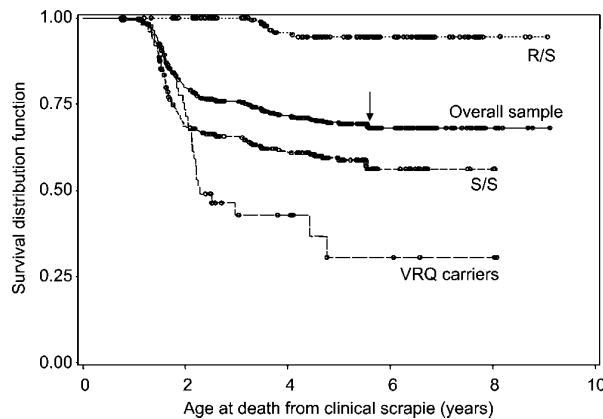


Fig. 2. Kaplan–Meier survival distribution functions for deaths from scrapie in the studied birth cohorts from six flocks of Manech red-faced sheep. Numbers in each group were: 604 (overall sample), 212 (R/S genotype group); 343 (S/S genotype group) and 49 (VRQ allele carrier group). Survival times are measured from date of birth. Censored events (i.e. culling or deaths from causes other than scrapie) are shown as open circles. The last observed clinical case (5.54 years old) is indicated by an arrow.

72.12 % reached 3.57 years old, which represented the 90th percentile of the age distribution of scrapie clinical cases.

Because of the implementation of French TSE legislation at the beginning of 2003, breeders had to remove VRQ carrier animals from scrapie-affected flocks. Consequently, VRQ animals, mainly ARR/VRQ, were eliminated earlier than expected and had a statistically shorter follow-up than R/S animals (analysis of variance; $F_{431,2} = 13.13, P < 10^{-4}$).

Active detection of subclinically infected sheep

Of the 471 clinically healthy sheep (with no clinical scrapie) eliminated during the study, 220 (46.71 %) were submitted

to post-mortem for PrP^{Sc} detection (mean age 4.72 years; 95 % confidence interval 1.87–7.57 years). Sampled animals represented 70 % of VRQ/x and 72.04 % of S/S but only 24.13 % of R/S and 13.51 % of R/R (Table 5).

From the tested animals, only four ARQ/ARQ sheep (aged 1.89, 2.25, 2.32 and 4.39 years) and one ARQ/VRQ sheep (3.04 years) were found to be positive. Fifty-nine animals initially included in the study were still alive at the time of writing (5 ARR/ARR, 42 ARR/ARQ and 12 ARQ/ARQ). No PrP^{Sc} was detected in any of these animals using tonsil biopsy. According to the hypergeometric law, the one-sided 95 % confidence interval for the number of apparently healthy but infected animals eliminated from the flocks was 5–25 (Table 5).

Results from the mixture cure model analysis

The deviance statistic test ($\chi^2_{01} = 30.93; P < 10^{-4}$) indicated that incorporating a scrapie-free fraction provided a better fit to the data than the traditional Cox PH model and that the estimated effects were more relevant.

Proportion of infected animals

ARR/ARR animals were not included in the analysis, because there were no confirmed clinical cases of this genotype. According to the proposed model, the predicted number of subclinically infected animals was 20.75 (respectively 3.65 in the R/S group, 14.83 in the S/S group and 3.22 in the VRQ carrier group; point estimate minus number of clinical cases) (Table 5). Strikingly, these predicted numbers were in close agreement with those obtained by the active detection of subclinical cases.

Genotype influence on incidence and incubation duration

Results from the Cox mixture cure model indicated that ARR heterozygote animals were at lower risk of infection than S/S animals (Table 6). Conversely, VRQ allele carriers (excluding ARR/VRQ) were at higher risk of being infected.

Table 5. Number and percentage of infected animals by PrP genotype group according to clinical cases and laboratory test findings in clinically healthy animals and the logistic Cox PH mixture cure model in the studied birth cohorts

PrP genotype group	Clinical cases (n)	Clinically healthy sheep [positive/tested (total)]	Estimated number (%) of infected sheep	
			Clinical + subclinical cases (95 % CI)*	Logistic Cox PH mixture cure model (point estimate)
R/R	0	0/5 (37)	–	–
R/S	9	0/49 (203)	9–20 (4.24–9.43)	12.65 (5.97)
S/S	132	4/152 (211)	136–142 (39.35–41.40)	146.83 (42.81)
VRQ/x	29	1/14 (20)	30–33 (61.22–67.34)	31.22 (63.71)
Total	170	5/220 (471)	175–195 (28.97–32.28)	190.70 (31.58)

*The lower bound is the observed proportion of SI sheep based on clinical survey and laboratory tests; the upper bound is derived from the hypergeometric law to account for clinically healthy sheep without laboratory test (n=251). The large confidence interval for R/S animals was due to the small proportion of animals sampled.

Table 6. Effects of PrP genotype and flock origin on incidence and incubation duration in the studied birth cohorts from six scrapie-infected flocks according to the logistic Cox PH mixture cure model

OR, Adjusted odds ratio; HR, adjusted hazard risk.

Variable	Incidence			Incubation duration		
	OR	CI _{95%} *	P†	HR	CI _{95%} *	P†
Genotype vs S/S						
R/S	0.08	0.04–0.14	<0.0001	0.28	0.14–0.56	0.0003
VRQ/x	2.71	1.41–5.21	0.003	0.84	0.55–1.28	0.4117
Flock vs flock C (ARQ/ARQ animals only)						
A	0.35	0.17–0.72	0.0045	1.75	0.95–3.22	0.0723
B	0.72	0.36–1.46	0.3695	2.49	1.46–4.25	0.0008
D	0.40	0.20–0.81	0.0113	1.92	1.08–3.43	0.0264
E	0.20	0.09–0.43	<0.0001	1.33	0.67–2.61	0.4144
F	1.74	0.76–3.89	0.1796	0.41	0.23–0.72	0.0021

*95% confidence interval for OR and HR estimated by the bias-corrected accelerated bootstrap method over 5000 resamples.

†Level of significance of the variable (*P* value). The S/S genotype group and flock C were used as the baseline. The influence of flock could not be investigated in the R/S and VRQ allele carriers groups due to the small numbers of clinical cases in some flocks.

Age at clinical onset was significantly greater in R/S animals when compared with S/S animals. No significant difference was found between S/S and VRQ allele carriers (Table 6).

Flock effect on incidence and incubation length

Because of insufficient numbers of ARQ/VRQ or ARR/ARQ clinical scrapie cases in some flocks, the analysis was restricted to the ARQ/ARQ animals (Table 6). The risk for ARQ/ARQ animals of being infected was significantly lower in flocks A, D and E than in flock C (Table 6). No significant difference was observed between flocks C, B and F. Interestingly, age at clinical onset was significantly less in some flocks with a lower infection risk (flock D compared with flock C). Meanwhile, age at clinical onset could be significantly different in flocks harbouring a similar infection risk: infected animals from flock F had delayed clinical onset compared with flock C while, in flock B, it was shortened. Taken together, these results strongly suggest that age at clinical onset and infection risk are not associated parameters.

DISCUSSION

Working hypothesis

Only a few studies of scrapie outbreaks based on longitudinal monitoring have been published (Elsen *et al.*, 1999; Diaz *et al.*, 2005). Using survival analysis, they aimed to determine the influence of PrP genotype, rearing type and dam clinical status on individual risk of developing clinical scrapie. Survival analysis assumes intrinsically that if, in an exposed population, a long enough and complete surveillance of individuals is possible, each would experience the event of interest, i.e. clinical scrapie. For TSE, this hypothesis is obviously inadequate. Indeed, under natural exposure,

ARR/ARR animals and a large proportion of susceptible-genotype animals will remain uninfected (Elsen *et al.*, 1999).

Because a mixed population of susceptible and non-susceptible individuals is considered, 'mixture cure models' appear to be an attractive approach for scrapie epidemiological analysis. However, several conditions must be fulfilled for their sound application. Such constraints require hypotheses about scrapie pathogenesis and biology. Amongst the basic hypotheses we considered were that animals born in an infected flock, if not infected in their early life, would remain negative. Under natural exposure conditions, scrapie contamination is considered to occur preferentially around birth (Andréoletti *et al.*, 2000; Heggebo *et al.*, 2000; van Keulen *et al.*, 2000). We therefore hypothesized that age at death from scrapie (clinical stage) was a relevant approximation of the incubation period. Moreover, animal susceptibility seems to decrease dramatically with age (Hourrigan *et al.*, 1979; Andréoletti *et al.*, 2000). Clinical cases have been reported in young and adult susceptible animals introduced to infected flocks (Hourrigan *et al.*, 1979; Ryder *et al.*, 2004) but the importance, relative to neonatal contamination, of such lateral transmission in adult sheep could not be estimated. The other main hypothesis we made was that very few (if not zero) infected individuals would be alive at the end of the study. The observed survival distribution plots were consistent with this hypothesis. However, existence of long-term subclinical carriers remains a major question of scrapie epidemiology. Currently, it is impossible to assume that an apparently healthy animal (whatever the test used to establish infectious status) is not incubating scrapie. Recent description of atypical cases or Nor98 cases in old and apparently healthy animals, and difficulties in assessing the diagnosis, sustain the 'long-term subclinical carriers'

hypothesis (Benestad *et al.*, 2003; Le Dur *et al.*, 2005). However, atypical scrapie occurs at a very low detected prevalence level (3–11 cases per 10 000 examined) and, in most cases, only one to three cases could be detected in stamped-out affected flocks (De Bosschere *et al.*, 2004; Onnasch *et al.*, 2004; Orge *et al.*, 2004). This implies that approximately 0.2–0.8 sheep could have been infected with atypical scrapie in the considered flocks, which is negligible when compared with the number of classical scrapie cases in the studied cohorts ($n=170$). At the population level, the influence of an atypical case on the model outputs was considered to be negligible.

Finally, even if the hypothesis of some adult lateral transmission and long-term subclinical carriers could not definitely be ruled out, both phenomena seemed marginal enough in our population to avoid major transgression from application of the model. From simulations, major biases were observed only when the monitoring length was shorter than the median (theoretical) incubation duration. Similar trends were obtained by Yu *et al.* (2004) when studying the influence of the follow-up length on the cure fraction estimation for several human cancers. The monitoring length in the studied sheep was long enough to ensure small biases for the estimates of PrP genotype and flock effects.

Asymptomatic culled animals

The mixture cure model approach enabled us to estimate the number of infected individuals and included those eliminated while incubating the disease. Model outputs and laboratory findings were in close agreement and indicated that a very small number of sheep were removed while incubating scrapie.

This result is consistent with observations from another study based on a longitudinal survey in a Texel flock (Baylis *et al.*, 2002). It contrasts, however, with other publications in which large numbers of scrapie-incubating animals were reported (Thorgeirsdottir *et al.*, 2002; Billinis *et al.*, 2004). Similarly, the modelling of a scrapie outbreak in a Cheviot flock predicted a high ratio of infections to cases (2.2:1) (Matthews *et al.*, 2001).

Discrepancies between these results certainly lay in the data collection plan. Studies reporting a large proportion of asymptomatic animals were based on cross-sectional designs, with data collected at stamping-out. In our study, most sheep were culled after a long individual monitoring period which allowed scrapie clinical onset. As indicated by our simulations, shorter monitoring lengths, as modelled by Matthews *et al.* (2001) (median length 3.00 years), would have resulted in the observation of fewer clinical cases and a larger number of subclinical cases.

Genetic susceptibility to scrapie and incubation period

Comparison of the fit provided by the mixture cure model and the traditional Cox PH model indicated that our

approach was more relevant when analysing PrP genotype and flock effects. According to our results, with ARQ homozygote animals as the baseline, VRQ carriers were at higher risk of infection and ARR heterozygotes at lower risk. This is consistent with most published studies based on data collected from culled flocks (Thorgeirsdottir *et al.*, 1999; Tranulis *et al.*, 1999; Acin *et al.*, 2004; Billinis *et al.*, 2004).

Incubation length is a major feature of TSE phenotype. In our population, while clinical signs were delayed in ARR heterozygotes compared with ARQ homozygotes, no difference could be observed between ARQ/VRQ carriers and ARQ homozygotes. A similar phenomenon was observed in an Irish flock (O'Doherty *et al.*, 2002). However, it differed from estimations obtained in a French Romanov flock (Elsen *et al.*, 1999) and in a Texel flock (Baylis *et al.*, 2002). In both these naturally affected scrapie flocks, significant differences in age at death were reported between ARQ/ARQ and ARQ/VRQ.

In sheep, experimental challenge has indicated that incubation period depends on both sheep genotype and TSE isolate. While most scrapie isolates will produce shorter incubation periods in VRQ allele carriers, other TSE agents such as BSE behave differently (Foster *et al.*, 2001; Jeffrey *et al.*, 2006). In this context, difference in agent (strain) could be a possible explanation for the observed variability.

In rodent scrapie models, it has been demonstrated that variations in attack rate and incubation length can be observed according to the infectious dose. Low infectious dose could lead to lengthening of the incubation period and decreased infection efficiency (Kimberlin & Wilesmith, 1994; Jacquemot *et al.*, 2005). In natural scrapie, there is no available estimation of the actual infectious pressure. Because of differences in the observed incidences, infection pressure is usually considered to be different according to the cohort considered within a flock or between flocks (Baylis *et al.*, 2002; Touzeau *et al.*, 2005). In our study, duration of scrapie incubation appeared not to be associated with the infection rate. Age at clinical onset in ARQ/ARQ infected animals also clearly differed (shorter incubation period in our study) from values reported previously in animals bearing the same genotype (Woolhouse *et al.*, 1998; Elsen *et al.*, 1999; O'Doherty *et al.*, 2002; Redman *et al.*, 2002; Baylis *et al.*, 2004). Taken together, these observations could suggest that biologically different scrapie agents are involved in the different flocks or that other factors linked to flock management could influence incidence and incubation.

To take these studies further, evaluation of agent biodiversity in the studied cohorts from these flocks is ongoing through biochemical studies and bioassays. Meanwhile, the effect of increasing infectious dose on incubation length in animals bearing similar PrP genotypes and which were orally contaminated at birth is under investigation.

The mixture cure model presented here has provided an interesting tool to analyse data collected from longitudinal

surveys in naturally affected scrapie flocks. Its main constraint is the requirement for a sufficiently long individual monitoring period. Finally, because such models allow for the combinatory analysis of several covariate effects, they should be considered as a potentially powerful tool for epidemiological analysis in animal diseases.

ACKNOWLEDGEMENTS

Financial support of this work was provided by EU project QLRT-2000-01733 'Scrapiefreesheep'. The authors wish to thank the technical staff of the Manech breeders' organization (Centre Départemental d'Élevage Ovins) and the Groupement de Défense Sanitaire des Pyrénées Atlantiques for their involvement in the longitudinal survey of scrapie-affected flocks which started in the mid-1990s.

REFERENCES

- Acin, C., Martin-Burriel, I., Goldmann, W., Lyahyai, J., Monzon, M., Bolea, R., Smith, A., Rodellar, C., Badiola, J. J. & Zaragoza, P. (2004). Prion protein gene polymorphisms in healthy and scrapie-affected Spanish sheep. *J Gen Virol* **85**, 2103–2110.
- Andréoletti, O., Berthon, P., Marc, D., Sarradin, P., Grosclaude, J., van Keulen, L., Schelcher, F., Elsen, J. M. & Lantier, F. (2000). Early accumulation of PrP^{Sc} in gut-associated lymphoid and nervous tissues of susceptible sheep from a Romanov flock with natural scrapie. *J Gen Virol* **81**, 3115–3126.
- Andréoletti, O., Berthon, P., Levassesseur, E., Marc, D., Lantier, F., Monks, E., Elsen, J. M. & Schelcher, F. (2002). Phenotyping of protein-prion (PrP^{Sc})-accumulating cells in lymphoid and neural tissues of naturally scrapie-affected sheep by double-labeling immunohistochemistry. *J Histochem Cytochem* **50**, 1357–1370.
- Baylis, M., Goldmann, W., Houston, F., Cairns, D., Chong, A., Ross, A., Smith, A., Hunter, N. & McLean, A. R. (2002). Scrapie epidemic in a fully PrP-genotyped sheep flock. *J Gen Virol* **83**, 2907–2914.
- Baylis, M., Chihota, C., Stevenson, E., Goldmann, W., Smith, A., Sivam, K., Tongue, S. & Gravenor, M. B. (2004). Risk of scrapie in British sheep of different prion protein genotype. *J Gen Virol* **85**, 2735–2740.
- Begara-McGorum, I., Clark, A. M., Martin, S. & Jeffrey, M. (2000). Prevalence of vacuolar lesions consistent with scrapie in the brains of healthy cull sheep of the Shetland Islands. *Vet Rec* **147**, 439–441.
- Benestad, S. L., Sarradin, P., Thu, B., Schonheit, J., Tranulis, M. A. & Bratberg, B. (2003). Cases of scrapie with unusual features in Norway and designation of a new type, Nor98. *Vet Rec* **153**, 202–208.
- Billinis, C., Psychas, V., Leontides, L., Spyrou, V., Argyroudou, S., Vlemmas, I., Leontides, S., Sklaviadis, T. & Papadopoulos, O. (2004). Prion protein gene polymorphisms in healthy and scrapie-affected sheep in Greece. *J Gen Virol* **85**, 547–554.
- Bohning, D. & Seidel, W. (2003). Editorial: recent developments in mixture models. *Comput Stat Data Anal* **41**, 349–357.
- Cloucard, C., Beaudry, P., Elsen, J. M., Milan, D., Dussaucy, M., Bounneau, C., Schelcher, F., Chatelain, J., Launay, J. M. & Laplanche, J. L. (1995). Different allelic effects of the codons 136 and 171 of the prion protein gene in sheep with natural scrapie. *J Gen Virol* **76**, 2097–2101.
- Cox, D. R. (1972). Regression models and life-tables. *J R Stat Soc Ser B Methodol* **34**, 187–220.
- Davison, A. C. & Hinkley, D. V. (1997). *Bootstrap Methods and their Application*. Cambridge: Cambridge University Press.
- De Bosschere, H., Roels, S., Benestad, S. L. & Vanopdenbosch, E. (2004). Scrapie case similar to Nor98 diagnosed in Belgium via active surveillance. *Vet Rec* **155**, 707–708.
- DEFRA (2003). *National Scrapie Plan for Great Britain*. London: Department for Food, Environment and Rural Affairs. <http://www.defra.gov.uk/animalh/bse/otherses/scrapie/nsp/publicatsrpts/tables.htm>
- Detwiler, L. A. & Baylis, M. (2003). The epidemiology of scrapie. *Rev Sci Tech* **22**, 121–143.
- Diaz, C., Vitezica, Z. G., Rupp, R., Andréoletti, O. & Elsen, J. M. (2005). Polygenic variation and transmission factors involved in the resistance/susceptibility to scrapie in a Romanov flock. *J Gen Virol* **86**, 849–857.
- Eglin, R. D., Warner, R., Gubbins, S., Sivam, S. K. & Dawson, M. (2005). Frequencies of PrP genotypes in 38 breeds of sheep sampled in the National Scrapie Plan for Great Britain. *Vet Rec* **156**, 433–437.
- Elsen, J. M., Amigues, Y., Schelcher, F., Ducrocq, V., Andréoletti, O., Eychenne, F., Khang, J. V., Poivey, J. P., Lantier, F. & Laplanche, J. L. (1999). Genetic susceptibility and transmission factors in scrapie: detailed analysis of an epidemic in a closed flock of Romanov. *Arch Virol* **144**, 431–445.
- Farewell, V. T. (1982). The use of mixture models for the analysis of survival data with long-term survivors. *Biometrics* **38**, 1041–1046.
- Foster, J. D., Parnham, D., Chong, A., Goldmann, W. & Hunter, N. (2001). Clinical signs, histopathology and genetics of experimental transmission of BSE and natural scrapie to sheep and goats. *Vet Rec* **148**, 165–171.
- Fraser, H. (1976). The pathology of a natural and experimental scrapie. *Front Biol* **44**, 267–305.
- Hagenaars, T. J., Donnelly, C. A., Ferguson, N. M. & Anderson, R. M. (2003). Dynamics of a scrapie outbreak in a flock of Romanov sheep – estimation of transmission parameters. *Epidemiol Infect* **131**, 1015–1022.
- Heggebo, R., Press, C. M., Gunnes, G., Lie, K. I., Tranulis, M. A., Ulvund, M., Groschup, M. H. & Landsverk, T. (2000). Distribution of prion protein in the ileal Peyer's patch of scrapie-free lambs and lambs naturally and experimentally exposed to the scrapie agent. *J Gen Virol* **81**, 2327–2337.
- Hopp, P., Webb, C. R. & Jarp, J. (2003). Monte Carlo simulation of surveillance strategies for scrapie in Norwegian sheep. *Prev Vet Med* **61**, 103–125.
- Hourigan, J., Klingsporn, A. L., Clark, W. W. & de Camp, M. (1979). Epidemiology of scrapie in the United States. In *Slow Transmissible Diseases of the Nervous System*, pp. 331–356. Edited by S. B. Prusiner & W. J. Hadlow. New York: Academic Press.
- Hunter, N., Foster, J. D., Goldmann, W., Stear, M. J., Hope, J. & Bostock, C. (1996). Natural scrapie in a closed flock of Cheviot sheep occurs only in specific PrP genotypes. *Arch Virol* **141**, 809–824.
- Hunter, N., Goldmann, W., Foster, J. D., Cairns, D. & Smith, G. (1997). Natural scrapie and PrP genotype: case-control studies in British sheep. *Vet Rec* **141**, 137–140.
- Jacquemot, C., Cuche, C., Dormont, D. & Lazarini, F. (2005). High incidence of scrapie induced by repeated injections of subinfectious prion doses. *J Virol* **79**, 8904–8908.
- Jeffrey, M., Gonzalez, L., Chong, A., Foster, J., Goldmann, W., Hunter, N. & Martin, S. (2006). Ovine infection with the agents of scrapie (CH1641 isolate) and bovine spongiform encephalopathy: immunochemical similarities can be resolved by immunohistochemistry. *J Comp Pathol* **134**, 17–29.

- Kimberlin, R. H. & Wilesmith, J. W. (1994). Bovine spongiform encephalopathy. Epidemiology, low dose exposure and risks. *Ann N Y Acad Sci* 724, 210–220.
- Kuk, A. Y. C. & Chen, C.-H. (1992). A mixture model combining logistic regression with proportional hazards regression. *Biometrika* 79, 531–541.
- Lam, K. F., Fong, D. Y. & Tang, O. Y. (2005). Estimating the proportion of cured patients in a censored sample. *Stat Med* 24, 1865–1879.
- Le Dur, A., Beringue, V., Andréoletti, O., Reine, F., Lai, T. L., Baron, T., Bratberg, B., Vilotte, J. L., Sarradin, P. & other authors (2005). A newly identified type of scrapie agent can naturally infect sheep with resistant PrP genotypes. *Proc Natl Acad Sci U S A* 102, 16031–16036.
- Maller, R. A. & Zhou, X. (1996). *Survival Analysis with Long-Term Survivors*. Chichester: Wiley.
- Matthews, L., Coen, P. G., Foster, J. D., Hunter, N. & Woolhouse, M. E. (2001). Population dynamics of a scrapie outbreak. *Arch Virol* 146, 1173–1186.
- McKinley, M. P., Bolton, D. C. & Prusiner, S. B. (1983). A protease-resistant protein is a structural component of the scrapie prion. *Cell* 35, 57–62.
- O'Doherty, E., Healy, A., Aherne, M., Hanrahan, J. P., Weavers, E., Doherty, M., Roche, J. F., Gunn, M. & Sweeney, T. (2002). Prion protein (PrP) gene polymorphisms associated with natural scrapie cases and their flock-mates in Ireland. *Res Vet Sci* 73, 243–250.
- Onnasch, H., Gunn, H. M., Bradshaw, B. J., Benestad, S. L. & Bassett, H. F. (2004). Two Irish cases of scrapie resembling Nor98. *Vet Rec* 155, 636–637.
- Orge, L., Galo, A., Machado, C., Lima, C., Ochoa, C., Silva, J., Ramos, M. & Simas, J. P. (2004). Identification of putative atypical scrapie in sheep in Portugal. *J Gen Virol* 85, 3487–3491.
- Palhiere, I., François, D., Elsen, J. M., Barillet, F., Amigues, Y., Perret, G. & Bouix, J. (2002). Allele frequencies of the PrP gene in 29 French sheep breeds. Possible use in selection for resistance to scrapie. In *Proceedings of the Seventh World Congress on Genetics Applied to Livestock Production*, Session 13, pp. 13–16. Montpellier, France.
- Peng, Y. & Dear, K. B. (2000). A nonparametric mixture model for cure rate estimation. *Biometrics* 56, 237–243.
- Peng, Y. & Carriere, K. C. (2002). An empirical comparison of parametric and semiparametric cure models. *Biom J* 44, 1002–1014.
- Race, R., Raines, A., Raymond, G. J., Caughey, B. & Chesebro, B. (2001). Long-term subclinical carrier state precedes scrapie replication and adaptation in a resistant species: analogies to bovine spongiform encephalopathy and variant Creutzfeldt-Jakob disease in humans. *J Virol* 75, 10106–10112.
- Redman, C. A., Coen, P. G., Matthews, L., Lewis, R. M., Dingwall, W. S., Foster, J. D., Chase-Topping, M. E., Hunter, N. & Woolhouse, M. E. (2002). Comparative epidemiology of scrapie outbreaks in individual sheep flocks. *Epidemiol Infect* 128, 513–521.
- Ryder, S. J., Spencer, Y. I., Bellerby, P. J. & March, S. A. (2001). Immunohistochemical detection of PrP in the medulla oblongata of sheep: the spectrum of staining in normal and scrapie-affected sheep. *Vet Rec* 148, 7–13.
- Ryder, S., Dexter, G., Bellworthy, S. & Tongue, S. (2004). Demonstration of lateral transmission of scrapie between sheep kept under natural conditions using lymphoid tissue biopsy. *Res Vet Sci* 76, 211–217.
- Thorgeirsdottir, S., Sigurdarson, S., Thorisson, H. M., Georgsson, G. & Palsdottir, A. (1999). PrP gene polymorphism and natural scrapie in Icelandic sheep. *J Gen Virol* 80, 2527–2534.
- Thorgeirsdottir, S., Georgsson, G., Reynisson, E., Sigurdarson, S. & Palsdottir, A. (2002). Search for healthy carriers of scrapie: an assessment of subclinical infection of sheep in an Icelandic scrapie flock by three diagnostic methods and correlation with PrP genotypes. *Arch Virol* 147, 709–722.
- Tongue, S. C., Pfeiffer, D. U., Warner, R., Elliott, H. & Del Rio Vilas, V. (2006). Estimation of the relative risk of developing clinical scrapie: the role of prion protein (PrP) genotype and selection bias. *Vet Rec* 158, 43–50.
- Touzeau, S., Chase-Topping, M. E., Matthews, L., Lajous, D., Eychenne, F., Hunter, N., Foster, J. D., Simm, G., Elsen, J. M. & Woolhouse, M. E. (2005). Modelling the spread of scrapie in a sheep flock: evidence for increased transmission during lambing seasons. *Arch Virol* 151, 735–751.
- Tranulis, M. A., Osland, A., Bratberg, B. & Ulvund, M. J. (1999). Prion protein gene polymorphisms in sheep with natural scrapie and healthy controls in Norway. *J Gen Virol* 80, 1073–1077.
- van Keulen, L. J., Schreuder, B. E., Vromans, M. E., Langeveld, J. P. & Smits, M. A. (2000). Pathogenesis of natural scrapie in sheep. *Arch Virol* 16, 57–71.
- Woolhouse, M. E. J., Stringer, S. M., Matthews, L., Hunter, N. & Anderson, R. M. (1998). Epidemiology and control of scrapie within a sheep flock. *Proc R Soc Lond B Biol Sci* 265, 1205–1210.
- Yu, B., Tiwari, R. C., Cronin, K. A. & Feuer, E. J. (2004). Cure fraction estimation from the mixture cure models for grouped survival data. *Stat Med* 23, 1733–1747.

6.3 Analyse par vraisemblance pénalisée

Dans l'article présenté précédemment, l'analyse a été effectuée à l'aide d'un modèle de mélange semi-paramétrique (logistique/risques conditionnels proportionnels). Nous présentons ici les résultats obtenus par l'approche par vraisemblance pénalisée. Un modèle de mélange avec fraction non à risque (logistique/risques conditionnels proportionnels) et un modèle de promotion sont appliqués aux données avec les mêmes variables explicatives (génotype PRP, élevage). Dans les deux cas, le paramètre de lissage est choisi par la méthode automatique basée sur la maximisation du score approché de validation croisée. Le nombre de noeuds est fixé à 12, les tracés des fonctions estimées de risque étant très semblables avec un nombre de noeuds plus important.

Pour comparer les modèles de survie avec fraction non à risque et le modèle de survie classique nous avons utilisé le critère approché de validation croisé fourni par $LCV_a = l(.) - Tr(H_{pl}^{-1}H_l)$, où $l(.)$ est la vraisemblance complète du modèle non pénalisé, H_l la matrice Hessienne correspondante et H_{pl} la matrice Hessienne basée sur la vraisemblance pénalisée (Commenges et al., 2007). Pour le modèle de survie classique, nous avons utilisé le programme PHPML (Joly et al., 1999). Sous ce modèle qui ne prend pas en compte de fraction non à risque, la valeur du critère de validation est $LCV_{a_c} = 1.12436$. Pour le modèle de mélange avec fraction non à risque nous obtenons $LCV_{a_m} = 0.53256$, et pour le modèle de promotion $LCV_{a_p} = 0.53987$. Les modèles prenant en compte une fraction non à risque sont, selon ce critère, "meilleurs" (au sens du risque de Kullback-Leibler) que le modèle de survie classique. En revanche, l'adéquation aux données fournie par le modèle de mélange et le modèle de promotion est équivalente puisque la "distance" entre les deux modèles est négligeable $\Delta LVC_a = LCV_{a_m} - LCV_{a_p} = -0.0073$. Cependant, une préférence est donnée dans l'absolu, au modèle de mélange puisque $LCV_{a_m} < LCV_{a_p}$.

6.3.1 Effets des variables explicatives

Pour le modèle de mélange (tableau 6.1), les résultats obtenus par vraisemblance pénalisée sont très proches de ceux obtenus par une approche semi-paramétrique (voir article précédent, tableau 6). Les paramètres de régression du modèle de promotion n'ont pas d'interprétation simple (en terme d'odds ratio ou de risque relatif) et sont présentés dans le tableau 6.2. Cependant nous pouvons constater qu'ils vont dans le même sens que ceux obtenus avec le modèle de mélange : le risque d'être infecté est significativement accru pour les animaux porteurs de l'allèle VRQ (groupe VRQ/x) et à l'inverse significativement réduit pour les animaux hétérozygotes ARR (groupe R/S), comparés aux animaux ARQ/ARQ (groupe S/S). Les durées d'incubation sont significativement allongées pour les animaux R/S, mais ne diffèrent pas pour les animaux VRQ/x. L'effet élevage va aussi dans le même sens, que ce soit pour la fraction infectée, ou pour la distribution des temps de survie des animaux infectés.

TAB. 6.1 – Influence du génotype au gène PrP sur l'incidence et la latence : résultats de l'analyse par vraisemblance pénalisée, modèle de mélange.

Variable	Incidence			Latence		
	OR	IC _{95%}	p	RR	IC _{95%}	p
Génotype PrP (versus groupe S/S)			<0.001			<0.001
Groupe R/S	0.07	0.03 - 0.14	<0.001	0.33	0.16 - 0.66	<0.001
Groupe VRQ/x	2.70	1.31 - 5.52	0.007	0.91	0.54 - 1.55	0.743
Élevage* (versus élevage C)			<0.001			<0.001
Elev. A	0.35	0.16 - 0.76	0.001	1.69	0.91 - 3.13	0.093
Elev. B	0.74	0.35 - 1.56	0.441	2.61	1.42 - 4.83	0.001
Elev. D	0.41	0.19 - 0.87	0.001	2.17	1.22 - 3.85	0.008
Elev. E	0.21	0.09 - 0.47	<0.001	1.23	0.65 - 2.32	0.517
Elev. F	1.66	0.74 - 3.92	0.887	0.41	0.22 - 0.73	0.003

OR = odds ratio ; RR = risque relatif.

* analyse réalisée sur les animaux S/S uniquement.

TAB. 6.2 – Influence du génotype au gène PrP sur l'incidence et la latence : résultats de l'analyse par vraisemblance pénalisée, modèle de promotion.

Variable	Long terme			Court terme		
	Estimateur β	ET	p	Estimateur γ	ET	p
Génotype PrP (versus groupe S/S)			<0.001			<0.001
Groupe R/S	-2.2952	0.4088	<0.001	-1.0575	0.3827	0.005
Groupe VRQ/x	0.6767	0.2444	0.006	-0.2940	0.3387	0.385
Élevage* (versus élevage C)			<0.001			<0.001
Elev. A	-0.8233	0.2772	0.003	0.6215	0.3460	0.072
Elev. B	-0.3267	0.2699	0.226	1.1192	0.3399	0.001
Elev. D	-0.8156	0.2721	0.003	0.9652	0.3198	0.002
Elev. E	-1.1011	0.2816	<0.001	0.4354	0.3500	0.213
Elev. F	0.1539	0.3650	0.673	-0.9099	0.3367	0.006

ET = écart-type.

* analyse réalisée sur les animaux S/S uniquement.

6.3.2 Fractions à risque estimées

Le tableau (6.3) présente les fractions à risque (animaux infectés) estimées par l'approche par vraisemblance pénalisée et par le modèle de mélange semi-paramétrique. Nous constatons que pour chacun des groupes de géotypes R/S, S/S et VRQ/x, ces proportions sont très proches. Les estimateurs du modèle de promotion sont cependant très légèrement supérieurs à ceux obtenus par les deux autres modèles.

TAB. 6.3 – Pourcentage d'animaux infectés par groupe de géotype.

	Géotype		
	R/S	S/S	VRQ/x
Modèle de mélange semi-paramétrique	5.57	42.74	63.69
Modèle de mélange vraisemblance pénalisée	5.36	42.62	63.32
Modèle de promotion vraisemblance pénalisée	5.85	44.11	64.71

6.3.3 Fonctions de survie et de risque

L'existence des relations entre le modèles de mélange et le modèle de promotion (chapitre 2.3.3) permet de comparer les estimateurs des fonctions de survie et de risque obtenues sous les deux modèles. La figure (6.1) présente les fonctions $\hat{S}_{0_{pop}}(t)$ (fonctions de survie marginale de base), $\hat{S}_0(t|U = 1)$ (fonction de risque conditionnelle de base), $\hat{\lambda}_{0_{pop}}(t)$ (fonction de risque marginale de base) et $\hat{\lambda}_0(t|U = 1)$ (fonction de risque conditionnelle de base), estimées par vraisemblance pénalisée pour le modèle de mélange et le modèle de promotion. Ces fonctions correspondent aux brebis du groupe S/S de l'élevage C. Les fonctions de survies marginales et conditionnelles estimées par les deux modèles sont superposées. La durée de survie médiane des animaux contaminés est de 1.9 années (premier quartile 1.5 ans, troisième quartile 2.5 ans).

De manière intéressante, la fonction de risque conditionnelle de base estimée

présente un premier pic aux environs de 2.2 ans, puis une nouvelle augmentation après 3 ans, suggérant une potentielle hétérogénéité au sein des animaux contaminés. De même les courbes de risque marginales indiquent qu'à l'échelle de la population, le risque de mourir de tremblante est maximal aux environs de deux ans. On retrouve un second pic aux environs de 4 ans, beaucoup moins marqué. Cette forme bimodale, plus marquée pour l'estimateur obtenu par le modèle de mélange, est conservée lorsque l'on ne considère que les animaux S/S dans l'analyse, et n'est donc pas due au choix du modèle à risques proportionnels. Plusieurs hypothèses peuvent être avancées pour expliquer ce résultat. Un déterminisme génétique autre que celui pris en compte dans cette analyse avec le gène majeur PRP pourrait participer au contrôle de la durée d'incubation de la tremblante. Plusieurs souches d'agent infectieux, au sein d'un même troupeau, pourraient avoir des durées d'incubation différentes. De même on peut supposer que certains animaux puissent se contaminer en dehors de la période post-natale immédiate, et donc avoir une courbe de survie décalée.

Le risque marginal estimé par le modèle de promotion est, en tout temps, légèrement supérieur au risque marginal estimé par le modèle de de mélange. Cette différence est à l'origine des estimateurs plus élevés des fractions contaminées présentés dans le tableau 6.3.

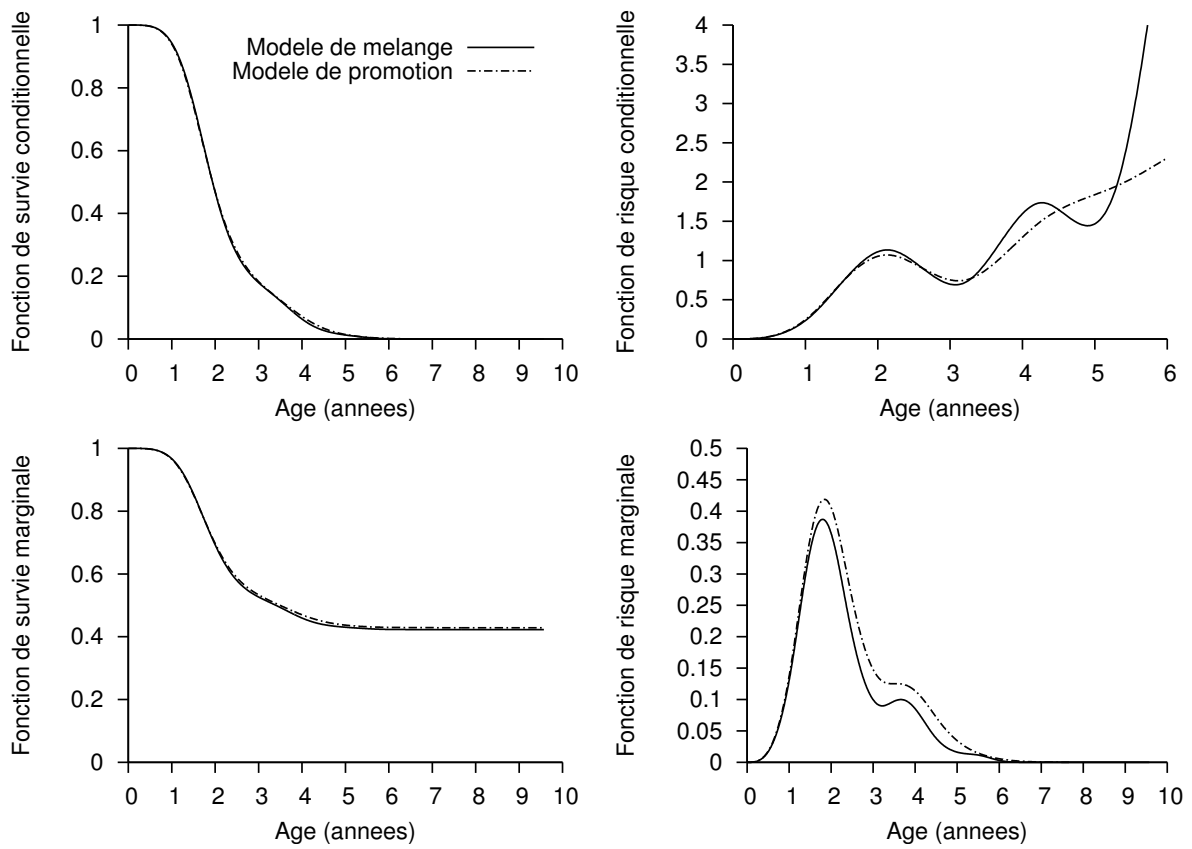


FIG. 6.1 – Fonctions (de base) de survie et de risque, marginales et conditionnelles, estimées par vraisemblance pénalisée.

Chapitre 7

Application à la tremblante du mouton : une analyse plus large

Dans le chapitre précédent, nous avons présenté une première application des modèles de survie avec fraction non risque à l'étude de la tremblante ovine classique pour un sous-échantillon des données récoltées dans les Pyrénées Atlantiques. L'étendue du suivi des élevages autorisait cependant une exploitation plus large des données, en incorporant un plus grand nombre d'élevages et de cohortes de naissance au sein de chaque élevage. En se fondant sur un échantillon d'étude nettement plus important, l'analyse présentée dans le présent chapitre avait pour objectif d'évaluer l'influence de facteurs de risque non pris en compte dans l'article précédent et d'obtenir des estimateurs plus robustes de l'influence du génotype au gène PrP sur le risque de contamination et la durée d'incubation. En particulier, nous nous sommes intéressés à l'influence de l'âge au premier contact avec un milieu contaminant et aux relations mères-agneaux. Au travers de la prise en compte de différentes mesures de la structure génétique des troupeaux, nous avons aussi cherché à expliquer les différentes dynamiques de l'infection entre les troupeaux et les variations d'incidence clinique au sein d'un même troupeau. Enfin, par l'utilisation de modèles à effets aléatoires, nous avons cherché à tenir compte de la dépendance statistique

pouvant exister entre les divers groupe d'animaux d'un même élevage.

7.1 Matériels et méthodes

7.1.1 Schéma d'étude

Il s'agit d'une cohorte prospective et rétrospective, basée sur le suivi de plusieurs élevages adhérents aux programmes de recherche Sheeprion, Projet 64 et Scrapie-FreeSheep.

7.1.2 Population d'étude

Sélection des élevages

Pour être inclus dans l'étude, les élevages devaient être adhérents au Contrôle Laitier Officiel (CLO) et avoir présenté plus de 10 cas de tremblante depuis cette adhésion. La date d'entrée au CLO peut être antérieure ou postérieure à l'apparition de la tremblante dans le troupeau. Le suivi du troupeau, dans le cadre des programmes Sheeprion, Projet64 et ScrapieFreeSheep, devait avoir été réalisé de façon complète et satisfaisante (génotypage de l'ensemble des animaux présents à partir de 1999, enregistrement des cas de tremblante, des dates d'euthanasie). En pratique, seuls des élevages ayant adhéré au programme initial Sheeprion et qui ont été suivis par la suite dans les programmes Projet64, puis ScrapieFreeSheep, ont été retenus. De plus seuls les élevages dans lesquels la tremblante présentait un caractère enzootique (plusieurs cas de tremblante dans plusieurs cohortes de naissance) ont été inclus.

Sélection des cohortes de naissance

Dans chacun des élevages retenus, un certain nombre de millésimes (cohortes) de naissance ont été sélectionnés. Le choix de ces cohortes reposait sur :

- L’existence d’un inventaire exhaustif des animaux nés dans le millésime considéré. L’exhaustivité de cet inventaire est assurée pour les millésimes nés après l’adhésion de l’élevage au Contrôle Laitier Officiel (CLO). Toutefois, dans certains élevages, un inventaire exhaustif a pu être reconstruit pour 2 à 3 millésimes nés avant cette adhésion.
- Une connaissance de l’information génétique au gène PrP suffisamment importante. Seuls les millésimes dans lesquels plus de 30 % des animaux avaient été génotypés ont été retenus.
- La majorité des animaux d’un millésime considéré devait être présente à l’apparition de la tremblante dans l’élevage. Cette condition était remplie pour toutes les cohortes de naissance nées après la date supposée d’apparition de la tremblante dans les troupeaux. Pour les cohortes nées avant cette date, nous n’avons retenu que les millésimes nés au plus tôt 3 ans avant l’apparition supposée de la tremblante. Le fait de limiter l’inclusion de vieux millésimes permettait de s’assurer, pour chaque élevage, que la durée de suivi n’était pas différente entre les millésimes nés avant et ceux nés après l’apparition de la tremblante (test global du log-rank : $\chi^2 = 0.8397$, $p=0.3595$).
- Les animaux des millésimes sélectionnés devaient être nés avant la mise en place des plans de lutte fondés sur la stratégie génétique, c’est à dire l’utilisation systématique de béliers ARR/ARR pour la reproduction.

