



HAL
open science

De la surveillance à la détection des problèmes de santé au travail potentiellement émergents dans un contexte de multi-exposition : développement de méthodes appliquées au réseau national de vigilance et de prévention des pathologies professionnelles RNV3P

Laurie Faisandier

► To cite this version:

Laurie Faisandier. De la surveillance à la détection des problèmes de santé au travail potentiellement émergents dans un contexte de multi-exposition : développement de méthodes appliquées au réseau national de vigilance et de prévention des pathologies professionnelles RNV3P. Médecine humaine et pathologie. Université de Grenoble, 2011. Français. NNT : 2011GRENS004 . tel-00582939

HAL Id: tel-00582939

<https://theses.hal.science/tel-00582939>

Submitted on 4 Apr 2011

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.

THÈSE

Pour obtenir le grade de

DOCTEUR DE L'UNIVERSITÉ DE GRENOBLE

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

Arrêté ministériel : 7 août 2006

Présentée par

Laurie FAISANDIER

Thèse dirigée par **Dominique BICOUT**

Co-dirigée par **Régis DE GAUDEMARIS**

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

De la surveillance à la détection des Problèmes de Santé au Travail potentiellement émergents dans un contexte de multi-exposition : Développement de méthodes appliquées au Réseau National de Vigilance et de Prévention des Pathologies Professionnelles

Thèse soutenue publiquement le **17 Février 2011**

devant le jury composé de :

Pr Jacques DEMONGEOT

Directeur-adjoint du laboratoire AGIM FRE CNRS 3405, Président

Pr Marc ARTOIS

Professeur d'infectiologie, VetAgro Sup Lyon, Rapporteur

Pr Christophe PARIS

Professeur de médecine, CHU et Inserm Nancy, Rapporteur

Pr Raymond AGIUS

Professeur de médecine, Université de Manchester, Membre

Pr Mirta GORDON

Directrice de recherche, AMA-LIG Grenoble, Membre

Pr Gérard LASFARGUES

Directeur général adjoint, ANSES Maisons-Alfort, Membre

Dr Dominique J BICOUT

Physicien Modélisateur, EPSP-TIMC VetAgro Sup Lyon, Directeur

Pr Régis DE GAUDEMARIS

Professeur de médecine, EPSP-TIMC et CHU Grenoble, Directeur



Remerciements

Je remercie les membres du jury qui ont accepté de consacrer du temps pour évaluer ce travail :

- Pr Jacques DEMONGEOT en qualité de président, professeur de médecine, directeur du Pôle de Santé Publique au CHU de Grenoble et directeur adjoint du laboratoire AGIM FRE CNRS 3405,
- Pr Marc ARTOIS en qualité de rapporteur, docteur vétérinaire et professeur d'infectiologie et d'épidémiologie au Campus vétérinaire de VetAgro-sup, chercheur à l'Unité 1233 INRA / ENVL spécialisé dans l'analyse des risques infectieux de la faune sauvage,
- Pr Christophe PARIS en qualité de rapporteur, professeur de médecine spécialiste des pathologies professionnelles et environnementales au CHU de Nancy, chercheur au sein de l'équipe Inserm U.954 dans l'équipe Nutrition, génétique et Exposition aux risques environnementaux et président du groupe de travail Système d'Information du RNV3P,
- Pr Raymond AGIUS en qualité d'examinateur, professeur de médecine spécialiste des pathologies professionnelles et environnementales à l'Université de Manchester et responsable du réseau THOR (The Health and Occupational Reporting System) au Royaume-Uni,
- Pr Mirta GORDON en qualité d'examinatrice, professeur et directrice de Recherches au sein du laboratoire LIG, responsable de l'équipe AMA (Apprentissage : Modèles et Algorithmes),
- Pr Gérard LASFARGUES en qualité d'examinateur, professeur de médecine spécialiste des maladies professionnelles et environnementales, directeur général adjoint scientifique de l'Agence Nationale de Sécurité Sanitaire de l'Alimentation, de l'Environnement et du Travail (ANSES).

Je remercie particulièrement mes directeurs de recherche pour la qualité de leur encadrement et pour m'avoir fait confiance pendant toute la durée de ce travail. Qu'ils trouvent ici les marques de ma reconnaissance et de mon respect :

- Dr Dominique J BICOUT, physicien et modélisateur au sein de l'équipe EPSP-TIMC, Université de Grenoble, pour avoir dirigé ce travail, pour sa disponibilité, sa patience et l'intérêt qu'il m'a portée tout au long du projet. Je retiendrais surtout ses conseils et son optimisme à toute épreuve qui m'ont permis de garder confiance dans les moments délicats.
- Pr Régis De GAUDEMARIS, professeur de médecine et santé au travail au CHU de Grenoble, également affilié à l'équipe EPSP-TIMC, Université de Grenoble, épidémiologiste et fondateur du RNV3P. Son expérience et ses conseils ont également contribué à l'aboutissement de ce travail.

... concernant la thèse

Je remercie l'Agence Nationale de Sécurité Sanitaire de l'Alimentation, de l'Environnement et du Travail (ANSES) pour m'avoir fait confiance et accordé le financement de ce projet, notamment Pr Gérard LASFARGUES, pour le suivi des travaux.

Merci aux acteurs et partenaires du RNV3P pour leurs contributions aux données du réseau, sans qui ce travail ne pourrait exister.

Merci à Lynda LARABI, animatrice informatique du RNV3P (CHU Grenoble/ANSES), baptisée la "grande prêtresse des données", qui assure la qualité des données RNV3P et qui nous fait partager ses connaissances.

Je remercie toute l'équipe EPSP - Environnement et Prédiction de la Santé des Populations, notamment Pr Anne MAITRE, Pascale PALENZUELA pour sa constante bienveillance sur les formalités administratives, Franck BALDUCCI pour son aimable contribution et sa participation à la réalisation de mes cours de master, Sylvette LIAUDY pour sa patience et ses encouragements lors du choix des revues et pour la révision des références bibliographies de ce document.

Mes remerciements se tournent vers les doctorants et post-doctorants de l'équipe ; Jennifer PRADEL pour ses encouragements, Adeline TARANTINI pour tous les moments de décompression, Damien BARBEAU pour son éternel optimisme et Renaud PERSOONS à qui je passe le relais pour la prochaine soutenance de thèse.

Un grand merci à Marjo et Kib pour la relecture de ce document, sans oublier ma maman et Danièle pour leurs nombreux encouragements dans la dernière ligne droite.

... autour de la thèse

Je remercie chaleureusement l'équipe de Médecine et Santé au Travail du CHU de Grenoble que j'ai côtoyée tout au long de la thèse : Dr Vincent BONNETERRE, Dr Cyril BERNARDET, Christine BUISSON, Chantal HEYER, Francette VINAY. Je remercie également les capacitaires, internes et stagiaires du service pour leur soutien et leur gentillesse.

Un grand merci à Alain PERDRIX pour ses fabuleux récits et contes qui ravissent et émerveillent à chaque récréation "la gamine" que je suis.

Je remercie l'équipe d'HyperTension Artérielle pour leurs encouragements et tous les bons moments passés en leur compagnie au rez-de-chaussée haut.

Mes sincères remerciements se tournent aussi à mes parents et mon frère Rémi qui m'ont apporté, malgré la distance et les moments difficiles, leur soutien tout au long de ce travail.

Merci à mes amis pour leur bienveillance et leurs encouragements nombreux, en particulier Patrick, JP, Thibault, Connexion, Océ, Marjo, Sophie et tous ceux qui ont suivi (ou subi) le déroulement de ma thèse.

Je ne remerciais jamais assez Antoine (du latin *antonius* : inestimable) qui a largement contribué à ma motivation de tous les jours, pour son soutien sans faille et son enthousiasme contagieux à l'égard de mes travaux... et de la vie en général.