Sélection des animaux

Seuls les animaux ayant été en contact avec un environnement contaminant ont été retenus pour l’analyse. Les animaux qui avaient quitté leur élevage avant l’apparition supposée de la tremblante ont été exclus.

Concernant les animaux achetés, par opposition aux brebis nées dans l’élevage contaminé suivi, le statut sanitaire vis à vis de la tremblante de l’élevage d’origine (naisseur) a été recherché. Les animaux issus d’élevages atteints de tremblante (avant ou après la vente de l’animal considéré) ou dont l’élevage d’origine n’était pas

connu ont été exclus de l'échantillon d'étude. En effet, une contamination précoce dans l'élevage d'origine ne pouvait être écartée pour ces animaux. Ces informations sont issues du Centre Département d'Élevage Ovin (CDEO) et de la Direction des Services Vétérinaires des Pyrénées Atlantiques (DSV 64).

Dans l'analyse portant sur l'influence des relations mères-agneaux et de la pression d'infection, nous avons exclu les animaux pour lesquels la filiation maternelle et/ou la taille de la portée n'étaient pas connue.

Remarques

Les brebis incluses dans l'échantillon d'étude sont issues d'élevages privés. Tous les animaux nés dans les élevages ne sont pas conservés. Chaque année, les éleveurs choisissent un certain nombre d'animaux pour le renouvellement de leur cheptel et en réforment (éliminent) d'autres sur des critères de production ou d'âge avancé. La quasi totalité des agneaux mâles et une majorité des jeunes femelles ne sont pas conservés au delà du sevrage (arrêt de l'allaitement), c'est-à-dire au-delà d'un mois d'âge. Seuls les brebis conservées pour le renouvellement des troupeaux, soit environ 20 à 35% des agnelles nées chaque année (campagne) ont fait l'objet d'un suivi. Aucune information n'est disponible pour les animaux non conservés.

7.1.3 Nature et origines des informations recueillies

Diagnostic de tremblante

Le diagnostic de tremblante a été établi sur des bases cliniques, histologiques et immunohistochimiques (Laboratoire de physiopathologie des ruminants, École Nationale Vétérinaire de Toulouse). Jusqu'à l'automne 2002, seules les suspicions cliniques ont fait l'objet d'analyses. Cependant, pour des raisons logistiques, toutes n'ont pas pu faire l'objet d'une confirmation par le laboratoire. En cas de suspicion clinique, l'éleveur contactait son vétérinaire, qui confirmait le diagnostic clinique et procédait à l'euthanasie. Dans le cas où un examen complémentaire était réalisé,

l'encéphale de l'animal était prélevé et envoyé avec un double du bordereau d'euthanasie au Laboratoire Départemental d'Analyses Vétérinaires (EPLD 64), qui centralisait les prélèvements, avant de les transmettre au laboratoire de physiopathologie des ruminants de l'ENVV. Là, chaque prélèvement reçu faisait l'objet d'un enregistrement (manuscrit initialement, puis informatique à partir de 2000) comprenant la date de réception, l'identification de l'animal et son élevage d'origine. Les bordereaux d'euthanasie étaient communiqués à la Direction Départementale des Services Vétérinaires des Pyrénées Atlantiques (DDSV 64).

Date de sortie des élevages

La date de sortie des animaux a été recueillie pour chaque brebis. Plusieurs modalités étaient possibles :

- Euthanasie en ferme pour cause de tremblante clinique : l'euthanasie suit de quelques semaines l'apparition des premiers symptômes cliniques, et avait généralement lieu en phase avancée de la maladie. La date d'euthanasie est celle fournie par le vétérinaire ayant réalisé l'acte.
- Euthanasie pour cause d'élimination systématique des animaux de génotypes sensibles et très sensibles : ces abattages, débutés en 2002-2003, ont été réalisés au clos d'équarrissage, sous le contrôle des Services Vétérinaires. La date d'euthanasie est celle fournie par le bordereau d'abattage transmis au Laboratoire Départemental d'Analyses Vétérinaires des Pyrénées Atlantiques. Les informations d'euthanasie ont été recueillies jusqu'à fin mai 2007. Les animaux euthanasiés ont systématiquement fait l'objet d'une recherche de PrPres dans l'encéphale par Western-Blot ou ELISA.
- Sortie de l'élevage pour raisons zootechniques : cette dernière modalité concerne les brebis sans signes cliniques et ne tombant pas sous le coup de la police sanitaire. Les dates de sortie sont celles fournies par les données du contrôle laitier.

Une confrontation de la date d'euthanasie fournie avec les autres informations disponibles par ailleurs (date d'analyse des prélèvements en laboratoire, date du dernier contrôle laitier ou du dernier agnelage) a été réalisée chaque fois que possible pour rechercher les incohérences et les corriger.

Génotype au gène PrP

Depuis 1999, avec la mise en place du programme Sheeprion, poursuivi en suite par le programme ScrapieFreeSheep, le génotype des brebis au gène PrP a été établi systématiquement pour l'ensemble des animaux présents de plus de six mois par analyse directe à partir d'un prélèvement sanguin. Pour trois élevages, les données étaient disponibles à partir de 1995. Pour les animaux n'ayant pas fait l'objet d'une analyse biologique, les informations ont été recherchées de manière rétrospective : le génotype au gène PrP de ces animaux a été prédit, à partir des connaissances issues de leur généalogie et des typages de leurs apparentés.

Génotypage direct Les prélèvements ont été très majoritairement traités par méthode RFLP-PCR par le laboratoire Labogena (Jouy-en-Josas) ou parfois par le Laboratoire Vétérinaire Départemental des Pyrénées Atlantiques. Les analyses pratiquées ne permettaient pas de distinguer les allèles H et Q au codon 171, de sorte que les 4 allèles suivants sont identifiés ARR, AHQ, ARQ et VRQ.

Prédictions Une partie des animaux inclus dans l'étude n'a pas fait l'objet d'une analyse biologique. Le génotype de ces animaux a été prédit en utilisant la connaissance des généalogies, sur la base des règles de transmission mendéliennes. Les probabilités d'appartenir à chacun des 10 génotypes observés ont été calculées en utilisant le programme GP (Vitezica, 2003) basé sur une technique d'"alternative-peeling" (Fernando et al., 1993) et déjà utilisé pour l'analyse des facteurs de risque associés à la tremblante du mouton (Vitezica, 2003 ; Vitezica et al., 2005 ; Diaz et al., 2005).

Regroupement Compte tenu du déséquilibre très marqué dans la répartition des 10 génotypes observés (voir résultats), nous avons choisi de regrouper les génotypes selon 5 classes de susceptibilité à la tremblante, d'après la classification anglaise du National Scrapie Plan (NSP) : groupe RR (NSP I) : animaux de génotype ARR/ARR; groupe RS (NSP II) : animaux de génotypes ARR/ARQ, ARR/AHQ; groupe SS (NSP IV) : animaux de génotypes ARQ/ARQ, ARQ/AHQ, AHQ/AHQ; groupe RH¹ (NSP IV) : animaux de génotype ARR/VRQ et groupe SH (NSP V) : animaux de génotypes ARQ/VRQ, AHQ/VRQ, VRQ/VRQ.

Age au contact avec un milieu contaminant

L'appréciation de l'influence de l'âge au premier contact avec un milieu contaminant s'est faite de manière indirecte en s'appuyant sur le fait que certains animaux n'étaient pas nés dans les élevages suivis et que d'autres étaient nés avant l'apparition de la tremblante dans les troupeaux.

- **Achats** : (variable **ACHAT**) traditionnellement, les achats de brebis dans les élevages Manech Tête Rousse des Pyrénées Atlantiques suivent 2 modalités :

- achat de très jeunes brebis (ou agnelles), généralement au sevrage, âgées d'environ 30 à 60 jours. Cette modalité est la plus fréquente en Pyrénées Atlantiques.
- achat de brebis ayant déjà commencé une carrière laitière dans un élevage précédent. Cette deuxième modalité est nettement plus rare.

Le très faible nombre de cas observés chez les animaux achetés à l'âge adulte (1 seul cas) nous a conduit à ne considérer que deux modalités pour la variable **ACHAT** :

- naissance dans l'élevage suivi (pas d'achat).
- achat jeune à environ 2 mois, ou achat à l'âge adulte.

- **Période de naissance** (variable **PÉRIODE**) : la date d'apparition de la tremblante dans les troupeaux n'est connue avec précision pour aucun des troupeaux.

⁰Dans l'analyse précédente, les animaux RH étaient confondus avec les animaux RS

Cependant, les dires des éleveurs et l'analyse des premiers cas de tremblante décrits dans chaque élevage ont permis de définir une année à partir de laquelle la présence l'agent infectieux dans le troupeau était certaine. Pour chaque élevage, cette année a été définie comme l'année n de naissance du premier cas observé de tremblante. Cet animal sert de cas index. Reprenant la définition de la campagne, la date d'apparition de la tremblante a été arbitrairement fixée au 1er septembre de la campagne n , c'est à dire juste avant la période des mises bas. Cette date ne correspond pas à l'introduction réelle de l'agent infectieux dans l'élevage, mais a permis de définir la variable **PERIODE** qui distingue 2 périodes de naissance pour chaque élevage :

- année de naissance antérieure à la campagne n , c'est-à-dire antérieure à l'année de naissance du cas index,
- année de naissance égale ou postérieure à celle du cas index.

Comme cette définition de l'année de naissance est conditionnée, pour chaque élevage, à la campagne de naissance du cas index, le terme d'année de naissance "relative" sera employé par la suite.

Relations mères-agneaux

- **Statut de la mère vis à vis de la tremblante** (variable **SCRAPm**) : la connaissance des ascendants et descendants a permis de définir le statut "tremblante clinique" ou "sans symptômes" de la mère de chaque animal inclus dans l'étude.
- **Génotype PRP de la mère** (variable **GENMERE**) : nous avons défini deux catégories, selon que la mère était porteuse d'au moins un allèle ARR (appartenant aux groupes RR, RS ou RH), ou de génotype sensible (SS ou SH).
- **Mode de naissance** (variable **MODENAISS**) : nous avons différencié les filles nées simples de celles nées de portées multiples (au moins 2 agneaux).

Estimation de la pression d'infection autour de l'agnelage

À travers les variables décrites ci-dessous, nous avons cherché à traduire le degré de “contamination” de l'environnement autour de la mise bas. Nous avons fondé notre raisonnement sur le fait que le placenta des brebis de génotype sensible en incubation de la maladie accumule de la PrPres. A l'inverse pour les mères porteuses d'un allèle ARR, ou dont le foetus est porteur d'un allèle ARR, le placenta ne semble pas accumuler de PrPres (Andreoletti et al., 2002, Lacroux et al., 2007). Le niveau de portage de l'allèle ARR, au sein de chaque cohorte de naissance ou à l'échelle des mères mettant bas à chaque campagne, peut renseigner sur le niveau de résistance.

- **Niveau de résistance du millésime de naissance** (variable **pARR**) : pour chaque millésime de naissance des filles au sein de chaque troupeau, nous avons calculé le pourcentage d'agnelles conservées pour le renouvellement porteuses d'au moins un allèle de résistance (animaux homozygotes ou hétérozygotes ARR). Cette variable, établie sur la base des génotypes PRP connus et imputés, a été utilisée sous forme continue ou catégorielle dans le cas où un effet linéaire n'était pas attendu.

- **Niveau de résistance du troupeau adulte** (variable **pARRm**) : dans chaque élevage, pour chaque campagne correspondant au millésime de naissance des agnelles incluses dans l'étude, nous avons calculé le pourcentage des brebis âgées de plus d'un an (brebis adultes) présentes pendant la période d'agnelage et porteuses d'au moins un allèle de résistance (brebis homozygotes ou hétérozygotes ARR). Cette variable, établie sur la base des génotypes PRP connus et imputés, a été utilisée sous forme continue ou catégorielle dans le cas où un effet linéaire n'était pas attendu.

- **Pourcentage de mères malades** (variable **pSCRAPm**) : pour chaque millésime de naissance des agnelles suivies, le pourcentage de mères de génotypes sensibles présentes pendant la période d'agnelage et qui sont mortes de tremblante (quel que soit le moment) a été calculé.

7.1.4 Recueil et circuit des informations

Les informations zootechniques et de généalogie étaient issues de la base de données génétiques nationale gérée au CTIG (département de génétique animale de l'INRA). Les données relatives aux euthanasies (dates d'euthanasie et identification des animaux sacrifiés) ont été recueillies auprès des éleveurs (carnets de suivi sanitaire) et de la DSV64 (bordereaux d'euthanasie) par le technicien en charge du suivi des élevages adhérents aux programmes de recherche du Centre Départemental d'Élevage ovins (CDEO) qui en assurait la saisie informatique et transmettait les informations à la Station d'Amélioration Génétique des Animaux (SAGA, INRA Centre de Toulouse). Les résultats des génotypages ont été transmis par le laboratoire ayant réalisé l'analyse, soit directement à la SAGA, soit au CDEO qui répercutait les informations. Les résultats d'examens complémentaires ont été recueillis auprès de l'École Nationale Vétérinaire de Toulouse, sous forme de fichiers informatiques saisis par l'ingénieur de recherche qui réalisait et interprétait les analyses. Les vérifications des identifications animales, des résultats et des dates d'analyses ont été réalisées à partir des registres manuscrits et informatiques de réception des prélèvements. L'ensemble de ces informations est conservé sous la forme de fichiers de données ACCESS® et SAS® à la SAGA, pilote global de la gestion de l'ensemble de ces informations avec le CDEO.

7.1.5 Analyse statistique

L'événement d'intérêt était la mort par tremblante clinique. Le temps de survie considéré était le délai entre l'exposition à un milieu contaminé et la mort par tremblante (mort naturelle ou euthanasie en phase clinique avancée). Sous l'hypothèse que la contamination a lieu précocement après la mise en contact avec un milieu infecté, la date d'origine était la naissance pour les animaux nés après l'apparition de la tremblante dans les troupeaux et la date d'arrivée dans le troupeau pour les animaux achetés. Pour les animaux nés avant l'apparition de la tremblante dans

les troupeaux, nous avons arbitrairement fixé la date d'origine au 1^{er} septembre de l'année supposée d'apparition de la tremblante dans le troupeau. Les temps de survie des animaux morts ou sortis des élevages sans symptômes de tremblante ont été considérés comme censurés à droite. Les durées de survie des animaux de génotypes sensibles euthanasiés dans le cadre des mesures de police sanitaire ont été censurées à droite.

Stratégie d'analyse

L'étude des facteurs de risque liés aux relations mères-agneaux n'étant pas pertinente pour les animaux achetés et pour ceux nés avant l'apparition de la tremblante, l'analyse a été conduite en plusieurs étapes. Dans un premier temps, nous avons étudié l'effet de l'âge au premier contact sur la probabilité de contamination et sur les durées d'incubation des animaux infectés en prenant en considération la totalité des animaux (Analyse 1). Nous avons ensuite restreint l'échantillon d'étude aux seuls animaux de filiation maternelle connue, non achetés et nés la même année ou après le cas index, afin d'évaluer l'influence des variables indicatrices des relations mères-agneaux, c'est à dire le statut clinique de la mère, le mode de naissance et la pression d'infection autour de la mise bas (Analyse 2).

Dans un premier temps, l'analyse a été réalisée par l'utilisation des modèles semi-paramétriques de survie avec fraction non à risque sans effets aléatoires à l'aide de la macro SAS présentée dans le chapitre 3. L'analyse par vraisemblance pénalisée n'a pas pu être finalisée à la date d'écriture et ne sera donc pas présentée. Compte-tenu de leur interprétation plus simple, l'approche par modèles de mélange a été choisie. La partie incidence a été modélisée par un modèle logistique et la partie latence par un modèle semi-paramétrique à risques proportionnels. Ces modèles ont été ajustés sur le génotype PRP et l'élevage, qui sont supposés avoir une forte influence. Compte-tenu de la corrélation existante entre plusieurs variables explicatives (pARR et pARRm, pSCRAPm et SCRAPm), plusieurs modèles ont été évalués. Le choix du modèle final le plus informatif s'est fait sur la base du critère d'Akaike.

Pour chacune des analyses (analyse 1 et analyse 2), nous avons ensuite pris en compte la corrélation entre les observations des animaux d'un même élevage, par l'utilisation des modèles de survie de mélange avec fraction non à risque et deux effets aléatoires non-indépendants. Dans ces modèles, la partie incidence a été modélisée par un modèle logistique et la distribution des temps de survie des animaux infectés a été modélisée par un modèle exponentiel par morceaux. Ce choix était fondé sur les résultats de l'analyse par vraisemblance pénalisée présentée au chapitre 6. La forme bimodale de la fonction de risque conditionnelle ne pouvant pas être capturée par un modèle paramétrique classique, l'utilisation d'un modèle plus souple paraissait préférable. L'axe des temps a été divisé en intervalles de taille inégale, de sorte que chaque intervalle contiennent suffisamment d'événements. Le choix de ce découpage était basé sur la distribution des temps de survie. Dix intervalles ont été définis : [0-1.2 an],[1.2-1.5],[1.5-2],[2-2.5],[2.5-3],[3-4],[4-5],[5-6],[6-7],[>7 ans]. La concentration des premiers intervalles s'explique par le nombre important de cas cliniques observés avant 3 ans. A l'inverse, compte-tenu du faible nombre de cas cliniques âgés de plus de 7 ans, un seul intervalle est défini entre 7 ans et la fin de la période d'observation. Dans chaque élevage, l'unité de regroupement était constituée par les animaux appartenant au même millésime de naissance. L'opportunité d'introduire un effet aléatoire dans la modélisation de la fraction non à risque et/ou dans la modélisation des temps de survie des animaux infectés a été évaluée par la comparaison des vraisemblances entre les modèles avec ou sans effets aléatoires.

Pour chacun des modèles testés, les variables explicatives ont été introduites à la fois dans la modélisation de la partie incidence et la modélisation de la partie latence. La significativité globale de chaque variable a été évaluée par le test du rapport de vraisemblance entre le modèle contenant la variable évaluée et le modèle ne la contenant pas. Le seuil d'erreur $\alpha = 0.05$ a été retenu pour considérer une variable comme ayant un effet significatif.

Imputation multiple

Du fait de données manquantes concernant les génotypes au gène PrP d'une partie des animaux inclus dans l'étude, trois approches ont été suivies. La première consistait à n'inclure dans l'analyse que les animaux de génotype PRP connu. Dans la seconde approche, les animaux de génotype inconnu ont été considérés dans un même groupe et l'analyse a été conduite avec 5 classes de génotypes (classes "RR", "RS", "SS", "RH", "SH" et "Inconnu"). La dernière approche a utilisé les probabilités prédites d'appartenance à chacun des génotypes, regroupées par classes de susceptibilité. Diaz et al. (2005) et Vitezica et al. (2005) utilisent les probabilités comme valeurs des variables explicatives dans un modèle de survie classique. Cette approche a l'avantage de ne nécessiter l'analyse que d'un seul jeu de données et fournit des estimateurs comparables à une analyse par imputation multiple (Vitezica, 2003). Elle n'a cependant pas pu être utilisée dans le cadre du modèle de survie avec fraction non à risque. En effet, l'absence de cas clinique observé chez les animaux de génotype ARR/ARR (groupe "RR") conduit à des estimateurs infinis des paramètres de régression et à des résultats aberrants. Nous avons fait l'hypothèse, confortée par la littérature, que ces animaux n'étaient pas à risque de se contaminer par la tremblante classique dans les conditions naturelles, ce qui nous a conduit à les exclure de l'analyse. Aussi, nous avons opté pour une approche par imputation multiple dans laquelle les animaux de génotype ARR/ARR connu ou imputé ont été supprimés dans chacun des jeux de données générés. Dans l'analyse par imputation multiple, M jeux de données sont générés, et pour chacun des M échantillons, le génotype des animaux non génotypés est imputé aléatoirement selon la distribution des probabilités d'appartenance à chacun des G_j , $j = 1, \dots, 10$ génotypes. Les estimateurs des paramètres de régression et de leur variance sont alors estimés séparément pour chaque échantillon imputé. Soit \hat{Q}_i le vecteur des estimateurs de paramètres et \hat{U}_i la matrice de variance-covariance des estimateurs pour le i^{ieme} échantillon, ($i = 1, \dots, M$). Un estimateur global du vecteur de paramètres Q est

donné par

$$\bar{\mathbf{Q}} = \frac{1}{M} \sum_{i=1}^M \hat{\mathbf{Q}}_i$$

Un estimateur de la matrice de variance-covariance des paramètres de régression est fourni par

$$\mathbf{T} = \bar{\mathbf{U}} + \left(1 + \frac{1}{M}\right) \mathbf{B}$$

où $\bar{\mathbf{U}}$ est la matrice de variance-covariance intra-imputation, correspondant à la moyenne des $\hat{\mathbf{U}}_i$,

$$\bar{\mathbf{U}} = \frac{1}{M} \sum_{i=1}^M \hat{\mathbf{U}}_i$$

et \mathbf{B} est la matrice de variance-covariance inter-imputation, donnée par

$$\mathbf{B} = \frac{1}{M-1} \sum_{i=1}^M (\hat{\mathbf{Q}}_i - \bar{\mathbf{Q}})(\hat{\mathbf{Q}}_i - \bar{\mathbf{Q}})'$$

Lorsque M est grand en regard du nombre de paramètres estimés, l'inférence basée sur la statistique de Wald $\bar{\mathbf{Q}}/\sqrt{\mathbf{T}}$ ou le rapport de vraisemblance reste valide (Meng et Rubin, 1992). Nous avons choisi $M=100$ dans toutes les analyses conduites par imputation multiple.

7.2 Résultats : analyse descriptive des données

7.2.1 Échantillon d'étude

Pour l'analyse 1, le processus de sélection a conduit à retenir 6687 brebis dans l'échantillon d'étude (figure 7.1). Le nombre médian de brebis retenues par élevage est de 613 (minimum = 377, maximum=768) (tableau 7.1). Entre 6 et 9 millésimes sont retenus par élevage, la plupart (67 sur 83 soit 80.72%) correspondant à des brebis nées après l'apparition supposée de la tremblante dans les troupeaux. Dans

3 élevages, pour lesquels la tremblante est apparue au tout début des années 1990 ou avant, aucune brebis nées avant cette date n'a été retenue.

Sur les 6687 brebis retenues dans l'échantillon d'étude, 1166 (17.44%) sont mortes de tremblante au stade clinique. La répartition par élevage, présentée dans le tableau 7.1 est très hétérogène, puisque selon les élevages de 33 à 185 cas cliniques sont observés (médiane 89), représentant respectivement 6.00% à 32.23% (médiane 15.45%) des animaux inclus par élevage.

Pour l'analyse 2, l'échantillon d'étude initial était constitué de 4712 brebis nées la même année ou après le cas index de tremblante (figure 7.1). Nous avons, de plus, exclu 226 animaux pour lesquels les informations sur la filiation maternelle et sur le mode de naissance n'étaient pas connues. Ces animaux représentaient, pour 170 d'entre eux (75.22%), l'ensemble des animaux de 3 cohortes de naissance des élevages F et G, nées avant l'adhésion de ces élevages au contrôle laitier. Nous avons donc exclu ces 3 cohortes de l'échantillon d'étude. Les 56 autres animaux sans filiation connue étaient répartis dans l'ensemble des autres élevages. Notons que les animaux exclus ne différaient pas des animaux à filiation connue en terme de proportion d'animaux cliniquement atteints ($\chi_1^2 = 0.0192$, $p=0.8889$), de temps de suivi (test du log rank $\chi_1^2=0.0818$, $p=0.7749$) ou de génotype PRP ($\chi_4^2=7.4977$, $p=0.1118$). Au final, l'échantillon d'étude retenu pour l'analyse 2 comptait 4486 brebis réparties dans 64 cohortes de naissance, dont 975 cas cliniques.

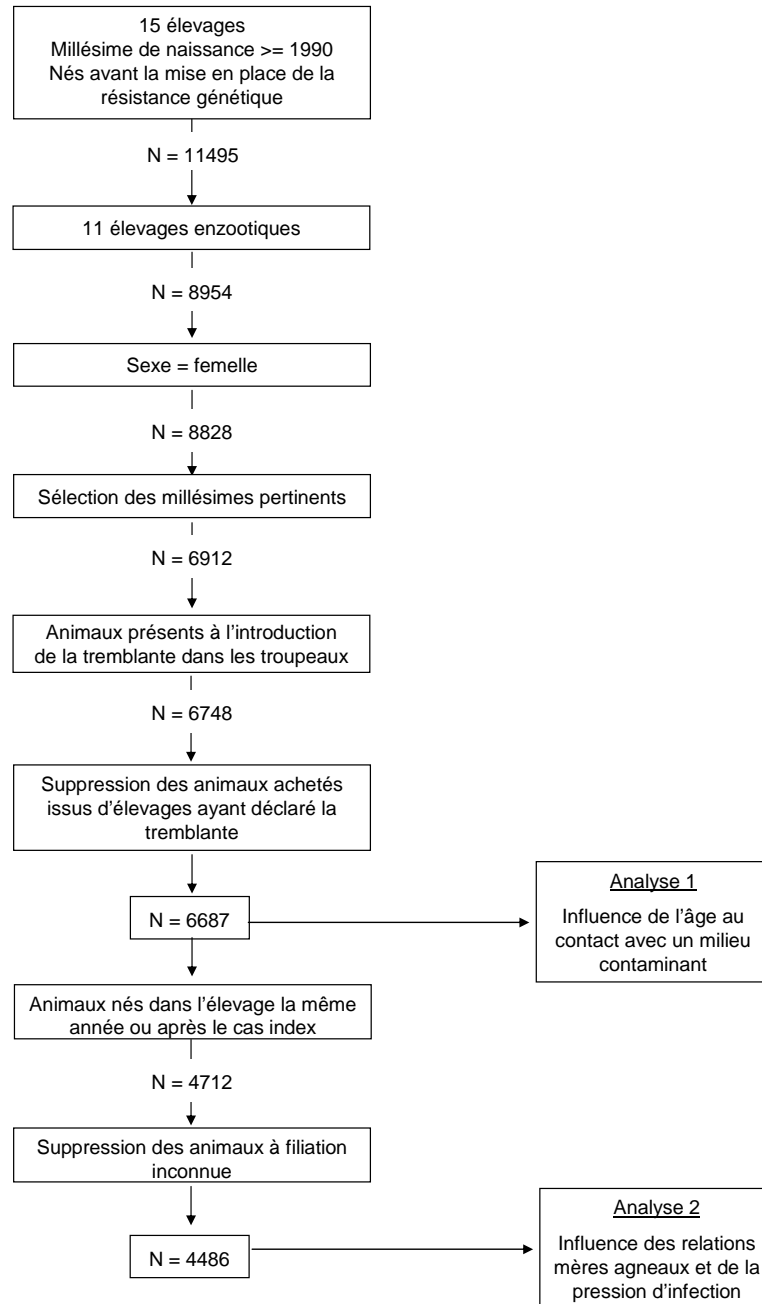


FIG. 7.1 – Sélection de l'échantillon d'étude.

TAB. 7.1 – Nombre de brebis (millésimes) incluses dans l'échantillon d'étude par élevage et par période de naissance par rapport à l'apparition de la tremblante dans les troupeaux.

Élevage	Début ^a	Naissance par rapport au cas index ^b				Total	
		Millésimes antérieurs		Même millésime ou millésimes postérieurs			
		# cas ^c	effectif (# mill) ^d	# cas	effectif (# mill)	# cas	effectif (# mill)
A	1994	4	169 (2)	65	599 (7)	69	768 (9)
B	1998	2	259 (3)	31	290 (3)	33	549 (6)
C	1993	7	53 (1)	134	576 (6)	141	629 (7)
D	1995	5	120 (3)	32	257 (5)	37	377 (8)
E	1992	-	-	89	576 (8)	89	576 (8)
F	1991	4	52 (1)	167	593 (7)	171	645 (8)
G	1995	6	98 (1)	136	646 (5)	142	744 (6)
H	1988	-	-	142	557 (9)	142	557 (9)
I	1996	9	224 (3)	80	431 (4)	89	655 (7)
J	1995	31	142 (2)	154	432 (5)	185	574 (7)
K	1993	-	-	68	613 (8)	68	613 (8)
Total		68	1117 (16)	1098	5570 (67)	1166	6687 (83)

^a Campagne supposée d'apparition de la tremblante dans les troupeaux.

^b Naissance avant ou après l'apparition supposée de la tremblante dans les troupeaux.

^c Nombre de cas cliniques de tremblante.

^d Nombre de millésimes de naissance.

7.2.2 Temps de suivi

Les cas cliniques ont été observés dès l'âge de 9.3 mois et jusqu'à 8.5 ans et étaient le plus souvent âgés d'environ 2 ans (médiane du délai à la mort chez les cas cliniques 1.97 ans, intervalle interquartile : 1.56 - 3.02). Le tracé de Kaplan-Meier de la fonction de survie marginale (figure 7.2), présente un plateau stable après le dernier événement observé, indiquant la présence potentielle d'une fraction d'animaux non infectés. La fraction d'animaux infectés, fournie par l'estimateur non paramétrique $1 - \hat{S}_{6687}(t_\kappa)$ avec $t_\kappa = 8.51$ le dernier temps d'événement, est estimée à 25.49% (IC_{95%} = 23.86 - 27.21 %).

La fonction de survie de Kaplan-Meier inverse (figure 7.3) permet, au travers du temps de suivi des animaux au sein des troupeaux, d'apprécier le phénomène de censure. Dans cette figure, la sortie du troupeau pour raison autre que tremblante clinique est considérée comme l'événement d'intérêt. Les animaux morts de tremblante clinique sont censurés à droite. Le temps de suivi médian des 6687 animaux était de 4.55 années et 25% des animaux était toujours présents après 6.52 années. Le phénomène de censure s'apparente à une loi uniforme U[0,10] (ligne pointillé) avec une censure légèrement moins importante avant 5 ans et légèrement plus importante ensuite. Dans l'échantillon d'étude, seuls 266 animaux (3.98%) avaient des temps de suivi (censurés à droite) supérieurs au temps du dernier événement (8.51 ans), mais 2272 (33.98%) ont été suivis au delà de 5.46 années, qui correspond au 95^{ième} percentile des temps de décès des cas cliniques. Conformément aux simulations réalisées dans l'article présenté au chapitre 6, ces conditions de suivi sont suffisamment longues pour pouvoir utiliser un modèle de survie avec fraction non à risque. Une situation encore plus favorable était observée pour le sous échantillon d'animaux retenus pour l'analyse 2, puisque 43.66 % des animaux cliniquement sains étaient encore en vie au delà du 95^{ième} percentile des temps de décès des cas cliniques.

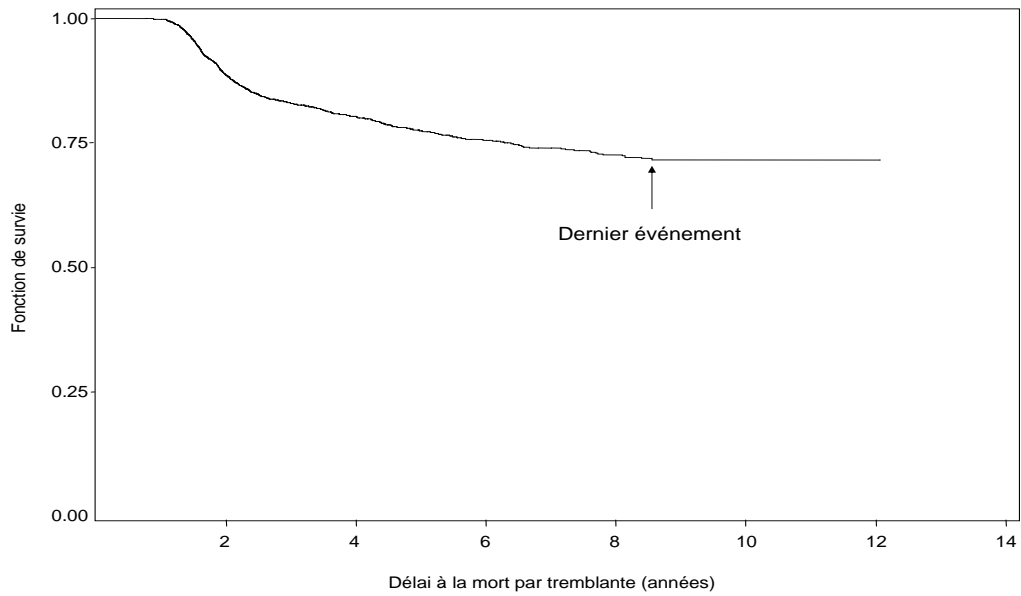


FIG. 7.2 – Fonction de survie marginale.

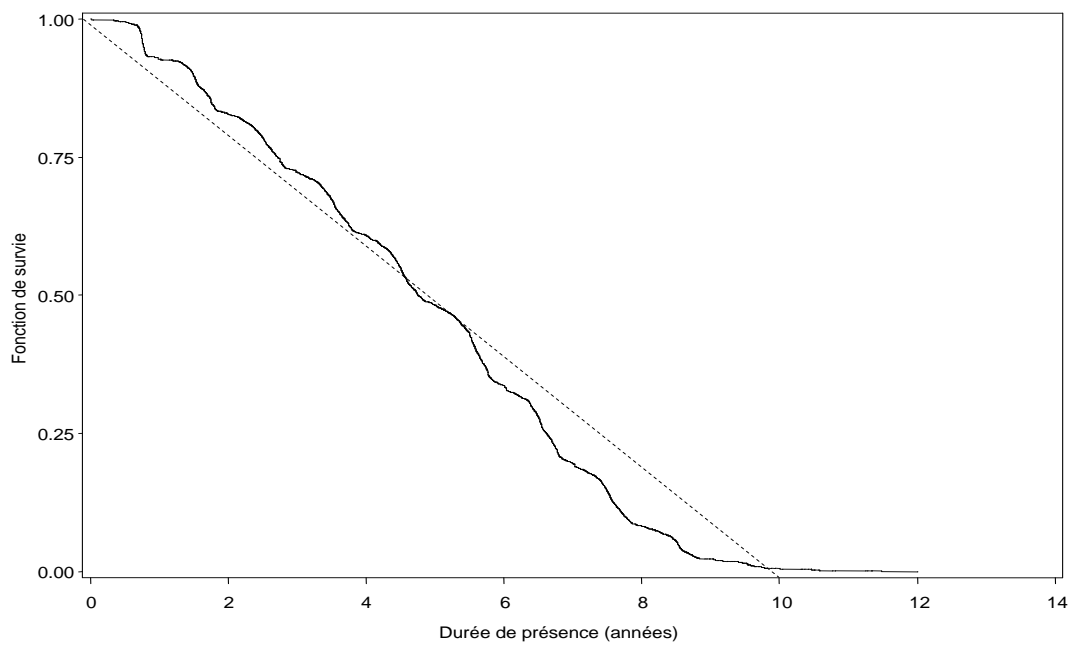


FIG. 7.3 – Durée de présence dans les troupeaux.

7.2.3 Génotype au gène PrP

L'information de génotype au gène PrP est disponible pour 4356 brebis (65.14%) (tableau 7.2). La connaissance de l'information génétique était hétérogène entre élevages ($\chi^2_{14}=156.1$, $p < 10^{-4}$) et variait de 54.91% (élevage D) à 79.78% (élevage B) avec une valeur médiane de 63.67%. Chez les animaux génotypés, les fréquences des allèles ARQ, ARR, VRQ et AHQ étaient respectivement de 70.26%, 25.09%, 3.80% et 0.85%. Les animaux de génotype ARQ/ARQ (groupe SS) étaient majoritaires et représentaient 49.70% des animaux de génotypes connus. Les animaux porteurs de l'allèle VRQ étaient peu nombreux ($n=329$, soit 7.55%) et 30.70% d'entre eux étaient hétérozygotes ARR/VRQ. Les fréquences alléliques étaient significativement différentes entre les troupeaux étudiés ($\chi^2_{40}=278.90$, $p < 10^{-4}$), certains présentant des fréquences d'allèle AHQ (élevages F et G) ou VRQ (élevages E et J) élevées pour la race (figure 7.4).

La connaissance des génotypes PRP était significativement plus faible chez les cas cliniques de tremblante (45.56%) par rapport aux animaux sans signes cliniques (73.04%) ($\chi^2=274.18$, $p < 10^{-4}$).

L'utilisation des prédictions d'appartenance à chaque génotype a permis d'imputer les données manquantes. Les résultats obtenus, présentés dans le tableau 7.3, ont permis de reproduire une situation proche des génotypes observés, avec une fréquence légèrement augmentée pour les génotypes des groupes SS et SH et une diminution respective pour le groupe RS. On se souvient que les génotypages systématiques des ovins des élevages suivis ont été mis en place à partir de 1999 (hormis pour 3 élevage, depuis 1995), grâce aux financements des contrats européens "Sheeprion" puis "ScrapieFreeSheep". Les présents résultats illustrent logiquement que les brebis mortes de tremblante avant les génotypages de 1999 appartenaient majoritairement aux génotypes sensibles.

Les autres variables explicatives sont décrites dans le tableau 7.4.

TAB. 7.2 – Distribution des génotypes au gène PrP : génotypes observés par groupe (typés ou inconnus)

Génotype	Effectif	Pourcentage/total	Pourcentage/effectif génotypé
Groupe RR	223	3.33	5.12
<i>ARR/ARR</i>	223	3.33	5.12
Groupe RS	1639	24.51	37.62
<i>ARR/ARQ</i>	1615	24.15	37.07
<i>ARR/AHQ</i>	24	0.36	0.55
Groupe SS	2165	32.37	49.70
<i>ARQ/ARQ</i>	2118	31.67	48.62
<i>ARQ/AHQ</i>	47	0.70	1.08
Groupe RH	101	1.51	2.32
<i>ARR/VRQ</i>	101	1.51	2.32
Groupe SH	228	3.41	5.24
<i>ARQ/VRQ</i>	223	3.33	5.12
<i>AHQ/VRQ</i>	3	0.04	0.07
<i>VRQ/VRQ</i>	2	0.03	0.05
Groupe Inconnu	2331	34.86	–
Total	6687	100.00	100.00

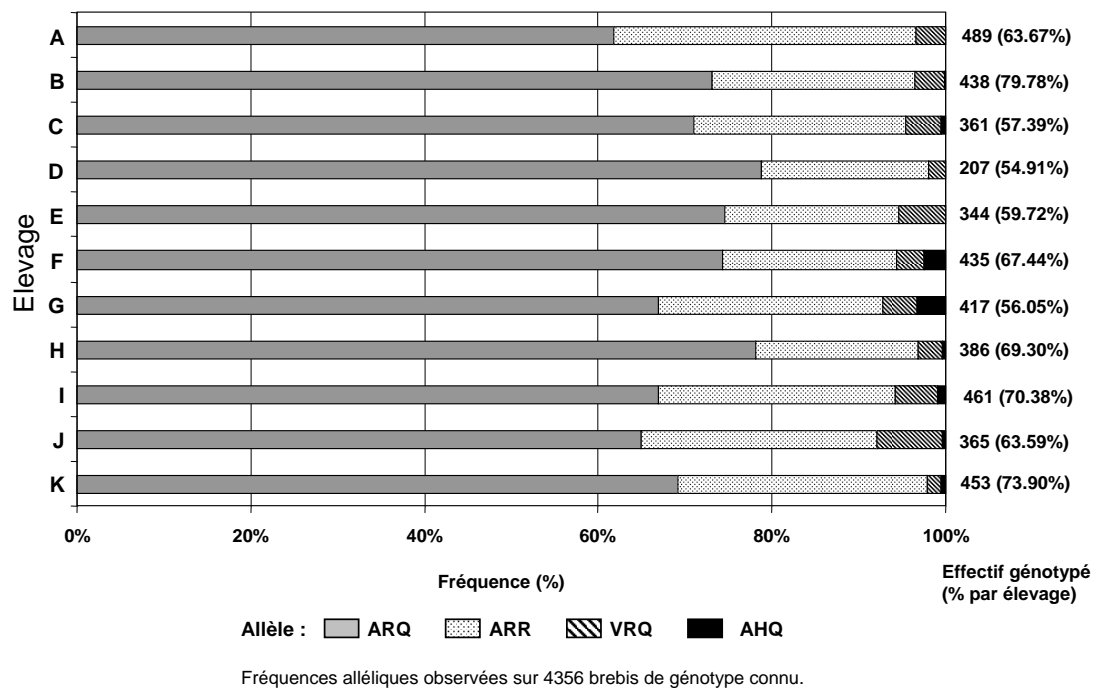


FIG. 7.4 – Fréquences alléliques observées.

TAB. 7.3 – Distribution des génotype au gène PrP : génotypes imputés (connus ou prédits)

Génotype	Effectif moyen	Pourcentage moyen
Groupe RR	328.4	4.91
<i>ARR/ARR</i>	328.4	4.80
Groupe RS	2271.4	33.97
<i>ARR/ARQ</i>	2239.4	33.49
<i>ARR/AHQ</i>	32.0	0.48
Groupe SS	3475.2	51.97
<i>ARQ/ARQ</i>	3406.2	50.94
<i>ARQ/AHQ</i>	68.8	1.03
<i>AHQ/AHQ</i>	0.2	0.00
Groupe RH	144.7	2.16
<i>ARR/VRQ</i>	144.7	2.16
Groupe SH	467.3	6.99
<i>ARQ/VRQ</i>	453.9	6.79
<i>AHQ/VRQ</i>	5.1	0.08
<i>VRQ/VRQ</i>	8.3	0.12

Effectifs et pourcentages moyens observés sur 100 échantillons imputés

TAB. 7.4 – Description des variables explicatives principales

Variable	Effectif	Pourcentage
<u>Applicables à tous les individus (Analyse 1, n=6687)</u>		
Période de naissance / cas index (PÉRIODE)		
<i>Millésimes antérieurs</i>	1117	16.70
<i>Même millésime ou postérieurs</i>	5570	83.30
Achat		
<i>Pas d'achat</i>	5673	84.84
<i>Achat agnelle</i>	872	13.04
<i>Achat adulte</i>	142	2.12
 <u>Applicables aux animaux nés dans l'élevage</u>		
<u>et la même année ou après le cas index (Analyse 2, n=4486)</u>		
Statut clinique de la mère (SCRAPm)		
<i>Sans symptômes</i>	4027	89.77
<i>Tremblante clinique</i>	459	10.23
Mode de naissance (MODENAISS)		
<i>Naissance simple</i>	2399	53.48
<i>Naissance multiple</i>	2087	46.52

	1 ^{er} quartile	Médiane	3 ^e quartile
Pourcentage de filles porteuses de l'allèle ARR (pARR)*	32.25	40.26	47.37
Pourcentage de mères porteuses de l'allèle ARR (pARRm)*	35.50	42.27	47.19
Pourcentage observé de mères malades (pSCRAPm)	6.63	14.23	20.88

* en moyenne, sur 100 fichiers imputés

7.3 Résultats de l'analyse multifactorielle

7.3.1 Description des modèles évalués

Le tableau 7.5 présente les différents modèles utilisés, la significativité globale des variables au sein de chaque modèle et le critère d'Akaike correspondant.

Analyse 1 : Le modèle M1 étant basé sur un nombre plus faible d'individus (animaux de génotypes connus uniquement), il n'est pas comparable, en termes de critère d'Akaike, au modèle M2 et M3, qui prennent en considération l'ensemble des animaux. Le modèle M3, fondé sur l'imputation multiple des génotypes inconnus, a été retenu, car il a été considéré comme plus informatif que le modèle M2, dans lequel les animaux de génotype inconnu étaient regroupés. On pouvait supposer que l'interaction ELEVAGE*PERIODE serait significative mais elle n'a pas pu être réellement testée, en raison du trop faible nombre de cas cliniques parmi les animaux nés avant l'apparition de la tremblante dans les troupeaux. Cependant, dans l'élevage J, un grand nombre de cas cliniques a été observé parmi ces animaux (tableau 7.1). Pour tenir compte de cette spécificité, nous avons considéré les animaux de cet élevage nés avant l'apparition de la tremblante dans une classe séparée. Cette approche fournissait une adéquation supérieure, en termes d'AIC, à un modèle ne tenant pas compte de cette "singularité" (non montré).