Sommaire

Abréviations	11
Figures	13
Tableaux	17
Glossaire	19
Contexte	21
Partie 1 : Introduction	25
Introduction	27
1. Méthodes et outils pour la recherche en émergence	29
1.1. Les réseaux de surveillance	29
1.1.1. Les réseaux de médecins.....	29
1.1.2. Les réseaux de vigilance	30
1.1.3. Les programmes de surveillance et de détection des alertes sanitaires.....	30
1.2. Investigations des outils et méthodes	31
1.2.1. Les méthodes de seuillage	31
1.2.2. Les cartes de contrôle	32
1.2.3. Les méthodes de clustering.....	33
1.2.4. Les méthodes de régression	34
1.3. Approche par modélisation	35
2. Présentation du RNV3P	39
2.1. Le RNV3P	39
2.1.1. Historique et missions.....	39
2.1.2. Fonctionnement	41
2.1.2.1. Principales étapes de l'interrogatoire professionnel.....	41
2.1.2.2. Codage et exportation des données locales au niveau national.....	44
2.1.2.3. Procédures qualité	45
2.2. Travaux de recherche en émergence	45
2.2.1. Application des méthodes de pharmacovigilance	46
2.2.1.1. Objectif	46
2.2.1.2. Méthodologie	46
2.2.1.3. Résultats obtenus.....	47
2.2.1.4. Conclusion générale.....	47
2.2.2. Veille des émergences cliniques	47
2.2.2.1. Objectif	48
2.2.2.2. Méthodologie	48
2.2.2.3. Résultats obtenus.....	48
2.2.2.4. Conclusion générale.....	48
3. Définition du concept d'émergence au sens du RNV3P	51
3.1. Le concept d'émergence dans la littérature	51
3.1.1. Un risque sanitaire nouveau	52
3.1.2. Un risque sanitaire émergent.....	53
3.1.3. Risque sanitaire ré-émergent ou réurgent	53
3.1.4. Une évolution incertaine du risque	54

3.2. Le concept d'émergence au sens du RNV3P	54
3.2.1. Récurrence dans l'utilisation des codes	54
3.2.1.1. Le thésaurus des pathologies	55
3.2.1.2. Le thésaurus des nuisances.....	56
3.2.1.3. Les thésaurus des métiers et secteurs d'activité.....	57
3.2.2. Le cumul des connaissances.....	58
3.3. Définition de l'émergence au sens du RNV3P.....	60
4. De la mono à la multi-exposition	63
4.1. Introduction : la méthode PRR pour l'étude de la mono-exposition.....	63
4.1.1. Découpage des associations composites	63
4.1.2. Mesure de disproportions	64
4.1.3. Résultats	65
4.2. Objectifs	66
4.3. Défis méthodologiques.....	66
4.3.1. La multiplicité des expositions professionnelles	67
4.3.2. Taille des échantillons	67
4.3.3. Population-mère et population-cible	68
Partie 2 : Exposomes	71
Introduction.....	73
5. Introduction au concept des réseaux.....	75
5.1. Le concept des réseaux	75
5.1.1. Structure des données.....	75
5.1.2. Composantes des réseaux	76
5.2. Les réseaux dans la littérature scientifique.....	77
5.2.1. Les réseaux sociaux	77
5.2.2. Les réseaux en biologie moléculaire	77
5.2.3. Les réseaux en santé	78
5.3. Topologies des réseaux	79
6. Formalisation des Exposomes	81
6.1. Définition des Exposomes.....	81
6.2. Construction des Exposomes	84
7. Caractérisation des Exposomes	107
7.1. A l'échelle des nœuds	107
7.1.1. Degré de connectivité	107
7.1.2. Coefficient d'agglomération	108
7.1.3. Similarité	109
7.2. Identification de groupes d'exposition	110
7.2.1. Groupe d'exposition simple	110
7.2.2. Groupe d'exposition multiple	111
7.2.2.1. Groupe d'exposition hybride.....	111
7.2.2.2. Recouvrement des groupes d'exposition	111
7.2.2.3. Coefficient d'agglomération des groupes d'exposition	112
7.3. Application aux données RNV3P.....	113
7.3.1. Topologie des nœuds	113
7.3.2. Topologie des groupes d'exposition	115
Partie 3 : Fonctions Expertes	119
Introduction.....	121
8. Formalisation des Fonctions Expertes	123
8.1. Formalisation du système	123
8.1.1. Variables	123
8.1.1.1. La pathologie	123
8.1.1.2. Les nuisances.....	123
8.1.1.3. L'activité professionnelle	124

8.1.1.4. Formalisation d'une séquence	124
8.1.2. Descripteurs.....	124
8.1.2.1. Longueur des séquences	124
8.1.2.2. Imputabilité des couples pathologie - nuisance	124
8.1.2.3. Reconnaissance des couples pathologie – nuisance	125
8.1.3. Paramètres	127
8.1.3.1. Le Centre de Consultation.....	127
8.1.3.2. L'activité professionnelle	127
8.2. Configuration des séquences.....	128
8.2.1. Estimation des séquences possibles.....	128
8.2.2. Propriétés des séquences observées	130
8.3. Modélisation du processus.....	131
8.3.1. Construction de l'espace des nuisances associables H.....	132
8.3.2. Construction des seuils.....	134
8.3.3. Fonctions Expertes	135
8.3.3.1. Fonction itérative	137
8.3.3.2. Fonction par blocs	138
9. Détection des associations potentiellement émergentes	143
9.1. Critères statistiques.....	143
9.1.1. Test 0 : Susceptibilité émergentielle	144
9.1.1.1. Condition sur les connaissances acquises	144
9.1.1.2. Condition des fréquences minimales	144
9.1.2. Test 1 : Equiprobabilité des fréquences	145
9.1.2.1. Hypothèse nulle.....	145
9.1.2.2. Expérience.....	145
9.1.2.3. Statistique du test.....	145
9.1.2.4. Règle de décision.....	146
9.1.3. Test 2 : Réussite émergentielle	147
9.1.3.1. Hypothèse nulle.....	147
9.1.3.2. Expérience.....	147
9.1.3.3. Statistique du test.....	148
9.1.3.4. Règle de décision.....	148
9.2. Stratégie de détection des PST potentiellement émergents.....	149
9.3. Application aux données RNV3P.....	150
9.3.1. Echantillon des Sclérodermies Systémiques.....	152
9.3.1.1. Résultats obtenus avec la méthode PRR	152
9.3.1.2. Résultats obtenus avec les Fonctions Expertes	153
9.3.1.3. Bilan.....	155
9.3.2. Echantillon des Tumeurs Malignes de la Vessie.....	156
9.3.3. Echantillon des Lymphomes Non Hodgkiniens	158
10. Exploration de l'émergence au sein de l'exposome	161
10.1. Projection des PST potentiellement émergents.....	162
10.2. Caractérisation des PST potentiellement émergents	163
10.2.1. A l'échelle des nœuds.....	163
10.2.2. A l'échelle des groupes d'exposition.....	164
10.3. Dynamique temporelle des PST potentiellement émergents.....	166
10.3.1. A l'échelle des nœuds.....	166
10.3.2. A l'échelle des groupes d'exposition.....	167
Partie 4 : Discussion - Perspectives.....	169
Synthèse - Discussion	171
Perspectives	179
Conclusion	185
Références	187
Logiciels	197