Analyse 2 : Les variables représentant le niveau de résistance des mères (pARR_m) et des filles conservées (pARR) étant basées sur l'information du génotype PRP, l'analyse 2 a seulement été conduite uniquement à partir de modèles prenant en compte les génotypes connus et prédits. Le choix de modèles, sur la base du critère d'Akaike, nous a conduit à retenir le modèle M8. Ce modèle, prenant en compte le niveau de contamination des femelles agnelant chaque année (variable pSCRAPM), était meilleur, en terme d'AIC, que ceux tenant compte de l'information individuelle du statut clinique de la mère (variable SCRAPM, modèles M5 et M6). De même,

la prise en compte du niveau de résistance des mères (variable pARR_m) offrait une meilleure adéquation que la prise en compte du niveau de résistance des agnelles conservées (variable pARR, modèles M5 et M7).

Nous avons cherché à catégoriser les variables traduisant la pression d'infection, pour s'affranchir de l'hypothèse de linéarité. Cependant, de tels modèles ne pouvaient plus être ajustés sur l'élevage, certaines classes n'étant représentées que par les millésimes appartenant à un seul élevage. Les AIC des différents modèles testés étaient toujours inférieurs à ceux des modèles comprenant la variable élevage. De plus, la variabilité induite par l'utilisation de l'imputation multiple rendait les effectifs par classes instables. La variance inter-imputation des estimateurs des paramètres de régression des variables définissant ces classes devenait très supérieure à la variance intra-imputation, rendant les conclusions de ces modèles peu satisfaisantes. Ces modèles n'ont donc pas été retenus et les variables caractérisant la pression d'infection ont été considérées de façon continue.

TAB. 7.5 – Description des modèles, degré de significativité global des variables (p value), et critère d’Akaike

	Analyses								
	Analyse 1 ^a (N=6687)			Analyse 2 ^b (N= 4486)					
Variables explicatives	M1*	M2**	M3***	M4	M5	M6	M7	M8	M9
Génotype PRP	< 10 ⁻⁴	< 10 ⁻⁴	< 10 ⁻⁴	< 10 ⁻⁴	< 10 ⁻⁴	< 10 ⁻⁴	< 10 ⁻⁴	< 10 ⁻⁴	< 10 ⁻⁴
Élevage	< 10 ⁻⁴	< 10 ⁻⁴	< 10 ⁻⁴	< 10 ⁻⁴	< 10 ⁻⁴	< 10 ⁻⁴	< 10 ⁻⁴	< 10 ⁻⁴	< 10 ⁻⁴
Période	< 10 ⁻⁴	< 10 ⁻⁴	< 10 ⁻⁴	< 10 ⁻⁴					
Millésime (aléatoire)				< 10 ⁻⁴					< 10 ⁻⁴
Achat	< 10 ⁻⁴	< 10 ⁻⁴	< 10 ⁻⁴	< 10 ⁻⁴					
Scrapm					< 10 ⁻⁴	0.0001			
Gemmere					0.2173	0.2347	0.2471	0.2591	0.5757
Modenaiss					0.4382	0.4210	0.4604	0.4871	0.5623
pARR (variable continue)					0.0416		0.1613		
pARRm (variable continue)						< 10 ⁻⁴		< 10 ⁻⁴	0.0016
pScrapm (variable continue)							< 10 ⁻⁴	< 10 ⁻⁴	0.0034
AIC		15092.18	14942.20		12112.82	12037.54	12123.20	12006.10	
Nombre de paramètres		33	33		37	37	37	37	37

^a Tous les animaux.^b Brebis nées dans l’élevage la même année ou après le cas index, de généalogie connue.

* Analyse excluant les animaux de génotype inconnu.

** Analyse considérant les animaux de génotype inconnu dans une classe séparée.

*** Analyse avec imputation multiple des génotypes inconnus.

7.3.2 Résultats de l'analyse 1

Modèle sans effets aléatoires

Nous présentons dans cette partie les résultats de l'analyse portant sur l'influence de l'âge au premier contact avec un milieu contaminant sur le risque d'infection et sur la durée d'incubation des animaux infectés. Le tableau 7.6 présente les résultats du modèle M3 obtenus à partir des données imputées pour les animaux dont le génotype PRP n'était pas connu. Les résultats des analyses ne prenant en compte que les animaux dont le génotype était connu (modèle M1) ou regroupant les animaux de génotype inconnu dans une classe séparée (modèle M2) sont présentés en annexes. La plupart des estimateurs de la partie incidence sont relativement semblables entre les 3 approches, mais ils diffèrent de façon plus importante pour la partie latence.

Les résultats indiquent que le génotype au gène PRP influençait de manière très significative le risque de contamination ($p < 10^{-4}$). Comparés aux animaux de génotypes sensibles (groupe SS), les animaux du groupe RS ($OR=0.05$, $p < 10^{-4}$), et les animaux de génotype RH ($OR=0.41$, $p=0.013$), étaient moins à risque de se contaminer, tandis que les animaux de génotypes hypersensibles (groupe SH) étaient beaucoup plus à risque ($OR = 3.33$, $p < 10^{-4}$). Le génotype PRP influençait également de manière significative le délai à la mort des ovins infectés ($p=0.005$). Ainsi les animaux des groupes RS et RH infectés mouraient plus tardivement que les animaux de génotype SS (respectivement $RR=0.71$, $p=0.038$ et $RR=0.32$, $p=0.031$), alors qu'il n'y avait pas de différence significative pour les animaux SH ($p=0.718$). Ces résultats sont en accord avec ceux obtenus dans la première analyse présentée dans le chapitre 6. Notons que les résultats des modèles M1 et M2 suggéraient que les animaux de génotype ARR/VRQ (groupe RH) avaient un risque de contamination semblable aux animaux de génotype SS et que les durées de survie des animaux RS infectés ne différaient plus de celles des animaux SS ($p=0.655$).

L'effet de l'âge au premier contact avec un milieu contaminant était fortement significatif. Comparés aux animaux nés dans les élevages, les animaux achetés étaient

moins fréquemment contaminés ($OR = 0.25$, $p < 10^{-4}$) et avaient des durées d'incubation allongées ($RR=0.47$, $p < 10^{-4}$). De même, à l'exception de l'élevage J, les animaux nés avant l'apparition de la tremblante dans les troupeaux étaient moins à risque de se contaminer ($OR = 0.57$, $p < 10^{-4}$) et mouraient plus tardivement de tremblante ($RR=0.22$, $p < 10^{-4}$) que ceux nés après. En revanche les animaux de millésime relatif négatif de l'élevage J étaient plus à risque de se contaminer ($OR=2.22$, $p=0.025$) mais avaient des durées d'incubation semblables.

Un effet élevage marqué était observé. Par rapport à l'élevage de référence C, les risques de contamination étaient significativement augmentés dans l'élevage J et réduits dans tous les autres élevages à l'exception de l'élevage F¹. De même les âges à la mort différaient de manière significative entre élevages (p global $< 10^{-4}$), les animaux mourant plus tardivement de tremblante dans les élevages A et J et plus précocement dans les élevages D, G et I que dans l'élevage C.

Modèle à 2 effets aléatoires non-indépendants

Le tableau 7.7 présente les résultats du modèle M4. Ce modèle correspondant, en termes de variables explicatives, au modèle M3, mais deux effets aléatoires sont ajoutés, l'un dans la partie incidence, l'autre dans la partie latente. Les estimateurs ne sont pas strictement comparables entre les deux modèles M3 et M4, car nous avons ici modélisé la distribution des temps de survie par un modèle exponentiel par morceaux en divisant l'axe des temps en 10 intervalles. Ce modèle est donc beaucoup moins souple que le modèle semi-paramétrique. L'objectif principal était de tester l'existence d'une corrélation entre les animaux d'un même millésime de naissance au sein d'un élevage donné.

Les estimateurs des rapports de cotes et des risques relatifs sont globalement comparables à ceux obtenus par le modèle M3. Notons cependant que, dans l'ensemble, les intervalles de confiance sont plus larges, traduisant une augmentation de la variance des estimateurs par la prise en compte du regroupement des données.

¹note : la dénomination des élevages ne correspond pas à celle de l'analyse du chapitre 6.

Nous n'avons pas estimé de modèle ne comprenant qu'un effet aléatoire, soit dans la partie survie, soit dans la partie incidence. Cependant un test du rapport de vraisemblance conservatif à 3 degrés de liberté confirme la supériorité du modèle à deux effets aléatoires par rapport au même modèle sans effets aléatoires ($\chi_3^2 = 35.254$, $p < 10^{-4}$). Ces résultats indiquent l'existence d'une forte corrélation entre les animaux d'un même groupe. La covariance des effets aléatoires n'était pas significativement différente de 0 ($p=0.5402$), suggérant que l'effet "millésime" influençait de manière indépendante le risque de contamination et les durées d'incubation.

TAB. 7.6 – Résultats du modèle M3 (Analyse 1, n=6687, modèles sans effets aléatoires)

Variable explicative	Partie Incidence			Partie Latence		
	OR	IC _{95%}	p	RR	IC _{95%}	p
GENOTYPE PRP			< 0.0001			0.005
Groupe SS	1.00	-		1.00	-	
Groupe RS	0.05	0.04 – 0.07	< 0.0001	0.71	0.51 – 0.98	0.0382
Groupe RH	0.41	0.20 – 0.83	0.0132	0.32	0.11 – 0.95	0.0308
Groupe SH	3.33	2.58 – 4.30	< 0.0001	0.96	0.79 – 1.17	0.7176
ACHAT			< 0.0001			< 0.0001
Non achetés	1.00	-		1.00	-	
Achetés	0.25	0.20 – 0.31	< 0.0001	0.47	0.37 – 0.60	< 0.0001
PÉRIODE DE NAISSANCE			< 0.0001			< 0.0001
Millésime relatif ≥ 0	1.00	-		1.00		
Millésime relatif < 0						
Élevage J	2.22	1.11 – 4.45	0.0246	0.16	0.11 – 0.23	< 0.0001
Autres élevages	0.57	0.43 – 0.76	< 0.0001	0.22	0.14 – 0.33	< 0.0001
ÉLEVAGE			< 0.0001			< 0.0001
Élevage A	0.37	0.27 – 0.50	< 0.0001	0.66	0.49 – 0.89	0.0063
Élevage B	0.16	0.11 – 0.23	< 0.0001	1.19	0.80 – 1.75	0.3904
Élevage C	1.00	-		1.00	-	
Élevage D	0.25	0.17 – 0.36	< 0.0001	2.37	1.61 – 3.45	< 0.0001
Élevage E	0.34	0.25 – 0.46	< 0.0001	0.96	0.73 – 1.26	0.7634
Élevage F	1.23	0.93 – 1.64	0.1496	0.99	0.78 – 1.26	0.9423
Élevage G	0.58	0.43 – 0.77	0.0001	2.06	1.61 – 2.62	< 0.0001
Élevage H	0.74	0.55 – 0.99	0.0408	0.98	0.77 – 1.24	0.8594
Élevage I	0.42	0.31 – 0.57	< 0.0001	1.51	1.15 – 1.98	0.0003
Élevage J	1.78	1.28 – 2.47	0.0005	0.82	0.65 – 0.98	0.0433
Élevage K	0.44	0.32 – 0.60	< 0.0001	0.80	0.59 – 1.08	0.1405

OR = odds ratio; RR = risque relatif

TAB. 7.7 – Résultats du modèle M4 (Analyse 1, n=6687, modèles avec 2 effets aléatoires indépendants)

Variable explicative	Partie Incidence			Partie Latence		
	OR	IC _{95%}	p	RR	IC _{95%}	p
GENOTYPE PRP			< 0.0001			0.003
Groupe SS	1.00	-		1.00	-	
Groupe RS	0.05	0.04 – 0.10	< 0.0001	0.43	0.24 – 0.74	0.0047
Groupe RH	0.34	0.15 – 0.66	0.0055	0.60	0.45 – 1.02	0.0755
Groupe SH	3.60	2.49 – 5.06	< 0.0001	0.95	0.77 – 1.14	0.6524
ACHAT			< 0.0001			< 0.0001
Non achetés	1.00	-		1.00	-	
Achetés	0.26	0.17 – 0.39	< 0.0001	0.43	0.27 – 0.68	0.0001
PÉRIODE DE NAISSANCE			< 0.0001			< 0.0001
Millésime relatif ≥ 0	1.00	-		1.00		
Millésime relatif < 0						
Élevage J	2.46	1.02 – 4.60	0.0431	0.17	0.11 – 0.33	< 0.0001
Autres élevages	0.50	0.39 – 0.86	< 0.0001	0.18	0.11 – 0.29	< 0.0001
ÉLEVAGE			< 0.0001			< 0.0001
Élevage A	0.26	0.15 – 0.45	< 0.0001	0.34	0.29 – 0.76	0.0023
Élevage B	0.20	0.18 – 0.33	< 0.0001	1.22	0.75 – 1.84	0.4914
Élevage C	1.00	-		1.00	-	
Élevage D	0.35	0.22 – 0.46	< 0.0001	2.32	1.60 – 3.44	< 0.0001
Élevage E	0.31	0.24 – 0.56	< 0.0001	1.02	0.70 – 1.36	0.8712
Élevage F	1.34	0.88 – 2.11	0.3496	0.99	0.78 – 1.26	0.9423
Élevage G	0.62	0.33 – 0.86	0.0044	2.16	1.81 – 2.82	< 0.0001
Élevage H	0.65	0.56 – 1.10	0.0854	0.95	0.75 – 1.32	0.8794
Élevage I	0.32	0.25 – 0.45	< 0.0001	1.54	1.22 – 2.05	0.0001
Élevage J	1.88	1.34 – 2.87	< 0.0001	0.72	0.64 – 1.03	0.0634
Élevage K	0.55	0.31 – 0.62	< 0.0001	0.95	0.71 – 1.22	0.4405

Effet aléatoire incidence $\hat{\theta}_v = 0.6348$ Effet aléatoire survie $\hat{\theta}_w = 0.6885$ Covariance $cov(\hat{\theta}_v, \hat{\theta}_w) = 0.0945$ $var(cov(\hat{\theta}_v, \hat{\theta}_w)) = 0.1543$ $p = 0.5402$

7.3.3 Résultats de l'analyse 2

L'analyse a été conduite sur les seuls animaux nés la même année ou après le cas index, en excluant les animaux achetés et les animaux de filiation maternelle inconnue.

Modèle sans effets aléatoires

Les résultats, présentés dans le tableau 7.8, sont sensiblement équivalents à ceux obtenus précédemment en ce qui concerne l'effet du génotype PRP. Notons, cependant, que, cette fois, l'influence du génotype RH sur l'âge à la mort des animaux infectés n'est plus significative au seuil de 5% (RR=0.46, p=0.099).

Le fait d'être issu d'une portée multiple n'avait pas d'influence, ni sur le risque de contamination (p=0.382), ni sur la durée de survie des animaux contaminés (p=0.222). Le génotype de la mère n'avait pas non plus d'influence significative (p=0.274). L'interaction entre le génotype de la mère et le statut clinique de la mère n'a pas été testée, compte-tenu du très faible nombre de brebis issues de mère malades et porteuses de l'allèle ARR (n=12 en moyenne sur 100 imputations).

Un effet protecteur significatif du niveau de résistance des mères (variable pARRm) était observé à la fois sur le risque de contamination (p=0.0080) et sur l'âge à la mort de tremblante (p=0.0007). A l'inverse, le risque de contamination était augmenté (p=0.013) et les durées d'incubation étaient réduites ($p < 10^{-4}$) quand le pourcentage de cas cliniques observés parmi les mères était élevé (variable pSCRAPm). Lorsque l'information du statut clinique de la mère est considérée individuellement (modèle M6), les résultats du modèle M6 indiquaient que les animaux dont la mère était morte de tremblante étaient plus à risque de se contaminer (OR=1.34, IC_{95%}=1.04 – 1.72, p=0.025) et mourraient plus précocement (RR=1.52, IC_{95%}=1.27 – 1.82, $p < 10^{-4}$) que ceux dont la mère n'était pas malade. Globalement, l'effet du niveau de résistance des agnelles conservées dans le millésime (variable pARR) n'était que faiblement significatif (p=0.041) quand le statut clinique de la mère était considéré

individuellement (modèle M5). Un effet protecteur significatif était uniquement observé sur la partie latence du modèle (RR=0.86, IC_{95%}= 0.77 – 0.97, p=0.012, pour une augmentation de 10% du niveau de résistance). Cet effet n'était plus significatif après ajustement sur le le pourcentage de mères malades (modèle M7, p global=0.161).

L'effet de la variable élevage reste très significatif, mais les rapports de cotes et les risques relatifs sont fortement modifiés pour certains troupeaux, indiquant qu'une partie de l'information relative à l'effet de l'élevage est prise en compte par le niveau de résistance et de contamination des mères dans le modèle M8 comparativement au modèle M3. Ainsi, par exemple, les élevages I, J et K, considérés, dans l'analyse 1, comme présentant un risque de contamination réduit par rapport à l'élevage de référence C, ne diffèrent plus de manière significative à même niveau de résistance et de contamination des mères. Le même phénomène est observé pour la partie latence pour les élevages A, G, I et J. A l'inverse, une fois ajustées sur la pression d'infection estimée, les durées d'incubation deviennent significativement allongées dans l'élevage K.

Modèle à 2 effets aléatoires non-indépendants

Le tableau 7.9 présente les résultats du modèle M9. Ce modèle incorpore les même variables explicatives que le modèle M8, mais deux effets aléatoires sont ajoutés, l'un pour la partie incidence, l'autre pour la partie latente. Comme pour l'analyse 1, l'objectif principal de cette analyse était de tester la présence d'une corrélation entre les animaux d'un même millésime de naissance au sein d'un élevage donné.

De même que dans l'analyse 1, les estimateurs des rapports de cotes et des risques relatifs sont globalement assez peu modifiés par rapport au modèle sans effets aléatoires. Le test du rapport de vraisemblance conservatif à 3 degrés confirme, là aussi, l'existence d'une forte corrélation entre les animaux d'un même millésime de naissance ($\chi^2_3 = 20.254$, p=0.0003), même après la prise en compte du niveau de résistance et du statut clinique des mères. La covariance des effets aléatoires n'était

pas non plus significativement différente de 0 ($p=0.2076$).

La figure 7.5 représente la fonction conditionnelle de risque de base et de la fonction marginale de risque de base, estimées par le modèle M9. Un premier pic aux alentours de 2 ans, puis une nouvelle augmentation à partir de 4 ans sont visibles sur le tracé de la fonction de risque conditionnelle. A l'échelle de la population (fonction de risque marginale), le risque de mourir de tremblante est maximal autour de 2 ans, et décroît fortement après 2.5 à 3 ans. Un pic beaucoup moins marqué est de plus observé vers 4 ans. Bien que moins précis en raison de l'utilisation d'un nombre limité d'intervalles dans la fonction exponentielle par morceaux, ces résultats sont comparables, en terme de forme et d'intensité du risque, à ceux obtenus dans l'analyse par vraisemblance pénalisée à partir d'un échantillon plus faible d'animaux.

TAB. 7.8 – Résultat du modèle M8 (Analyse 2, n=4486, modèles sans effets aléatoires)

Variable explicative	Partie Incidence			Partie Latence		
	OR	IC _{95%}	p	RR	IC _{95%}	p
GÉNOTYPE PRP			< 0.0001			0.002
Groupe SS	1.00	-		1.00	-	
Groupe RS	0.05	0.04 – 0.07	< 0.0001	0.62	0.43 – 0.88	0.0073
Groupe RH	0.31	0.16 – 0.60	0.0006	0.46	0.18 – 1.16	0.0993
Groupe SH	2.89	2.15 – 3.90	< 0.0001	1.01	0.82 – 1.24	0.9371
MODE DE NAISSANCE			0.3818			0.2220
Naiss. simple	1.00	-		1.00	-	
Naiss. multiple	1.08	0.91 – 1.27	0.3818	0.92	0.81 – 1.05	0.2220
GÉNOTYPE DE LA MÈRE			0.2744			0.2179
Mère sensible	1.00	-		1.00		
Mère porteuse “ARR”	1.14	0.90 – 1.43	0.2744	1.09	0.92 – 1.42	
pARRm (versus 40 %)						
Augmentation de 10%	0.44	0.24 – 0.81	0.0080	0.37	0.20 – 0.66	0.0007
pSCRAPm (versus 10 %)						
Augmentation de 10%	1.21	1.04 – 1.41	0.0132	1.51	1.23 – 1.85	< 0.0001
ÉLEVAGE			< 0.0001			0.003
Élevage A	3.66	0.88 – 4.34	0.0563	0.64	0.37 – 1.13	0.1276
Élevage B	0.43	0.24 – 0.77	0.0044	0.93	0.44 – 1.96	0.8508
Élevage C	1.00	-		1.00	-	
Élevage D	0.24	0.11 – 0.52	0.0004	5.49	2.46 – 12.27	< 0.0001
Élevage E	0.34	0.20 – 0.58	< 0.0001	1.92	1.07 – 3.44	0.0284
Élevage F	1.15	0.74 – 1.80	0.5258	1.43	0.94 – 2.17	0.0940
Élevage G	0.87	0.50 – 1.52	0.6164	1.72	0.99 – 2.99	0.0552
Élevage H	0.39	0.22 – 0.70	0.0016	1.41	0.83 – 2.40	0.2002
Élevage I	0.83	0.50 – 1.37	0.4632	1.06	0.61 – 1.85	0.8363
Élevage J	3.50	1.49 – 8.23	0.0040	0.24	0.10 – 0.59	0.0019
Élevage K	0.82	0.44 – 1.55	0.5447	0.37	0.18 – 0.73	0.0043

OR = odds ratio; RR = risque relatif

TAB. 7.9 – Résultat du modèle M9 (Analyse 2, n=4486, modèles à 2 effets aléatoires non-indépendants)

Variable explicative	Partie Incidence			Partie Latence		
	OR	IC _{95%}	p	RR	IC _{95%}	p
GÉNOTYPE PRP			< 0.0001			0.002
Groupe SS	1.00	-		1.00	-	
Groupe RS	0.07	0.04 – 0.15	< 0.0001	0.56	0.38 – 0.88	0.0039
Groupe RH	0.29	0.13 – 0.67	0.0034	0.44	0.20 – 1.01	0.1407
Groupe SH	2.85	1.83 – 4.45	< 0.0001	1.20	0.76 – 1.91	0.7847
MODE DE NAISSANCE			0.8472			0.5754
Naiss. simple	1.00	-		1.00	-	
Naiss. multiple	1.03	0.79 – 1.34	0.8472	0.92	0.67 – 1.25	0.5754
GÉNOTYPE DE LA MÈRE			0.2744			0.5623
Mère sensible	1.00	-		1.00	-	
Mère porteuse “ARR”	1.10	0.78 – 1.49	0.3744	1.14	0.82 – 1.44	0.5623
pARRm (versus 40 %)						
Augmentation de 10%	0.29	0.10 – 0.82	0.01918	0.55	0.37 – 0.88	0.0017
pSCRAPm (versus 10 %)						
Augmentation de 10%	1.35	1.07 – 1.71	0.0126	1.77	1.38 – 2.28	< 0.0001
ÉLEVAGE			< 0.0001			0.0213
Élevage A	1.14	0.21 – 6.23	0.8807	0.91	0.58 – 1.43	0.6785
Élevage B	0.58	0.36 – 0.92	0.0219	1.39	0.74 – 2.58	0.8654
Élevage C	1.00	-		1.00	-	
Élevage D	0.14	0.06 – 0.30	< 0.0001	4.74	1.88 – 11.90	< 0.0001
Élevage E	0.32	0.11 – 0.75	0.0085	1.92	1.15 – 3.20	0.0124
Élevage F	0.89	0.49 – 1.63	0.7188	1.85	1.00 – 3.43	0.0498
Élevage G	0.48	0.19 – 1.22	0.1260	2.02	0.66 – 6.20	0.2224
Élevage H	0.19	0.06 – 0.64	0.0078	1.70	0.88 – 3.29	0.1136
Élevage I	0.81	0.32 – 2.08	0.6599	1.66	0.78 – 3.18	0.2040
Élevage J	2.51	1.15 – 5.47	0.0203	0.28	0.10 – 0.76	0.0121
Élevage K	0.63	0.29 – 1.37	0.2449	0.24	0.12 – 0.53	0.0003

Effet aléatoire incidence $\hat{\theta}_v = 0.5196$ Effet aléatoire survie $\hat{\theta}_w = 0.4885$ Covariance $côv(\hat{\theta}_v, \hat{\theta}_w) = 0.1345$ $var(côv(\hat{\theta}_v, \hat{\theta}_w)) = 0.1066$ $p=0.2076$

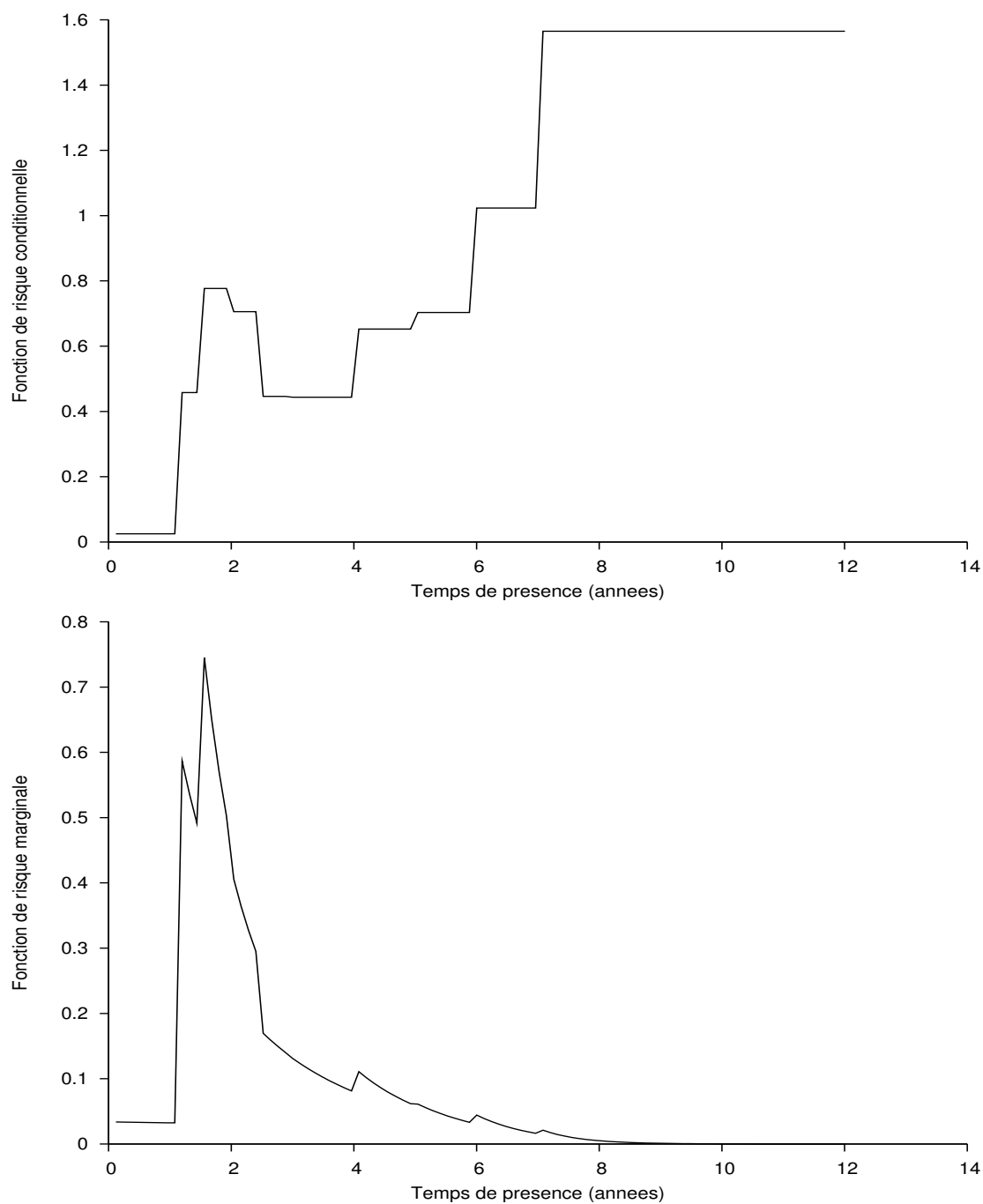


FIG. 7.5 – Estimateurs de la fonction de risque conditionnelle de base et de la fonction de risque marginale de base (modèle M9).

7.3.4 Estimation des fractions infectées

A partir des estimateurs des paramètres de régression $\hat{\Theta}$ et des estimateurs de la fonction de survie conditionnelle de base $\hat{S}_0(t|U = 1)$, nous pouvons calculer, pour chaque animal i , la probabilité *a posteriori* $\hat{\pi}_i$ d'être infecté, sachant ses différentes caractéristiques (vecteur \mathbf{x}_i des variables explicatives), le fait qu'il soit observé comme cas clinique (indicateur de censure δ_i) et son temps de survie t_i . Cette probabilité est définie par :

$$\begin{aligned}\hat{\pi}_i &= Pr(U = 1 | \mathbf{x}_i, (t_i, \delta_i), \hat{\Theta}, \hat{S}_0(t|U = 1)) \\ &= \delta_i + (1 - \delta_i) \frac{\hat{p}_i(\mathbf{x}_i, \hat{\Theta}) \hat{S}(t|U = 1, \mathbf{x}_i, \hat{\Theta})}{1 - \hat{p}_i(\mathbf{x}_i, \hat{\Theta}) + \hat{p}_i(\mathbf{x}_i, \hat{\Theta}) \hat{S}(t|U = 1, \mathbf{x}_i, \hat{\Theta})}\end{aligned}$$

où $\hat{p}_i(\cdot)$ est fourni par le modèle logistique et $\hat{S}(t|U = 1, \cdot)$ par le modèle à risque proportionnel.

En sommant les probabilités individuelles, nous avons calculé, pour chaque élevage, le nombre prédit d'animaux infectés pour les différents modèles. Les résultats sont présentés, dans le tableau 7.10, selon les modèles semi-paramétriques de l'analyse 2. Nous n'avons pas calculé les fractions à risque estimées à partir des modèles à effets aléatoires car le programme Fortran utilisé ne permet pas, pour le moment, d'estimer la valeur des effets aléatoires pour une prédiction individuelle. Nous avons comparé ces valeurs à l'estimateur non paramétrique de la fraction à risque fourni, pour chaque élevage j , par $\hat{F}_j(t_{n_j})$, qui est la valeur de l'estimateur de Kaplan-Meier de la fonction de distribution des temps de survie au dernier temps de survie t_{n_j} dans l'élevage j .

Les différents modèles évalués prédisent, dans l'ensemble, des fractions à risque très semblables. Les disparités les plus marquées entre modèles sont observées pour les élevages C (de 196.3 à 207.9 animaux infectés prédits) et H (de 170.5 à 183.1). En moyenne, la fraction à risque estimée est supérieure de 26.20% (modèle M7) à 27.55% (modèle M5) au nombre de cas cliniques observé. Selon les élevages entre 13% (élevage G) et 42% (élevage A) des animaux infectés auraient ainsi été réformés avant l'apparition des signes cliniques.

Les temps de suivi des animaux différaient significativement entre élevages (test du log-rank, $\chi^2_{10} = 104.17$, $p < 10^{-4}$). Nous avons cherché à vérifier si la variabilité de l'écart entre le nombre de cas cliniques et le nombre prédit d'animaux infectés pouvait être liée à l'hétérogénéité des temps de présence. Nous avons classé les élevages en trois classes selon la valeur de cet écart : $< 21\%$ (élevages F, G, I et J), entre 21% et 30% (élevages B, D, E et H) et supérieure à 30% (élevages A, C et K). Nous avons ensuite tracé la fonction de survie de Kaplan-Meier inverse en fonction de ces 3 classes. Les temps de suivi ne différaient pas entre les 3 groupes (test du log-rank, $\chi^2_2 = 0.966$, $p = 0.6168$). Les différentes politiques de réforme des éleveurs, qui conditionnaient les différences de temps de présence entre élevages, ne permettaient donc pas d'expliquer la variabilité observée entre nombre de cas cliniques observés et nombre prédit d'animaux infectés.

Les valeurs prédites par les différents modèles sont très proches de celles fournies par l'estimateur de Kaplan-Meier (tableau 7.10), excepté pour les élevages A et C, pour lesquels des différences supérieures à 3% sont notées. Malgré tout, les valeurs prédites par les modèles de régression sont toujours incluses dans l'intervalle de confiance à 95% de l'estimateur de Kaplan-Meier de la fraction à risque. Ces résultats suggèrent que les modèles de régression utilisés fournissent une adéquation relativement bonne aux données.

TAB. 7.10 – Fractions à risque estimées par élevage selon les différents modèles évalués

Élevage	Eff.	Cas observés		Nombre prédit (%) d'animaux infectés par le modèle											
		#	(%)	M5		M6		M7		M8		$\hat{F}_j(t_{n_j})$ [IC _{95%}]			
A	594	65	(10.9)	113.4	(19.1)	114.1	(19.2)	112.9	(19.0)	114.3	(19.2)		15.1	[11.9-19.0]	
B	289	31	(10.7)	42.2	(14.6)	44.1	(15.2)	43.7	(15.1)	41.8	(14.5)	13.8	[9.8-19.3]		
C	503	129	(25.6)	213.0	(42.3)	202.0	(40.2)	196.3	(39.0)	207.9	(41.3)	36.7	[31.5-42.4]		
D	214	32	(14.9)	44.7	(20.9)	45.0	(21.0)	44.5	(20.8)	44.2	(20.7)	20.9	[15.0-28.6]		
E	510	87	(17.1)	120.6	(23.7)	120.7	(23.7)	119.0	(23.3)	119.3	(23.4)	23.0	[18.6-28.3]		
F	354	131	(37.0)	165.0	(46.6)	166.5	(47.0)	162.2	(45.8)	162.7	(46.0)	46.7	[40.4-53.5]		
G	384	97	(25.3)	111.9	(29.1)	111.9	(29.1)	112.0	(29.2)	112.3	(29.2)	29.1	[24.5-34.4]		
H	474	128	(27.0)	176.6	(37.3)	170.5	(36.0)	173.8	(36.7)	183.1	(38.6)	38.3	[32.1-45.2]		
I	397	80	(20.1)	98.9	(24.9)	99.9	(25.2)	99.1	(25.0)	98.5	(24.8)	23.0	[18.9-27.9]		
J	320	132	(41.2)	159.3	(49.8)	158.2	(49.4)	156.7	(49.0)	158.9	(49.7)	47.5	[41.4-54.0]		
K	447	63	(14.1)	100.3	(22.4)	102.7	(23.0)	101.0	(22.6)	101.6	(23.7)	22.7	[15.8-31.9]		
Total	4486	975	(21.7)	1345.8	(30.0)	1335.6	(29.8)	1321.1	(29.5)	1344.6	(30.0)	29.4	[27.6-31.3]		

7.4 Discussion

Dans ce chapitre 7, nous avons présenté les résultats de l'analyse des données d'un suivi longitudinal de plusieurs troupeaux atteints de tremblante classique dans les Pyrénées Atlantiques. Cette analyse complète celle présentée dans le chapitre 6, et dans laquelle, seuls les effets du génotype au gène PRP et de l'élevage étaient considérés. L'étude d'un jeu de données élargi, incorporant un plus grand nombre d'élevages et de millésimes de naissance par élevage nous a permis de prendre en considération plusieurs variables pertinentes d'un point de vue biologique, telles que l'âge au premier contact avec un milieu contaminant, les relations mères-agneaux et la pression d'infection estimée autour de l'agnelage. L'utilisation d'un modèle de survie de mélange avec fraction non à risque nous a permis d'évaluer l'influence de ces différents facteurs de risque à la fois sur la probabilité de contamination et sur la durée d'incubation des animaux infectés. Enfin nous nous sommes intéressés au nombre estimé d'animaux infectés par élevage et avons confronté les résultats au nombre de cas cliniques observés. Nous discutons ici quelques points qui nous semblent importants.

Typages PRP manquants et imputation multiple. L'analyse a été compliquée par la nécessité de tenir compte de l'information manquante concernant le génotype PRP, déterminant majeur de la maladie. Plusieurs éléments suggèrent que les données manquantes ne l'étaient pas aléatoirement (MCAR = "Missing Completely At Random"). En effet, parmi les animaux appartenant aux millésimes de naissance les plus anciens, seuls ceux qui ont vécu suffisamment longtemps ont été génotypés. Ainsi les durées de présence des animaux non génotypés étaient très significativement inférieures à celles des animaux génotypés (test du log-rank, $\chi_1^2=2354.3395$, $p < 10^{-4}$). Dans le présent suivi, une sélection des animaux les plus "résistants" à la maladie était donc possible et les animaux génotypés ne constituaient probablement pas un échantillon représentatif de l'ensemble des animaux

vivant dans les élevages suivis. Un phénomène de troncature à gauche était donc présent quand l'analyse était conduite sur les seuls animaux de génotype connu (modèle M1). Ceci se traduisait, dans notre analyse, par une modification des estimateurs de l'influence du génotype PRP, comparés à l'analyse conduite par imputation multiple. Lorsque les animaux de génotype inconnu étaient considérés dans une classe séparée (modèle M2), ils présentaient un sur-risque de contamination très marqué et des durées d'incubation plus courtes, comparés aux animaux de génotype SS connu. Ce résultat s'explique par le nombre important de cas cliniques parmi les animaux non génotypés.

Dans un modèle de survie classique appliqué aux données du troupeau expérimental de Langlade, Vitezica et al. (2005) obtiennent des risques relatifs comparables entre l'analyse réalisée en excluant les animaux de génotype inconnu et celle fondée sur la prédiction des génotypes. Ils constatent cependant une modification majeure des courbes de survie estimées entre les deux approches. Ce phénomène est lié à la présence d'un nombre important d'animaux de génotype inconnu qui ne sont pas morts de tremblante et dont les temps de survie sont censurées à droite. Dans notre étude, la proportion de données manquantes est comparable (34.89% pour l'analyse 1 et 31.65% pour l'analyse 2) à celle observée dans l'analyse des données de Langlade (32.55%). L'information relative aux temps de survie des animaux est une composante importante de notre approche par modèle de survie avec fraction non à risque et l'estimateur de la fonction de survie conditionnelle influence les résultats de la partie incidence. L'analyse excluant les animaux de génotype inconnu ne prend pas en compte l'information relative aux durées de survie de ces animaux. La modification des estimateurs des paramètres de régression observée dans notre étude entre l'analyse conduite par imputation multiple et celles excluant les animaux de génotype inconnu ou les considérant dans une classe séparée pourrait aussi être expliquée par la prise en compte de cette information supplémentaire. L'analyse fondée sur la prédiction des génotypes manquants a donc été préférée car elle prenait en compte l'ensemble de l'information disponible.

Année d'apparition de la tremblante dans les troupeaux. Une des difficultés rencontrées dans cette analyse a été la définition de l'année d'apparition de la tremblante dans les troupeaux suivis. Cette information ne peut généralement pas être établie avec précision en raison des longues durées d'incubation caractérisant la tremblante (Hoinville, 1996 ; Baylis et al., 2002) et est sujette à discussion, même dans les troupeaux expérimentaux (Elsen et al., 1999 ; Redman et al., 2002). L'année réelle d'apparition de la tremblante dans les troupeaux est souvent bien antérieure à celle de l'observation des premiers cas cliniques (Baylis et al., 2000). Dans notre étude, nous avons tenté de définir cette information en confrontant les dires des éleveurs, les données relatives au nombre et à l'âge des premiers cas cliniques observés et à leur année de naissance. Nous nous sommes de plus fondés sur les connaissances de physiopathologie qui suggèrent une contamination néonatale précoce et des durées d'incubation minimales d'environ 1 an en conditions expérimentales. Cette approche nous a permis d'établir avec un certain degré de certitude l'année d'apparition de la tremblante dans la majorité des troupeaux, à l'exception des élevages C et J. Pour l'élevage C, l'incertitude était de 1 an. Dans l'élevage J, un nombre important de cas cliniques a été observé parmi les animaux considérés comme nés avant l'apparition de la tremblante (campagne 1995). Ces observations sont compatibles avec une apparition réelle de la tremblante beaucoup plus précoce. Cependant, ces cas cliniques étaient tous beaucoup plus âgés que ceux nés après 1995, ce qui était compatible avec une contamination à l'âge adulte. Un défaut de détection des premiers cas cliniques dans l'élevage J pourrait expliquer ces observations. Pour tenir compte de la singularité de cet élevage, nous avons choisi de considérer séparément les animaux nés avant la date supposée d'apparition de la tremblante dans ce troupeau.

Résultats biologiques. Les résultats de cette étude confirment le rôle déterminant du génotype PRP sur le risque de contamination des ovins par la tremblante. L'effet protecteur majeur de l'allèle ARR, déjà largement connu, est retrouvé. En revanche les études épidémiologiques portant sur l'influence de cet allèle sur la durée d'in-

incubation de la maladie sont rares, voire inexistantes. Nos résultats suggèrent que les animaux hétérozygotes ARR infectés ont des durées de survie plus longues que les animaux de génotypes sensibles. Dans l'analyse présentée au chapitre 6, les animaux de génotype ARR/VRQ étaient confondus avec les animaux ARR/ARQ et ARR/AHQ (groupe RS) en raison de leur très faible nombre (3 animaux). Dans la présente étude, l'utilisation d'un échantillon d'étude de taille nettement plus importante nous a permis d'évaluer spécifiquement l'influence de ce génotype codifié RH. Les résultats indiquent que, comparés aux animaux SS, les animaux de génotype ARR/VRQ ont un risque réduit de contamination. L'influence de ce génotype sur les durées d'incubation n'est pas clairement établie, les modèles concluant soit à un effet protecteur, soit à l'absence d'effet.

À travers le suivi d'animaux ne naissant pas dans un milieu contaminé (animaux achetés et animaux nés avant l'apparition de la tremblante dans les troupeaux), nous avons cherché à évaluer l'influence de l'âge sur le risque de contamination. La contamination à l'âge adulte reste possible, puisque des cas cliniques ont été observés chez des animaux nés avant l'apparition de la tremblante dans l'élevage. Cependant ces animaux étaient nettement moins à risque de se contaminer que les animaux nés ensuite dans un milieu fortement contaminé. De même, les animaux achetés, même très jeunes, étaient fortement protégés. Ces résultats semblent confirmer l'hypothèse d'une contamination néonatale précoce prépondérante (Andreoletti et al., 2000 ; van Keulen et al., 2002). Les études expérimentales montrent, qu'en conditions naturelles de contamination, les plaques de Peyer intestinales (appartenant au GALT) sont l'une des portes d'entrée principales de l'agent infectieux dans l'organisme. Ces structures s'atrophient avec l'âge, pour disparaître quasi-complètement chez l'animal de plus de 1 an. Une réduction de la perméabilité de la barrière intestinale pourrait donc expliquer le risque de contamination moindre observé chez les animaux entrant en contact à l'âge adulte avec le milieu contaminé. Une augmentation des âges à la mort chez les animaux infectés parmi les ovins achetés ou nés avant l'apparition de la tremblante dans les troupeaux a aussi été mise en évidence. Ces

résultats sont en accord avec les observations réalisées par Hourrigan et al. (1979) et Ryder et al. (2004) qui notent une réduction de l'incidence et l'allongement de la durée d'incubation chez les animaux exposés tardivement à un milieu contaminé. En pratique, les agnelles achetées à environ 2 mois d'âge ne sont pas mises en contact avec les femelles adultes ayant mis bas, mais sont élevées avec les agnelles de renouvellement dans des parcs séparés. Elles ne seraient donc pas exposées à l'agent infectieux au même moment que les agnelles nées dans l'élevage et se contamineraient plus tardivement. Ainsi, une contamination plus tardive pourrait expliquer la différence d'âge à la mort observée entre les animaux achetés ou non. Une augmentation du délai entre l'exposition à l'agent infectieux et la contamination, liée à une relation dose-effet ou à une réduction de la susceptibilité chez les ovins adultes, est aussi possible, comme le suggère l'augmentation des âges à la mort des animaux nés avant l'apparition de la tremblante dans les troupeaux.