Valorisation des travaux	199
Annexes.....	237
Annexes A. Recensement des dispositifs contribuant à la surveillance temporelle et à la détection des événements sanitaires émergents	239
Annexe A-1. Les réseaux de recherche fondés sur les pratiques.....	239
Annexe A-2. Les réseaux de vigilance des produits de santé.....	240
Annexes B. Recensement des programmes pour la surveillance et la détection des alertes sanitaires	241
Annexe B-1. Les dispositifs contribuant à l'alerte sanitaire hors France.....	241
Annexe B-2. Les programmes de surveillance et de détection des alertes sanitaire en France.	243
Annexes C. Bordereaux de saisie.....	245
Annexe C-1. Bordereau PATIENT	245
Annexe C-2 : Bordereau CONSULTATION.....	246
Annexe C-3 : Bordereau PROBLEME	247
Annexes D. Résultats détaillés du test 2 - Réussite émergentielle.	249
Annexe D-1. Echantillon des Sclérodermies Systémiques.	249
Annexe D-2. Echantillon des Tumeurs Malignes de la Vessie.....	251
Annexe D-3. Echantillon des Lymphomes Non Hodgkiniens.	257
Annexe E. " Occupational Exposome: a network-based approach for characterizing occupational health problems" - Supplementary Files.	258

Abréviations

Afssa	Agence Française de Sécurité Sanitaire des Aliments
Afsset	Agence Française de Sécurité Sanitaire de l'Environnement et du Travail
ANoVA	ANalysis of VAriance
ANSES	Agence Nationale de Sécurité Sanitaire de l'Alimentation, l'Environnement et du Travail
BCPNN	Bayesian Confidence Propagating Neural Network
BIT	Bureau International du Travail
CCPP	Centres de Consultations des Pathologies Professionnelles
CHU	Centre Hospitalier Universitaire
CITP	Classification Internationale des Types de Professions
CIM	Classification Internationale des Maladies et des problèmes de santé connexes
CNAM	Caisse Nationale d'Assurance Maladie
CNAM-TS	Caisse Nationale d'Assurance Maladie des Travailleurs Salariés
CRRMP	Comité Régional de Reconnaissance des Maladies Professionnelles
CTA	Cellule de Traitement des Alertes
CUSUM	CUmulative SUM control chart
EWMA	Exponentially Weighted Moving Average
GLM	Generalized Linear Mixed Model
KF	Kalman Filter
LQE	Linear Quadratic Estimation
Modernet	Monitoring Trends in Occupational Diseases and New and Emerging Occupation Risks
MPI	Maladies Professionnelles Indemnisables
MSP	Maîtrise Statistique des Procédés
NAF	Nomenclature des Activités Françaises
NAP	Nomenclature des Activités et des Produits
NCvB	Netherlands Center for Occupational Diseases
OIE	Office international des épizooties, désormais Organisation mondiale de la santé animale
OMS	Organisation Mondiale de la Santé
PRR	Proportional Reporting Ratio
PST	Problème de Santé au Travail
RNV3P	Réseau National de Vigilance et de Prévention des Pathologies Professionnelles
RNVPP	Réseau National de Vigilance des Pathologies Professionnelles
ROR	Reporting Odd Ratio
RRFP	Réseaux de Recherche Fondés sur les Pratiques
SFMT	Société Française de Médecine du Travail
SI	Système d'Information
SPRT2	Sequential Probability Ratio Test
THOR	The Health and Occupational Reporting Network