Le niveau de contamination environnementale est une composante majeure de la dynamique de la tremblante dans les troupeaux infectés, mais n'est pas mesurable directement. La dissémination de l'infectiosité à travers les placentas contaminés des mères malades ou en incubation (Andreoletti et al., 2002 ; Lacroux et al., 2007) et la persistance de l'agent infectieux dans l'environnement (Brown et Gajdusek, 1991 ; Georgsson et al., 2006) semblent jouer un rôle majeur dans la dynamique de l'infection au sein des troupeaux. La prise en compte du statut sanitaire des mères, soit au niveau individuel (variable SCRAPM), soit à l'échelle du millésime de naissance de leurs filles (variable pSCRAPm), ainsi que du pourcentage des mères portant l'allèle ARR présentes pendant la période d'agnelage (variable pARRm) nous renseignent cependant indirectement sur ce niveau de contamination. De même, la structure génétique au gène PRP des filles (variable pARR) conservées pour le renouvellement peut aussi être un marqueur intéressant. En fait, nos résultats indiquent que la structure génétique des filles conservées n'avait pas d'influence, ni sur le risque individuel de contamination, ni sur les durées d'incubation. Il est vrai que les agnelles conservées pour le renouvellement ne représentent qu'une faible

partie des agneaux nés chaque année dans le troupeau. L'absence d'information concernant le génotype PRP des agneaux mâles et femelles non conservés n'autorise en fait qu'une vision partielle de la structure génétique de chaque millésime. En revanche, la structure génétique des mères agnelant à chaque campagne et leur statut sanitaire influençaient significativement le risque de contamination des filles et leurs durées d'incubation. Ainsi, un niveau de résistance élevé des mères au gène PRP était associé à un risque de contamination réduit et des durées d'incubation allongées. A l'inverse, plus le nombre de mères malades ou en incubation était important, plus le risque de contamination des filles était élevé et leurs durées d'incubation réduites. Un effet dose-réponse pourrait expliquer ces résultats, compatibles avec les essais expérimentaux de transmission, qui mettent en évidence une incidence plus faible de la maladie et un allongement de la période d'incubation lorsque l'inoculum est dilué (Kimberlin et Walker, 1978b). Cependant, le fort degré de significativité de la variable élevage et des effets aléatoires "millésime de naissance", même après ajustement, indiquent que les niveaux de résistance et de contamination des mères n'expliquent cependant qu'une partie des différences observées entre élevages et entre millésimes au sein d'un même élevage. Enfin, la prise en compte du statut clinique des mères était plus explicatif des données observées lorsqu'il était considéré à l'échelle du millésime de naissance (variable pSCRAPm) plutôt qu'à l'échelle individuelle (variable SCRAPM). Ces résultats suggèrent que la contamination environnementale était importante dans les troupeaux suivis, de sorte que le contact avec le milieu contaminé influençait plus le risque d'infection que le lien entre une mère malade et son (ses) agneaux.

L'utilisation de modèles de survie avec fraction non à risque nous a permis d'estimer la proportion d'animaux infectés parmi les animaux réformés sans symptômes de la maladie. Nos résultats indiquent qu'entre 15 et 40% des animaux infectés ne sont pas détectés par la seule surveillance clinique, en raison d'une sortie d'élevage trop précoce. Ces estimations sont très inférieures à celles fournies par les modèles dynamiques proposés par Matthews et al. (2001) et Hagenars et al. (2006). D'après

ces auteurs, entre 70% et 80% des animaux infectés seraient réformés avant l'apparition des signes cliniques : ces résultats sont cependant obtenus en considérant des durées moyennes de présence de 3 ans dans les troupeaux. Dans notre échantillon d'étude, la durée de présence médiane était plus élevée, supérieure à 4.5 ans. Le fait que les animaux vivent plus longtemps dans notre étude a permis l'expression des signes cliniques pour une proportion plus importante d'animaux infectés, réduisant par conséquent la part des animaux réformés en incubation de la maladie.

Chapitre 8

Discussion & perspectives

Ce travail de thèse était motivé par l'étude des facteurs de risque liés à la tremblante ovine classique. Bien qu'utiles à la compréhension de la dynamique de la maladie au sein des troupeaux infectés, ces facteurs de risque demeurent mal connus. Les enquêtes cas-témoins ou les études transversales existantes se sont principalement concentrées sur les facteurs génétiques. L'exploration des facteurs environnementaux et individuels autres que génétiques par le biais de ces schémas d'étude est rendue difficile par les longues durées d'incubation de cette maladie. Pour ce travail nous nous sommes fondés sur un suivi longitudinal conduit pendant de nombreuses années dans des troupeaux fortement atteints en Pyrénées Atlantiques. Ces conditions de suivi avantageuses, rarement rencontrées en dehors de quelques troupeaux expérimentaux dédiés à l'étude de cette maladie, ont permis l'observation d'un nombre important de cas cliniques de tremblante classique dans des élevages privés. Une telle situation épidémiologique unique a résulté d'une collaboration pérenne et confiante entre des éleveurs de ce département et leurs organismes techniques, et la recherche et l'enseignement agronomique et vétérinaire. Ce suivi, initialement fondé principalement sur la surveillance clinique, devait de fait se plier aux contraintes économiques et zootechniques qui limitent la longévité des animaux dans de tels troupeaux commerciaux. Par conséquent, les données collectées ne permettaient d'accéder qu'à une connaissance partielle du statut sanitaire des animaux. Leur analyse par des

approches classiques n'étant pas satisfaisante d'un point de vue statistique, de nouvelles méthodes devaient être employées. En particulier, les modèles de survie, classiquement employés dans l'analyse du délai de survenue d'un événement constituaient une approche intéressante, mais il fallait pouvoir tenir compte aussi du fait qu'une partie seulement des animaux, *a priori* inconnue, est contaminée par l'agent de la tremblante classique.

L'utilisation des modèles de survie avec fraction non à risque a été initiée dans le cadre des essais thérapeutiques de cancérologie (Boag, 1949; Berkson et Gage, 1952). Les avancées médicales et thérapeutiques permettent en effet de considérer que, pour certaines pathologies cancéreuses, des patients peuvent être considérés comme définitivement guéris après la mise en place d'un traitement chirurgical et/ou médicamenteux. L'objectivation de cette rémission complète ne peut cependant généralement se faire qu'après une longue période de surveillance. Ces patients ne peuvent en toute logique plus être considérés comme à risque de connaître l'événement d'intérêt (récidive ou décès des suites du cancer) du fait de leur guérison. Les modèles de survie classiques, nous l'avons vu, ne tiennent pas compte de la possibilité qu'une fraction de la population ne soit pas à risque et ne sont pas applicables, en toute rigueur, à l'analyse de telles données. Une des particularités majeures des modèles de survie avec fraction non à risque est de considérer que la durée de survie d'un individu apporte une information pertinente sur son risque de connaître l'événement. Ainsi, la probabilité qu'un patient soit en rémission complète est d'autant plus forte que son temps de survie est important. Le développement de ces modèles était aussi motivé par l'évaluation des différentes thérapeutiques (Taylor et Kim, 1993; Gamel et al., 1994; Ibrahim et al., 2001; Sposto, 2002). En effet, en permettant d'estimer la proportion de patients guéris et la distribution des temps de survie des patients non guéris, ils permettent d'évaluer si un traitement influence la guérison complète ou s'il retarde la récurrence. Depuis, l'utilisation des modèles de survie avec fraction non à risque a dépassé le cadre de la seule cancérologie et a été appliquée aux domaines de la criminologie, du marketing, de l'ingénierie ou de

la médecine vétérinaire (Broadhurst et Maller, 1991 ; Yamaguchi, 1992 ; Meeker et LuValle, 1995 ; Hanson et al., 2003).

L'utilisation de ces modèles de survie connaît cependant des limites. En premier lieu, elle est de fait conditionnée à l'existence d'une fraction non à risque. Des éléments biologiques sont souvent nécessaires pour juger de la pertinence d'une telle hypothèse. L'une des limites importantes est la nécessité de disposer d'une période de suivi suffisamment longue pour que la majorité des individus à risque connaissent l'événement. L'existence d'un plateau sur le tracé de Kaplan-Meier de la fonction de survie marginale, nous l'avons vu, constitue un indicateur simple permettant de s'assurer que de tels modèles sont applicables. Une des autres limites des modèles de survie avec fraction non à risque réside dans leurs faibles performances lorsque la fraction non à risque est très faible ou à l'inverse très élevée. Les études de simulation que nous avons menées indiquent en effet que les biais sur les estimateurs peuvent être importants dans ces conditions. De même, nos résultats montrent qu'un phénomène de censure précoce peut conduire à une perte d'information réduisant les performances de telles approches. Enfin l'utilisation de ces approches implique que les temps d'événement se concentrent sur une période de temps limitée au delà de laquelle la probabilité de connaître l'événement est réduite voire nulle. L'application de ces méthodes doit donc se faire à bon escient et n'est pas possible dans un nombre important de phénomènes pour lesquels l'événement peut avoir lieu même après une longue période de suivi.

L'utilisation de ces modèles à l'analyse des données des programmes Sheeprion et ScrapieFreeSheep nécessitait de vérifier que les conditions de suivi requises pour l'obtention d'estimateurs non biaisés étaient respectées. Via des études de simulation, nous avons montré que ces modèles étaient applicables aux données disponibles. Dans ce cadre, les résultats que nous obtenons dans notre application confirment un certain nombre d'observations réalisées par d'autres auteurs. En particulier, l'effet très protecteur sur le risque de contamination de l'allèle ARR au gène PRP et celui délétère de l'allèle VRQ sont confirmés en Race Manech Tête Rousse. De

même, l'influence de l'âge sur le risque de contamination, mise en évidence dans le cadre de d'études expérimentales (Ryder et al., 2004; Hourrigan et al., 1979) est retrouvée pour la première fois dans une étude à grande échelle, réalisée dans les conditions de terrain. Nos résultats sont aussi en faveur d'une contamination néonatale prépondérante, et sont en accord avec les données de physiopathologie (van Keulen et al., 2000; Andreoletti et al., 2000). L'originalité de l'approche statistique utilisée et la qualité des données exploitées ont, de plus, permis d'analyser l'influence de ces facteurs sur les durées d'incubation des animaux infectés. Nous montrons ainsi que l'allèle ARR est associé à des durées d'incubation allongées, ce qui conforte les résultats des études expérimentales. A l'inverse l'absence de différence entre les durées d'incubation des ovins de génotype ARQ/VRQ (groupe SH) et celles des brebis ARQ/ARQ (groupe SS) va à l'encontre des observations réalisées dans le troupeau expérimental de Langlade par Elsen et al. (1999). Ces résultats, ainsi que la mortalité précoce des cas cliniques ARQ/ARQ suggèrent l'existence d'une souche ayant un tropisme particulier pour ce génotype. Des études par transmission expérimentale devraient venir étayer cette hypothèse. Enfin un effet dose-réponse est suggéré par l'augmentation des âges à la mort des animaux entrant tardivement en contact avec un milieu contaminé.

L'approche utilisée nous a aussi permis d'estimer pour la première fois la proportion d'animaux infectés dans des élevages ovins atteints de tremblante classique. Nos résultats indiquent qu'une partie seulement des animaux infectés développe les signes cliniques de la maladie en raison d'une sortie précoce liée à des contraintes zootechniques. Dans certains élevages fortement atteints, plus de 40% des ovins pourraient ainsi être infectés. Les cas cliniques observés n'apparaissent être que la partie émergée de l'iceberg. Des proportions similaires sont obtenues par des études transversales menées à l'occasion de l'abattage total d'élevages infectés (Ersdal et al., 2005; Billinis et al., 2004) ou dans des troupeaux expérimentaux (Elsen et al., 1999). Ces résultats suggèrent aussi que la réalisation d'une analyse en cas témoins fondée sur la seule information clinique conduit à un biais de classification majeur des

animaux sans symptômes.

De manière intéressante, nos résultats sont en faveur d'une influence très significative de la pression d'infection sur le risque de contamination et sur les durées d'incubation. Cependant une forte hétérogénéité entre élevages demeure après ajustement sur les variables décrivant le niveau de résistance du troupeau adulte et le pourcentage de mères malades ou en incubation. Ceci illustre la difficulté de caractériser et de modéliser le niveau de contamination environnementale, mais aussi la spécificité de chaque élevage. Différentes hypothèses peuvent être évoquées pour expliquer l'effet résiduel fort de l'élevage. Des pratiques contribuant à réduire le risque de contamination chez les jeunes ovins (retrait des placentas, élevages des agnelles de renouvellement dans des bâtiments séparés, curage régulier des bergeries, élevages sur caillebotis plutôt que sur litière paillée...), ont été adoptées précocement par certains éleveurs. L'utilisation de méthodes biochimiques (cinétique de résistance à la protéinase K et profils de glycosylation) a permis de montrer, à partir de prélèvements d'un certain nombre des cas cliniques, l'existence de souches différentes d'agent infectieux dans les élevages suivis (Andréoletti, communication personnelle). La mise en relation de ces données de laboratoire et des résultats épidémiologiques pourrait contribuer à mieux comprendre la dynamique de l'infection dans ces troupeaux.

L'hétérogénéité entre les animaux appartenant à des groupes différents a aussi été mise en évidence par l'utilisation de modèles de survie avec fraction à risque et fragilités partagées. Largement utilisées dans le cadre des modèles linéaires ou dans celui des modèles de survie classiques (Rondeau et Commenges, 2002), ces méthodes sont encore peu développées pour les modèles de survie avec fraction non à risque. Les approches proposées sont pour la plupart fondées sur des modèles paramétriques simples (Yau et Ng, 2001 ; Yin et Ibrahim, 2005). Les difficultés liées à l'estimation des paramètres de la fonction de survie dans les modèles semi-paramétriques ont conduit certains auteurs à proposer des approches marginales, d'estimation plus simple (Peng et al., 2007). Nous avons développé des modèles paramétriques com-

prenant deux effets aléatoires indépendants ou non, afin de tenir compte du regroupement des animaux par élevages. Les performances des modèles proposés ont été évaluées par des études de simulation. L'objectif initial de généralisation de l'approche par vraisemblance pénalisée aux modèles à fragilités partagées n'a pas pu être atteint dans le cadre de ce travail de thèse, en raison de problèmes numériques pour le moment non résolus. Le modèle exponentiel par morceaux proposé offrait cependant une flexibilité suffisante pour modéliser la fonction de risque conditionnelle. Dans notre application, le regroupement à deux niveaux des animaux (millésime de naissance et troupeaux) nous a amené à faire un choix pour l'utilisation des modèles à effets aléatoires. Nous avons considéré le millésime de naissance au sein d'un élevage comme l'unité de regroupement plutôt que l'élevage lui-même. En ajustant les modèles utilisés sur la variable élevage, cette approche nous a permis de tenir compte du regroupement à deux niveaux des animaux. Un modèle à effets aléatoires emboîtés pourrait être envisagé. Les méthodes d'estimation basées sur le calcul numérique des intégrales pourraient cependant s'avérer très coûteuses en temps.

En épidémiologie, la fonction de risque est d'un intérêt particulier car elle est une description probabiliste du futur immédiat du sujet "encore à risque". Elle peut de plus s'interpréter en termes d'incidence lorsque le temps de base est l'âge ou la durée de présence. Dans le cadre des modèles de survie avec fraction non à risque, la fonction de risque conditionnelle permet de représenter l'incidence de l'événement d'intérêt chez les individus "à risque" de connaître cet événement. A l'inverse, le risque à l'échelle de la population est représenté à travers la fonction de risque marginale. La volonté d'obtenir un estimateur lisse et flexible de la fonction de risque a motivé le développement d'une approche d'estimation par vraisemblance pénalisée. Cette méthode permet, contrairement aux modèles paramétriques classiques, de ne pas faire d'hypothèses sur la distribution des temps d'événement. Cependant des problèmes d'identifiabilité ont été rencontrés dans certains cas, en raison de l'absence de contrainte sur la forme de la fonction de survie conditionnelle après le dernier

événements observé. Une solution envisagée pour faire face à ces difficultés est de réduire le choix des solutions possibles pour l'estimateur de la fonction de survie, en lui imposant des valeurs faibles voire nulles après le dernier événement. Cette méthode est déjà employée dans le cadre des modèles semi-paramétriques mais elle n'est satisfaisante en termes biologiques que lorsque les temps de suivi sont longs, comparativement à la distribution des temps d'événement. Elle demeure cependant envisageable dans notre application.

Les estimateurs des fonctions conditionnelle et marginale de risque obtenues par vraisemblance pénalisée dans notre application permettent de soulever plusieurs questions biologiques. En effet, la fonction de risque conditionnelle estimée présente un premier pic à l'âge de 2 ans puis une nouvelle augmentation aux alentours de 4 ans. Ces résultats sont retrouvés, de manière logiquement moins nette mais sur un échantillon d'étude plus important, par l'utilisation d'une fonction exponentielle par morceaux. Cette forme bimodale suggère une hétérogénéité du risque au sein des animaux infectés et pourrait traduire l'existence d'une contamination en dehors de la période néonatale de certains animaux nés dans un milieu infecté. D'autres processus biologiques, tels qu'un déterminisme génétique autre que lié au gène PRP, ou à l'existence de souches d'agent infectieux à durées d'incubation plus longues, peuvent là aussi être évoqués.

Dans notre application nous n'avons pas pris en compte les résultats d'exams complémentaires qu'une partie des ovins avait subi. Dans certains des élevages suivis, les animaux de génotypes sensibles et hypersensibles éliminés à l'équarrissage dans le cadre des mesures de police sanitaire ont fait l'objet d'une recherche de PrPres dans l'encéphale par l'utilisation de tests rapides. De plus, depuis 2000, nous avons réalisé des biopsies d'amygdales sur un grand nombre d'animaux. Ces exams complémentaires fournissent une information non négligeable sur le statut sanitaire des ovins testés. La prise en compte de cette information supplémentaire dans les modèles de survie pourrait permettre d'affiner l'estimation des fractions infectées. Pour cela, nous pourrions imposer une probabilité nulle d'être à risque pour les ani-

maux testés négatifs. A l'inverse les animaux testés positifs mais sortis des élevages avant l'apparition des signes cliniques, pourraient être considérés comme à risque et leur temps de survie censurés à droite. Cette approche devrait de plus tenir compte de la sensibilité non parfaite des outils diagnostique utilisés.

Notre étude des facteurs de risque liés à la tremblante ovine classique s'est concentrée sur une période où la stratégie de lutte contre la maladie fondée sur l'utilisation systématique de béliers ARR/ARR n'était pas encore mise en place dans les élevages suivis. La mise en place de ces mesures de lutte à partir de 1998-2001 dans ces élevages a profondément modifié leur structure génétique au gène PRP. De même l'euthanasie dans un délai de 2 ans des animaux de génotypes sensibles, imposée alors par les arrêtés ministériels de 2002 et 2003, a considérablement réduit le nombre de brebis susceptibles de développer la maladie. En conséquence, dans les élevages étudiés, une disparition des cas cliniques a suivi dans les deux ans à trois ans la mise en place de ce plan d'action. L'absence de nouveaux cas cliniques et les résultats des examens complémentaires réalisés dans le cadre du programme ScrapieFreeSheep et des mesures de police sanitaire semblent indiquer l'efficacité de la stratégie de lutte fondée sur la sélection de l'allèle ARR et l'élimination des animaux de génotypes sensibles. Cependant, ce plan d'action, adopté depuis 2003 à l'échelle européenne, n'a pour le moment pas fait l'objet d'étude d'évaluation à partir de données réelles. Le très faible nombre de cas cliniques observés après l'utilisation systématique de béliers ARR/ARR ne permet plus d'appliquer directement l'approche par modèle de survie avec fraction non à risque aux données relatives à cette nouvelle période. Cependant, une des perspectives que nous envisageons est d'utiliser les estimateurs fournis dans la présente analyse par ces modèles et les résultats des examens complémentaires disponibles afin d'évaluer le nombre d'animaux potentiellement infectés parmi les animaux nés dans ces élevages suite la mise en place des plans de lutte contre la tremblante ovine classique. L'analyse du temps nécessaire à l'éradication de la tremblante classique dans ces troupeaux sera d'autant plus pertinente que nous avons pu montré que la pression d'infection était

auparavant particulièrement élevée dans ces élevages privés en suivi longitudinal.

Chapitre 9

Bibliographie

- Aalen, O. O. (1988). Heterogeneity in survival analysis. *Stat. Med.* **7**, 1121–1137.
- Acin, C., Martin-Burriel, I., Goldmann, W., Lyahyai, J., Monzon, M., Bolea, R., Smith, A., Rodellar, C., Badiola, J. J. et Zaragoza, P. (2004). Prion protein gene polymorphisms in healthy and scrapie-affected spanish sheep. *J. Gen. Virol.* **85**, 2103–2110.
- Acin, C., Martin-Burriel, I., Monleon, E., Rodellar, C., Badiola, J. J. et Zaragoza, P. (2004). PrP polymorphisms in spanish sheep affected with natural scrapie. *Vet. Rec.* **155**, 370–372.
- Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. *International Symposium on Information Theory, 2 nd, Tsahkadsor, Armenian SSR* pages 267–281.
- Andreoletti, O., Berthon, P., Marc, D., Sarradin, P., Grosclaude, J., van Keulen, L., Schelcher, F., Elsen, J. M. et Lantier, F. (2000). Early accumulation of PrP(sc) in gut-associated lymphoid and nervous tissues of susceptible sheep from a romanov flock with natural scrapie. *J. Gen. Virol.* **81**, 3115–3126.
- Andreoletti, O., Lacroux, C., Chabert, A., Monnereau, L., Tabouret, G., Lantier, F., Berthon, P., Eychenne, F., Lafond-Benestad, S., Elsen, J. M. et Schelcher, F. (2002). PrP(sc) accumulation in placentas of ewes exposed to natural scrapie : influence of foetal PrP genotype and effect on ewe-to-lamb transmission. *J. Gen. Virol.* **83**, 2607–2616.
- Andreoletti, O., Morel, N., Lacroux, C., Rouillon, V., Barc, C., Tabouret, G., Sarradin, P.,

- Berthon, P., Bernardet, P., Mathey, J., Lugan, S., Costes, P., Corbiere, F., Espinosa, J. C., Torres, J. M., Grassi, J., Schelcher, F. et Lantier, F. (2006). Bovine spongiform encephalopathy agent in spleen from an ARR/ARR orally exposed sheep. *J. Gen. Virol.* **87**, 1043–1046.
- Arsac, J. N., Andreoletti, O., Bilheude, J. M., Lacroux, C., Benestad, S. L. et Baron, T. (2007). Similar biochemical signatures and prion protein genotypes in atypical scrapie and nor98 cases, france and norway. *Emerg. Infect. Dis.* **13**, 58–65.
- Barillet, F. (2003). Monitoring the effect of scrapie control policies that use genetics in different countries. european contract QLKS-CT 2002 01733 (acronym "scrapie-free-sheep"). Technical report.
- Barillet, F., Andreoletti, O., Palhière, I., Aguerre, X., Arranz, J., Minery, S., Soulas, C., Belloc, J., Briois, M., Frégeat, P., Teinturier, P., Amigues, Y., Astruc, J., Boscher, M. et Schelcher, F. (2002). Breeding for scrapie resistance using PrP genotyping in the French dairy sheep breeds. In *7th World Congress on Genetics Applied to Livestock Production*, August 19-23, Montpellier, France (CD-rom communication 12-20).
- Basler, K., Oesch, B., Scott, M., Westaway, D., Walchli, M., Groth, D. F., McKinley, M. P., Prusiner, S. B. et Weissmann, C. (1986). Scrapie and cellular PrP isoforms are encoded by the same chromosomal gene. *Cell* **46**, 417–428.
- Baylis, M., Chihota, C., Stevenson, E., Goldmann, W., Smith, A., Sivam, K., Tongue, S. et Gravenor, M. B. (2004). Risk of scrapie in british sheep of different prion protein genotype. *J. Gen. Virol.* **85**, 2735–2740.
- Baylis, M., Goldmann, W., Houston, F., Cairns, D., Chong, A., Ross, A., Smith, A., Hunter, N. et McLean, A. R. (2002). Scrapie epidemic in a fully PrP-genotyped sheep flock. *J. Gen. Virol.* **83**, 2907–2914.
- Baylis, M., Houston, F., Goldmann, W., Hunter, N. et McLean, A. R. (2000). The signature of scrapie : differences in the PrP genotype profile of scrapie-affected and scrapie-free uk sheep flocks. *Proc. Biol. Sci.* **267**, 2029–2035.
- Beekes, M. et McBride, P. A. (2000). Early accumulation of pathological PrP in the enteric nervous system and gut-associated lymphoid tissue of hamsters orally infected with scrapie. *Neurosci. Lett.* **278**, 181–184.
- Beekes, M. et McBride, P. A. (2007). The spread of prions through the body in naturally

- acquired transmissible spongiform encephalopathies. *Febs J.* **274**, 588–605.
- Bell, J. E., Gentleman, S. M., Ironside, J. W., McCardle, L., Lantos, P. L., Doey, L., Lowe, J., Fergusson, J., Luthert, P., McQuaid, S. et Allen, I. V. (1997). Prion protein immunocytochemistry—uk five centre consensus report. *Neuropathol. Appl. Neurobiol.* **23**, 26–35.
- Bellworthy, S. J., Hawkins, S. A., Green, R. B., Blamire, I., Dexter, G., Dexter, I., Lockey, R., Jeffrey, M., Ryder, S., Berthelin-Baker, C. et Simmons, M. M. (2005). Tissue distribution of bovine spongiform encephalopathy infectivity in romney sheep up to the onset of clinical disease after oral challenge. *Vet. Rec.* **156**, 197–202.
- Benestad, S. L., Sarradin, P., Thu, B., Schonheit, J., Tranulis, M. A. et Bratberg, B. (2003). Cases of scrapie with unusual features in norway and designation of a new type, nor98. *Vet. Rec.* **153**, 202–208.
- Berkson, J. et Gage, R. P. (1952). Survival curve for cancer patients following treatment. *J. Am. Stat. Assoc.* **47**, 501–515.
- Billinis, C., Psychas, V., Leontides, L., Spyrou, V., Argyroudis, S., Vlemmas, I., Leontides, S., Sklaviadis, T. et Papadopoulos, O. (2004). Prion protein gene polymorphisms in healthy and scrapie-affected sheep in greece. *J. Gen. Virol.* **85**, 547–554.
- Boag, J. W. (1949). Maximum likelihood estimates of the proportion of patients cured by cancer therapy. *J. R. Stat. Soc. Ser. B Met.* **11**, 15–53.
- Bohning, D. et Seidel, W. (2003). Editorial : recent developments in mixture models. *Comput. Stat. Data Anal.* **41**, 349–357.
- Breslow, N. E. (1972). Contribution to the discussion of the paper by dr cox. *J. Roy. Stat. Soc. B* **34**, 216–217.
- Broadhurst, R. et Maller, R. (1991). Estimating the numbers of prison terms in criminal careers from one-step probabilities of recidivism. *Journal of Quantitative Criminology* **7**, 275–290.
- Broet, P., De Rycke, Y., Tubert-Bitter, P., Lellouch, J., Asselain, B. et Moreau, T. (2001). A semiparametric approach for the two-sample comparison of survival times with long-term survivors. *Biometrics* **57**, 844–852.
- Brotherston, J. G., Renwick, C. C., Stamp, J. T., Zlotnik, I. et Pattison, I. H. (1968). Spread and scrapie by contact to goats and sheep. *J. Comp. Pathol.* **78**, 9–17.

- Brown, P. et Bradley, R. (1998). 1755 and all that : a historical primer of transmissible spongiform encephalopathy. *BMJ* **317**, 1688–1692.
- Brown, P., Cervenakova, L., McShane, L. M., Barber, P., Rubenstein, R. et Drohan, W. N. (1999). Further studies of blood infectivity in an experimental model of transmissible spongiform encephalopathy, with an explanation of why blood components do not transmit creutzfeldt-jakob disease in humans. *Transfusion* **39**, 1169–1178.
- Brown, P. et Gajdusek, D. C. (1991). Survival of scrapie virus after 3 years' interment. *Lancet* **337**, 269–270.
- Bruce, M. (1979). Biological stability of different classes of scrapie agent. In Prusiner, S. B., editor, *Slow transmissible diseases of the nervous system*, pages 71–86. Academic Press, New York.
- Bruce, M. E., Boyle, A., Cousens, S., McConnell, I., Foster, J., Goldmann, W. et Fraser, H. (2002). Strain characterization of natural sheep scrapie and comparison with bse. *J. Gen. Virol.* **83**, 695–704.
- Bruce, M. E. et Fraser, H. (1991). Scrapie strain variation and its implications. *Curr. Top. Microbiol. Immunol.* **172**, 125–138.
- Bruce, M. E., McConnell, I., Will, R. G. et Ironside, J. W. (2001). Detection of variant creutzfeldt-jakob disease infectivity in extraneural tissues. *Lancet* **358**, 208–209.
- Bruce, M. E., Will, R. G., Ironside, J. W., McConnell, I., Drummond, D., Suttie, A., McCardle, L., Chree, A., Hope, J., Birkett, C., Cousens, S., Fraser, H. et Bostock, C. J. (1997). Transmissions to mice indicate that 'new variant' cjd is caused by the bse agent. *Nature* **389**, 498–501.
- Buschmann, A., Luhken, G., Schultz, J., Erhardt, G. et Groschup, M. H. (2004). Neuronal accumulation of abnormal prion protein in sheep carrying a scrapie-resistant genotype (PrParr/arr). *J. Gen. Virol.* **85**, 2727–2733.
- Cantor, A. B. et Shuster, J. J. (1992). Parametric versus non-parametric methods for estimating cure rates based on censored survival data. *Stat. Med.* **11**, 931–937.
- Caramelli, M., Ru, G., Casalone, C., Bozzetta, E., Acutis, P. L., Calella, A. et Forloni, G. (2001). Evidence for the transmission of scrapie to sheep and goats from a vaccine against mycoplasma agalactiae. *Vet. Rec.* **148**, 531–536.
- Carp, R. I. (1982). Transmission of scrapie by oral route : effect of gingival scarification.

- Lancet* **1**, 170–171.
- Carp, R. I., Callahan, S. M., Sersen, E. A. et Moretz, R. C. (1984). Preclinical changes in weight of scrapie-infected mice as a function of scrapie agent-mouse strain combination. *Intervirology* **21**, 61–69.
- Carpenter, J. et Bithell, J. (2000). Bootstrap confidence intervals : when, which, what ? a practical guide for medical statisticians. *Stat. Med.* **19**, 1141–1164.
- Chandler, R. L. (1961). Encephalopathy in mice produced by inoculation with scrapie brain material. *Lancet* **1**, 1378–1379.
- Chen, M. H., Ibrahim, J. G. et Sinha, D. (1999). A new bayesian model for survival data with a surviving fraction. *J. Am. Stat. Assoc.* **94**, 909–910.
- Choi, K. C. et Zhou, X. (2002). Large sample properties of mixture models with covariates for competing risks. *J. Multivariate Anal.* **82**, 331–366.
- Clayton, D. G. (1978). A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence. *Biometrika* **65**, 141.
- Commenges, D., Joly, P., Gégout-Petit, A. et Liqueur, B. (2007). Choice between semi-parametric estimators of markov and non-markov multi-state models from coarsened observations. *Scand. J. Stat.* **34**, 33–52.
- Cox, D. R. (1972). Regression models and life-tables (with discussion). *J. R. Stat. Soc. Ser B Met* **34**, 187–220.
- Cox, D. R. (1984). *Analysis of Survival Data*. CRC Press.
- Cuillé, J. et Chelle, P. (1936). La maladie dite tremblante du mouton est-elle inoculable ? *C. R. Acad. Sci. Paris.* **203**, 1552–1554.
- Dawson, M., Hoinville, L. J., Hosie, B. D. et Hunter, N. (1998). Guidance on the use of PrP genotyping as an aid to the control of clinical scrapie. scrapie information group. *Vet. Rec.* **142**, 623–625.
- Del Rio Vilas, V. J., Hopp, P., Nunes, T., Ru, G., Sivam, K. et Ortiz-Pelaez, A. (2007). Explaining the heterogeneous scrapie surveillance figures across europe : a meta-regression approach. *BMC Vet. Res.* **3**, 13.
- Detwiler, L. A. et Baylis, M. (2003). The epidemiology of scrapie. *Rev. Sci. Tech.* **22**, 121–143.

- Diaz, C., Vitezica, Z. G., Rupp, R., Andreoletti, O. et Elsen, J. M. (2005). Polygenic variation and transmission factors involved in the resistance/susceptibility to scrapie in a romanov flock. *J. Gen. Virol.* **86**, 849–857.
- Dickinson, A. G. (1976). Scrapie in sheep and goats. *Front. Biol.* **44**, 209–241.
- Dickinson, A. G., Fraser, H. et Outram, G. W. (1975). Scrapie incubation time can exceed natural lifespan. *Nature* **256**, 732–733.
- Dickinson, A. G. et Outram, G. W. (1988). Genetic aspects of unconventional virus infections : the basis of the virino hypothesis. *Ciba Found Symp* **135**, 63–83.
- Dickinson, A. G., Stamp, J. T. et Renwick, C. C. (1974). Maternal and lateral transmission of scrapie in sheep. *J. Comp. Pathol.* **84**, 19–25.
- Dickinson, A. G. et Taylor, D. M. (1978). Resistance of scrapie agent to decontamination. *N. Engl. J. Med.* **299**, 1413–1414.
- Diringer, H. (1984). Sustained viremia in experimental hamster scrapie. brief report. *Arch. Virol.* **82**, 105–109.
- Diringer, H. (1991). Transmissible spongiform encephalopathies (tse) virus-induced amyloidoses of the central nervous system (cns). *Eur. J. Epidemiol.* **7**, 562–566.
- Donnelly, C. A. (2000). Likely size of the french bse epidemic. *Nature* **408**, 787–788.
- Donnelly, C. A., Ferguson, N. M., Ghani, A. C., Woolhouse, M. E., Watt, C. J. et Anderson, R. M. (1997). The epidemiology of bse in cattle herds in great britain. i. epidemiological processes, demography of cattle and approaches to control by culling. *Philos. Trans. R. Soc. Lond. B Biol Sci* **352**, 781–801.
- Efron, B. et Tibshirani, R. J. (1993). *An introduction to the bootstrap*. Chapman and Hall New York.
- EFSA (2007). European commission : Report on the monitoring and testing of ruminants for the presence of transmissible spongiform encephalopathy (tse) in the eu in 2006. Technical report.
- Elsen, J. M., Amigues, Y., Schelcher, F., Ducrocq, V., Andreoletti, O., Eychenne, F., Khang, J. V., Poivey, J. P., Lantier, F. et Laplanche, J. L. (1999). Genetic susceptibility and transmission factors in scrapie : detailed analysis of an epidemic in a closed flock of romanov. *Arch. Virol.* **144**, 431–445.
- Ersdal, C., Ulvund, M. J., Espenes, A., Benestad, S. L., Sarradin, P. et Landsverk, T.

- (2005). Mapping PrPsc propagation in experimental and natural scrapie in sheep with different PrP genotypes. *Vet. Pathol.* **42**, 258–274.
- Espenes, A., Press, C. M., Landsverk, T., Tranulis, M. A., Aleksandersen, M., Gunnes, G., Benestad, S. L., Fuglesteit, R. et Ulvund, M. J. (2006). Detection of PrP(sc) in rectal biopsy and necropsy samples from sheep with experimental scrapie. *J. Comp. Pathol.* **134**, 115–125.
- Farewell, V. T. (1982). The use of mixture models for the analysis of survival data with long-term survivors. *Biometrics* **38**, 1041–1046.
- Farewell, V. T. (1986). Mixture models in survival analysis : Are they worth the risk? *Can. J. Stat.* **14**, 257–262.
- Ferguson, N. M., Ghani, A. C., Donnelly, C. A., Hagensaaers, T. J. et Anderson, R. M. (2002). Estimating the human health risk from possible bse infection of the british sheep flock. *Nature* **415**, 420–424.
- Fernando, R. L., Stricker, C. et Elston, R. C. (1993). An efficient algorithm to compute the posterior genotypic distribution for every member of a pedigree without loops. *Theor. Appl. Gen.* **87**, 89–93.
- Foote, W. C., Clark, W., Maciulis, A., Call, J. W., Hourrigan, J., Evans, R. C., Marshall, M. R. et de Camp, M. (1993). Prevention of scrapie transmission in sheep, using embryo transfer. *Am. J. Vet. Res.* **54**, 1863–1868.
- Foster, J., McKenzie, C., Parnham, D., Drummond, D., Chong, A., Goldman, W. et Hunter, N. (2006). Lateral transmission of natural scrapie to scrapie-free new zealand sheep placed in an endemically infected uk flock. *Vet. Rec.* **159**, 633–634.
- Foster, J. D. et Dickinson, A. G. (1988). The unusual properties of ch1641, a sheep-passaged isolate of scrapie. *Vet. Rec.* **123**, 5–8.
- Foster, J. D. et Dickinson, A. G. (1989). Age at death from natural scrapie in a flock of suffolk sheep. *Vet. Rec.* **125**, 415–417.
- Foster, J. D., Hunter, N., Williams, A., Mylne, M. J., McKelvey, W. A., Hope, J., Fraser, H. et Bostock, C. (1996). Observations on the transmission of scrapie in experiments using embryo transfer. *Vet. Rec.* **138**, 559–562.
- Foster, J. D., McKelvey, W. A., Mylne, M. J., Williams, A., Hunter, N., Hope, J. et Fraser, H. (1992). Studies on maternal transmission of scrapie in sheep by embryo transfer.

- Vet. Rec.* **130**, 341–343.
- Fraser, H. (1976). The pathology of a natural and experimental scrapie. *Front. Biol.* **44**, 267–305.
- Fraser, H. et Dickinson, A. G. (1973). Scrapie in mice. agent-strain differences in the distribution and intensity of grey matter vacuolation. *J. Comp. Pathol.* **83**, 29–40.
- Gajdusek, D. C. (1977). Unconventional viruses and the origin and disappearance of kuru. *Science* **197**, 943–960.
- Gamel, J. W., McLean, I. W. et Rosenberg, S. H. (1990). Proportion cured and mean log survival time as functions of tumour size. *Stat. Med.* **9**, 999–1006.
- Gamel, J. W., Vogel, R. L., Valagussa, P. et Bonadonna, G. (1994). Parametric survival analysis of adjuvant therapy for stage ii breast cancer. *Cancer* **74**, 2483–2490.
- Georgsson, G., Sigurdarson, S. et Brown, P. (2006). Infectious agent of sheep scrapie may persist in the environment for at least 16 years. *J. Gen. Virol.* **87**, 3737–3740.
- Ghitany, M. E., Maller, R. et Zhou, S. (1995). Estimating the proportion of immunes in censored samples : a simulation study. *Stat. Med.* **14**, 39–49.
- Gieser, P. W., Chang, M. N., Rao, P. V., Shuster, J. J. et Pullen, J. (1998). Modelling cure rates using the gompertz model with covariate information. *Stat. Med.* **17**, 831–839.
- Goldman, A. I. (1984). Survivorship analysis when cure is a possibility : a monte carlo study. *Stat. Med.* **3**, 153–163.
- Goldmann, W., Hunter, N., Smith, G., Foster, J. et Hope, J. (1994). PrP genotype and agent effects in scrapie : change in allelic interaction with different isolates of agent in sheep, a natural host of scrapie. *J. Gen. Virol.* **75 (Pt 5)**, 989–995.
- Gonzalez, L., Dagleish, M. P., Bellworthy, S. J., Siso, S., Stack, M. J., Chaplin, M. J., Davis, L. A., Hawkins, S. A., Hughes, J. et Jeffrey, M. (2006). Postmortem diagnosis of preclinical and clinical scrapie in sheep by the detection of disease-associated PrP in their rectal mucosa. *Vet. Rec.* **158**, 325–331.
- Gonzalez, L., Jeffrey, M., Siso, S., Martin, S., Bellworthy, S. J., Stack, M. J., Chaplin, M. J., Davis, L., Dagleish, M. P. et Reid, H. W. (2005). Diagnosis of preclinical scrapie in samples of rectal mucosa. *Vet. Rec.* **156**, 846–847.
- Grassi, J. (2003). Pre-clinical diagnosis of transmissible spongiform encephalopathies using rapid tests. *Transfus. Clin. Biol.* **10**, 19–22.

- Grassi, J., Comoy, E., Simon, S., Creminon, C., Frobert, Y., Trapmann, S., Schimmel, H., Hawkins, S. A., Moynagh, J., Deslys, J. P. et Wells, G. A. (2001). Rapid test for the preclinical postmortem diagnosis of bse in central nervous system tissue. *Vet. Rec.* **149**, 577–582.
- Gravenor, M. B., Cox, D. R., Hoinville, L. J., Hoek, A. et McLean, A. R. (2000). Scrapie in britain during the bse years. *Nature* **406**, 584–585.
- Gray, R. J. et Tsiatis, A. A. (1989). A linear rank test for use when the main interest is in differences in cure rates. *Biometrics* **45**, 899–904.
- Greig, J. R. (1950). Scrapie in sheep. *J. Comp. Pathol.* **60**, 263–266.
- Griffith, J. S. (1967). Self-replication and scrapie. *Nature* **215**, 1043–1044.
- Groschup, M. H., Lacroux, C., Buschmann, A., Luhken, G., Mathey, J., Eiden, M., Lugan, S., Hoffmann, C., Espinosa, J. C., Baron, T., Torres, J. M., Erhardt, G. et Andreoletti, O. (2007). Classic scrapie in sheep with the ARR/ARR prion genotype in germany and france. *Emerg. Infect. Dis.* **13**, 1201–1207.
- Gruner, L. (2003). Final consolidated report of the european union contract of acronym "sheeprion" (FAIR CT 98-7023). Technical report.
- Gubbins, S., Clark, A. M., Eglin, R. D. et Sivam, S. K. (2006). Results of a postal survey of scrapie in the shetland islands in 2003. *Vet. Rec.* **158**, 255–260.
- Hadlow, W. J., Kennedy, R. C. et Race, R. E. (1982). Natural infection of suffolk sheep with scrapie virus. *J. Infect. Dis.* **146**, 657–664.
- Hadlow, W. J., Kennedy, R. C., Race, R. E. et Eklund, C. M. (1980). Virologic and neurohistologic findings in dairy goats affected with natural scrapie. *Vet. Pathol.* **17**, 187–199.
- Hagenaars, T. J., Donnelly, C. A. et Ferguson, N. M. (2006). Epidemiological analysis of data for scrapie in great britain. *Epidemiol. Infect.* **134**, 359–367.
- Hagenaars, T. J., Donnelly, C. A., Ferguson, N. M. et Anderson, R. M. (2000). The transmission dynamics of the aetiological agent of scrapie in a sheep flock. *Math. Biosci.* **168**, 117–135.
- Hagenaars, T. J., Donnelly, C. A., Ferguson, N. M. et Anderson, R. M. (2003). Dynamics of a scrapie outbreak in a flock of romanov sheep—estimation of transmission parameters. *Epidemiol. Infect.* **131**, 1015–1022.

- Hagenaars, T. J., Ferguson, N. M., Donnelly, C. A. et Anderson, R. M. (2001). Persistence patterns of scrapie in a sheep flock. *Epidemiol. Infect.* **127**, 157–167.
- Hanson, T., Bedrick, E. J., Johnson, W. O. et Thurmond, M. C. (2003). A mixture model for bovine abortion and foetal survival. *Stat. Med.* **22**, 1725–1739.
- Harris, D. A. (1999). Cellular biology of prion diseases. *Clin. Microbiol. Rev.* **12**, 429–444.
- Head, M. W., Kouverianou, E., Taylor, L., Green, A. et Knight, R. (2005). Evaluation of urinary PrP^{sc} as a diagnostic test for sporadic, variant, and familial cjd. *Neurology* **64**, 1794–1796.
- Healy, A. M., Hannon, D., Morgan, K. L., Weavers, E., Collins, J. D. et Doherty, M. L. (2004). A paired case-control study of risk factors for scrapie in irish sheep flocks. *Prev. Vet. Med.* **64**, 73–83.
- Heggebo, R., Gonzalez, L., Press, C. M., Gunnes, G., Espenes, A. et Jeffrey, M. (2003). Disease-associated PrP in the enteric nervous system of scrapie-affected suffolk sheep. *J. Gen. Virol.* **84**, 1327–1338.
- Heggebo, R., Press, C. M., Gunnes, G., Gonzalez, L. et Jeffrey, M. (2002). Distribution and accumulation of PrP in gut-associated and peripheral lymphoid tissue of scrapie-affected suffolk sheep. *J. Gen. Virol.* **83**, 479–489.
- Heggebo, R., Press, C. M., Gunnes, G., Lie, K. I., Tranulis, M. A., Ulvund, M., Groschup, M. H. et Landsverk, T. (2000). Distribution of prion protein in the ileal peyer's patch of scrapie-free lambs and lambs naturally and experimentally exposed to the scrapie agent. *J. Gen. Virol.* **81**, 2327–2337.
- Hill, A. F., Butterworth, R. J., Joiner, S., Jackson, G., Rossor, M. N., Thomas, D. J., Frosh, A., Tolley, N., Bell, J. E., Spencer, M., King, A., Al-Sarraj, S., Ironside, J. W., Lantos, P. L. et Collinge, J. (1999). Investigation of variant creutzfeldt-jakob disease and other human prion diseases with tonsil biopsy samples. *Lancet* **353**, 183–189.
- Hill, A. F., Desbruslais, M., Joiner, S., Sidle, K. C., Gowland, I., Collinge, J., Doey, L. J. et Lantos, P. (1997). The same prion strain causes vcjd and bse. *Nature* **389**, 448–450, 526.
- Hill, A. F., Zeidler, M., Ironside, J. et Collinge, J. (1997). Diagnosis of new variant creutzfeldt-jakob disease by tonsil biopsy. *Lancet* **349**, 99–100.
- Hilton, D. A., Fathers, E., Edwards, P., Ironside, J. W. et Zajicek, J. (1998). Prion

- immunoreactivity in appendix before clinical onset of variant creutzfeldt-jakob disease. *Lancet* **352**, 703–704.
- Hoinville, L. J. (1996). A review of the epidemiology of scrapie in sheep. *Rev. Sci. Tech.* **15**, 827–852.
- Hoinville, L. J., Hoek, A., Gravenor, M. B. et McLean, A. R. (2000). Descriptive epidemiology of scrapie in great britain : results of a postal survey. *Vet. Rec.* **146**, 455–461.
- Hopp, P., Ulvund, M. J. et Jarp, J. (2001). A case-control study on scrapie in norwegian sheep flocks. *Prev. Vet. Med.* **51**, 183–198.
- Hopp, P., Webb, C. R. et Jarp, J. (2003). Monte carlo simulation of surveillance strategies for scrapie in norwegian sheep. *Prev. Vet. Med.* **61**, 103–125.
- Hougaard, P. (1984). Life table methods for heterogeneous populations : Distributions describing the heterogeneity. *Biometrika* **71**, 75.
- Hougaard, P. (1986). Survival models for heterogeneous populations derived from stable distributions. *Biometrika* **73**, 387.
- Hourrigan, J., Klingsporn, A. L., Clark, W. W. et de Camp, M. (1979). Epidemiology of scrapie in the united states. In Prusiner, S. B. et Hadlow, W. J., editors, *Slow transmissible diseases of the nervous system, Vol.1*, pages 331–356. Academic Press, Nex York.
- Hourrigan, J. L. (1990). Experimentally induced bovine spongiform encephalopathy in cattle in mission, tex, and the control of scrapie. *J. Am. Vet. Med. Assoc.* **196**, 1678–1679.
- Houston, F., Foster, J. D., Chong, A., Hunter, N. et Bostock, C. J. (2000). Transmission of bse by blood transfusion in sheep. *Lancet* **356**, 999–1000.
- Houston, F., Goldmann, W., Chong, A., Jeffrey, M., Gonzalez, L., Foster, J., Parnham, D. et Hunter, N. (2003). Prion diseases : Bse in sheep bred for resistance to infection. *Nature* **423**, 498.
- Hunter, N. (1997). PrP genetics in sheep and the applications for scrapie and bse. *Trends Microbiol.* **5**, 331–334.
- Hunter, N., Foster, J., Chong, A., McCutcheon, S., Parnham, D., Eaton, S., MacKenzie, C. et Houston, F. (2002). Transmission of prion diseases by blood transfusion. *J. Gen. Virol.* **83**, 2897–2905.

- Hunter, N., Foster, J. D., Goldmann, W., Stear, M. J., Hope, J. et Bostock, C. (1996). Natural scrapie in a closed flock of cheviot sheep occurs only in specific PrP genotypes. *Arch. Virol.* **141**, 809–824.
- Hunter, N., Foster, J. D. et Hope, J. (1992). Natural scrapie in british sheep : breeds, ages and PrP gene polymorphisms. *Vet. Rec.* **130**, 389–392.
- Hunter, N., Goldmann, W., Smith, G. et Hope, J. (1994). The association of a codon 136 PrP gene variant with the occurrence of natural scrapie. *Arch. Virol.* **137**, 171–177.
- Hunter, N., Moore, L., Hosie, B. D., Dingwall, W. S. et Greig, A. (1997). Association between natural scrapie and PrP genotype in a flock of suffolk sheep in scotland. *Vet. Rec.* **140**, 59–63.
- Ibrahim, J. G., Chen, M. H. et Sinha, D. (2001). Bayesian semiparametric models for survival data with a cure fraction. *Biometrics* **57**, 383–388.
- Ikeda, T., Horiuchi, M., Ishiguro, N., Muramatsu, Y., Kai-Uwe, G. D. et Shinagawa, M. (1995). Amino acid polymorphisms of PrP with reference to onset of scrapie in suffolk and corriedale sheep in japan. *J. Gen. Virol.* **76 (Pt 10)**, 2577–2581.
- Jeffrey, M., Begara-McGorum, I., Clark, S., Martin, S., Clark, J., Chaplin, M. et Gonzalez, L. (2002). Occurrence and distribution of infection-specific PrP in tissues of clinical scrapie cases and cull sheep from scrapie-affected farms in shetland. *J. Comp. Pathol.* **127**, 264–273.
- Jeffrey, M., Martin, S., Thomson, J. R., Dingwall, W. S., Begara-McGorum, I. et Gonzalez, L. (2001). Onset and distribution of tissue PrP accumulation in scrapie-affected suffolk sheep as demonstrated by sequential necropsies and tonsillar biopsies. *J. Comp. Pathol.* **125**, 48–57.
- Jeffrey, M., Ryder, S., Martin, S., Hawkins, S. A., Terry, L., Berthelin-Baker, C. et Bellworthy, S. J. (2001). Oral inoculation of sheep with the agent of bovine spongiform encephalopathy (bse). 1. onset and distribution of disease-specific PrP accumulation in brain and viscera. *J. Comp. Pathol.* **124**, 280–289.
- Joly, P., Letenneur, L., Alioum, A. et Commenges, D. (1999). PHMPL : a computer program for hazard estimation using a penalized likelihood method with interval-censored and left-truncated data. *Comput. Methods Programs Biomed.* **60**, 225–231.
- Jones, D. R., Powles, R. L., Machin, D. et Sylvester, R. J. (1981). On estimating the

- proportion of cured patients in clinical studies. *Biometrie-Praximetrie* **21**, 1–11.
- Kalbfleisch, J. D. et Prentice, R. (1980). *The statistical analysis of failure time data*. Wiley New York.
- Kao, R. R., Houston, F., Baylis, M., Chihota, C. M., Goldmann, W., Gravenor, M. B., Hunter, N. et McLean, A. R. (2003). Epidemiological implications of the susceptibility to bse of putatively resistant sheep. *J. Gen. Virol.* **84**, 3503–3512.
- Kascsak, R. J., Rubenstein, R., Merz, P. A., Tonna-DeMasi, M., Fersko, R., Carp, R. I., Wisniewski, H. M. et Diringier, H. (1987). Mouse polyclonal and monoclonal antibody to scrapie-associated fibril proteins. *J. Virol.* **61**, 3688–3693.
- Kimberlin, R. H., Cole, S. et Walker, C. A. (1987). Temporary and permanent modifications to a single strain of mouse scrapie on transmission to rats and hamsters. *J. Gen. Virol.* **68** (Pt 7), 1875–1881.
- Kimberlin, R. H. et Walker, C. A. (1978a). Evidence that the transmission of one source of scrapie agent to hamsters involves separation of agent strains from a mixture. *J. Gen. Virol.* **39**, 487–496.
- Kimberlin, R. H. et Walker, C. A. (1978b). Pathogenesis of mouse scrapie : effect of route of inoculation on infectivity titres and dose-response curves. *J. Comp. Pathol.* **88**, 39–47.
- Kimberlin, R. H., Walker, C. A. et Fraser, H. (1989). The genomic identity of different strains of mouse scrapie is expressed in hamsters and preserved on reisolation in mice. *J. Gen. Virol.* **70** (Pt 8), 2017–2025.
- Kimberlin, R. H., Walker, C. A., Millson, G. C., Taylor, D. M., Robertson, P. A., Tomlinson, A. H. et Dickinson, A. G. (1983). Disinfection studies with two strains of mouse-passaged scrapie agent. guidelines for creutzfeldt-jakob and related agents. *J. Neurol. Sci.* **59**, 355–369.
- Klein, J. P. et Moeschberger, M. L. (1997). *Survival analysis : techniques for censored and truncated data*. New York : Springer.
- Konold, T., Davis, A., Bone, G. E., Simmons, M. M., Kahn, J., Blake-Dyke, M. C., Bracegirdle, J. et Shimwell, C. J. (2006). Atypical scrapie cases in the uk. *Vet. Rec.* **158**, 280.
- Kuk, A. Y. C. et Chen, C.-H. (1992). A mixture model combining logistic regression with

- proportional hazards regression. *Biometrika* **79**, 531–541.
- Lacroux, C., Corbiere, F., Tabouret, G., Lugan, S., Costes, P., Mathey, J., Delmas, J. M., Weisbecker, J. L., Foucras, G., Cassard, H., Elsen, J. M., Schelcher, F. et Andreoletti, O. (2007). Dynamics and genetics of PrPsc placental accumulation in sheep. *J. Gen. Virol.* **88**, 1056–1061.
- Lam, K. F., Fong, D. Y. et Tang, O. Y. (2005). Estimating the proportion of cured patients in a censored sample. *Stat. Med.* **24**, 1865–1879.
- Lambert, P., Collett, D., Kimber, A. et Johnson, R. (2004). Parametric accelerated failure time models with random effects and an application to kidney transplant survival. *Stat. Med.* **23**, 3177–3192.
- Langeveld, J. P., Jacobs, J. G., Erkens, J. H., Bossers, A., van Zijderveld, F. G. et van Keulen, L. J. (2006). Rapid and discriminatory diagnosis of scrapie and bse in retropharyngeal lymph nodes of sheep. *BMC Vet. Res.* **2**, 19.
- Laplanche, J. L., Chatelain, J., Westaway, D., Thomas, S., Dussaucy, M., Brugere-Picoux, J. et Launay, J. M. (1993). PrP polymorphisms associated with natural scrapie discovered by denaturing gradient gel electrophoresis. *Genomics* **15**, 30–37.
- Laska, E. M. et Meisner, M. J. (1992). Nonparametric estimation and testing in a cure model. *Biometrics* **48**, 1223–1234.
- Lee, J. W. (1995). Two-sample rank tests for acceleration in cure models. *Stat. Med.* **14**, 2111–2118.
- Lefrere, J. J. (2007). [the risk of prion contamination through the transfusion of cellular products]. *Transfus. Clin. Biol.* **14**, 25–34.
- Li, C. S. et Taylor, J. M. (2002). A semi-parametric accelerated failure time cure model. *Stat. Med.* **21**, 3235–3247.
- Li, C.-S., Taylor, J. M. G. et Sy, J. P. (2001). Identifiability of cure models. *Stat. Probabil. Lett.* **54**, 389–395.
- Ligos, C., Cancedda, M. G., Madau, L., Santucci, C., Maestrone, C., Agrimi, U., Ru, G. et Di Guardo, G. (2006). PrP(sc) deposition in nervous tissues without lymphoid tissue involvement is frequently found in ARQ/ARQ sarda breed sheep preclinically affected with natural scrapie. *Arch. Virol.* **151**, 2007–2020.
- Ligos, C., Sigurdson, C. J., Santucci, C., Carcassola, G., Manco, G., Basagni, M., Maes-

- trale, C., Cancedda, M. G., Madau, L. et Aguzzi, A. (2005). PrPsc in mammary glands of sheep affected by scrapie and mastitis. *Nat. Med.* **11**, 1137–1138.
- Longini, I. M. et Halloran, M. E. (1996). A frailty mixture model for estimating vaccine efficacy. *Appl. Stat.-J. Roy. St. C* **45**, 165–173.
- Louis, T. A. (1982). Finding the observed information matrix when using the em algorithm. *J. Roy. Stat. Soc. B* **44**, 226–233.
- Luhken, G., Buschmann, A., Brandt, H., Eiden, M., Groschup, M. H. et Erhardt, G. (2007). Epidemiological and genetical differences between classical and atypical scrapie cases. *Vet. Res.* **38**, 65–80.
- Maignien, T., Lasmezas, C. I., Beringue, V., Dormont, D. et Deslys, J. P. (1999). Pathogenesis of the oral route of infection of mice with scrapie and bovine spongiform encephalopathy agents. *J. Gen. Virol.* **80** (Pt 11), 3035–3042.
- Maller, R. et Zhou, X. (1996). *Survival analysis with long-term survivors*. Wiley, New York.
- Maller, R. A. et Zhou, S. (1995). Testing for the presence of immune or cured individuals in censored survival data. *Biometrics* **51**, 1197–1205.
- Manuelidis, L. (2007). A 25 nm virion is the likely cause of transmissible spongiform encephalopathies. *J. Cell. Biochem.* **100**, 897–915.
- Manuelidis, L., Yu, Z. X., Barquero, N. et Mullins, B. (2007). Cells infected with scrapie and creutzfeldt-jakob disease agents produce intracellular 25-nm virus-like particles. *Proc. Natl. Acad. Sci. U S A* **104**, 1965–1970.
- Marquardt, D. W. (1963). An algorithm for least-squares estimation of nonlinear parameters. *SIAM J. Appl. Math.* **11**, 431–441.
- Mathiason, C. K., Powers, J. G., Dahmes, S. J., Osborn, D. A., Miller, K. V., Warren, R. J., Mason, G. L., Hays, S. A., Hayes-Klug, J., Seelig, D. M., Wild, M. A., Wolfe, L. L., Spraker, T. R., Miller, M. W., Sigurdson, C. J., Telling, G. C. et Hoover, E. A. (2006). Infectious prions in the saliva and blood of deer with chronic wasting disease. *Science* **314**, 133–136.
- Matthews, L., Coen, P. G., Foster, J. D., Hunter, N. et Woolhouse, M. E. (2001). Population dynamics of a scrapie outbreak. *Arch. Virol.* **146**, 1173–1186.
- Matthews, L., Woolhouse, M. E. et Hunter, N. (1999). The basic reproduction number for

- scrapie. *Proc. Biol. Sci.* **266**, 1085–1090.
- McGilchrist, C. A. (1993). Reml estimation for survival models with frailty. *Biometrics* **49**, 221–225.
- McGilchrist, C. A. et Yau, K. K. W. (1995). The derivation of blup, ml, reml estimation methods for generalised linear mixed models. *Commun. Stat. Theory* **24**, 2963–2980.
- McIntyre, K. M., Gubbins, S., Sivam, S. K. et Baylis, M. (2006). Flock-level risk factors for scrapie in great britain : analysis of a 2002 anonymous postal survey. *BMC Vet. Res.* **2**, 25.
- McLean, A. R., Hoek, A., Hoinville, L. J. et Gravenor, M. B. (1999). Scrapie transmission in britain : a recipe for a mathematical model. *Proc. Biol. Sci.* **266**, 2531–2538.
- Meeker, W. et LuValle, M. (1995). An accelerated life test model based on reliability kinetics. *Technometrics* **37**, 133–146.
- Meng, X. L. I. et Rubin, D. B. (1992). Performing likelihood ratio tests with multiply-imputed data sets. *Biometrika* **79**, 103.
- Merz, P. A., Rohwer, R. G., Kascsak, R., Wisniewski, H. M., Somerville, R. A., Gibbs, C. J., J. et Gajdusek, D. C. (1984). Infection-specific particle from the unconventional slow virus diseases. *Science* **225**, 437–440.
- Merz, P. A., Somerville, R. A., Wisniewski, H. M. et Iqbal, K. (1981). Abnormal fibrils from scrapie-infected brain. *Acta Neuropathol (Berl)* **54**, 63–74.
- Merz, P. A., Somerville, R. A., Wisniewski, H. M., Manuelidis, L. et Manuelidis, E. E. (1983). Scrapie-associated fibrils in creutzfeldt-jakob disease. *Nature* **306**, 474–476.
- Meyer, R. K., McKinley, M. P., Bowman, K. A., Braunfeld, M. B., Barry, R. A. et Prusiner, S. B. (1986). Separation and properties of cellular and scrapie prion proteins. *Proc. Natl. Acad. Sci. U S A* **83**, 2310–2314.
- Miller, M. W., Hobbs, N. T. et Tavener, S. J. (2006). Dynamics of prion disease transmission in mule deer. *Ecol. Appl.* **16**, 2208–2214.
- Monleon, E., Monzon, M., Hortells, P., Bolea, R., Acin, C., Vargas, F. et Badiola, J. J. (2005). Approaches to scrapie diagnosis by applying immunohistochemistry and rapid tests on central nervous and lymphoreticular systems. *J. Virol. Methods* **125**, 165–171.
- Moreno, C. R., Moazami-Goudarzi, K., Laurent, P., Cazeau, G., Andreoletti, O., Chadi, S., Elsen, J. M. et Calavas, D. (2007). Which PrP haplotypes in a french sheep population

- are the most susceptible to atypical scrapie? *Arch. Virol.* **152**, 1229–1232.
- Morgan, K. L., Nicholas, K., Glover, M. J. et Hall, A. P. (1990). A questionnaire survey of the prevalence of scrapie in sheep in Britain. *Vet. Rec.* **127**, 373–376.
- Moum, T., Olsaker, I., Hopp, P., Moldal, T., Valheim, M., Moum, T. et Benestad, S. L. (2005). Polymorphisms at codons 141 and 154 in the ovine prion protein gene are associated with scrapie nor98 cases. *J. Gen. Virol.* **86**, 231–235.
- Muramoto, T., Scott, M., Cohen, F. E. et Prusiner, S. B. (1996). Recombinant scrapie-like prion protein of 106 amino acids is soluble. *Proc. Natl. Acad. Sci. U S A* **93**, 15457–15462.
- Nentwig, A., Oevermann, A., Heim, D., Botteron, C., Zellweger, K., Drogemüller, C., Zurbriggen, A. et Seuberlich, T. (2007). Diversity in neuroanatomical distribution of abnormal prion protein in atypical scrapie. *PLoS Pathog.* **3**, e82.
- O'Doherty, E., Healy, A., Aherne, M., Hanrahan, J. P., Weavers, E., Doherty, M., Roche, J. F., Gunn, M. et Sweeney, T. (2002). Prion protein (PrP) gene polymorphisms associated with natural scrapie cases and their flock-mates in Ireland. *Res. Vet. Sci.* **73**, 243–250.
- Oesch, B., Westaway, D., Walchli, M., McKinley, M. P., Kent, S. B., Aebersold, R., Barry, R. A., Tempst, P., Teplow, D. B., Hood, L. E. et al. (1985). A cellular gene encodes scrapie PrP 27-30 protein. *Cell* **40**, 735–746.
- Orge, L., Galo, A., Machado, C., Lima, C., Ochoa, C., Silva, J., Ramos, M. et Simas, J. P. (2004). Identification of putative atypical scrapie in sheep in Portugal. *J. Gen. Virol.* **85**, 3487–3491.
- O'Rourke, K. I., Baszler, T. V., Besser, T. E., Miller, J. M., Cutlip, R. C., Wells, G. A., Ryder, S. J., Parish, S. M., Hamir, A. N., Cockett, N. E., Jenny, A. et Knowles, D. P. (2000). Preclinical diagnosis of scrapie by immunohistochemistry of third eyelid lymphoid tissue. *J. Clin. Microbiol.* **38**, 3254–3259.
- Pattison, I. H., Hoare, M. N., Jebbett, J. N. et Watson, W. A. (1974). Further observations on the production of scrapie in sheep by oral dosing with foetal membranes from scrapie-affected sheep. *Br Vet J* **130**, lxx–lxxvii.
- Pattison, I. H. et Millson, G. C. (1961). Scrapie produced experimentally in goats with special reference to the clinical syndrome. *J. Comp. Pathol.* **71**, 101–109.

- Pattison, I. H. et Millson, G. C. (1962). Distribution of the scrapie agent in the tissues of experimentally inoculated goats. *J. Comp. Pathol.* **72**, 233–244.
- Peng, Y. (2003). Fitting semiparametric cure models. *Comput. Stat. Data An.* **41**, 481–490.
- Peng, Y. et Carriere, K. (2002). An empirical comparison of parametric and semiparametric cure models. *Biom. J.* **8**, 1002–1014.
- Peng, Y. et Dear, K. B. (2000). A nonparametric mixture model for cure rate estimation. *Biometrics* **56**, 237–243.
- Peng, Y., Dear, K. B. et Carriere, K. C. (2001). Testing for the presence of cured patients : a simulation study. *Stat. Med.* **20**, 1783–1796.
- Peng, Y., Dear, K. B. et Denham, J. W. (1998). A generalized f mixture model for cure rate estimation. *Stat. Med.* **17**, 813–830.
- Peng, Y., Taylor, J. M. et Yu, B. (2007). A marginal regression model for multivariate failure time data with a surviving fraction. *Lifetime Data Anal.* **13**, 351–369.
- Philippe, S., Ducrot, C., Roy, P., Remontet, L., Jarrige, N. et Calavas, D. (2005). Sheep feed and scrapie, france. *Emerg. Infect. Dis.* **11**, 1274–1279.
- Pinheiro, J. et Bates, D. (1995). Approximations to the log-likelihood function in the nonlinear mixed-effects model. *J. Comput. Graph. Stat.* **4**, 12–35.
- Porter, D. D., Porter, H. G. et Cox, N. A. (1973). Failure to demonstrate a humoral immune response to scrapie infection in mice. *J. Immunol.* **111**, 1407–1410.
- Post, K., Riesner, D., Walldorf, V. et Mehlhorn, H. (1999). Fly larvae and pupae as vectors for scrapie. *Lancet* **354**, 1969–1970.
- Price, D. L. et Manatunga, A. K. (2001). Modelling survival data with a cured fraction using frailty models. *Stat. Med.* **20**, 1515–1527.
- Prusiner, S. B. (1982). Novel proteinaceous infectious particles cause scrapie. *Science* **216**, 136–144.
- Prusiner, S. B. (1991). Molecular biology of prion diseases. *Science* **252**, 1515–1522.
- Prusiner, S. B., McKinley, M. P., Bowman, K. A., Bolton, D. C., Bendheim, P. E., Groth, D. F. et Glenner, G. G. (1983). Scrapie prions aggregate to form amyloid-like birefringent rods. *Cell* **35**, 349–358.
- Race, R. E., Raines, A., Baron, T. G., Miller, M. W., Jenny, A. et Williams, E. S. (2002).

- Comparison of abnormal prion protein glycoform patterns from transmissible spongiform encephalopathy agent-infected deer, elk, sheep, and cattle. *J. Virol.* **76**, 12365–12368.
- Reckzeh, C., Hoffmann, C., Buschmann, A., Buda, S., Budras, K. D., Reckling, K. F., Bellmann, S., Knobloch, H., Erhardt, G., Fries, R. et Groschup, M. H. (2007). Rapid testing leads to the underestimation of the scrapie prevalence in an affected sheep and goat flock. *Vet. Microbiol.* **123**, 320–327.
- Redman, C. A., Coen, P. G., Matthews, L., Lewis, R. M., Dingwall, W. S., Foster, J. D., Chase-Topping, M. E., Hunter, N. et Woolhouse, M. E. (2002). Comparative epidemiology of scrapie outbreaks in individual sheep flocks. *Epidemiol. Infect.* **128**, 513–521.
- Rondeau, V. et Commenges, D. (2002). Modélisation de la fragilité en survie. *Journal de la Société française de statistique* **143**, 103–119.
- Ryder, S., Dexter, G., Bellworthy, S. et Tongue, S. (2004). Demonstration of lateral transmission of scrapie between sheep kept under natural conditions using lymphoid tissue biopsy. *Res. Vet. Sci.* **76**, 211–217.
- Ryder, S. J., Spencer, Y. I., Bellerby, P. J. et March, S. A. (2001). Immunohistochemical detection of PrP in the medulla oblongata of sheep : the spectrum of staining in normal and scrapie-affected sheep. *Vet. Rec.* **148**, 7–13.
- Schreuder, B. E., de Jong, M. C., Pekelder, J. J., Vellema, P., Broker, A. J. et Betcke, H. (1993). Prevalence and incidence of scrapie in the netherlands : a questionnaire survey. *Vet. Rec.* **133**, 211–214.
- Schreuder, B. E., van Keulen, L. J., Vromans, M. E., Langeveld, J. P. et Smits, M. A. (1996). Preclinical test for prion diseases. *Nature* **381**, 563.
- Schreuder, B. E., van Keulen, L. J., Vromans, M. E., Langeveld, J. P. et Smits, M. A. (1998). Tonsillar biopsy and PrPsc detection in the preclinical diagnosis of scrapie. *Vet. Rec.* **142**, 564–568.
- Scott, A. C., Wells, G. A., Chaplin, M. J. et Dawson, M. (1992). Bovine spongiform encephalopathy : detection of fibrils in the central nervous system is not affected by autolysis. *Res. Vet. Sci.* **52**, 332–336.
- Seeger, H., Heikenwalder, M., Zeller, N., Kranich, J., Schwarz, P., Gaspert, A., Seifert, B., Miele, G. et Aguzzi, A. (2005). Coincident scrapie infection and nephritis lead to

- urinary prion excretion. *Science* **310**, 324–326.
- Seuberlich, T., Botteron, C., Benestad, S. L., Brunisholz, H., Wyss, R., Kihm, U., Schwermer, H., Friess, M., Nicolier, A., Heim, D. et Zurbriggen, A. (2007). Atypical scrapie in a swiss goat and implications for transmissible spongiform encephalopathy surveillance. *J. Vet. Diagn. Invest.* **19**, 2–8.
- Shaked, G. M., Shaked, Y., Kariv-Inbal, Z., Halimi, M., Avraham, I. et Gabizon, R. (2001). A protease-resistant prion protein isoform is present in urine of animals and humans affected with prion diseases. *J. Biol. Chem.* **276**, 31479–31482.
- Sigurdson, C. J., Williams, E. S., Miller, M. W., Spraker, T. R., O'Rourke, K. I. et Hoover, E. A. (1999). Oral transmission and early lymphoid tropism of chronic wasting disease PrPres in mule deer fawns (*odocoileus hemionus*). *J. Gen. Virol.* **80** (Pt 10), 2757–2764.
- Simmons, M. M., Konold, T., Simmons, H. A., Spencer, Y. I., Lockey, R., Spiropoulos, J., Everitt, S. et Clifford, D. (2007). Experimental transmission of atypical scrapie to sheep. *BMC Vet. Res.* **3**, 20.
- Sivam, S. K., Baylis, M., Gravenor, M. B. et Gubbins, S. (2006). Descriptive analysis of the results of an anonymous postal survey of the occurrence of scrapie in great britain in 2002. *Vet. Rec.* **158**, 501–506.
- Sivam, S. K., Baylis, M., Gravenor, M. B., Gubbins, S. et Wilesmith, J. W. (2003). Results of a postal survey in 2002 into the occurrence of scrapie in great britain. *Vet. Rec.* **153**, 782–783.
- Somerville, R. A., Birkett, C. R., Farquhar, C. F., Hunter, N., Goldmann, W., Dornan, J., Grover, D., Hennion, R. M., Percy, C., Foster, J. et Jeffrey, M. (1997). Immunodetection of PrPsc in spleens of some scrapie-infected sheep but not bse-infected cows. *J. Gen. Virol.* **78** (Pt 9), 2389–2396.
- Sposto, R. (2002). Cure model analysis in cancer : an application to data from the children's cancer group. *Stat. Med.* **21**, 293–312.
- Sposto, R., Sather, H. N. et Baker, S. A. (1992). A comparison of tests of the difference in the proportion of patients who are cured. *Biometrics* **48**, 87–99.
- Stack, M. J., Scott, A. C., Done, S. H. et Dawson, M. (1993). Scrapie associated fibril detection on decomposed and fixed ovine brain material. *Res. Vet. Sci.* **55**, 173–178.

- Stringer, S. M., Hunter, N. et Woolhouse, M. E. (1998). A mathematical model of the dynamics of scrapie in a sheep flock. *Math. Biosci.* **153**, 79–98.
- Supervie, V. et Costagliola, D. (2004). The unrecognised french bse epidemic. *Vet. Res.* **35**, 349–362.
- Supervie, V. et Costagliola, D. (2007). Estimating incidence of the french bse infection using a joint analysis of both asymptomatic and clinical bse surveillance data. *Math. Biosci.* **209**, 90–107.
- Sutton, J. M., Dickinson, J., Walker, J. T. et Raven, N. D. (2006). Methods to minimize the risks of creutzfeldt-jakob disease transmission by surgical procedures : where to set the standard? *Clin. Infect. Dis.* **43**, 757–764.
- Sy, J. P. et Taylor, J. M. G. (2000). Estimation in a cox proportional hazards cure model. *Biometrics* **56**, 227–236.
- Sy, J. P. et Taylor, J. M. G. (2001). Standard errors for the cox proportional hazards cure model. *Math. Comput. Model.* **33**, 1237–1251.
- Taylor, D. M. (2004). Resistance of transmissible spongiform encephalopathy agents to decontamination. *Contrib. Microbiol.* **11**, 136–145.
- Taylor, D. M., McConnell, I. et Fraser, H. (1996). Scrapie infection can be established readily through skin scarification in immunocompetent but not immunodeficient mice. *J. Gen. Virol.* **77** (Pt 7), 1595–1599.
- Taylor, J. M. (1995). Semi-parametric estimation in failure time mixture models. *Biometrics* **51**, 899–907.
- Taylor, J. M. et Kim, D. K. (1993). Statistical models for analysing time-to-occurrence data in radiobiology and radiation oncology. *Int. J. Radiat. Biol.* **64**, 627–640.
- Terry, L. A., Marsh, S., Ryder, S. J., Hawkins, S. A., Wells, G. A. et Spencer, Y. I. (2003). Detection of disease-specific PrP in the distal ileum of cattle exposed orally to the agent of bovine spongiform encephalopathy. *Vet. Rec.* **152**, 387–392.
- Thorgeirsdottir, S., Sigurdarson, S., Thorisson, H. M., Georgsson, G. et Palsdottir, A. (1999). PrP gene polymorphism and natural scrapie in icelandic sheep. *J. Gen. Virol.* **80** (Pt 9), 2527–2534.
- Tongue, S. C., Wilesmith, J. W. et Cook, C. J. (2004). Frequencies of prion protein (PrP) genotypes and distribution of ages in 15 scrapie-affected flocks in great britain. *Vet.*

- Rec.* **154**, 9–16.
- Tranulis, M. A. (2002). Influence of the prion protein gene, *prnp*, on scrapie susceptibility in sheep. *APMIS* **110**, 33–43.
- Tranulis, M. A., Osland, A., Bratberg, B. et Ulvund, M. J. (1999). Prion protein gene polymorphisms in sheep with natural scrapie and healthy controls in Norway. *J. Gen. Virol.* **80** (Pt 4), 1073–1077.
- Tsodikov, A. (1998). A proportional hazards model taking account of long-term survivors. *Biometrics* **54**, 1508–1516.
- Tsodikov, A. (2001). Estimation of survival based on proportional hazards when cure is a possibility. *Math. Comput. Model.* **33**, 1227–1236.
- Tsodikov, A. (2002). Semi-parametric models of long- and short-term survival : an application to the analysis of breast cancer survival in Utah by age and stage. *Stat. Med.* **21**, 895–920.
- Tsodikov, A. D., Ibrahim, J. G. et Yakovlev, A. Y. (2003). Estimating cure rates from survival data : An alternative to two-component mixture models. *J. Am. Stat. Assoc.* **98**, 1063–1079.
- Tuo, W., O'Rourke, K. I., Zhuang, D., Cheevers, W. P., Spraker, T. R. et Knowles, D. P. (2002). Pregnancy status and fetal prion genetics determine PrP^{Sc} accumulation in placentomes of scrapie-infected sheep. *Proc. Natl. Acad. Sci. U S A* **99**, 6310–6315.
- van Keulen, L. J., Schreuder, B. E., Meloen, R. H., Mooij-Harkes, G., Vromans, M. E. et Langeveld, J. P. (1996). Immunohistochemical detection of prion protein in lymphoid tissues of sheep with natural scrapie. *J. Clin. Microbiol.* **34**, 1228–1231.
- van Keulen, L. J., Schreuder, B. E., Vromans, M. E., Langeveld, J. P. et Smits, M. A. (1999). Scrapie-associated prion protein in the gastrointestinal tract of sheep with natural scrapie. *J. Comp. Pathol.* **121**, 55–63.
- van Keulen, L. J., Schreuder, B. E., Vromans, M. E., Langeveld, J. P. et Smits, M. A. (2000). Pathogenesis of natural scrapie in sheep. *Arch. Virol. Suppl.* pages 57–71.
- van Keulen, L. J., Vromans, M. E. et van Zijderveld, F. G. (2002). Early and late pathogenesis of natural scrapie infection in sheep. *APMIS* **110**, 23–32.
- Vascellari, M., Aufiero, G. M., Nonno, R., Agrimi, U., Vaccari, G., Basilicata, L., Falcaro, C., Mancin, M., Marcon, S. et Mutinelli, F. (2005). Diagnosis and PrP genotype target

- of scrapie in clinically healthy sheep of massese breed in the framework of a scrapie eradication programme. *Arch. Virol.* **150**, 1959–1976.
- Vascellari, M., Nonno, R., Mutinelli, F., Bigolaro, M., Di Bari, M. A., Melchiotti, E., Marcon, S., D'Agostino, C., Vaccari, G., Conte, M., De Grossi, L., Rosone, F., Giordani, F. et Agrimi, U. (2007). PrPsc in salivary glands of scrapie-affected sheep. *J. Virol.* **81**, 4872–4876.
- Vaupel, J. W., Manton, K. G. et Stallard, E. (1979). The impact of heterogeneity in individual frailty on the dynamics of mortality. *Demography* **16**, 439–454.
- Vitezica, Z. G. . (2003). *Analysis of complex pedigrees for estimating genotypes. Application to the study of genetic resistance to scrapie*. PhD thesis, Institut National Agronomique Paris-Grignon, France.
- Vitezica, Z. G., Elsen, J. M., Rupp, R. et Diaz, C. (2005). Using genotype probabilities in survival analysis : a scrapie case. *Genet. Sel. Evol.* **37**, 403–415.
- Wang, S., Foote, W. C., Sutton, D. L., Maciulis, A., Miller, J. M., Evans, R. C., Holyoak, G. R., Call, J. W., Bunch, T. D., Taylor, W. D. et Marshall, M. R. (2001). Preventing experimental vertical transmission of scrapie by embryo transfer. *Theriogenology* **56**, 315–327.
- Wells, G. A., Dawson, M., Hawkins, S. A., Green, R. B., Dexter, I., Francis, M. E., Simmons, M. M., Austin, A. R. et Horigan, M. W. (1994). Infectivity in the ileum of cattle challenged orally with bovine spongiform encephalopathy. *Vet. Rec.* **135**, 40–41.
- Wells, G. A., Hawkins, S. A., Green, R. B., Austin, A. R., Dexter, I., Spencer, Y. I., Chaplin, M. J., Stack, M. J. et Dawson, M. (1998). Preliminary observations on the pathogenesis of experimental bovine spongiform encephalopathy (bse) : an update. *Vet. Rec.* **142**, 103–106.
- Wells, G. A., Scott, A. C., Wilesmith, J. W., Simmons, M. M. et Matthews, D. (1994). Correlation between the results of a histopathological examination and the detection of abnormal brain fibrils in the diagnosis of bovine spongiform encephalopathy. *Res. Vet. Sci.* **56**, 346–351.
- Wilesmith, J. W., Ryan, J. B. et Atkinson, M. J. (1991). Bovine spongiform encephalopathy : epidemiological studies on the origin. *Vet. Rec.* **128**, 199–203.
- Williams, E. S. et Young, S. (1980). Chronic wasting disease of captive mule deer : a

- spongiform encephalopathy. *J. Wildl. Dis.* **16**, 89–98.
- Wineland, N. E., Detwiler, L. A. et Salman, M. D. (1998). Epidemiologic analysis of reported scrapie in sheep in the united states : 1,117 cases (1947-1992). *J. Am. Vet. Med. Assoc.* **212**, 713–718.
- Wisniewski, H. M., Sigurdarson, S., Rubenstein, R., Kascsak, R. J. et Carp, R. I. (1996). Mites as vectors for scrapie. *Lancet* **347**, 1114.
- Woolhouse, M. E., Stringer, S. M., Matthews, L., Hunter, N. et Anderson, R. M. (1998). Epidemiology and control of scrapie within a sheep flock. *Proc. Biol. Sci.* **265**, 1205–1210.
- Yakovlev, A., Goot, R. E. et Osipova, T. T. (1994). The choice of cancer treatment based on covariate information. *Stat. Med.* **13**, 1575–1581.
- Yakovlev, A., Tsodikov, A. D. et Bass, L. (1993). A stochastic model of hormesis. *Math. Biosci.* **116**, 197–219.
- Yamaguchi, K. (1992). Accelerated failure-time regression models with a regression model of surviving fraction : An application to the analysis of "permanent employment" in japan. *J. Am. Stat. Assoc.* **87**, 284–292.
- Yau, K. K. et Ng, A. S. (2001). Long-term survivor mixture model with random effects : application to a multi-centre clinical trial of carcinoma. *Stat. Med.* **20**, 1591–1607.
- Yau, K. K. W. et McGilchrist, C. A. (1998). Ml and reml estimation in survival analysis with time dependent correlated frailty. *Stat. Med.* **17**, 1201–1213.
- Yin, G. (2005). Bayesian cure rate frailty models with application to a root canal therapy study. *Biometrics* **61**, 552–558.
- Yin, G. et Ibrahim, J. G. (2005). A general class of bayesian survival models with zero and nonzero cure fractions. *Biometrics* **61**, 403–412.
- Yu, B., Tiwari, R. C., Cronin, K. A. et Feuer, E. J. (2004). Cure fraction estimation from the mixture cure models for grouped survival data. *Stat. Med.* **23**, 1733–1747.
- Yu, B., Tiwari, R. C., Cronin, K. A., McDonald, C. et Feuer, E. J. (2005). Cansurv : A windows program for population-based cancer survival analysis. *Comput. Methods Programs Biomed.* **80**, 195–203.
- Zhang, J. et Peng, Y. (2007). A new estimation method for the semiparametric accelerated failure time mixture cure model. *Stat. Med.* **26**, 3157–3171.

Chapitre 10

Annexes

10.1 Liste des publications et communications scientifiques

Publications

Corbière, F., Joly, P. (2007). A SAS macro for parametric and semiparametric mixture cure models. *Computer Methods and Programs in Biomedicine*, **85**, 173 – 180.

Corbière, F., Barillet, F., Andréoletti, O., Fidelle, F., Laphitz-Bordet, N., Schelcher, F. et Joly, P. (2007). Advanced survival models for risk factor analysis in scrapie, *Journal of General Virology*, **88**, 66 – 705.

Corbière, F., Taylor, J.M.G., Commenges, D. et Joly, P.. A penalized likelihood for mixture cure models. *Statistics in Medicine*. Article en révision.

Communications Orales

Corbière, F., Barillet, F., Joly, P. (2004). Estimation of parameters associated with scrapie infection and time to death from scrapie. *Réunion annuelle du programme Scrapie-Free-Sheep*, 23 – 24 février, Amsterdam (Pays-Bas).

Corbière, F., Fidelle, F., Andréoletti, F., Schelcher, F. et Barillet, F. (2005). Genetic control of scrapie disease using ARR/ARR rams in affected flocks of Manech red faced sheep, *Ruminants TSE network meetings*, 5 – 8 mai Biarritz (France).

Corbière, F., Barillet, F., Andréoletti, O., Fidelle, F. et Schelcher, F. (2006). The use of ARR/ARR rams in scrapie infected Manech faced Flocks in Pyrénées Atlantiques : evolution of genetic structures and scrapie incidence. *Réunion annuelle du programme Scrapie-Free-Sheep*, 27 – 28 janvier, Toulouse, (France).

Corbière, F., Barillet, F., Andréoletti, O., Fidelle, F., Schelcher, F. et Joly, P. (2006). Advanced survival models for risk factors analysis in Scrapie. Some results from Manech

Red Faced Flocks in Pyrénées Atlantiques. *Réunion annuelle du programme Scrapie-Free-Sheep*, 27 – 28 janvier, Toulouse (France).

Joly, P., Corbière, F. (2006). A penalized likelihood approach in a mixture cure model. *International Biometrics Congress*, 17 – 21 juillet, Montreal (Canada).

Corbière, F., Fidelle, F., Lacroux, C., Andréoletti, O., Schelcher, F. et Barillet, F. (2006). The use of ARR/ARR rams in scrapie infected Manech faced Flocks in Pyrénées Atlantiques : evolution of genetic structures and scrapie incidence. *Réunion finale du programme Scrapie-Free-Sheep*, 23 – 24 novembre, Sassari, (Italie).

Corbière, F., Joly, P. (2007). A penalized likelihood approach for cure rate models. *Journées de la Société Française de Statistique*, 11 – 15 juin, Angers (France).

10.2 Autres publications liées à la tremblante

Nous présentons dans cette section les trois articles cités dans le rapport de thèse.

Barillet, F., Palhière, I., Astruc, J.M., Brochard, M., Baelden, M., Aguerre, X., Fidelle, F., Arranz, J.M., Belloc, J.P., Briois, M., Frégeat, G., Soulas, C., Andréoletti, O., Corbière F. et Schelcher, F. (2004) Le programme français d'éradication de la tremblante du cheptel ovin fondé sur l'utilisation de la génétique. INRA Production Animales, numéro hors série, 87 – 100.

Andréoletti, O., Morel, N., Lacroux, C., Rouillon, V., Barc, C., Tabouret, G., Sarradin, P., Berthon, P., Bernardet, P., Mathey, J., Lugan, S., Costes, P., Corbière, F., Espinosa, J., C. Torres, J. M., Grassi, J., Schelcher, F. et Lantier, F. (2006). Bovine spongiform encephalopathy agent in spleen from an ARR/ARR orally exposed sheep. *Journal of General Virology*, **87**, 1043 – 1046.