Figures

- 1 Détection de l'alerte canicule avec la méthode de seuillage.
- 2 Carte géographique des Centres de Consultations des Pathologies Professionnelles.
- 3 Alimentation des bases de données locales.
- 4 Schéma de l'acquisition des informations à la production des connaissances.
- 5 Schéma illustratif des informations retenues à la suite de l'interrogatoire professionnel.
- 6 Structure d'un Problème de Santé au Travail.
- 7 Répartition des natures des Problèmes de Santé au Travail (Données émergence RNV3P 2001-2009 ; N = 81 132 PST).
- 8 Représentation schématique des stades d'évolution d'un risque.
- 9 Répartition de l'espace des codes utilisés et des codes par rapport à l'espace des possibles (Données émergence RNV3P 2001-2009 ; N = 81 132 PST).
- 10 Répartition de l'espace disponible, utilisé et récurrent des codes pathologie (3 digits) (Données émergence RNV3P 2001-2009 ; N = 81 132 PST).
- 11 Répartition de l'espace utilisé, non utilisé et récurrent des codes nuisances (5 digits) (Données émergence RNV3P 2001-2009 ; N = 81 132 PST).
- 12 Répartition de l'espace utilisé, non utilisé et récurrent des associations secteur-métier supposées responsables dans la survenue de la pathologie (2 x 2 digits) classés par catégorie de secteur d'activité (Données émergence RNV3P 2001-2009 ; N = 81 132 PST).
- 13 Dynamique temporelle des effectifs des troubles mentaux et du comportement (code F), des maladies de la peau et du tissu cellulaire sous-cutané (code L) et des tumeurs malignes (code C) (Données émergence RNV3P 2001-2009 ; N = 81 132 PST).
- 14 Représentation schématique du découpage des séquences composites en couples pathologie - nuisance.
- 15 Origine de la demande de consultation (Données émergence RNV3P 2001-2009 ; N = 81 132 PST).
- 16 Schéma illustratif de la structure des données.
- 17 Trois types de réseau (Illustration extraite de [Barabási 2004]).
- 18 Approche conceptuelle des Exposomes - Superposition des plans pathologie, nuisances et activité professionnelle.
- 19 Structure emboîtée du réseau avec des exposomes à différentes échelles [Faisandier 2009].
- 20 Illustration du degré de connectivité.
- 21 Illustration du coefficient d'agglomération.

- 22 Illustration de la similarité.
- 23 Illustration des groupes d'exposition. A) Groupe d'exposition simple : Les nœuds sont reliés entre eux par le même élément de l'exposition. B) Groupes d'exposition hybride : Les nœuds partagent une ou plusieurs expositions de nature différente.
- 24 Illustration du recouvrement des groupes d'exposition.
- 25 Illustration du recouvrement "élaboré" des groupes d'exposition. Illustration extraite de [Onnella 2007].
- 26 1₁-Exposome et 1₂-Exposome des Sclérodermies Systémiques (Données émergence RNV3P 2001-2009, N = 77 913 PST hors "99999 - Pas de Nuisance Identifiée").
- 27 1₂- Exposome et 1₃-Exposome des groupes d'expositions associés aux Sclérodermies Systémiques. Entre parenthèses, le nombre de PST dans chaque groupe d'exposition.
- 28 Présence / absence des nuisances dans les groupes d'exposition d'ordre 1 à 5. En vert, les nuisances reconnues au titre des tableaux MPI, en rouge les nuisances non reconnues.
- 29 Echelle ordinale du degré d'imputabilité entre une pathologie et une nuisance.
- 30 Probabilité moyenne des longueurs des PST, notée P_L (moyenne +/- écart type). La répartition est décrite par la loi : $P(L) \sim 0.6832L^{-2}$ (Pearson = 0.99) (Données émergence RNV3P 2001-2009 ; N = 81 132 PST).
- 31 Estimations des configurations de nuisances possibles.
- 32 Diversité des séquences observées par pathologie en fonction du spectre des nuisances (Données émergence RNV3P 2001-2009 ; N = 81 132 PST).
- 33 Paramétrage des Fonctions Expertes.
- 34 Espace des nuisances associables à une pathologie et activité données.
- 35 Espace des nuisances associables aux codes-pathologies enregistrés dans la base RNV3P (Données émergence RNV3P 2001-2009 ; N = 81 132 PST).
- 36 Schéma illustratif du seuil α .
- 37 Distribution de la loi de probabilité empirique des activités professionnelles associées aux sclérodermies systémiques (code M34, N = 171 PST hors "99999 - Pas de nuisance identifiée").
- 38 Organigramme de la fonction itérative, avec s-m l'activité professionnelle, h_1 à h_5 les nuisances, L les longueurs des séquences, et P les probabilités d'association à chaque niveau de L.
- 39 Illustration des lois de probabilité empirique associée aux sclérodermies systémiques (code M34) et à l'activité ouvriers et artisans des métiers de l'extraction et du bâtiment dans la construction (code 45-71).
- 40 Organigramme de la fonction par bloc, avec s-m l'activité professionnelle, L les longueurs des séquences, l le nombre de nuisances reconnues, h_1 à h_5 les nuisances associées, et P les probabilités d'association à chaque niveau de L.
- 41 Organigramme de la Fonction par blocs correspondant aux sclérodermies systémiques associées au métier d'ouvriers et artisans des métiers de l'extraction et du bâtiment dans le secteur dans le secteur de la construction.