Lacroux, C., Corbière, F., Tabouret, G., Lugan, S., Costes, P., Mathey, J., Delmas, J. M.,

Weisbecker, J. L., Foucras, G., Cassard, H., Elsen, J. M., Schelcher, F. et Andréoletti, O. (2007). Dynamics and genetics of PrPsc placental accumulation in sheep. *Journal of General Virology*, **88**, 1056 – 1061.

INRA Prod. Anim.,
2004, Numéro
hors série, 87-100

F. BARILLET¹, I. PALHIÈRE¹,
J.M. ASTRUC², M. BROCHARD³,
M. BAELDEN⁴, X. AGUERRE⁵,
F. FIDELLE⁶, J.M. ARRANZ⁷,
J.P. BELLOC⁸, M. BRIOIS⁶,
G. FREGEAT⁷, C. SOULAS⁸,
O. ANDREOLETTI⁸, F. CORBIÈRE⁸,
F. SCHELCHER⁸

¹ INRA, Station d'Amélioration
Génétique des Animaux, BP 27,
F-31326 Castanet-Tolosan Cedex

² Institut de l'Élevage, INRA-SAGA, BP
27, F-31326 Castanet-Tolosan Cedex

³ France UPra Sélection (FUS), 16 rue
Claude Bernard, F-75231 Paris Cedex 05

⁴ CDEO & UPra des races ovines lai-
tières des Pyrénées, Quartier Ahetzia,
F-64130 Ordiarp

⁵ Ovitest, Les Balquières, route
d'Espalion, F-12850 Onet-le-Château

⁶ Confédération Générale de Roquefort,
BP 348, F-12103 Millau Cedex

⁷ UPra Lacaune, carrefour de
l'Agriculture, F-12006 Rodez Cedex 9

⁸ UMR INRA-ENVT, 23 Chemin des
Capelles, F-31076 Toulouse

Courriel : barillet@germinal.toulouse.inra.fr

Le programme français d'éradication de la tremblante du cheptel ovin fondé sur l'utilisation de la génétique

Sont membres du Comité de suivi du
programme national de sélection des
ovins pour la résistance à la tremblante :
J.M. Astruc, F. Barillet, B. Bed'Hom,
B. Bibé, J. Bouix, M. Brochard,
O. Catrou, F. Dion, J.M. Elsen,
D. François, L. Griffon, E. Jullien,
M. Orlianges, I. Palhière, G. Perret,
P. Tribon.

Introduction

A partir de juin 1996, la tremblante ovine est devenue une maladie à déclaration obligatoire en France et un réseau de type passif, fondé sur la surveillance clinique des ovins présentant des troubles nerveux, a été mis en place au plan national. Parallèlement, les mesures de police sanitaire relatives à la tremblante ovine ont fortement évolué depuis 1997 : le dispositif initial de police sanitaire, mis en place en mars 1997, reposait sur la mise sous surveillance de l'exploitation lors de suspicion de tremblante chez un animal de plus de 12 mois et, en cas de confirmation de la suspicion, sur l'euthanasie et la destruction par le service public d'équarissage de tous les animaux présentant des symptômes en situation de tremblante sporadique, ou de tout le cheptel en situation de tremblante enzootique.

Depuis 2002, le Ministère de l'Agriculture a mis en place un programme global d'éradication de la tremblante, fondé sur les quatre dispositifs complémentaires suivants :

En matière de dépistage et surveillance

- le maintien du réseau de type passif fondé sur la surveillance clinique des ovins présentant des troubles cliniques évocateurs de tremblante.

Résumé

Les travaux conduits en ovins laitiers dès 1993 dans des élevages ovins atteints de tremblante ont permis de connaître la fréquence des allèles sensibles du gène PrP selon les races considérées, vérifier le risque accru de tremblante pour ces génotypes, et tester la faisabilité d'une sélection sur le gène PrP et son efficacité sur le risque de tremblante dans les troupeaux. Ils ont ainsi contribué à l'émergence de l'outil génétique pour éradiquer la tremblante dans les élevages atteints. Depuis 2002, le programme d'éradication de la tremblante mis en place en France par le Ministère de l'Agriculture, est fondé sur le génotypage du gène PrP : il vise à éradiquer à court et moyen terme la tremblante dans les élevages atteints, tout en renforçant la résistance génétique à moyen et long terme de l'ensemble du cheptel national, compte tenu des délais de renouvellement des cheptels femelles. Avoir choisi d'asseoir le programme national de sélection sur le dispositif existant d'amélioration génétique du cheptel national, organisé dans le cadre de la loi de l'Élevage de 1966, est manifestement un point clé pour expliquer la mise en œuvre rapide et efficace du programme dès la première année, confirmant la mobilisation massive de tous les maîtres d'œuvre des schémas de sélection des ovins en France : les évolutions conjointes constatées pour les fréquences alléliques en faveur de l'allèle ARR, les index de sélection pour les caractères de production et les indicateurs de gestion de la variabilité génétique sont une bonne illustration de l'application des objectifs assignés aux programmes de sélection. L'encadrement national, conduit par l'INRA, l'Institut de l'Élevage et France UPra Sélection (Unité de Promotion et de sélection de Race), s'attachera, dans les prochaines années, à vérifier et à aider à la mise en œuvre des quatre objectifs du programme national de sélection, en particulier la fourniture de reproducteurs résistants pour les élevages atteints, la gestion de la variabilité génétique dans les noyaux de sélection et la diffusion vers les élevages de production. Ce programme repose sur les points clés d'universalité de la résistance génétique à la tremblante en situation de contamination naturelle et d'absence de porteurs sains, qu'il faudra en permanence continuer de valider.

- la mise en place, à l'initiative de l'Union Européenne, d'un programme de surveillance active, fondé sur des tests rapides à l'abattoir et à l'équarissage. Il s'agit pour l'instant de sondages dont l'objectif premier est de fournir une estimation de la prévalence nationale de la tremblante des petits ruminants.

En matière de lutte et de contrôle

- un nouveau dispositif de police sanitaire, qui prévoit, en cas de confirmation de la suspicion de la tremblante, de génotyper au gène PrP tous les ovins du cheptel atteint. Suite à ce génotypage, est prévu un abattage sélectif : les ovins de génotypes sensibles et très sensibles sont marqués, euthanasiés et détruits ; les ovins résistants (non marqués) sont conservés et le repeuplement du cheptel est effectué en utilisant des béliers homozygotes résistants ARR/ARR et des brebis au moins hétérozygotes ARR. De plus, les ovins marqués sensibles peuvent, sous certaines conditions, être conservés pendant deux ans pour contribuer au repeuplement, à la condition expresse de mettre en place un programme génétique dans l'élevage, fondé sur l'usage de béliers homozygotes résistants de génotype ARR/ARR.

- un programme de sélection sur le gène PrP de toutes les races ovines exploitées en France.

Il s'agit donc d'un programme global faisant largement appel à la stratégie génétique, tant pour éradiquer la tremblante dans les élevages atteints, que pour renforcer la résistance de l'ensemble du cheptel national.

Le dispositif de recherche-développement, qui avait été mis en place dès 1993 en ovins laitiers, en particulier en Pyrénées-Atlantiques, département français le plus touché par la tremblante ovine, a vraisemblablement contribué à l'évolution nationale de la réglementation de 1996 à 2002, sans oublier bien sûr la réglementation européenne. Le présent article décrira dans une première partie le dispositif et les principaux résultats du projet conduit en ovins laitiers. Puis la deuxième partie sera consacrée à la présentation du programme national de sélection sur le gène PrP, mis en place à partir de 2002 pour toutes les races ovines françaises.

1 / Le dispositif de recherche-développement en ovins laitiers (depuis 1993)

Dans les Pyrénées-Atlantiques, des épisodes de tremblante ont évolué de façon enzootique depuis une trentaine d'années, alors que la situation épidémiologique du rayon de Roquefort correspondait plus à des formes sporadiques (quelques animaux touchés dans les troupeaux atteints). En outre, la prévalence de la tremblante dans les Pyrénées-Atlantiques apparaît comme particulièrement élevée, puisque, de juin 1996 à août 2002, 60 % des élevages français ayant été déclarés atteints de tremblante sont localisés dans ce département et que 90 % d'entre

eux exploitent des ovins de race Manech tête rousse (Calavas *et al* 2002). Compte-tenu de la situation, les responsables des unités de sélection des races Lacaune (lait) et pyrénéennes ont sollicité dès 1993 la station d'amélioration génétique des animaux de l'INRA de Toulouse et l'École Nationale Vétérinaire (ENV) de Toulouse pour étudier diverses modalités d'éradication de la tremblante. Un dispositif de recherche-développement a ainsi été instauré dès 1993, visant (i) à caractériser la structure génétique des races ovines laitières au gène PrP ; (ii) à conduire des études cas-témoin de génétique de la sensibilité à la tremblante ; (iii) à étudier le portage des ovins selon leur génotype PrP ; (iv) à réaliser des études épidémiologiques sur les facteurs de transmission de la tremblante intra-troupeau ; (v) à conduire enfin des actions d'utilisation de la génétique pour éradiquer la tremblante dans des élevages infectés.

1.1 / Le gène PrP et la sensibilité à la tremblante des ovins laitiers français

L'impact de l'insémination animale (IA) est important en France en ovins laitiers, avec quelque 40 % des brebis inséminées et donc plus de 50 % du renouvellement issu de béliers d'IA. Ainsi dès 1995, un typage systématique du gène PrP des béliers d'IA a été pratiqué, afin de caractériser les races concernées et de disposer d'une base de travail pour organiser ensuite la sélection pour la résistance à la tremblante pour ces dernières. Simultanément, une étude génétique cas-témoin était mise en place, afin de comparer le génotype PrP de brebis atteintes de tremblante avec celui d'un échantillon d'animaux sains.

a / Fréquences alléliques initiales au gène PrP des races ovines laitières françaises

Les génotypages ont été réalisés par le laboratoire LABOGENA. Les résultats du tableau 1, qui ne distinguent pas les allèles H et Q au codon 171, concernent les béliers d'IA nés avant 1999 et les mères à béliers (femelles élites des noyaux de sélection destinées à la procréation des mâles de la nouvelle génération) présentes en 1999 en race Manech tête rousse ou 2001 pour les races Basco-Béarnaise et Manech tête noire. Ces millésimes reflètent bien la situation initiale de chaque race, puisque la pression de sélection sur le gène PrP n'est devenue réellement effective qu'à partir de 2000 ou 2001 selon la race laitière considérée.

Pour chaque race, les fréquences alléliques observées chez les béliers et brebis sont comparables. Il existe à la fois des similitudes et des différences entre les cinq races : dans tous les cas, les fréquences alléliques de l'allèle AHQ (plutôt résistant) et de l'allèle VRQ (hyper sensible) sont très faibles. En revanche, la distribution des allèles ARR (résistant) et ARQ (sensible) diffère nettement entre les cinq races : la race Manech tête

Tableau 1. Fréquences alléliques initiales au gène PrP pour les races ovines laitières françaises (source UPRa (Unité de Promotion et de sélection de Race) et INRA) (pas de distinction entre H et Q au codon 171, noté Q dans le tableau).

Race	Echantillon (béliers nés avant 1999)	Effectif	Fréquence allélique			
			ARR	AHQ	ARQ	VRQ
Lacaune (lait)	Béliers d'IA	3394	0,603	0,016	0,368	0,013
Manech tête noire	Béliers d'IA	195	0,508	0,008	0,482	0,003
	Mères à béliers	793	0,458	0,006	0,532	0,004
Corse	Béliers d'IA	100	0,505	0,040	0,450	0,005
Basco-Béarnaise	Béliers d'IA	242	0,384	0	0,616	0
	Mères à béliers	1130	0,328	0,002	0,669	0,001
Manech tête rousse	Béliers d'IA	631	0,169	0,003	0,810	0,018
	Mères à béliers	2123	0,173	0,004	0,794	0,029

rousse est très sensible (80 % d'ARQ), la race Basco-Béarnaise assez sensible (62 % d'ARQ), les races Corse et Manech tête noire mieux protégées (45 et 50 % d'ARQ respectivement), et la race Lacaune la plus résistante (37 % d'allèle ARQ). Ces différences raciales apparaissent cohérentes avec la forte incidence de tremblante observée en race Manech tête rousse, dans le cadre du réseau national d'épidémiologie-surveillance.

b / Etude cas-témoin en races Lacaune (lait) et Manech tête rousse

Cinquante-huit brebis Lacaune et 368 brebis Manech tête rousse atteintes de tremblante, respectivement dans 10 et 32 élevages, ont fait l'objet d'une confirmation en laboratoire du diagnostic clinique par histopathologie et immunohistochimie. Les témoins correspondaient à 937 béliers des centres d'IA en race Lacaune (lait) et 3149 mères à béliers de 152 élevages en sélection indemnes de tremblante en race Manech tête rousse (tableau 2).

Comparativement aux ovins ARQ/ARQ pris comme référence (risque de 1), le risque de déclarer la tremblante est 3,90 à 3,78 fois plus élevé, en race Lacaune ou Manech tête rousse, pour les ovins de génotype ARQ/VRQ ou VRQ/VRQ. En revanche, il apparaît comme extrêmement faible pour les hétérozygotes ARR/ARQ (risque de 0,02) en race Manech et les homozygotes ARR/ARR ne sont jamais affectés. Malgré les différences marquées de fréquences alléliques au gène PrP, les risques pour les ovins ARQ/VRQ ou VRQ/VRQ sont donc très comparables dans les deux races (Barillet *et al* 2002). Ces résultats sont en accord avec la littérature, revue par Hunter (1997) et Smits *et al* (2000), et les résultats d'une étude européenne portant sur plus de 2000 cas de tremblante confirmée (Elsen *et al* 2002).

1.2 / Le portage des ovins en race Manech tête rousse selon leur génotype PrP

Le suivi de 15 élevages Manech tête rousse atteints de tremblante a été mis en place dans

le cadre d'un contrat de recherche européen d'acronyme « sheeprion ». Tous les ovins de ces troupeaux ont été génotypés depuis 1999 (7346 animaux) et les suspicions cliniques de tremblante confirmées par histopathologie. L'étude du portage de tremblante lors de la mort ou de l'abattage (animal sans signe clinique dans lequel on trouve de la protéine prion pathogène) est fondée sur la recherche de l'accumulation de la forme anormale de la protéine « scrapie » ou PrPsc dans les tissus lymphoïdes ou nerveux, dont les amygdales explorables *in vivo* (Schreuder *et al* 1996).

Les résultats de 654 brebis adultes, réformées en 2000 ou 2001 et âgées de 2 à 6 ans, sont présentés dans le tableau 3 : seules des brebis de génotype ARQ/ARQ, ARQ/VRQ ou VRQ/VRQ ont été détectées comme porteuses, plus particulièrement dans les élevages qui présentaient une incidence de la tremblante supérieure à 2 % du troupeau durant les deux dernières années.

Inversement, en accord avec les études de cinétiques d'infection conduites en fermes expérimentales (van Keulen *et al* 1999, Andréoletti *et al* 2000), aucun ovin ARR/ARR n'a montré d'accumulation de PrPsc dans les amygdales, suggérant l'absence de porteurs sains.

1.3 / Analyse épidémiologique

L'étude a porté sur 3531 brebis dans 8 des 15 élevages de race Manech tête rousse atteints de tremblante du projet européen « sheeprion », suivis pendant sept années. Les brebis appartenaient, soit à 3 millièmes (cohorte de naissance) nés avant l'apparition de la tremblante dans l'élevage, soit à 4 millièmes nés après, et toujours avant l'utilisation de béliers résistants ARR/ARR, tels que décrits dans le paragraphe suivant. L'incidence moyenne de la tremblante pour les cohortes considérées était de 13,8 % (663 cas) dans ces huit élevages, indiquant une situation épidémiologique enzootique. Les résultats d'une analyse de facteurs de risque intra-troupeau par régression logistique sont présentés dans le tableau 4 (Barillet *et al* 2003).

Tableau 2. Etude cas-témoin : fréquences génotypiques aux codons 136, 154 et 171 du gène PrP d'animaux atteints de tremblante et sains en races Lacaune et Manech tête rousse.

Genotype PrP	Race Lacaune (lait)			Manech tête rousse		
	malades (n = 58) 10 élevages	Sains (n = 937)	Odds Ratio (OR) ^b	malades (n = 368) 32 élevages	Sains (n = 3149) 152 élevages	Odds Ratio (OR) ^b
ARR/ARR	0	38,7 % (363)		0	3,1 % (96)	
ARR/AHQ	0	2,2 % (21)		0	0,2 % (5)	
ARR/ARQ	0	43,9 % (411)		0,8 % (3)	27,9 % (879)	0,02*
ARR/VRQ	0	1,2 % (11)		0	1,2 % (38)	
ARQ/AHQ	0	1,0 % (9)		0	0,4 % (14)	
ARQ/ARQ	86,2 % (50)	12,4 % (116)	1,00	78,3 % (288)	62,8 % (1977)	1,00
ARQ/VRQ	13,8 % (8)	0,6 % (6)	3,90*	20,1 % (74)	4,4 % (137)	3,78*
VRQ/VRQ	0	0		0,8 % (3)	0,1 % (3)	3,78*
	P ^a 0,0054			P ^a 0,0001		

^a P = signification globale (statistique de Wald)

^b OR significativement différents de 1,00 (P < 0,01) sont identifiés par un astérisque

Tableau 3. Nombre de brebis porteuses (dans les amygdales) de la PrP^{sc} selon leur génotype PrP et l'incidence annuelle de la tremblante dans l'élevage.

Génotype PrP	Incidence de la tremblante dans les élevages (nombre d'élevages)				
	< 1 % (4)	1-2 % (3)	2-3 % (4)	> 3 % (4)	Total
ARR/ARR	0/4	0/2	0/23	0/7	0/36
ARR/AHQ	0/1		0/2	0/3	0/6
ARR/ARQ	0/36	0/25	0/91	0/88	0/240
ARR/VRQ		0/2	0/7	0/9	0/18
ARQ/AHQ	0/1		0/4	0/3	0/8
ARQ/ARQ	1/69	0/53	0/95	2/82	3/299
ARQ/VRQ	0/7	0/3	2/15	2/21	4/46
VRQ/VRQ				1/1	1/1

Comme attendu, le principal facteur de risque est associé au génotype PrP des ovins : les brebis de génotype inconnu devaient être principalement des ovins de génotype ARQ/VRQ ou VRQ/VRQ, mortes avant la mise en place des génotypages systématiques dans ces élevages en 1999 et présentant une période d'incubation plus courte (comme vérifiée par ailleurs dans l'étude cas-témoin présentée au paragraphe 1.2 ci-dessus). La référence (risque de 1) correspond aux ovins ARQ/ARQ, des quatre millésimes nés après l'apparition de la tremblante dans l'élevage, avec une pression d'infection supérieure à 9 %, une fréquence de l'allèle ARQ dans l'élevage supérieure à 0,62, dans l'élevage N°1.

Comparativement à cette référence, le risque des brebis de génotype ARQ/VRQ ou VRQ/VRQ, incluant aussi les brebis de génotype inconnu, est 3,6 fois plus élevé. Inversement, le risque des brebis de génotype ARR/ARQ ou ARR/VRQ est très faible (0,03), en accord avec les résultats de l'étude cas-témoin précédente (tableau 2).

Le risque des cohortes de naissance nées avant l'apparition de la tremblante dans l'éle-

vage est égal à 0,34 comparativement aux brebis des cohortes nées après l'apparition de la tremblante dans l'élevage (prises comme référence), qui apparaissent donc comme trois fois plus à risque de contracter la tremblante. La pression d'infection, définie comme la proportion de mères infectées durant la première année de vie de leurs filles gardées pour le renouvellement intra-troupeau, est également un facteur de risque hautement significatif (tableau 4). Ces résultats suggèrent donc une contamination précoce des filles dans leurs premiers mois de vie.

Enfin, même en tenant compte de la fréquence de l'allèle ARQ dans l'élevage (structure génétique moyenne du troupeau au gène PrP), qui constitue aussi un facteur de risque significatif, l'effet de l'élevage lui-même reste hautement significatif, illustrant indirectement l'existence de conduites d'élevage plus ou moins risquées, probablement en relation avec la gestion de l'agnelage que cette analyse pointe indirectement en relation avec les facteurs « cohortes de naissance » et « pression d'infection des mères lors de la première année de vie de leurs filles ». Au plan de la physio-pathologie, la présente analyse met en

Tableau 4. Facteurs de risque pour l'incidence de la tremblante exprimés en termes d'odds ratio (OR) (régression logistique).

Facteur	Niveau	P ^a	Odd-ratio ^b	Intervalle de confiance à 95 %
Génotype PrP	ARR/ARQ et ARR/VRQ (694)	< 0,0001	0,03*	0,01-0,08
	ARQ/ARQ (667)		1,00	-
	ARQ/VRQ et VRQ/VRQ (103)		2,34*	1,42-3,82
	Inconnu (2067)		4,22*	3,19-5,56
Cohortes de naissance	3 ans avant l'apparition de la tremblante (663)	< 0,0001	0,34*	0,26-0,45
	4 ans après (2868)		1,00	-
Pression d'infection (% de mères infectées)	< 1 % (239)	< 0,0001	0,69*	0,55-0,88
	1 à 9 % (1691)		0,45*	0,34-0,59
	> 9 % (1691)		1,0	-
Fréquence de l'allèle ARQ dans l'élevage	< 0,62 (1507)	< 0,0036	0,69*	0,55-0,88
	> = 0,62 (2024)		1,00	-
Effet élevage	Elevage N° 7	< 0,0001	5,31*	3,24-8,69
	Elevage N° 1		1,00	-

^a Signification globale de la variable (statistique de Wald)

^b OR significativement différents de 1,00 (P < 0,01) sont identifiés par un astérisque

évidence des facteurs de risque significatifs vraisemblablement liés aux facteurs maintenant identifiés de contamination des placentas, à savoir le statut infectieux de la mère et le génotype du foetus au gène PrP (Andréoletti *et al* 2003).

1.4 / Utilisation de la génétique dans les élevages infectés

Au fur et à mesure que les résultats en Pyrénées-Atlantiques venaient conforter l'ensemble de la littérature sur la notion de résistance universelle des ovins ARR/ARR à toutes les souches de tremblante en situation de contamination naturelle, ainsi que sur l'absence de porteurs sains, l'idée d'utiliser le levier génétique pour éradiquer la tremblante a fait son chemin dans ce département très concerné par la maladie.

En alternative à l'abattage des animaux marqués, alors en vigueur dans le cadre de la police sanitaire de l'époque, un projet, fondé sur l'utilisation de béliers ARR/ARR, a été proposé en 1999 à tous les éleveurs volontaires. Ce projet dit « 64 » soumis par la Direction des Services Vétérinaires (DSV) à la Direction Générale de l'Alimentation (DGAL) pour autorisation de mise en œuvre (Barillet *et al* 1999), a débuté dans les 15 élevages du contrat « sheeprion » atteints de tremblante. Il concerne 27 élevages à partir de 2000 et se poursuit actuellement dans le cadre d'un nouveau contrat européen « scrapiefreesheep ».

Toutefois, vu la faible fréquence initiale de l'allèle de résistance ARR en race Manech tête rousse (tableau 1), le nombre de béliers résistants disponibles en 1999 au Centre d'Insémination Animale (CIA) était insuffisant. Pour ne pas hypothéquer le travail de sélection laitière réalisé depuis 30 ans, les quelques béliers ARR/ARR alors présents au CIA ont été multipliés sous forme d'un cheptel « sanitaire » de béliers résistants, en plus

du cheptel classique de béliers d'insémination animale dit cheptel « génétique ». Ce cheptel de béliers « sanitaires » a été de plus en plus utilisé dans les élevages atteints du département : 5800 IA en 2000, 9800 en 2001, plus de 20000 en 2002. Mais ces béliers « sanitaires » disparaîtront à terme, quand la résistance génétique du cheptel « génétique » de béliers d'IA sera suffisante.

Le suivi des élevages du programme « sheeprion » montre que 80 % et 90 % des brebis exposées à l'agent infectieux dès la naissance et qui meurent de tremblante, meurent respectivement avant l'âge de 36 et 48 mois. Il faut donc disposer d'un recul d'au moins 3 à 4 ans, pour dresser un premier bilan. C'est pourquoi ne sont présentés que les résultats préliminaires des quatre premiers élevages, ayant commencé à utiliser des béliers « sanitaires », à partir de 1997 (élevage A), 1998 (élevage B), ou 1999 (élevages C et D).

A titre d'illustration, on peut considérer l'élevage B dont la quasi-totalité des brebis (99 à 100 %) sont nées de béliers ARR/ARR à partir du millésime 1999 (tableau 5).

Pour une race sensible telle la Manech tête rousse, le recours à des béliers ARR/ARR permet donc de remplacer des mères majoritairement sensibles ARQ/ARQ ou ARQ/VRQ (40 à 60 %), et pour partie hétérozygotes ARR (20 à 40 %), par des filles principalement hétérozygotes ARR (40 à 70 %), puis homozygotes résistantes (20 à 50 %) en proportion croissante au fil des nouvelles générations.

Pour les jeunes générations issues quasi uniquement de béliers ARR/ARR (par exemple à partir du millésime 1999 dans cet élevage B), il y a, pour l'instant, absence totale de cas de tremblante déclaré avec un recul de 4 à 5 ans (juin 2004). Or dans ces mêmes troupeaux, l'incidence de la tremblante était comprise entre 11,8 et 34,9 % pour les dernières générations sensibles (Minery *et al* 2002), avant l'introduction de la résistance (tableau 6).

Tableau 5. Utilisation de béliers d'IA de génotype ARR/ARR dans un élevage (élevage B) atteint de tremblante : conséquences sur la structure génotypique des brebis au gène PrP.

Millésime de naissance des brebis (nombre)	1997 (98)	1998 (92)	1999 (92)	2000 (73)	2001 (124)	2002 (125)
Brebis nées de béliers ARR/ARR (%)	3	4	99	100	100	100
Génotype des brebis (%) ARR/ARR	2	7,6	22,8	26,1	39,7	54,2
ARR/AHQ	0	0	0	4,1	2,6	1,4
ARR/ARQ	35,7	30,4	69,6	64,4	57,7	43,8
ARR/VRQ	2,0	1,0	6,5	5,4	0	0,6
ARQ/AHQ	0	2,2	0	0	0	0
ARQ/ARQ	48,0	43,5	0	0	0	0
ARQ/VRQ	2,0	7,6	0	0	0	0
VRQ/VRQ	0	0	0	0	0	0
inconnu	10,3	7,7	1,1	0	0	0

Tableau 6. Utilisation de béliers d'IA de génotype ARR/ARR dans quatre élevages atteints de tremblante : conséquences sur la structure génotypique des brebis au gène PrP et sur l'incidence de la tremblante selon le millésime.

Elevage	Millésime de naissance	1997	1998	1999	2000	2001	2002
A	Brebis nées de béliers ARR/ARR (%)	16	51	98	97	100	96
	Incidence de la tremblante (%)	27,8	1,3	0	0	0	0
B	Brebis nées de béliers ARR/ARR (%)	3	4	99	100	100	100
	Incidence de la tremblante (%)	29,6	21,7	0	0	0	0
C	Brebis nées de béliers ARR/ARR (%)	0	2	2	99	96	94
	Incidence de la tremblante (%)	34,9	18,2	11,8	0	0	0
D	Brebis nées de béliers ARR/ARR (%)	1	0	1	58	100	99
	Incidence de la tremblante (%)	32,7	28,8	32,9	5,7	0	0

Tableau 7. Utilisation de béliers d'IA de génotype ARR/ARR dans 4 élevages atteints de tremblante : conséquences sur l'incidence annuelle de la tremblante

Première année d'utilisation de béliers ARR/ARR	Elevage A (1997)/1998	Elevage B 1998	Elevage C 1999	Elevage D (1999)/2000
Campagne	Incidence annuelle de la tremblante (%)			
1997	10,2	15,2	8,4	16,4
1998	6,7	9,1	8,9	10,7
1999	4,2	10,4	10,7	5,6
2000	5,4	3,3	5,3	8,6
2001	1,4	0,9	2,9	4,5
2002	0	0,6	0,3	5,0
2003	0	0	0	3,6

Si l'introduction de la résistance n'est que partielle (51 à 58 % de filles nées de pères ARR/ARR, respectivement pour la génération 1998 chez l'éleveur A et 2000 chez l'éleveur D), l'incidence de la tremblante, avec un recul de 4 à 5 ans, pour ces générations semi-résistantes est plutôt faible (1,3 à 5,7% selon la cohorte) et concerne uniquement des brebis sensibles ARQ/ARQ ou ARQ/VRQ (tableau 6).

Si on considère maintenant les élevages à une campagne donnée (description transversale, toutes générations présentes confondues), on note le statut enzootique des quatre élevages, avec des incidences annuelles comprises entre 8,4 et 16,4 % avant l'introduction de la résistance (tableau 7). Puis, l'incidence annuelle diminue fortement deux ans après l'introduction de béliers ARR/ARR dans le

troupeau (entre 2,9 et 3,6 % pour les élevages A, B, C et D), et passe en dessous de 1 % à partir de la troisième année, pour tomber à 0 % à partir de la quatrième année (élevages A, B et C). Pour l'élevage D ayant utilisé le plus tardivement des béliers résistants, il faudra attendre 2005, pour confirmation des tendances constatées dans les trois autres élevages. Trois ans après l'introduction de la résistance, la mortalité résiduelle était imputable uniquement à des brebis sensibles ARQ/ARQ et ARQ/VRQ de vieilles générations nées avant l'introduction des béliers homozygotes résistants et qui n'avaient pas été éliminées dans ces élevages à statut dérogatoire depuis 1999, au titre du projet d'étude 64.

On peut donc penser que, dans ces élevages, l'application des nouvelles règles de police sanitaire (mises en œuvre depuis 2002), imposant l'utilisation de béliers ARR/ARR, l'élimination immédiate des ovins du troupeau porteurs de l'allèle VRQ et l'élimination des vieilles brebis ARQ/ARQ au plus tard deux ans après la mise sous surveillance officielle aurait permis de faire totalement disparaître la tremblante 2 ou 3 ans après l'introduction des béliers homozygotes résistants. La fin du suivi expérimental, programmé jusqu'en 2006 pour ces élevages, dans le cadre du projet européen « scrapie-freesheep », constitue donc un enjeu essentiel pour confirmer ces résultats partiels.

1.5 / La sélection des races ovines laitières des Pyrénées

Le typage PrP des béliers d'IA a débuté en 1995. L'année 1999 a marqué un tournant avec trois décisions déterminantes : la création du cheptel de béliers d'IA « sanitaire » pour protéger les élevages atteints et ne pas hypothéquer 30 ans de sélection laitière ; le début du typage PrP des mères à béliers pour accroître la sélection en faveur de l'allèle ARR en préservant la variabilité génétique, c'est-à-dire le nombre de familles sélectionnées ; le doublement de la taille du centre d'élevage de jeunes béliers pour augmenter la pression de sélection sur le gène PrP. La mise en place en 2002, par le Ministère de l'Agriculture, d'un plan national de sélection du cheptel ovin français sur le gène PrP (Palhière *et al* 2002) conforte le travail réalisé auparavant en Pyrénées-Atlantiques. Le nombre de typages PrP pour le programme de sélection a concerné en 2002 quelque 6200 brebis élites et béliers des noyaux de sélection des races ovines laitières pyrénéennes. Ainsi, des résultats très significatifs sont obtenus pour le cheptel génétique classique : la fréquence de l'allèle ARR des béliers d'IA Manech tête rousse s'est accrue de 19 % pour le millésime 1999 à 66 % pour le millésime 2003. Finalement l'objectif de l'UPRa est que tous les béliers du noyau de sélection ou du centre d'insémination artificielle soient porteurs de l'allèle ARR à partir de 2004.

L'ensemble des travaux conduits en ovins laitiers dès 1993 a contribué à l'émergence de l'outil génétique pour éradiquer la tremblante

dans les élevages atteints. Ils ont montré les différences de fréquence des allèles sensibles selon les races, vérifié le risque accru de tremblante pour ces génotypes, et testé la faisabilité d'une sélection sur le gène PrP et son efficacité sur le risque de tremblante dans les troupeaux. Les résultats actuellement disponibles confirment l'intérêt des mesures prises en 2002 en France en matière de lutte contre la tremblante ovine, dont le programme national de sélection pour la résistance à la tremblante, qui représente un investissement à moyen et long terme pour protéger tout le cheptel national.

2 / Le programme national de sélection des ovins pour la résistance à la tremblante (2002-2006)

Le programme d'amélioration génétique pour la résistance à la tremblante des populations ovines constitue un volet essentiel, partie prenante du nouveau dispositif de police sanitaire, contribuant au programme global de lutte contre la tremblante en France. Le principe général est que l'utilisation de béliers ARR/ARR donnerait rapidement une protection importante des élevages vis-à-vis de la tremblante, à l'instar des résultats présentés dans la première partie. Il protégerait aussi du portage tous les agneaux jusqu'à l'âge de l'abattage et, par conséquent, protégerait les consommateurs d'un risque potentiel de transmission d'une Encéphalopathie Spongiforme Transmissible (EST) qu'ils pourraient subir en consommant de la viande ovine. Ce principe a également été retenu par le Royaume-Uni (Dawson 2003), les Pays-Bas (Vellema 2003), ainsi que l'Union Européenne, qui envisagent de généraliser l'utilisation de béliers ARR/ARR dans les élevages ovins.

Pour l'instant, le programme français d'amélioration génétique pour la résistance à la tremblante est planifié pour cinq ans (2002-2006). Son financement est assuré par le Ministère de l'Agriculture, de l'Alimentation, de la Pêche et des Affaires Rurales (MAAPAR). La coordination générale est réalisée par un comité de pilotage, groupe technique du comité ovin et caprin de la Commission Nationale d'Amélioration Génétique (CNAG). Présidé par le MAAPAR, le comité de pilotage regroupe les différents représentants de la filière ovine (FUS, FNCBV, FNO, FNGDS, ANIO, CNBL...), le laboratoire de référence (LABOGENA) et les organismes scientifiques et techniques (INRA, Institut de l'Élevage) responsables de la conception et de l'encaissement national du dit programme (Palhière *et al* 2002).

2.1 / Cadre des applications

Le choix déterminant a consisté à asseoir ce programme d'amélioration génétique pour la résistance à la tremblante sur le dispositif national d'amélioration génétique des ani-

maux, organisé dans le cadre de la loi de l'Élevage de 1966. Ce dispositif fait appel à un ensemble de structures dont les UPRA ou les Unités de sélection, qui jouent un rôle moteur dans l'évolution génétique des races ovines françaises. Même si l'objectif final du programme est d'améliorer la résistance aux EST de l'ensemble de la population ovine française, la focalisation du programme chez les éleveurs sélectionneurs constituant les noyaux de sélection se justifie pour les raisons suivantes : bénéficier de la technicité de ces éleveurs (maîtrise des filiations, qualité de l'identification, enregistrement des performances, sélection de reproducteurs de qualité), capitaliser les efforts de sélection réalisés (le génotypage PrP pouvant être restreint au fil du temps), ne pas casser les circuits classiques de diffusion du progrès génétique, des noyaux de sélection vers les élevages de production.

Le tableau 8 montre l'importance des schémas de sélection ovins : 5 races laitières et 21 races allaitantes regroupent près de 600 000 brebis en bases de sélection, soient 9 % de l'ensemble des brebis de race pure. Des agneaux (7400), issus des meilleurs reproducteurs, entrent en centres d'élevage ou stations de contrôle individuel chaque année, et sont destinés à renouveler les béliers des centres d'insémination animale, les béliers de monte naturelle des élevages sélectionneurs, mais aussi des élevages de production. Ces reproducteurs mâles constituent donc le cœur du schéma de sélection. La diffusion vers les élevages de production est assurée par 680 000 IA et près de 10 000 béliers dont une partie est issue de centres d'élevage ou de stations de contrôle individuel, l'autre provenant directement des élevages sélectionneurs.

Ces chiffres globaux cachent une grande diversité des schémas de sélection ovins. Les principales différences concernent les objectifs de sélection (caractères laitiers pour les races laitières, caractères de reproduction et d'aptitudes maternelles pour les races allaitantes rustiques, caractères de reproduction et d'aptitudes bouchères pour les races à viande), l'organisation du schéma de sélection (par exemple, le taux d'IA varie de 90 % en race Lacaune à 0 % dans certaines races allaitantes), la taille relative du noyau de

sélection par rapport à la population à améliorer (de 5 % à 25 %) qui conditionne la capacité de diffusion hors base de sélection.

Le programme d'amélioration génétique pour la résistance à la tremblante a donc bénéficié d'outils déjà en place, primordiaux pour son efficacité : la capacité de génotypage à grande échelle était assurée, au démarrage du programme, par LABOGENA, puis aujourd'hui par neuf laboratoires supplémentaires agréés ; la qualité et la circulation des informations sont liées à l'existence d'un système fiable d'identification des animaux, d'une base de données génétiques nationale au Centre de Traitement de l'Information Génétique (CTIG) de l'INRA de Jouy-en-Josas, permettant d'organiser le stockage des données de typages PrP, et d'un dispositif de circulation rapide des informations génétiques entre les différents acteurs des programmes de sélection.

2.2 / Objectifs du programme

Le programme d'amélioration génétique pour la résistance à la tremblante vise quatre objectifs généraux, qui doivent être mis en œuvre de façon concomitante dans les UPRA ou unités de sélection :

- éliminer l'allèle VRQ, qui représente un facteur de risque important pour les animaux porteurs ;
- diffuser des béliers ARR/ARR dans les élevages atteints, en priorité, conformément aux exigences de la nouvelle police sanitaire ;
- sélectionner l'allèle ARR tout en maintenant la sélection pour les caractères de production et la gestion de la variabilité génétique des noyaux de sélection ;
- diffuser des béliers ARR/ARR dans les élevages de production.

Compte-tenu de la grande diversité des schémas de sélection, la mise en place de ces quatre objectifs a impliqué la définition de programmes raciaux spécifiques et adaptés à la situation de chaque race. Les principaux facteurs d'ajustement pris en compte sont les suivants :

- les fréquences alléliques initiales au gène PrP : la figure 1 montre les résultats d'une enquête européenne réalisée en 1999-2000 sur des échantillons représentatifs d'une

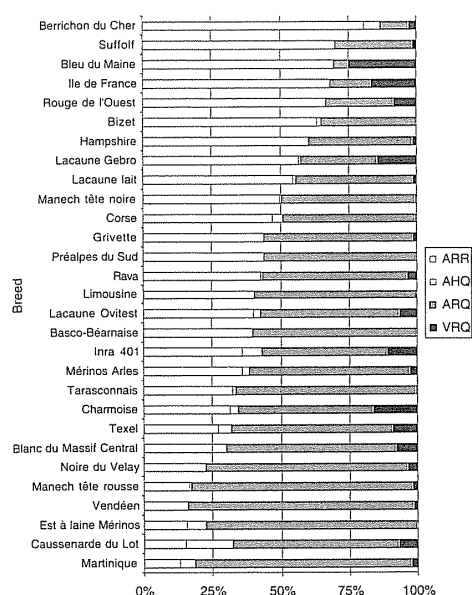
Tableau 8. Les schémas de sélection des ovins en France.

	Ovins laitiers	Ovins allaitants
Femelles de la population (élevages)	1 395 000 (5 250) ^a	8 000 000 ^b (87 900)
Femelles en base de sélection (élevages)	299 000 (821)	290 000 (1 217)
Races avec schéma de sélection	5	21
Béliers entrés en centres d'élevage ou stations	2 500	4 900
Béliers diffusés hors base de sélection	1 500	8 000
IA réalisées dans la base de sélection	200 000	68 000
IA diffusées hors base de sélection	360 000	320 000

^a en race pure

^b dont 5 200 000 en race pure

Figure 1. Fréquences alléliques initiales au gène PrP des races ovines françaises (ARQ = ARH).



trentaine de races ovines françaises. Les fréquences des allèles ARR, AHQ, ARQ, et VRQ sont très variables selon les races puisqu'elles sont comprises respectivement entre 0 et 80 %, 5 et 82 %, 0 et 25 %, 0 et 17 %. Les races ovines françaises présentent donc une situation de départ très variée avant toute sélection en faveur de la résistance à la tremblante, à l'instar du constat décrit auparavant en races ovines laitières.

- l'incidence de la tremblante dans chaque race : le fait que les cas de tremblante soient concentrés dans quelques bassins de production implique de mettre en place des actions particulières pour proposer des animaux résistants aux élevages atteints dans les races concernées (cf. section 1).
- le poids des schémas de sélection par rapport à l'ensemble de la population : celui-ci varie selon (i) le mode de diffusion des gènes : l'IA représente un moyen efficace et rapide de diffusion hors noyau de sélection alors que l'approvisionnement en mâles de monte naturelle est plus progressif ; (ii) l'importance de la population cible, permettant de distinguer les races « femelles » pour lesquelles la diffusion concerne les élevages de production en race pure, et les races qui sont utilisées en croisement terminal et devant fournir à la fois les élevages en race pure, mais aussi les élevages en croisement.
- les caractéristiques des schémas de sélection : utilisation de l'IA, type d'évaluation génétique,...

2.3 / Les animaux à génotyper

La principale action du programme d'amélioration génétique pour la résistance à la tremblante consiste à génotyper des animaux ciblés au sein des élevages sélectionneurs des noyaux de sélection. En outre, l'élimination et

le remplacement des animaux sensibles sont également financés, ainsi que l'augmentation de la taille des centres d'élevage et des stations de contrôle individuel, outils clés pour le renouvellement des béliers d'IA et de monte naturelle de toute la population.

La stratégie de génotypage, c'est-à-dire les catégories décrites ci-dessous d'animaux à génotyper dans les noyaux de sélection, a été adaptée aux grands ensembles (races ovines laitières, races bouchères allaitantes, races rustiques allaitantes), et race par race, en privilégiant le rapport qualité-prix, à nombre total de typages PrP fixé. Dans tous les cas, l'option a été prise de concentrer les actions sur la voie mâle qui est au cœur de la création et de la diffusion du progrès génétique.

a / Mâles adultes

Les mâles d'IA ont été génotypés très tôt, dès les années 95 en ovins laitiers (cf. section 1), en 2001 ou 2002 en ovins allaitants, afin de connaître la structure génétique des meilleurs reproducteurs de la race et ainsi programmer leur renouvellement en alliant amélioration de la résistance, maintien de la sélection pour les caractères de production et de la variabilité génétique. Le génotypage des mâles adultes de monte naturelle est utile pour identifier et éliminer les béliers les plus sensibles (porteurs de VRQ). Il permet également d'améliorer plus rapidement la structure PrP des femelles si le renouvellement est issu de pères plus résistants. Il a donc été entrepris systématiquement en 2002 lors de la première année de mise en œuvre du programme national (tableau 7).

b / Jeunes béliers candidats pour les centres d'élevages et stations de contrôle individuel

Les centres d'élevage de jeunes mâles ou stations de contrôle individuel regroupent des jeunes béliers au sevrage, à l'âge de un mois (ovins laitiers) à deux-trois mois (ovins allaitants), nés d'accouplements raisonnés entre mères à béliers et pères élités. Ils jouent un rôle clé pour la maîtrise génétique et sanitaire de chaque race. En effet, à la sortie de ces centres ou stations, les jeunes béliers sélectionnés approvisionnent les centres d'insémination animale pour le testage sur descendance ou la diffusion directe par IA (après sélection en station), ou contribuent au renouvellement des mâles de monte naturelle des noyaux de sélection et des élevages de production. Bref, ces centres ou stations sont au cœur à la fois de la création et de la diffusion du progrès génétique.

Par rapport à la situation *ante* sans sélection sur le gène PrP, si on veut à la fois sélectionner des jeunes mâles au moins hétérozygotes ARR tout en maintenant la sélection habituelle sur les caractères de production, il faut augmenter le nombre de candidats à l'entrée en centres d'élevages ou stations d'un facteur 2 à 4 dès que la fréquence initiale de l'allèle ARR est faible et comprise entre 0,2 et 0,3 (figure 2). Ainsi un nombre important de jeunes béliers est génotypé afin

de sélectionner des futurs reproducteurs de plus en plus résistants, à bon potentiel génétique pour les caractères de production, tout en préservant la variabilité génétique de la race. La plupart des races a fait le choix de génotyper précocément dans les élevages (avant l'entrée en centres d'élevage ou stations tout en augmentant le nombre de candidats génotypés), afin d'être plus efficace pour la sélection de ces jeunes béliers : en 2003, toutes races confondues, le facteur de multiplication pratiqué était de 4,3 illustrant l'effort important de sélection sur PrP tout en préservant la sélection sur les caractères de production.

c / Jeunes béliers pour les élevages atteints et pour les élevages de production

Si les jeunes béliers entrés en centres d'élevage ou stations de contrôle individuel ne suffisent pas à approvisionner les élevages atteints et les élevages de production, des jeunes béliers supplémentaires peuvent être génotypés dans les élevages de sélection avant d'être diffusés directement depuis ces élevages.

d / Mères à béliers

Le génotypage des mères à béliers permet d'organiser des accouplements raisonnés avec les mâles élites selon la résistance à la tremblante, afin d'accélérer la fixation de l'allèle ARR pour les futurs béliers élites. C'est un génotypage coûteux (15 à 20 % des femelles d'un noyau de sélection sont mères à béliers), car nécessitant beaucoup de travail (elles sont réparties dans tous les élevages). Il est donc déconseillé dans les situations suivantes : si la fréquence de l'allèle ARR est élevée dans la population femelle, si l'IA et donc les accouplements raisonnés sont peu utilisés, si le génotype d'un grand nombre de femelles peut être prédit (d'après les informations connues de leurs parents et/ou de leurs descendants). En revanche, si la pratique de l'IA est importante et la fréquence initiale de l'allèle ARR défavorable, le génotypage au gène PrP des mères à béliers s'avère particulièrement intéressant, pour organiser des

accouplements raisonnés permettant de remplacer les pères élites homozygotes sensibles par des fils hétérozygotes ARR, étape indispensable pour préserver la variabilité génétique de la race. On comprend dès lors pourquoi le typage des mères à béliers a été privilégié pour les races ovines laitières (tableau 1) hormis la race Lacaune (lait) initialement assez résistante pour éviter d'y recourir : un raisonnement rapport qualité-prix a ainsi été conduit pour chaque race.

e / Agnelles de renouvellement

Le génotypage des agnelles de renouvellement est un moyen pour éliminer l'allèle VRQ des bases de sélection. Il permet également de sélectionner les femelles résistantes. A terme, c'est un moyen pour connaître le génotype de l'ensemble des femelles des bases de sélection et donc de réduire le nombre de génotypages de jeunes béliers à réaliser. Comme pour les mères à béliers, le génotypage des agnelles est coûteux à mettre en œuvre, d'autant qu'il suppose une réalisation pluriannuelle pour son efficacité. En revanche, il est efficace dans toutes les configurations de schémas de sélection, puisqu'il fait appel uniquement à la sélection des femelles intra-troupeau pour le renouvellement. Ce typage des agnelles a donc été réservé aux races allaitantes, qui peuvent présenter une fréquence initiale de l'allèle VRQ élevée, outre le fait que leurs noyaux de sélection sont souvent de plus petite taille que ceux des races ovines laitières.

2.4 / Premiers résultats

a / Nombre de génotypages

Depuis la mise en place du programme national, environ 80 à 90 000 animaux ont été génotypés chaque année (tableau 9), ce qui représente autour de 15 % des animaux présents dans les noyaux de sélection et 1 % de l'ensemble des ovins en France. En comparaison, le National Scrapie Plan au Royaume-Uni génotype, chaque année, environ 2 % des brebis présentes (Dawson 2003) et le plan hollandais, 7 % (Vellema 2003). Les moyens en typages du gène PrP sont donc plus limités en France. Le travail d'optimisation race par race est donc indispensable : ainsi, le comité de suivi alloue un quota annuel de typages à chaque race, en fonction de ses besoins et de la capacité nationale de typage. Les besoins sont définis selon les spécificités et le degré d'organisation du schéma de sélection de chaque race : en tendance, 7 % des ovins laitières présents dans les noyaux de sélection sont typés au gène PrP, contre 38 % pour les races à viande allaitantes et 17 % pour les races rustiques allaitantes.

A l'échelle du programme global d'éradication de la tremblante en France, 133 000 génotypages ont été réalisés en 2002, si on ajoute les quelque 50 000 typages réalisés dans le cadre de la police sanitaire pour les élevages atteints.

Lors de la première année du programme

Figure 2. Facteur de multiplication du nombre de candidats pour ne retenir que des reproducteurs au moins hétérozygotes ARR avant la sélection sur les caractères de production.

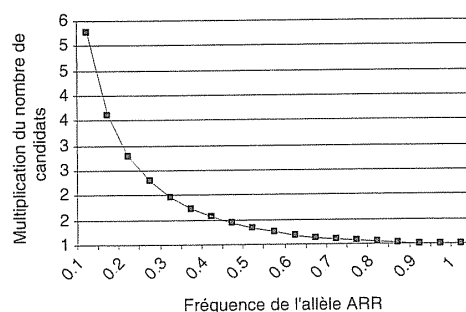


Tableau 9. Nombre de typages PrP dans le cadre du programme national de sélection.

Catégorie d'animaux	2002		2003	
	ovins laitiers	ovins allaitants	ovins laitiers	ovins allaitants
Béliers d'IA (adultes)	0	500	0	0
Béliers de monte naturelle (adultes)	3 590	5 910	0	0
Jeunes béliers	5 300	23 300	14 430	26 470
Mères à béliers	4 800	10 700	5 670	2 030
Agnelles	0	29 100	0	41 300
Total	83 200		89 900	

Tableau 10. Evolution des fréquences génotypiques en ovins allaitants.

Génotype PrP (simplifié)	Béliers actifs 2002 (adultes)	Béliers de millésime 2002		Béliers de millésime 2003	
		nés	entrés en station	nés	entrés en station
	Races à viande n = 4124	n = 11803	n = 1521	n = 16427	n = 1726
RR	26 %	29 %	52 %	44 %	77 %
RS	32 %	31 %	43 %	35 %	21 %
SS	24 %	20 %	0 %	11 %	0 %
VX	19 %	20 %	0 %	11 %	0 %
	Races rustiques n = 1568	n = 11803		n = 10949	
RR	17 %	16 %	28 %	22 %	52 %
RS	46 %	43 %	55 %	47 %	46 %
SS	32 %	34 %	16 %	25 %	2 %
VX	5 %	7 %	1 %	6 %	0 %

Nomenclature :

R = ARR ; S = ARQ ou ARH ou AHQ ; V = VRQ ; X = allèle indifférent

en 2002, des béliers adultes d'IA et de monte naturelle ont été typés. Puis à partir de 2003, le régime de croisière des typages ne porte que sur des jeunes béliers, des mères à béliers (surtout en ovins laitiers), et des agnelles (en ovins allaitants).

Le nombre de génotypes de chaque race peut varier ainsi de moins de 1000 à plus de 10 000, selon son importance et sa situation génétique. L'objectif est de donner proportionnellement plus de moyens de typages aux races présentant une situation initiale plus défavorable (en particulier selon les fréquences alléliques), pour offrir des chances équitables à toutes de progresser aussi rapidement que possible sur la sélection au gène PrP.

b / Evolution des fréquences génotypiques au gène PrP en ovins allaitants

Pour les races bouchères, la comparaison de l'évolution des fréquences génotypiques des 16 427 béliers à haut potentiel génétique nés en 2003 et des 4124 mâles adultes actifs en 2002 dans les noyaux de sélection montre que les maîtres d'œuvre des schémas de sélection ont éliminé ou fait fortement diminuer l'allèle VRQ (tableau 10). En même temps, la proportion de béliers résistants (RR

et RS selon la nomenclature simplifiée du tableau 10) a augmenté de 58 % à 79 %, et même à 100 % pour les béliers 2003 entrés en station.

Parallèlement, dans les noyaux de sélection de ces races bouchères fournissant des béliers de croisement terminal, le nombre de béliers qualifiés au moins hétérozygotes ARR a augmenté de 4425 pour les béliers nés en 2002 à 8225 pour les béliers nés en 2003, soit une augmentation de 86 % du nombre de béliers porteurs de l'allèle ARR, destinés à la diffusion vers les élevages de production.

Pour les *racés rustiques*, en comparant les béliers adultes actifs en 2002 et les béliers nés en 2003 dans les noyaux de sélection, la proportion de béliers résistants a augmenté de 63 à 69 % ; et elle atteint 98 % pour la fraction des mâles entrés en station en 2003.

Globalement, pour l'ensemble des races allaitantes (bouchères et rustiques), pratiquement tous les béliers entrés en station, qui fourniront les béliers élites des prochaines années, sont au moins hétérozygotes ARR en 2003, dont 77 % et 52 % sont homozygotes ARR/ARR respectivement en races bouchères et races rustiques (tableau 10). La sélection pour la résistance à la tremblante peut donc être rapidement efficace au bout de deux

années, quand elle s'appuie sur des schémas de sélection existants.

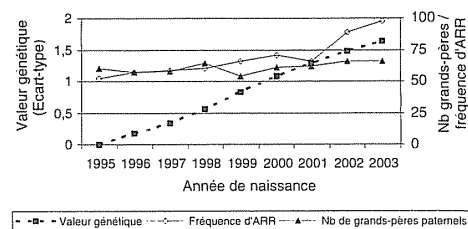
c / Evolution des fréquences génotypiques au gène PrP en ovins laitiers

Compte-tenu des différences de fréquences alléliques initiales, il y a lieu de distinguer la race Lacaune (lait), qui bénéficiait d'une structure initiale plus favorable au gène PrP (tableau 1), des autres races ovines laitières. La présentation concerne les béliers d'insémination animale qui procréent environ 85 % du renouvellement du noyau de sélection Lacaune et 70 % pour les races pyrénéennes. En Lacaune (lait), l'objectif que tous les béliers soient homozygotes résistants est quasi atteint pour la mise en testage (par insémination animale) des béliers de la génération 2003 (tableau 11). En revanche, pour les autres races ovines laitières, l'objectif intermédiaire obtenu est que tous les béliers d'IA du millésime 2003 soient porteurs de l'allèle ARR, dont 32 % d'homozygotes résistants et 68 % d'hétérozygotes, alors que la structure des béliers adultes actifs de 2002 était de 17 % de béliers ARR/ARR et 46 % de béliers hétérozygotes (tableau 11). Si l'allèle ARR est donc quasi fixé pour les béliers d'IA de la race Lacaune nés en 2003, toutefois la progression la plus spectaculaire est à l'actif de la race Manech tête rousse, dont la fréquence de l'allèle ARR s'est accrue, pour les béliers d'IA, de 19 % pour le millésime 1999 à 66 % pour le millésime 2003.

d / Evolution conjointe de la fréquence de l'allèle ARR, du niveau génétique pour les caractères de production et de la variabilité génétique.

La figure 3 présente la situation de la race Lacaune (lait) qui a démarré la sélection pour la résistance à la tremblante dans les années 95. L'évolution de la fréquence de l'allèle ARR reflète les modifications de stratégie pour cette race au fil du temps : dans les années

Figure 3. Evolution conjointe de la fréquence d'ARR, du niveau génétique moyen pour la production laitière et de la variabilité génétique en race Lacaune lait.



1995-2001, seuls les béliers entrant dans les centres d'IA (environ 500 par an) pour mise en testage étaient génotypés au gène PrP; l'objectif était plus d'éliminer les porteurs de l'allèle VRQ et d'utiliser des béliers ARR/ARR dans les élevages infectés que de sélectionner fortement à l'échelle de la population, conformément à la phase expérimentale de l'époque. En conséquence, dans cette période, la fréquence de l'allèle ARR a augmenté progressivement de 52 à 66 % pour les béliers d'IA des millésimes 1995 à 2001.

A partir de 2002, dans le cadre du programme national, quelque 3800 jeunes béliers nés d'accouplements raisonnés sont typés chaque année chez les sélectionneurs naisseurs avant même l'entrée en centres d'élevage, puis en centres d'insémination animale (tableau 11) : ainsi en deux ans, la fréquence de l'allèle ARR des béliers d'IA a augmenté de 66 à 98 % (millésime 2003), avec une réduction temporaire d'environ 15 % du gain génétique pour les caractères laitiers, qui a par ailleurs toujours progressé régulièrement pendant toute la période 1995-2003, tandis que le nombre de grands-pères paternels différents reste stable, témoin du maintien de la variabilité génétique. La fixation de l'allèle ARR est donc quasi acquise à partir du millésime 2003 pour les béliers Lacaune (lait) des deux centres d'insémination animale, sans perte de variabi-

Tableau 11. Evolution des fréquences génotypiques en ovins laitiers pour les béliers d'insémination animale.

Génotype PrP (simplifié)	Béliers actifs 2002 (adultes)	Béliers de millésime 2002		Béliers de millésime 2003	
		nés	mis en testage	nés	mis en testage
	Race Lacaune (lait)				
RR	47 %	n = 3755 47 %	n = 505 79 %	n = 3825 57 %	n = 529 96 %
RS	53 %	44 %	21 %	38 %	4 %
SS	0 %	8 %	0 %	4 %	0 %
VX	0 %	1 %	0 %	1 %	0 %
	Autres races laitières				
		n = 1211	n = 206	n = 1626	n = 208
RR	17 %	32 %	20 %	25 %	32 %
RS	46 %	43 %	66 %	56 %	68 %
SS	32 %	21 %	14 %	18 %	0 %
VX	5 %	4 %	0 %	1 %	0 %

Nomenclature :

R = ARR ; S = ARQ ou ARH ou AHQ ; V = VRQ ; X = allèle indifférent

lité génétique ni réduction notable du gain génétique laitier. En race Lacaune (lait), 85 % du renouvellement du noyau de sélection est procréé à partir des béliers d'IA, qui assurent aussi près de 50 % du renouvellement des élevages de production. La diffusion de l'allèle ARR à toute la population femelle sera donc assez rapide dans la décennie à venir.

Conclusion

Le programme d'éradication de la tremblante mis en place en France depuis 2002 par le Ministère de l'Agriculture est global, puisqu'il vise à la fois à éradiquer à court et moyen terme la tremblante dans les élevages atteints, tout en renforçant la résistance génétique à moyen et long terme de l'ensemble du cheptel national, compte tenu des délais de renouvellement des cheptels femelles. Autrement dit, la nouvelle police sanitaire et le programme national de sélection des ovins pour la résistance à la tremblante sont liés et font tous les deux largement appel à la stratégie génétique fondée sur le génotypage du gène PrP pour éradiquer la tremblante. Un point important est d'avoir choisi d'asseoir le programme national de sélection sur le dispositif existant d'amélioration génétique du cheptel national, organisé dans le cadre de la loi de l'Élevage de 1966. Le fait de s'appuyer sur les schémas et outils de sélection existants est manifestement un point clé pour expliquer à la fois une mise en œuvre rapide et efficace dès la première année, confirmant la mobilisation massive de tous les maîtres d'œuvre des schémas de sélection des ovins en France : les évolutions conjointes constatées pour les fréquences alléliques en faveur de l'allèle ARR, les index de sélection pour les caractères de production et les indicateurs de gestion de la variabilité génétique sont une bonne illustration de l'application des objectifs assignés aux programmes de sélection. L'encadrement national, conduit par l'INRA, l'Institut de l'Élevage et France UPRA Sélection, s'attachera, dans les prochaines années, à vérifier et à aider à la mise en œuvre des quatre objectifs du programme national de sélection, en particulier la fourniture de reproducteurs résistants pour les élevages atteints de tremblante, la gestion de la variabilité génétique dans les noyaux de

sélection et la diffusion vers les élevages de production. Pour ce dernier point, il faut mentionner les initiatives locales ou régionales qui se mettent actuellement en place à destination des élevages de production, le plus souvent pour typer les béliers présents dans ces élevages, ce qui fournit l'opportunité de renforcer les liens entre noyaux de sélection et reste de la population.

En tout état de cause, l'ensemble du programme repose sur les points clés d'universalité de la résistance en situation de contamination naturelle et d'absence de porteurs sains, illustrés dans la première partie de cet article, et qu'il faudra en permanence continuer de valider. La conduite du programme doit donc tenir compte des nouvelles connaissances scientifiques qui pourraient émerger en ce domaine. D'autres questions moins cruciales doivent aussi être abordées, telles que les relations génétiques éventuelles entre PrP et les autres caractères d'intérêt, qui font l'objet actuellement d'investigations systématiques (Barillet *et al* 2002, Prokopova *et al* 2002, Brandsma *et al* 2004) pour l'instant toutes favorables, c'est-à-dire d'absence de relations. Enfin, la poursuite du suivi épidémiologique d'élevages atteints en Pyrénées-Atlantiques et dans le Lot est prévu jusqu'en 2006, ce qui peut contribuer à valider le nouveau dispositif de police sanitaire adopté en 2002, tout en vérifiant l'existence ou non d'un effet de seuil sur la proportion nécessaire de géniteurs résistants, ce qui pourrait à terme infléchir à la fois le dispositif sanitaire et les objectifs de sélection finaux.

Remerciements

La démarche de recherche-développement en ovins laitiers a bénéficié du soutien de l'Union Européenne [contrats FAIR 97-3305 et CT98-7023 (sheepion), CT01-01733 (scrapie-freesheep)], du MAAPAR, ONILAIT et OFIVAL, et du Conseil Régional d'Aquitaine. Nous remercions les éleveurs, leurs techniciens et vétérinaires, et tout particulièrement la DDSV 64.

Le programme national de sélection des ovins pour la résistance à la tremblante est financé par le MAAPAR, via l'OFIVAL et l'ONILAIT.

Références

- Andréoletti O., Berthon P., Marc D., Sarradin P., Grosclaude J., van Keulen L.J.M., Schelcher F., Elsen J.M., Lantier F., 2000. Early accumulation of PrP^{Sc} in gut associated lymphoid and nervous tissues of susceptible lambs from a Romanov flock with natural scrapie. *Journal of General Virology*, 81, 3115-3126.
- Andréoletti O., Lacroux C., Chabert A., Monnereau L., Tabouret G., Lantier F., Berthon P., Eychenne F., Lafond-Benestad S., Elsen J.M., Schelcher F., 2003. PrP^{Sc} accumulation in placentas of ewes exposed to natural scrapie: influence of foetal PrP genotype and effect on ewe-to-lamb transmission. *Journal of General Virology*, 83, 2607-2616.
- Barillet F., Elsen J.M., Arranz J.M., Soulas C., Vial-Novella C., Astruc J.M., Andréoletti O., Schelcher F., Laphitz N., Grenouillet D., 1999. Projet d'étude de diverses modalités d'éradication de la tremblante dans des élevages de brebis laitières des Pyrénées-Atlantiques, 26 p.
- Barillet F., Andréoletti O., Palhière I., Aguerre X., Arranz J.M., Minery S., Soulas C., Belloc J.P., Briois M., Frégeat G., Teinturier P., Amigues Y., Astruc J.M., Boscher M.Y., Schelcher F., 2002. Breeding for scrapie resistance using PrP genotyping in the French dairy sheep breeds. *Proceeding 7th WCGALP, Montpellier, France, Session 13-20*.
- Barillet F., Schelcher F., Andréoletti O., Aguerre X., Arranz J.M., Duboucher C., Corbière F., Fidèle F., Soulas C., 2003. Scrapie epidemic in eight PrP-genotyped pedigree flocks of Manech red faced sheep and consequence of subsequent use of resistant ARR/ARR rams. *Conference on methods for control of scrapie, 15-16 May 2003, Oslo, Norway*, 37.
- Brandsma J.H., Janss L.L.G., Visscher A.H., 2004. Association between PrP genotypes and litter size and 135 days weight in Texel sheep. *Livestock Production Science*, 85, 59-64.

Calavas D., Ducrot C., Savey M., 2002. L'incidence de la tremblante des petits ruminants a-t-elle augmenté en France depuis l'apparition de l'ESB ? Analyse à partir du réseau de surveillance clinique. Bulletin des GTV, 13, 99-105.

Dawson M., 2003. The National Scrapie Plan: progress to date and future developments. State Veterinay Journal, 13, 40-46.

Elsen J.M., Barillet F., François D., Bouix J., Bibé B., Palhière I., 2002. Génétique de la sensibilité à la tremblante des ovins. Bulletin des GTV, 13, 123-128.

Hunter N., 1997. Molecular Biology and Genetics of Scrapie in Sheep. In : Piper L., Ruvinsky A. (Eds), The Genetics of Sheep, pp. 225-240. CAB International, Oxon, UK.

Minery S., Arranz J.M., Aguerre X., Fidelle F., Garrain C., Vial-Novela C., Soulas C., Palhière I., Astruc J.M., Andréoletti O., Schelcher F., Barillet F., 2002. Utilisation de la génétique pour éradiquer la tremblante dans les élevages de brebis laitières des Pyrénées-Atlantiques. Rencontres Recherches Ruminants, 9, 93-96.

Palhière I., Elsen J.M., Barillet F., Astruc J.M., Bibé B., Bouix J., Boscher M.Y., Catrou O., Dion F., François D., Griffon L., Jullien E., Orlianges M., Perret G., Valognes R., 2002. Génétique de la résistance à la tremblante des ovins :

état des connaissances et application pour l'amélioration génétique des populations ovines françaises. Rencontres Recherches Ruminants, 9, 3-9.

Prokopova L.R.M., Lewis W.S., Dingwall W.S., Simm G., 2002. Scrapie genotype: a correlation with lean growth rate? Proceeding 7th WCGALP, Montpellier, France, Session 13-44.

Schreuder B.C., van Keulen L.J.M., Vromans M.E.W., Langeveld J.P.M., Smits M.A., 1998. Tonsillar biopsy and PrPsc detection in the preclinical diagnosis of scrapie. Veterinary Record, 142, 564-568.

Smits M.A., Barillet F., Harders F., Boscher M., Vellema P., Aguerre X., Hellinga M., McLean A.R., Baylis M., Elsen J.M., 2000. Genetics of scrapie susceptibility and selection for resistance. 51th Annual Meeting of the European Association for Animal Production, Book of Abstracts, EAAP N° 6, 295.

Van Keulen L.J.M., Schreuder B.E.C., Vromans M.E.W., Langeveld J.P.M., Smits M.A., 1999. Scrapie-associated prion protein in the gastro-intestinal tract of sheep with natural scrapie. Journal of Comparative Pathology, 121, 55-63.

Vellema P., 2003. Breeding for scrapie resistance in the Netherlands. Conference on methods for control of scrapie, 15-16 May 2003, Oslo, Norway, 32.

Abstract

The French scrapie eradication programme for the entire national sheep flock based on a genetic strategy.

Studies carried out since 1993 on scrapie infected flocks, have allowed to describe the frequency of susceptible alleles of the PrP gene in French dairy sheep breeds, to verify the greater risk of susceptible animals, and to test the selection feasibility and efficiency based on PrP genotyping. The idea of using a genetic strategy to eradicate scrapie in infected flocks was born as a result. Since 2002, the programme to eradicate scrapie in French sheep livestock, implemented by the Ministry of Agriculture, is rooted in a genetic strategy based on the genotyping of the PrP gene: its objective is aimed both at eradicating scrapie in infected flocks and reinforcing the resistance of the entire national flock. The fact that it was launched within the framework of the livestock selection organisation, defined in 1966 by a French law, is obviously a key point to explain its efficient and rapid implementation from the

starting year. Moreover, it illustrates the mobilisation of all the people in charge of the sheep breeding programmes in France: the jointed evolutions observed for the ARR allelic frequency, the genetic merit for production traits and the maintenance of criteria describing the genetic variability are evidence of a good application of the objectives defined for the breeding programmes. The national supervision, carried out by INRA, the Institut de l'Élevage and France UPra selection, will set out, in the next years, to reach all the objectives of this national programme, particularly by supplying scrapie infected flocks with resistant sheep, managing the genetic variability of the selection flocks and diffusing resistant germplasm towards the commercial flocks. This plan is based on the following crucial points: resistance of ARR/ARR is universal to natural infections and ARR/ARR sheep are not healthy carriers. It will be necessary to always check these crucial points and include any new scientific knowledge in the breeding programmes.

Short Communication

Bovine spongiform encephalopathy agent in spleen from an ARR/ARR orally exposed sheep

Olivier Andréoletti,¹ Nathalie Morel,² Caroline Lacroux,¹ Virginie Rouillon,¹ Céline Barc,³ Guillaume Tabouret,¹ Pierre Sarradin,³ Patricia Berthon,³ Philippe Bernardet,³ Jacinthe Mathey,¹ Séverine Lukan,¹ Pierrette Costes,¹ Fabien Corbière,¹ Juan-Carlos Espinosa,⁴ Juan Maria Torres,⁴ Jacques Grassi,² François Schelcher¹ and Frédéric Lantier³

Correspondence

Olivier Andréoletti
o.andreoletti@envt.fr

¹UMR INRA ENVT 1225, Interactions Hôte-Agents Pathogènes, Ecole Nationale Vétérinaire de Toulouse, 23 Chemin des Capelles, 31076 Toulouse, France

²CEA, Service de Pharmacologie et d'Immunologie, CEA/Saclay, 91191 Gif sur Yvette cedex, France

³INRA, Pathologie Infectieuse et Immunologie, INRA Nouzilly, 37380 Nouzilly, France

⁴CISA, Instituto Nacional de Investigacion y Tecnologia Agraria y Alimentaria, 28130 Valdeolmos, Spain

Oral contamination with bovine spongiform encephalopathy (BSE) agent in susceptible PRNP genotype sheep results in widespread distribution of prion in the host. Because ARR homozygous sheep are considered to be resistant to transmissible spongiform encephalopathies, they have been selected to eradicate scrapie from sheep flocks and to protect the human food chain from small ruminant BSE risk. However, results presented here show that several months after an oral challenge with BSE agent, healthy ARR/ARR sheep can accumulate significant amounts of PrP^{Sc} in the spleen.

Received 5 July 2005

Accepted 14 December 2005

Several lines of evidence have shown that the variant Creutzfeldt–Jakob disease epidemic in humans is very probably due to the consumption of bovine spongiform encephalopathy (BSE) agent-contaminated bovine products (Bruce *et al.*, 1997). The possible presence of BSE agent in sheep has already led countries like Great Britain (National Scrapie Plan), France and The Netherlands to promote, through nationally-funded programmes and regulations, the breeding of ARR/ARR sheep (Dawson *et al.*, 2003), which are considered to be resistant to transmissible spongiform encephalopathies (TSEs). Recently, the first 'natural' case of BSE was reported in a goat (Eloit *et al.*, 2005).

TSE-free New Zealand Suffolk lambs from ARR/ARR and ARQ/ARQ ewes, kindly provided by H. Simmons (VLA, ADAS DEFRA, UK), were used for oral inoculation. All animals used in these experiments were treated according to EEC recommendations for animal welfare and under the supervision of the local INRA Ethics Committee. The PRNP genotypes at codons 136, 154 and 171 were routinely checked for each animal using RFLP-PCR (PII Laboratory, INRA-Tours, France). The infectious material was derived from ARQ/ARQ sheep intracerebrally (IC) inoculated with cattle BSE agent. Twenty-four-hour-old ARQ/ARQ ($n=8$) and ARR/ARR ($n=10$) lambs were exposed to a dose of 2.5 g infected brain through natural suckling. A second and

similar inoculation of those animals was performed at 14 days of age. Animals were housed in high-confinement (level 3) facilities dedicated to BSE in sheep studies, to avoid any risk of cross-contamination with scrapie. Animal groups ($n=3$) from each genotype were killed at 4 and 10 months post-challenge (p.c.). At 19 months, the remaining ARQ/ARQ ($n=2$) animals were clinically affected and euthanized. The remaining ARR/ARR ($n=4$) animals were still healthy at 29 months p.c.

PrP^{Sc} detection was carried out in collected samples using immunohistochemistry (IHC), ELISA and Western blot analyses. PrP^{Sc} IHC detection was performed as described previously (Andreoletti *et al.*, 2002) using mAb BAR221 (1/3000 diluted) antibody raised against recombinant ovine protein, which recognizes amino acid sequence 141–152. ELISA measurements were performed using the TeSeE sheep/goat kit (Bio-Rad), a test that has been recently validated by the European Community for the post-mortem diagnosis of TSE in small ruminants. Western blot analysis, preceded by immunopurification, was performed as previously described (Andreoletti *et al.*, 2004) using mAb Sha31, which binds to the 145–152 sequence of PrP (YEDRYRE).

In ARQ/ARQ animals, PrP^{Sc} was detected in various lymphoid tissues as early as 4 months p.c. and in the

central nervous system (obex and medulla) from 10 months p.c. At 10 and 19 months p.c., there was a high level of accumulation in secondary lymphoid organs. These findings are consistent with published data (Bellworthy *et al.*, 2005).

Strikingly, in one challenged ARR/ARR animal, a strong PrP^{Sc} ELISA-positive signal was observed in the spleen at 10 months p.c. This result was reproduced with 10 random samples taken from the same tissue. Following this finding, PRNP full open reading frame (ORF) sequences of all killed animals were checked by sequencing both strands of two overlapping PCR fragments covering the complete ovine PRNP ORF (Primer1F, GTGGGCATTTGATGCTGACAC; Primer1R, TGGTTGGGGTAACGGTACATG; Primer2F, TCAGCCCCATGGTGGTGGCT; Primer2R, CTGCAGGTAGACTCCCTCC). Sequencing analysis of the full PRNP ORF confirmed the ARR/ARR sample genotype and the absence of any other mutation in the PRNP gene.

Using these techniques, no PrP^{Sc} was detected in any other tissue from this animal or in any other challenged ARR/ARR animal. However, since it has been clearly established in rodents that consistent infectivity levels can accumulate in the absence of detectable PrP^{Sc} (Lasmezas *et al.*, 1997; Race *et al.*, 2001), a systematic bioassay in RIII mice of a tissue panel from these animals is ongoing.

Using Western blotting and Sha31 antibody, the observed apparent molecular mass of the non-glycosylated band was similar in spleen samples from BSE agent-exposed ARR/ARR and BSE agent-positive ARQ/ARQ animals, but clearly differed (lower molecular mass) from a control spleen from an ARQ/ARQ animal with 'natural' scrapie (Fig. 1a). As previously described (Race *et al.*, 1998), in both scrapie and BSE, a shift in the apparent molecular mass of proteinase K (PK)-resistant PrP was observed between brain and spleen (Fig. 1a).

Both Sha31 and 2A11 antibodies enabled detection of PrP^{Sc} in spleen and brain from scrapie- or BSE-affected ARQ/ARQ sheep (Fig. 1b, c). However, 2A11 (even after overexposure) was unable to detect abnormal PrP in brain tissue from an IC BSE agent-inoculated ARR/ARR sheep in the terminal stage of the disease, whereas strong labelling was observed with Sha31 (Fig. 1b, c; lane 3). Similarly, the PrP^{Sc} ELISA-positive ARR/ARR spleen was positive using Sha31, whereas no signal was observed using 2A11 (Fig. 1b, c; lane 5). Taken together, these data confirm unequivocally that PrP^{Sc} accumulating in this particular spleen sample is composed entirely of the ARR allelic variant. Finally, accumulated PrP^{Sc} can be considered to be the product of prion replication and not the consequence of a passive accumulation of the inoculum.

ELISA and Western blotting results showed that the total amounts of PrP^{Sc} in the ARR/ARR spleen samples were 5- to 10-fold lower than those in spleens of 10-month-old ARQ/ARQ BSE-infected sheep. Using IHC, abnormal PrP was observed in about 40 % of the ARR/ARR spleen lymphoid

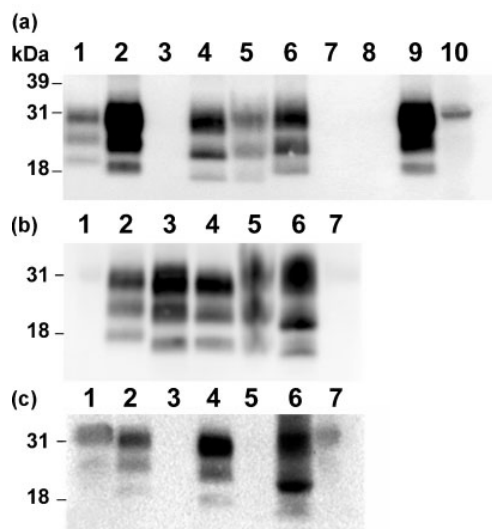


Fig. 1. Western blot PrP^{Sc} detection in spleen from BSE-affected ARR/ARR sheep. (a) Western blotting (with Sha31), preceded by immunopurification. Lanes: 1, brain [0.06 mg equivalent (mg equ⁻¹)] from ARQ/ARQ sheep with natural scrapie; 2 and 9 (identical), brain (0.15 mg equ⁻¹) from orally BSE-inoculated and clinically affected ARQ/ARQ sheep; 3 and 7, brain (40 mg equ⁻¹) and spleen (40 mg equ⁻¹), respectively, from a New Zealand TSE-free ARQ/ARQ sheep; 4 and 5, spleens from ARQ/ARQ (10 mg equ⁻¹) and ARR/ARR (40 mg equ⁻¹) sheep, respectively, that have been orally exposed to BSE 10 months p.c.; 6, spleen (10 mg equ⁻¹) from an ARQ/ARQ sheep with natural scrapie (the same animal as in lane 1); 8, loading buffer alone; and 10, molecular mass markers. (b, c) The same sample extracts were split equally and run in parallel for detection of PrP^{Sc} by Western blotting using either Sha31 (b) or 2A11(c, overexposed blot) antibodies. Lanes: 1 and 7, molecular mass markers; 2, brain from an ARQ/ARQ sheep with scrapie (0.1 mg equ⁻¹); 3, brain from ARR/ARR BSE sheep (IC inoculation – 0.6 mg equ⁻¹); 4, brain from ARQ/ARQ BSE sheep (IC inoculation – 0.2 mg equ⁻¹); and 5 and 6, spleen from ARR/ARR and ARQ/ARQ sheep, respectively, orally exposed to BSE 10 months p.c. (same source and amount loaded as in panel a).

formations, whereas in ARQ/ARQ sheep about 90–100 % were positive (Fig. 2a, b).

To confirm definitively the PK-resistance of the IHC-labelled PrP and characterize the distribution of abnormal prion protein, the paraffin-embedded tissue (PET) blot technique was used (Schulz-Schaeffer *et al.*, 2000) (Fig. 2c, d). Briefly, sections were collected on 0.45 µm nitrocellulose membranes, before drying and deparaffinization. PK (Roche) digestion (250 µg ml⁻¹, 2 h, 55 °C) was performed before denaturation in guanidium isothiocyanate solution (3 M, 10 min, at room temperature). Immunodetection was carried out using mAb Sha31 (4 µg ml⁻¹), followed by

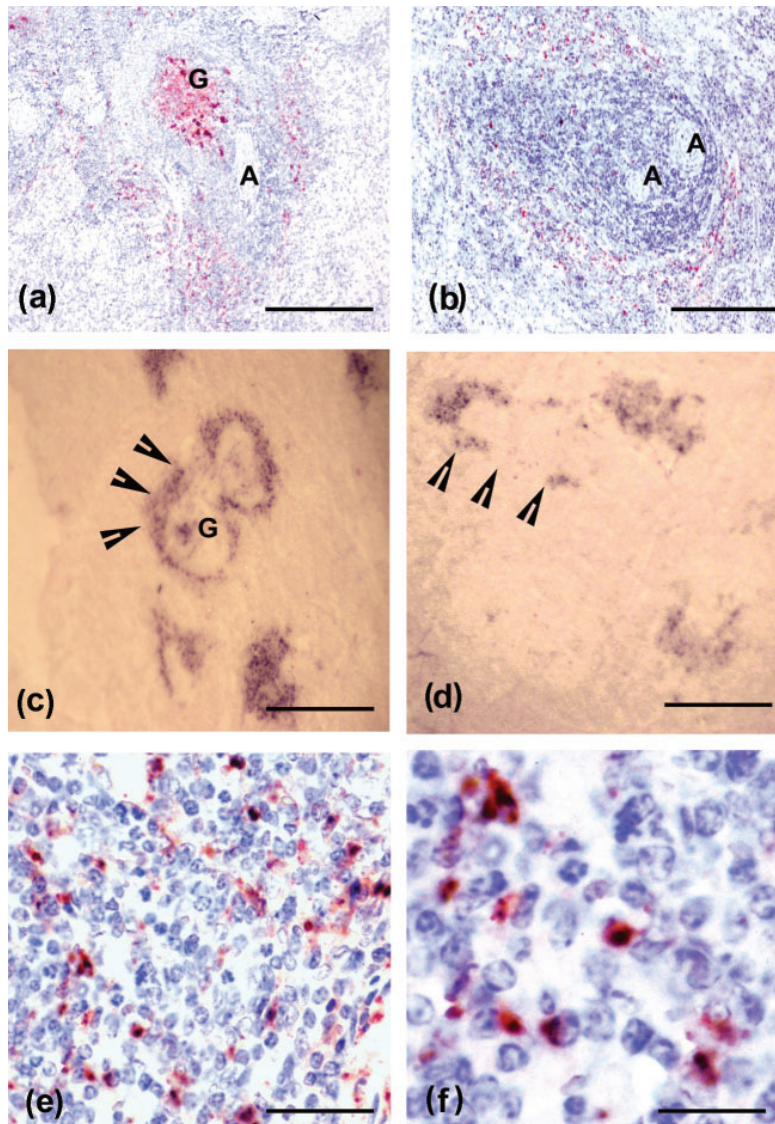


Fig. 2. IHC and PET blot identification of PrP^{Sc} accumulating in histological structures and cellular subsets in spleen from orally BSE-exposed ARQ/ARQ and ARR/ARR sheep. (a, b) *In situ* PrP^{Sc} detection performed using BAR221 mouse mAb (A, arteriole; G, germinal centre). (a) In ARQ/ARQ sheep 10 months after oral exposure to BSE agent, abnormal PrP (red) in spleen was identified in the marginal area of the white pulp and in the germinal centre. Bar, 12 μ m. (b) In ARR/ARR sheep 10 months after oral exposure to BSE agent, PrP^{Sc} was only observed in the marginal area. Bar, 10 μ m. (c, d) PET blots (Sha31 antibody) confirmed PK resistance and distribution (arrowheads point to the marginal area) of the IHC-labelled structures in ARQ/ARQ (c) and ARR/ARR (d) spleens. Bars, 150 μ m. (e, f) Double labelling for PrP^{Sc} (R521-7 rabbit polyclonal antibody, black) and CD68-positive cells (KiM6 mouse mAb, red) showing strong co-localization and accumulation level of PrP^{Sc} in phagocytic cells from the marginal area in white pulp (ARR/ARR spleen). Bars, 45 μ m (e); 20 μ m (f).

application of an alkaline phosphatase-coupled secondary antibody (diluted 1/500; Dako). Binding of secondary antibody was revealed using the nitro blue tetrazolium/5-bromo-4-chloro-3-indolyl phosphate substrate chromogen.

In ARQ/ARQ animals, strong labelling was observed in spleen B-cell follicle germinal centres and marginal zones, which demarcate the white from the red pulp (Fig. 2a–c). In ARR/ARR tissue samples, PrP^{Sc} was not observed in B-cell germinal centres, but was found only in the marginal zone (Fig. 2b–d). In this area, PrP^{Sc}-positive cells were distributed in a rim shape (Fig. 2c, d). To further investigate the nature of these cells, PrP^{Sc}/CD68 double-labelling was performed as previously described (Andreoletti *et al.*, 2002) using R521-7 rabbit anti-ovine PrP protein serum (1/1000 diluted; kindly provided by L. Van Keulen, CIDC, Lelystad, The Netherlands) and a mouse anti-CD68 antibody (1/150

diluted; Serotec). A strong positive double-labelling (Fig. 2e, f) was observed and demonstrated the ‘phagocyte’ nature of the PrP^{Sc}-positive cells (Holness *et al.*, 1993).

The marginal zone in the spleen is a highly efficient system specialized in blood-borne antigen trapping (Aichele *et al.*, 2003). Strong involvement of the CD68-positive cells in the marginal zone could thus suggest that blood circulation has occurred in ARR/ARR animals orally exposed to the BSE agent.

Successful transmission of BSE has been reported after IC challenge in ARR/ARR sheep (Houston *et al.*, 2003). The IC route is considered to be of low relevance to natural exposure, and PrP^{Sc} or infectivity have never been previously reported in tissue from ARR/ARR animals orally challenged with BSE (Bellworthy *et al.*, 2005; Jeffrey *et al.*, 2001).

However, the total number of animals orally inoculated with BSE included in both these published studies was small ($n=14$ for animals challenged and investigated post-mortem). Space constraints in high-security-level animal facilities is the probable explanation for the use of such a low number of animals. In our study, similar space constraints limited the size of the ARR/ARR animal group to 10. To date, from the six challenged and killed animals, only one was found to be positive, which indicates a low transmission rate of BSE agent in animals bearing the ARR/ARR genotype. Such a low transmission rate is a possible explanation for the negative results previously reported. Another point is that in both published studies, ARR/ARR animals were orally dosed at 6 months of age (Bellworthy *et al.*, 2005; Jeffrey *et al.*, 2001). In our study, lambs were challenged first at birth and then at 2 weeks of age. In lambs, the efficiency of infection is considered to decrease with age (Detwiler & Baylis, 2003) and, in nature, TSE infection in sheep seems to occur very soon after birth (Androletti *et al.*, 2000; Detwiler & Baylis, 2003). In this study, due to early exposure, the infectious challenge may have been more efficient than that shown in previous experiments. Moreover, it is worth noting that the challenge dose used in the present study was similar (5 g) to the dose used in previously published experiments, although sheep-derived BSE agent was used in this study instead of cattle BSE agent (Bellworthy *et al.*, 2005; Jeffrey *et al.*, 2001). A possible adaptation of BSE to the sheep could also be an explanation for its improved ability to infect ARR/ARR animals.

Our findings clearly demonstrate that ARR/ARR sheep orally exposed to BSE agent can replicate the agent in their spleen. It also suggests that ARR/ARR sheep are not absolutely resistant to a natural oral challenge of BSE. However, due to the low efficiency of transmission observed in our experimental conditions, even though highly infectious doses and sheep-adapted BSE were used, the BSE transmission risk in ARR/ARR sheep should be considered to be marginal.

Acknowledgements

The authors are greatly indebted to ADAS-DEFRA and VLA Weybridge for producing and breeding the animals, PFIE-INRA Tours for animal care and Bio-Rad for providing the TeSeE sheep/goat kits. 'GIS infections à prion' (French research ministry), EU (QLTR 2001-390), the region Centre (through the A3 animal facility) and the Midi-Pyrénées Region supported this work financially.

References

Aichele, P., Zinke, J., Grode, L., Schwendener, R. A., Kaufmann, S. H. & Seiler, P. (2003). Macrophages of the splenic marginal zone are essential for trapping of blood-borne particulate antigen but

dispensable for induction of specific T cell responses. *J Immunol* **171**, 1148–1155.

Androletti, O., Berthon, P., Marc, D., Sarradin, P., Grosclaude, J., van Keulen, L., Schelcher, F., Elsen, J. M. & Lantier, F. (2000). Early accumulation of PrP^{Sc} in gut-associated lymphoid and nervous tissues of susceptible sheep from a Romanov flock with natural scrapie. *J Gen Virol* **81**, 3115–3126.

Androletti, O., Berthon, P., Levavasseur, E., Marc, D., Lantier, F., Monks, E., Elsen, J. M. & Schelcher, F. (2002). Phenotyping of protein-prion (PrP^{Sc})-accumulating cells in lymphoid and neural tissues of naturally scrapie-affected sheep by double-labeling immunohistochemistry. *J Histochem Cytochem* **50**, 1357–1370.

Androletti, O., Simon, S., Lacroux, C. & 11 other authors (2004). PrP^{Sc} accumulation in myocytes from sheep incubating natural scrapie. *Nat Med* **10**, 591–593.

Bellworthy, S. J., Hawkins, S. A., Green, R. B. & 8 other authors (2005). Tissue distribution of bovine spongiform encephalopathy infectivity in Romney sheep up to the onset of clinical disease after oral challenge. *Vet Rec* **156**, 197–202.

Bruce, M. E., Will, R. G., Ironside, J. W. & 10 other authors (1997). Transmissions to mice indicate that 'new variant' CJD is caused by the BSE agent. *Nature* **389**, 498–501.

Dawson, M., Warner, R., Nolan, A., McKeown, B. & Thomson, J. (2003). 'Complex' PrP genotypes identified by the National Scrapie Plan. *Vet Rec* **152**, 754–755.

Detwiler, L. A. & Baylis, M. (2003). The epidemiology of scrapie. *Rev Sci Tech* **22**, 121–143.

Eloit, M., Adjou, K., Couplier, M. & 18 other authors (2005). BSE agent signatures in a goat. *Vet Rec* **156**, 523–524.

Holness, C. L., da Silva, R. P., Fawcett, J., Gordon, S. & Simmons, D. L. (1993). Macrosialin, a mouse macrophage-restricted glycoprotein, is a member of the lamp/lgp family. *J Biol Chem* **268**, 9661–9666.

Houston, F., Goldmann, W., Chong, A., Jeffrey, M., Gonzalez, L., Foster, J., Parnham, D. & Hunter, N. (2003). Prion diseases: BSE in sheep bred for resistance to infection. *Nature* **423**, 498.

Jeffrey, M., Ryder, S., Martin, S., Hawkins, S. A., Terry, L., Berthelin-Baker, C. & Bellworthy, S. J. (2001). Oral inoculation of sheep with the agent of bovine spongiform encephalopathy (BSE). 1. Onset and distribution of disease-specific PrP accumulation in brain and viscera. *J Comp Pathol* **124**, 280–289.

Lasmezas, C. I., Deslys, J. P., Robain, O. & 7 other authors (1997). Transmission of the BSE agent to mice in the absence of detectable abnormal prion protein. *Science* **275**, 402–405.

Race, R., Jenny, A. & Sutton, D. (1998). Scrapie infectivity and proteinase K-resistant prion protein in sheep placenta, brain, spleen, and lymph node: implications for transmission and antemortem diagnosis. *J Infect Dis* **178**, 949–953.

Race, R., Raines, A., Raymond, G. J., Caughey, B. & Chesebro, B. (2001). Long-term subclinical carrier state precedes scrapie replication and adaptation in a resistant species: analogies to bovine spongiform encephalopathy and variant Creutzfeldt–Jakob disease in humans. *J Virol* **75**, 10106–10112.

Schulz-Schaeffer, W. J., Tschoke, S., Kranefuss, N., Drose, W., Hause-Reitner, D., Giese, A., Groschup, M. H. & Kretzschmar, H. A. (2000). The paraffin-embedded tissue blot detects PrP^{Sc} early in the incubation time in prion diseases. *Am J Pathol* **156**, 51–56.

Dynamics and genetics of PrP^{Sc} placental accumulation in sheep

C. Lacroux,¹ F. Corbière,¹ G. Tabouret,¹ S. Lugañ,¹ P. Costes,¹ J. Mathey,¹ J. M. Delmas,¹ J. L. Weisbecker,² G. Foucras,¹ H. Cassard,¹ J. M. Elsen,³ F. Schelcher¹ and O. Andréoletti¹

Correspondence

O. Andréoletti
o.andreoletti@envt.fr

¹UMR INRA ENVT 1225, Interactions Hôte Agent Pathogène, Ecole Nationale Vétérinaire de Toulouse, 23 Chemin des Capelles, 31076 Toulouse, France

²INRA Domaine de Langlade, 31450 Pompertuzat, France

³INRA Station d'Amélioration Génétique des Animaux, 31326 Castanet Tolosan, France

Placentae from scrapie-affected ewes are an important source of contamination. This study confirmed that scrapie-incubating ewes bearing susceptible genotypes could produce both abnormal prion protein (PrP^{Sc})-positive and -negative placentae, depending only on the PRP genotype of the fetus. The results also provided evidence indicating that scrapie-incubating ARR/VRQ ewes may be unable to accumulate prions in the placenta, whatever the genotype of their progeny. Multinucleated trophoblast cells appeared to play a key role in placental PrP^{Sc} accumulation. PrP^{Sc} accumulation began in syncytiotrophoblasts before disseminating to uninucleated trophoblasts. As these result from trophoblast/uterine epithelial cell fusion, syncytiotrophoblast cells expressed maternal and fetal PrP^C, whilst uninucleated trophoblast cells only expressed fetal PrP^C. In ARR/VRQ scrapie-infected ewes, expression of the ARR allele by syncytiotrophoblasts appeared to prevent initiation of PrP^{Sc} placental deposition. The absence of prions in affected ARR/VRQ sheep placentae reinforces strongly the interest in ARR selection for scrapie control.

Received 20 May 2006

Accepted 12 November 2006

INTRODUCTION

Transmissible spongiform encephalopathies (TSEs) are neurodegenerative disorders occurring in sheep (scrapie), cattle [bovine spongiform encephalopathy (BSE)] and humans [Creutzfeldt–Jakob disease (CJD)] and share similar characteristics: long incubation periods and a progressive and chronic clinical course resulting in death (Fraser, 1976). The accumulation of an abnormal isoform (PrP^{Sc}) of a normal cellular protein (PrP^C) in tissues from infected individuals correlates with the presence of infectivity and is currently considered to be the only reliable biochemical disease marker (McKinley *et al.*, 1983; Race *et al.*, 2001). According to the prion hypothesis, PrP^{Sc} is an infectious protein, thought to be the causative agent of TSEs (Prusiner, 1982).

Ovine susceptibility to TSE (classical natural scrapie and experimental BSE) is controlled mainly by polymorphisms of the PRP gene encoding the PrP protein. The major mutations associated with susceptibility or resistance are located at codons 136 (A or V), 154 (R or H) and 171 (R, Q or H) (Cloucard *et al.*, 1995; Hunter *et al.*, 1996). V¹³⁶R¹⁵⁴Q¹⁷¹/VRQ, ARQ/VRQ and ARQ/ARQ PRP genotype animals are the most susceptible to scrapie, whereas homozygous or heterozygous AHQ and heterozygous ARR

animals show only marginal susceptibility to natural exposure. Similarly, under natural conditions, ARR/ARR sheep are considered to be highly resistant to classical scrapie (Hunter *et al.*, 1996, 1997).

Placentae from TSE-infected ewes are known to carry infectivity and PrP^{Sc} (Race *et al.*, 1998). Recently, it has been shown that PrP^{Sc} accumulation in placental structures seems to depend on the genotype of the fetus (Andréoletti *et al.*, 2002b; Tuo *et al.*, 2002; Alverson *et al.*, 2006). Indeed, fetal placentae obtained by mating VRQ/VRQ scrapie-incubating ewes with susceptible-genotype rams accumulate PrP^{Sc}, whilst placentae of heterozygous ARR fetuses (VRQ/VRQ scrapie-incubating ewes mated with ARR/ARR rams) remain PrP^{Sc} negative. This is especially important with regard to the dissemination of scrapie in infected flocks.

In this study, our main objective was to explore the mechanisms underlying the genetic control of PrP^{Sc} placental accumulation at a cellular level during the gestation period.

METHODS

Animals used in this study were born and bred in the INRA Langlade flock. This is a closed flock of 700 Romanov sheep in

which a high annual incidence (30%) of natural scrapie has been developing since 1993 (Elsen *et al.*, 1999). Animals were cared for following EU recommendations for animal welfare and under the supervision of the local INRA ethics committee. The PRP gene polymorphism at codons 136 (A/V), 154 (R/H) and 171 (R/Q) were determined for each animal using a single-nucleotide polymorphism minor groove binding probe technique (Labogena) (van Poucke *et al.*, 2005).

PrP-genotyped ewes (ARR/VRQ and VRQ/VRQ) were inseminated with either ARR/VRQ or VRQ/VRQ ram semen, according to their experimental groups. At no time did the rams have any contact with the ewes. As described previously, the presence of scrapie in these ewes was assessed by immunohistochemical (IHC) detection of PrP^{Sc} following palatine tonsil biopsy (Schreuder *et al.*, 1996, 1998).

VRQ/VRQ ewes that had positive tonsil biopsies were allocated randomly into one of two experimental groups, A ($n=12$) and B ($n=12$). A third group, C ($n=12$) comprised ARR/VRQ dams with negative tonsil biopsies. Group A dams were inseminated with semen from a VRQ/VRQ ram to produce VRQ/VRQ fetuses. Group B ewes were inseminated with ARR/VRQ ram semen to obtain either VRQ/VRQ or ARR/VRQ fetuses. The probability of obtaining multiple fetal gestation of each genotype in individual ewes was increased because of the high fecundity rate in the Romanov breed (Freetly & Leymaster, 2004). Group C dams were inseminated with semen from the same VRQ/VRQ ram as that used in group A.

Three dams from each group were euthanized (exsanguination following 10 mg pentobarbital sodium salt kg^{-1} by intravenous injection) at 2, 3, 4 and 5 months of gestation (i.e. up to 140 days after artificial insemination; normal duration of gestation in Romanov sheep is 145–150 days). For each ewe, lymphoid tissues (mesenteric lymph node, tonsil, spleen, prescapular lymph node), brain (obex) and reproductive tract tissues (ovary, intercaruncular uterine wall at three different locations) were collected. For each fetus, three cotyledons, chorion, amniotic fluid and umbilical cord were sampled carefully. Additionally, a fetal ear sample was collected for DNA extraction and genotyping.

Each sample was divided into two equal parts. One part was fixed for 10 days in a 10% formalin neutral buffered solution for immunohistochemical PrP^{Sc} detection using mouse monoclonal antibody 8G8 (IgG2a, raised against the human recombinant PrP protein and specifically recognizing the 95–108 aa sequence; kindly provided by J. Grassi, CEA Saclay, France) (Andréoletti *et al.*, 2002b). The second part was frozen at -80°C prior to PrP^{Sc} detection by ELISA (BSE detection test, Platelia BSE TM; Bio-Rad) (Andréoletti *et al.*, 2002b, 2004).

Cotyledons from VRQ/VRQ fetuses obtained from PrP^{Sc}-negative ARR/VRQ ewes (as assessed by ELISA and IHC in all tissues investigated) were used for PrP^C immunolabelling. Prior to this experiment, the absence of PrP^{Sc} in selected cotyledons was checked using both ELISA and Western blotting (not shown). PrP^C labelling *in situ* was carried out with mouse monoclonal antibody BAR224 (IgG2a, raised against the ovine recombinant PrP protein; kindly provided by J. Grassi, CEA Saclay, France), which has a high affinity for sheep PrP (Féraudet *et al.*, 2005), by incubation at a concentration of $4 \mu\text{g ml}^{-1}$ for 1 h at room temperature. Briefly, sections from paraffin-embedded cotyledons were treated for PrP^{Sc} detection (Andréoletti *et al.*, 2002b) but treatment with formic acid was excluded. The specificity of PrP^C labelling was assessed by controls in which the BAR224 antibody was replaced by isotype-matched mouse IgG2a.

Paraffin-embedded tissue blots were performed with sections from positive and control cotyledons (from ARR/VRQ PrP^{Sc}-negative ewes)

as described previously (Schulz-Schaeffer *et al.*, 2000; Andréoletti *et al.*, 2004). These techniques enabled the distribution of PrP^{Sc}-positive structures to be visualized on tissue sections.

RESULTS

PrP^{Sc} accumulation in VRQ/VRQ but not in ARR/VRQ placentae

PrP^{Sc} could not be detected using ELISA or IHC in VRQ/VRQ ewe tissues (ovary and uterine wall) or fetal annexes (chorion and umbilical cord).

In VRQ/VRQ ewes inseminated with VRQ/VRQ ram semen, PrP^{Sc} was detected in some fetal cotyledons as early as 2 months into gestation by ELISA, paraffin-embedded tissue blots and IHC (Figs 1a and 2a, d). From

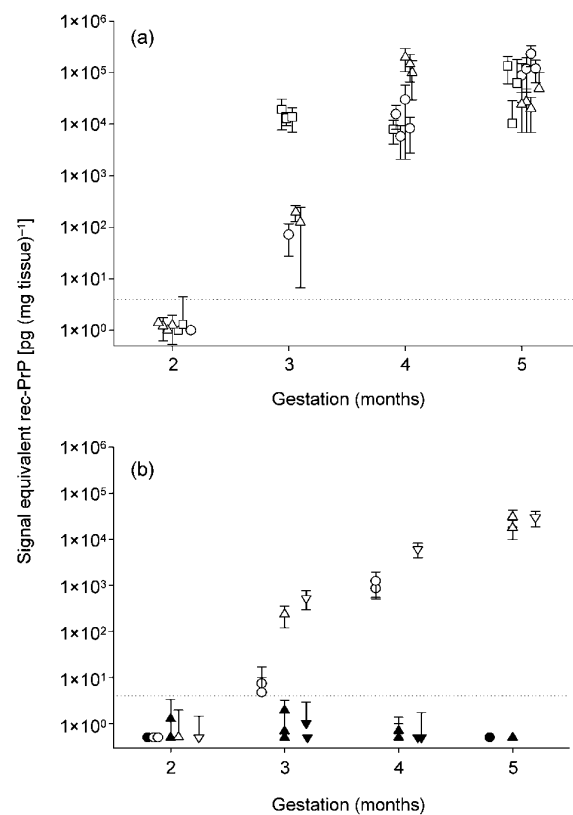


Fig. 1. PrP^{Sc} accumulation levels as assessed by ELISA (TeSeE test; Bio-Rad) in cotyledons from VRQ/VRQ dams inseminated with VRQ/VRQ (a) or ARR/VRQ (b) ram semen. Each symbol is for fetuses recovered from one ewe. Fetuses with ARR/VRQ genotypes are shown as closed symbols and fetuses with VRQ/VRQ genotypes as open symbols. Dotted lines represent the positive cut-off level, determined by testing cotyledons from negative ARR/VRQ ewes. Results are expressed as signal equivalent in positive tissue against an external peptide scale [$\text{pg peptide (mg fresh tissue)}^{-1}$].

3 months of gestation, all of the examined cotyledons were clearly positive and the level of PrP^{Sc}, as assessed by ELISA, increased exponentially (Fig. 1a).

In VRQ/VRQ ewes inseminated with ARR/VRQ ram semen, a clear segregation was observed according to the fetal genotype (Fig. 1b). Whilst all of the placentae of ARR/VRQ fetuses were negative, VRQ/VRQ fetal placentae were consistently positive at 3 months old. As expected, in some ewes (at 3, 4 and 5 months of gestation) both positive and negative placentae were detected, depending only on the fetal genotype (Fig. 1b).

The level of PrP^{Sc} accumulation in cotyledons from VRQ/VRQ fetuses in group A (Fig. 1a) and group B (Fig. 1b) at 3, 4 and 5 months was similar, indicating that the presence of an ARR/VRQ fetus in the same uterus did not interfere with PrP^{Sc} accumulation in VRQ/VRQ placentae.

PrP^{Sc} does not accumulate in placentae from VRQ/VRQ fetuses in ARR/VRQ scrapie-positive sheep

In the Langlade flock, the incidence of scrapie in ARR/VRQ sheep is extremely low (<5%) (Elsen *et al.*, 1999). Because of this low incidence, the ARR/VRQ ewes mated with VRQ/VRQ rams (group C) were initially included as controls in this study. However, at necropsy, three ARR/VRQ ewes were found to have accumulated low but consistent levels of abnormal PrP^{Sc} in various secondary lymphoid formations (ileal mesenteric lymph node, prescapular lymph node) and the enteric nervous system. PrP^{Sc} deposits were also detected in the obex of three animals, indicating that these ewes were in the later stages of scrapie incubation, but with an absence of clinical signs. Absence of PrP^{Sc} in ARR/VRQ scrapie-infected sheep tonsil has already been described on several occasions and explains the negative tonsil biopsies in these three animals (van Keulen *et al.*, 1996a; Andréoletti *et al.*, 2002a).

One of the ARR/VRQ scrapie-positive ewes was euthanized at 2 months of gestation and two at 4 months of gestation. From these last two ewes, eight VRQ/VRQ fetuses and one ARR/VRQ fetus were recovered (four fetuses from the first ewe and five from the second). Whilst group A and B ewes in late gestation (5 months) had associated high levels of PrP^{Sc} accumulation in the cotyledons of VRQ/VRQ fetuses (Fig. 1a, b), no PrP^{Sc} was observed in the placentae from the scrapie-incubating ARR/VRQ ewes, regardless of fetal genotype.

PrP^{Sc} accumulation in the placentome

In group A ewes, focal PrP^{Sc} deposits were observed in the cotyledons of two VRQ/VRQ fetuses at 2 months of gestation (Fig. 2a). By 3 months, the deposits were multifocal in the placentae from all fetuses. By 4 months, PrP^{Sc} deposits had extended and begun to coalesce (Fig. 2b). In most cases, about 50–60% of the tissue surface was positive.

By 5 months of gestation, about 90–100% of the epithelial caruncular interface was positive (Fig. 2c).

In VRQ/VRQ fetuses from infected VRQ/VRQ ewes early in gestation, isolated multinucleated trophoblasts (maternal and trophoblastic hybrid cells) were PrP^{Sc} positive (Fig. 2d). This observation suggested strongly that PrP^{Sc} accumulation could originate in these cells. At 2 and 3 months of gestation, PrP^{Sc} began to accumulate in small uninucleated trophoblastic cell foci surrounding positive multinucleated trophoblasts (Fig. 2e).

In cotyledons collected from PrP^{Sc}-negative ARR/VRQ ewes (group C – the absence of PrP^{Sc} was assessed by ELISA and Western blotting from various tissues including cotyledon), and whatever the genotype of the lamb, a strong and specific PrP^C labelling, using BAR224 monoclonal antibody, was observed only in multinucleated trophoblasts, suggesting overexpression of PrP^C by this cellular subtype (Fig. 3a). No signal was observed in the controls where primary anti PrP antibody was replaced by an isotype-matched mouse Ig, thus confirming the specificity of the labelling (Fig. 3b).

DISCUSSION

Our study confirms and extends previous results indicating that PrP^{Sc} accumulation in the placentae of susceptible scrapie incubating ewes is controlled by the genotype of the fetus (Andréoletti *et al.*, 2002b; Tuo *et al.*, 2002). In the same uterus, two different conceptuses, exposed to a similar risk of contamination, accumulated levels of PrP^{Sc} depending solely on their PrP genotype. Similar findings were recently reported in four ARQ/ARQ ewes mated with heterozygous ARR rams (Alverson *et al.*, 2006). Our study substantially consolidates this result, using a large number of animals and different experimental conditions.

Interestingly, our data also indicated a total lack of interference in PrP^{Sc} accumulation dynamics of VRQ/VRQ fetal placentae linked to the presence of ARR heterozygous fetuses. These observations suggest that both conceptuses behave as independent organisms in term of scrapie contamination.

ARR animals are unable to accumulate PrP^{Sc} in the placenta

In scrapie-incubating heterozygous ARR ewes (at an advanced stage of the disease), PrP^{Sc} could not be detected in fetal placentae, even those with the VRQ/VRQ genotype. This observation has major implications for the genetic control of scrapie and the understanding of the epidemiology of scrapie infection. ARR/ARR genotype sheep are considered to be highly resistant to classical scrapie under natural conditions of exposure and are selected in order to eradicate scrapie from sheep flocks (Dawson *et al.*, 2003).

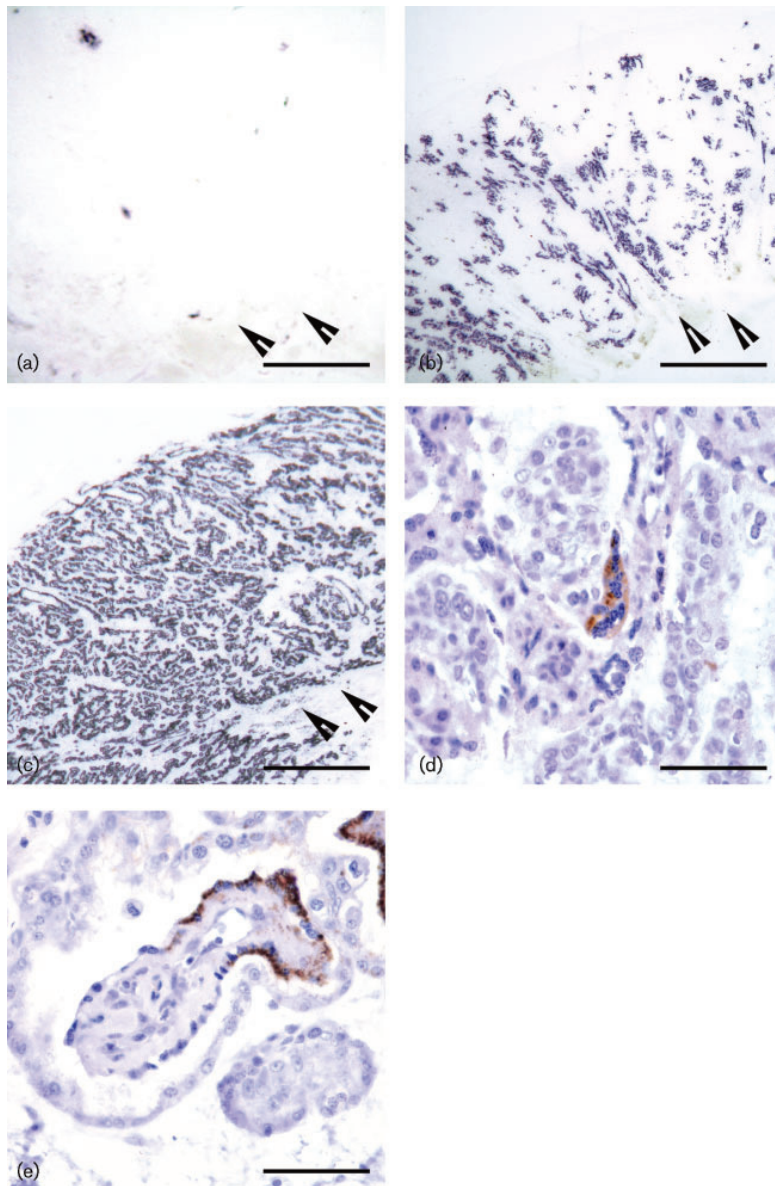


Fig. 2. Paraffin-embedded tissue blot and IHC study of PrP^{Sc} accumulation in the cotyledons of VRQ/VRQ fetuses from VRQ/VRQ scrapie-incubating ewes. At 2 months of gestation (a), faint and multifocal PrP^{Sc} deposits (blue) were observed at different levels in the cotyledons. By 4 months (b) and 5 months (c) of gestation, PrP^{Sc} deposits had extended and coalesced. Arrowheads indicate the base of the placentome in (a)–(c). At 2 months of gestation, initial PrP^{Sc} deposits were observed using IHC in multinucleated trophoblasts (d) before extending to surrounding uninucleate trophoblasts (e). Bars, 200 μ m (a); 500 μ m (b); 1 mm (c); 50 μ m (d); 100 μ m (e).

Introducing the ARR allele into scrapie-infected flocks is designed to reduce scrapie genetic susceptibility by producing, in the short-term, heterozygous ARR animals and then predominantly ARR/ARR animals. These genetic programmes have reduced considerably the dissemination of the scrapie agent through the placenta of incubating ewes with highly susceptible genotypes. Our study extends and reinforces these results by indicating that heterozygous ARR ewes seem unable to disseminate the agent through their placentae and hence participate in lateral or environmental contamination at lambing. However, because of the number of scrapie-incubating ARR heterozygous ewes in our study and the type of TSE agent involved, further experiments are needed before drawing definitive conclusions from this observation.

Contamination of trophoblastic cells

Ovine placenta is of synepitheliochorial type and results from the imbrications of fetal chorionic villi (cotyledons) with maternal preformed endometrial crypts (caruncles). In the placentome, binucleated trophoblastic cells are known to fuse with the maternal caruncular epithelial cell membrane to form a feto-maternal hybrid trinucleate trophoblastic cell. The continuous binucleated trophoblastic cell migration and fusion into trinucleate trophoblastic cells produces hybrid feto-maternal syncytial plaques (Wooding, 1983, 1984).

In VRQ/VRQ dams carrying VRQ/VRQ fetuses, the syncytial trophoblast is the initial cellular subset accumulating PrP^{Sc}.

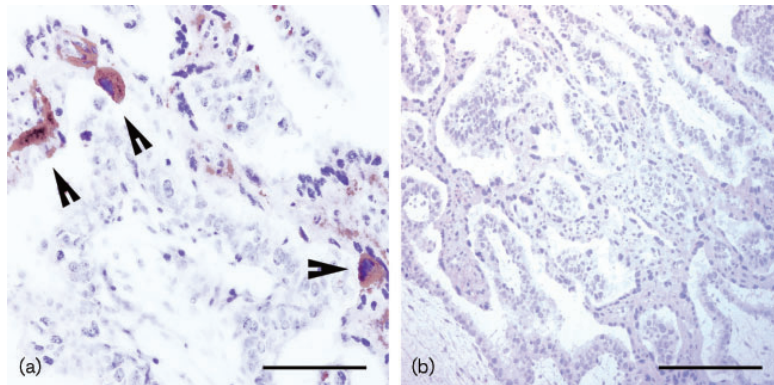


Fig. 3. IHC PrP^C detection (using anti-PrP monoclonal antibody BAR224) in VRQ/VRQ fetus placentae from an ARR/VRQ scrapie-negative ewe. A strong PrP^C signal was detected in multinucleated trophoblasts (arrowheads) using this method and no PrP^C could be detected in uterine epithelial cells (a). Use of an isotype-matched control (mouse IgG2a) confirmed the specificity of this labelling (b). Bars, 300 μm (a); 900 μm (b).

Interestingly, using IHC, multinucleated trophoblast cells also appeared to overexpress PrP^C when compared with trophoblastic uninucleated cells or uterine epithelial cells. However, rapidly after accumulation in syncytial plaques, PrP^{Sc} also accumulated in uninucleated trophoblast cells with a centrifugal pattern originating from multinucleated trophoblastic cells.

Trinucleate trophoblastic cells and syncytial plates are chimeric cells harbouring the PrP gene from both the fetus and the dam. In ARR/VRQ ewes carrying VRQ/VRQ fetuses, syncytial plates express both the ARR and the VRQ alleles, whilst uninucleated trophoblast cells express only the fetal VRQ allele. Although in VRQ/VRQ ewes bearing VRQ/VRQ fetuses, the uninucleated trophoblastic cells were strongly positive, the lack of PrP^{Sc} accumulation in uninucleated trophoblastic cells from ARR/VRQ scrapie-infected ewes cannot be explained by genetic factors.

Taken together, these data suggest that multinucleated trophoblasts are key elements for uninucleated trophoblastic cell contamination. Expression of the ARR allele in that particular cellular subset could avoid contamination of uninucleated trophoblastic cells expressing only the VRQ allele.

Route of placental contamination

The lack of PrP^{Sc} accumulation in the placentae of VRQ/VRQ fetuses carried by scrapie-contaminated ARR/VRQ dams could be linked to inefficient spread of prions to the cotyledons of ARR/VRQ ewes. In scrapie-incubating ewes, the agent might spread through the organism following different pathways, such as blood or blood-circulating cells (Houston *et al.*, 2000). In ARR/VRQ scrapie-incubating animals, the limited involvement of secondary lymphoid tissues during disease pathogenesis (van Keulen *et al.*, 1996b; Andréoletti *et al.*, 2002a) could limit the circulation of the agent in blood from these tissues. In addition to this, the peripheral nervous system might also be an efficient dissemination pathway of the agent (Andréoletti *et al.*, 2004; Thomzig *et al.*, 2004). However, from the results of our experiment, it was not possible to draw conclusions on

the role of these two possible dissemination routes in placental trophoblastic cell contamination. Specially designed studies will be needed to answer this question.

Conclusion

Our study confirms that the way scrapie spreads at lambing in an infected flock is very complex. Agent release into the environment is not a clear-cut process; however, it is certainly governed by interdependent factors such as (i) the presence and number of ewes bearing susceptible genotypes, (ii) their infectious status and (iii) the number and PrP genotype of fetuses from each ewe. The assimilation of these factors into a scrapie epidemiological model remains a challenge.

Our observations also provide new insights into *in vivo* genetic-linked cellular permissiveness to prion infection.

ACKNOWLEDGEMENTS

The authors are greatly indebted to the INRA domain of Langlade for producing and breeding the animals, and to BioRad for providing the TeSeE kits. This work was financially supported by the EU (QLTR 2001-390) and the Midi-Pyrénées Region. The authors are thankful to Jim Foster (IAH, Edinburgh, UK) for his critical reading of the manuscript.

REFERENCES

- Alverson, J., O'Rourke, K. I. & Baszler, T. V. (2006). PrP^{Sc} accumulation in fetal cotyledons of scrapie-resistant lambs is influenced by fetus location in the uterus. *J Gen Virol* **87**, 1035–1041.
- Andréoletti, O., Berthon, P., Levavasseur, E., Marc, D., Lantier, F., Monks, E., Elsen, J.-M. & Schelcher, F. (2002a). Phenotyping of protein-prion (PrP^{Sc})-accumulating cells in lymphoid and neural tissues of naturally scrapie-affected sheep by double-labeling immunohistochemistry. *J Histochem Cytochem* **50**, 1357–1370.
- Andréoletti, O., Lacroux, C., Chabert, A., Monnereau, L., Tabouret, G., Lantier, F., Berthon, P., Eychenne, F., Lafond-Benestad, S. & other authors (2002b). PrP^{Sc} accumulation in placentas of ewes exposed to natural scrapie: influence of foetal PrP genotype and effect on ewe-to-lamb transmission. *J Gen Virol* **83**, 2607–2616.

- Andréoletti, O., Simon, S., Lacroux, C., Morel, N., Tabouret, G., Chabert, A., Lugan, S., Corbière, F., Ferre, P. & other authors (2004). PrP^{Sc} accumulation in myocytes from sheep incubating natural scrapie. *Nat Med* **10**, 591–593.
- Cloucard, C., Beaudry, P., Elsen, J. M., Milan, D., Dussaucy, M., Bounneau, C., Schelcher, F., Chatelain, J., Launay, J. M. & Laplanche, J.-L. (1995). Different allelic effects of the codons 136 and 171 of the prion protein gene in sheep with natural scrapie. *J Gen Virol* **76**, 2097–2101.
- Dawson, M., Warner, R., Nolan, A., McKeown, B. & Thomson, J. (2003). 'Complex' PrP genotypes identified by the National Scrapie Plan. *Vet Rec* **152**, 754–755.
- Elsen, J.-M., Amigues, Y., Schelcher, F., Ducrocq, V., Andréoletti, O., Eychenne, F., Khang, J. V., Poivey, J.-P., Lantier, F. & Laplanche, J.-L. (1999). Genetic susceptibility and transmission factors in scrapie: detailed analysis of an epidemic in a closed flock of Romanov. *Arch Virol* **144**, 431–445.
- Féraudet, C., Morel, N., Simon, S., Volland, H., Frobert, Y., Créminon, C., Vilette, D., Lehmann, S. & Grassi, J. (2005). Screening of 145 anti-PrP monoclonal antibodies for their capacity to inhibit PrP^{Sc} replication in infected cells. *J Biol Chem* **280**, 11247–11258.
- Fraser, H. (1976). The pathology of a natural and experimental scrapie. *Front Biol* **44**, 267–305.
- Freetly, H. C. & Leymaster, K. A. (2004). Relationship between litter birth weight and litter size in six breeds of sheep. *J Anim Sci* **82**, 612–618.
- Houston, F., Foster, J. D., Chong, A., Hunter, N. & Bostock, C. J. (2000). Transmission of BSE by blood transfusion in sheep. *Lancet* **356**, 999–1000.
- Hunter, N., Foster, J. D., Goldmann, W., Stear, M. J., Hope, J. & Bostock, C. (1996). Natural scrapie in a closed flock of Cheviot sheep occurs only in specific PrP genotypes. *Arch Virol* **141**, 809–824.
- Hunter, N., Goldmann, W., Foster, J. D., Cairns, D. & Smith, G. (1997). Natural scrapie and PrP genotype: case-control studies in British sheep. *Vet Rec* **141**, 137–140.
- McKinley, M. P., Bolton, D. C. & Prusiner, S. B. (1983). A protease-resistant protein is a structural component of the scrapie prion. *Cell* **35**, 57–62.
- Prusiner, S. B. (1982). Novel proteinaceous infectious particles cause scrapie. *Science* **216**, 136–144.
- Race, R., Jenny, A. & Sutton, D. (1998). Scrapie infectivity and proteinase K-resistant prion protein in sheep placenta, brain, spleen, and lymph node: implications for transmission and antemortem diagnosis. *J Infect Dis* **178**, 949–953.
- Race, R., Raines, A., Raymond, G. J., Caughey, B. & Chesebro, B. (2001). Long-term subclinical carrier state precedes scrapie replication and adaptation in a resistant species: analogies to bovine spongiform encephalopathy and variant Creutzfeldt–Jakob disease in humans. *J Virol* **75**, 10106–10112.
- Schreuder, B. E., van Keulen, L. J., Vromans, M. E., Langeveld, J. P. & Smits, M. A. (1996). Preclinical test for prion diseases. *Nature* **381**, 563.
- Schreuder, B. E., van Keulen, L. J., Vromans, M. E., Langeveld, J. P. & Smits, M. A. (1998). Tonsillar biopsy and PrP^{Sc} detection in the preclinical diagnosis of scrapie. *Vet Rec* **142**, 564–568.
- Schulz-Schaeffer, W. J., Tschoke, S., Kranefuss, N., Drose, W., Hause-Reitner, D., Giese, A., Groschup, M. H. & Kretzschmar, H. A. (2000). The paraffin-embedded tissue blot detects PrP^{Sc} early in the incubation time in prion diseases. *Am J Pathol* **156**, 51–56.
- Thomzig, A., Schulz-Schaeffer, W., Kratzel, C., Mai, J. & Beekes, M. (2004). Preclinical deposition of pathological prion protein PrP^{Sc} in muscles of hamsters orally exposed to scrapie. *J Clin Invest* **113**, 1465–1472.
- Tuo, W., O'Rourke, K. I., Zhuang, D., Cheevers, W. P., Spraker, T. R. & Knowles, D. P. (2002). Pregnancy status and fetal prion genetics determine PrP^{Sc} accumulation in placentomes of scrapie-infected sheep. *Proc Natl Acad Sci U S A* **99**, 6310–6315.
- van Keulen, L. J., Schreuder, B. E., Meloen, R. H., Mooij-Harkes, G., Vromans, M. E. & Langeveld, J. P. (1996a). Immunohistochemical detection of prion protein in lymphoid tissues of sheep with natural scrapie. *J Clin Microbiol* **34**, 1228–1231.
- van Keulen, L. J., Schreuder, B. E., Meloen, R. H., Mooij-Harkes, G., Vromans, M. E. & Langeveld, J. P. (1996b). Immunohistochemical detection of prion protein in lymphoid tissues of sheep with natural scrapie. *J Clin Microbiol* **34**, 1228–1231.
- van Poucke, M., Vandesompele, J., Mattheeuws, M., van Zeveren, A. & Peelman, L. J. (2005). A dual fluorescent multiprobe assay for prion protein genotyping in sheep. *BMC Infect Dis* **5**, 13.
- Wooding, F. B. (1983). Frequency and localization of binucleate cells in the placentomes of ruminants. *Placenta* **4**, 527–539.
- Wooding, F. B. (1984). Role of binucleate cells in fetomaternal cell fusion at implantation in the sheep. *Am J Anat* **170**, 233–250.

10.3 Résultats de l'analyse multifactorielle

Résultat du modèle M1 : Dans ce modèle, seuls les animaux de génotype PRP connu sont inclus dans l'échantillon d'étude.

Résultat du modèle M2 : Dans ce modèle, les animaux de génotype PRP inconnu sont regroupés dans une classe "Groupe Inconnu".

.

TAB. 10.1 – Résultats du modèle M1

Variable explicative	Partie Incidence			Partie Latence		
	OR	IC _{95%}	p	RR	IC _{95%}	p
Génotype PRP			< 0.0001			0.0143
Groupe SS	1.00	-		1.00	-	
Groupe RS	0.05	0.03 – 0.07	< 0.0001	0.91	0.61 – 1.36	0.6549
Groupe RH	0.81	0.49 – 1.33	0.4059	0.09	0.03 – 0.24	< 0.0001
Groupe SH	2.65	1.96 – 3.58	< 0.0001	0.96	0.75 – 1.22	0.7402
Achat			< 0.0001			< 0.0001
Non achetés	1.00	-		1.00	-	
Achetés	0.33	0.25 – 0.43	< 0.0001	0.50	0.37 – 0.69	< 0.0001
Période de naissance			< 0.0001		< 0.0001	
Millésime relatif ≥ 0	1.00			1.00	-	
Millésime relatif < 0	0.41	0.28 – 0.59	< 0.0001	0.16	0.10 – 0.26	< 0.0001
Élevage			< 0.0001			0.0019
Élevage A	0.35	0.23 – 0.53	< 0.0001	0.83	0.54 – 1.27	0.3969
Élevage B	0.21	0.13 – 0.33	< 0.0001	1.65	1.01 – 2.69	0.0458
Élevage C	1.00	-		1.00	-	
Élevage D	0.14	0.07 – 0.29	< 0.0001	3.39	1.68 – 6.85	0.0007
Élevage E	0.51	0.34 – 0.75	0.0007	1.13	0.78 – 1.64	0.5227
Élevage F	1.80	1.26 – 2.57	0.0012	1.16	0.86 – 1.58	0.3279
Élevage G	0.35	0.22 – 0.53	< 0.0001	2.46	1.62 – 3.72	< 0.0001
Élevage H	0.61	0.42 – 0.89	0.0107	0.68	0.47 – 0.98	0.0377
Élevage I	0.41	0.27 – 0.61	< 0.0001	1.43	0.96 – 2.13	0.0818
Élevage J	2.61	1.77 – 3.83	< 0.0001	0.60	0.43 – 0.83	0.0022
Élevage K	0.34	0.22 – 0.53	< 0.0001	0.47	0.28 – 0.79	0.0047

OR = odds ratio ; RR = risque relatif

TAB. 10.2 – Résultats du modèle M2

Variable explicative	Partie Incidence			Partie Latence		
	OR	IC _{95%}	p	RR	IC _{95%}	p
Génotype PRP			< 0.0001			< 0.0001
Groupe SS	1.00	-		1.00	-	
Groupe RS	0.05	0.03 – 0.05	< 0.0001	0.80	0.53 – 1.20	0.2756
Groupe RH	1.19	0.74 – 1.25	0.4658	0.06	0.02 – 0.17	< 0.0001
Groupe SH	2.98	2.20 – 3.07	< 0.0001	0.92	0.72 – 1.16	0.4776
Groupe Inconnu	6.42	5.53 – 6.51	< 0.0001	1.56	1.36 – 1.79	< 0.0001
Achat			< 0.0001			< 0.0001
Non achetés	1.00	-		1.00	-	
Achetés	0.32	0.26 – 0.32	< 0.0001	0.42	0.33 – 0.53	< 0.0001
Période de naissance			< 0.0001			< 0.0001
Millésime relatif ≥ 0	1.00	-		1.00		
Millésime relatif < 0	0.47	0.39 – 0.48	< 0.0001	0.13	0.10 – 0.17	< 0.0001
Élevage			< 0.0001			0.0023
Élevage A	0.33	0.25 – 0.34	< 0.0001	0.55	0.41 – 0.74	< 0.0001
Élevage B	0.19	0.14 – 0.20	< 0.0001	1.63	1.11 – 2.39	0.012
Élevage C	1.00	-		1.00	-	
Élevage D	0.16	0.12 – 0.17	< 0.0001	1.80	1.25 – 2.60	0.0016
Élevage E	0.20	0.15 – 0.21	< 0.0001	0.93	0.71 – 1.22	0.6067
Élevage F	1.51	1.13 – 1.55	0.005	0.91	0.73 – 1.14	0.4290
Élevage G	0.31	0.24 – 0.32	< 0.0001	1.94	1.53 – 2.45	< 0.0001
Élevage H	0.58	0.43 – 0.60	0.0003	1.09	0.86 – 1.38	0.4754
Élevage I	0.43	0.32 – 0.44	< 0.0001	1.67	1.28 – 2.18	0.0002
Élevage J	1.75	1.29 – 1.81	0.0003	1.00	0.80 – 1.24	0.9719
Élevage K	0.34	0.25 – 0.35	< 0.0001	0.67	0.50 – 0.90	0.0074

OR = odd ratio ; RR = risque relatif

Les facteurs de risque individuels et environnementaux associés à la contamination par l'agent de la tremblante classique du mouton et à la durée d'incubation des animaux infectés demeurent mal connus. En l'absence de diagnostic antemortem précoce, la longue durée d'incubation de la maladie et les durées de vie limitées des animaux conduisent à une connaissance partielle du statut sanitaire des animaux. De plus, l'analyse doit tenir compte du fait qu'une partie seulement des animaux se contamine.

Nous utilisons des modèles d'analyse des données de survie prenant en compte l'existence d'une fraction non à risque. Nous proposons une approche par vraisemblance pénalisée, qui allie les avantages des modèles paramétriques et semi paramétriques existants. Nous nous intéressons ensuite aux modèles paramétriques de survie avec fraction non à risque et effets aléatoires afin de tenir compte du regroupement des animaux dans les élevages. Ces différentes approches sont évaluées à l'aide d'études de simulations.

L'application des ces modèles aux données issues du suivi longitudinal d'élevages infectés des Pyrénées Atlantiques (France) confirme le rôle déterminant du génotype au gène PRP sur le risque de contamination et les durées d'incubation. Nos résultats suggèrent de plus que la contamination par l'agent infectieux a principalement lieu en période néonatale. Enfin la forte hétérogénéité des risques de contamination et des durées d'incubation mise en évidence entre troupeaux pourrait être partiellement expliquée par la prise en compte de la structure génétique des élevages et du nombre d'animaux infectés présents.

Mots clés : tremblante ovine classique, analyse de survie, modèles de mélange, vraisemblance pénalisée, données groupées.

Although they are of major interest in the understanding of the classical scrapie dynamic in infected flocks, the factors influencing the contamination by the infectious agent and incubation period remain poorly known. The absence of antemortem diagnostic tools and the confounding effects of long incubation periods and flock management practices yield a partial knowledge of the infectious status of animals. Moreover, we must take into account that only an unknown fraction of animals become infected, even in heavily affected flocks.

To deal with issues, we use a class of particular survival models, which take account for the presence of long term survivors. We propose a penalized likelihood approach, allowing for the estimation of a smooth risk function. We also develop some parametric models with shared frailties to deal with the presence of grouped data. These different models are evaluated through simulation studies.

These statistical approaches are then applied to the analysis of real data collected during the following-up of infected flocks in Pyrénées Atlantiques, France. The key influence of the PRP genotype on the contamination risk and incubation periods is confirmed. Our results also suggest that, at the individual level, the infection mainly takes place around birth. Finally, the strong heterogeneity in the contamination risk and incubation periods observed between flocks could be partially explained by their PRP genetic structure and the number of incubating animals.

Keywords : scrapie, survival analysis, cure models, penalized likelihood, mixture models, frailty models.