- 42 Schéma d'utilisation de la Fonction Experte pour le test 2.
- 43 Stratégie de détection des séquences potentiellement émergentes.
- 44 Schéma détaillé de la taille de l'ensemble des PST susceptibles d'émerger au cours de la période 2001-2009.
- 45 Répartition des séquences associées aux sclérodermies systémiques et enregistrées dans la base RNV3P au cours de la période 2001-2007 (N = 106 PST) et résultats du test composite.
- 46 Répartition des séquences associées aux sclérodermies systémiques et enregistrées dans la base RNV3P au cours de la période 2001-2009 (N = 171 PST) et résultats du test composite.
- 47 Répartition des séquences associées aux Tumeurs Malignes de la Vessie et enregistrées dans la base RNV3P au cours de la période 2001-2009 (N = 495 PST) et résultats du test composite.
- 48 Répartition des séquences associées aux Lymphomes Non Hodgkiniens enregistrés dans la base RNV3P au cours de la période 2001-2009 (N = 31 PST).
- 49 Projection des PST potentiellement émergents sur le 1₁-Exposome et identification des groupes d'exposition associés aux Sclérodermies Systémiques au cours de la période 2001-2009.
- 50 Distributions des métriques relatives aux 90 séquences associées aux Sclérodermies Systémiques et identification des séquences potentiellement émergentes.
- 51 Nœud E dans le 1₁-Exposome: 35110-36200 - Solvants et diluants organiques - Peintures Vernis Laques Mastics.
- 52 "Generalized Topological Overlap Measure" des groupes d'exposition associés aux Sclérodermies Systémiques au cours de la période 2001-2009.
- 53 Synthèse de l'utilisation de l'approche par Exposome.
- 54 Synthèse de détection des séquences potentiellement émergentes.
- 55 Organigramme de la Fonction Experte hybride, où la variable s-m définit l'activité professionnelle, h₁ la nuisance principale, L la longueur de la séquence, l le nombre de nuisance reconnues au titre des Tableaux MPI et h₂ à h₅ les co-nuisances associées.

Tableaux

- 1 Recensement des algorithmes de détection.
- 2 Tableau de contingence 2x2 servant de base pour le calcul des mesures de disproportions utilisées en pharmacovigilance [Bonneterre 2006-2010].
- 3 Tableau extrait et traduit de [Bonneterre 2010]. "Sclérose Systémique - exposition simple", associations générant un signal avec la méthode de pharmacovigilance PRR. Données issues des 30 centres hospitaliers.
- 4 Illustration d'un Tableau de Maladie Professionnelle Indemnisable.
- 5 Répartition des proportions des séquences en fonction de leurs longueurs et du nombre de nuisances reconnues indemnissables au titre des Tableaux des Maladies Professionnelles.
- 6 Exemple illustratif de la règle de décision du test 0.
- 7 Exemple illustratif de la règle de décision du test 1.
- 8 Exemple illustratif de la règle de décision du test 2.
- 9 Classement des pathologies selon leur susceptibilité émergentielle (Données émergence RNV3P 2001-2009 ; N = 77 913 PST hors "99999 - Pas de nuisance identifiée").
- 10 Couples pathologie-nuisance émettant un signal d'émergence valide (en bleu) selon la méthode PRR issues des méthodes de pharmacovigilance [Bonneterre 2010].
- 11 Dynamique temporelle des séquences potentiellement émergentes (T = (1,1,1) en rouge) associées aux sclérodermies systémiques au cours de la période 2001-2007 (N = 106 PST).
- 12 Dynamique temporelle des séquences potentiellement émergentes associées aux sclérodermies systémiques au cours de la période 2001-2009 (N = 171 PST).
- 13 Synthèse des couples émettant un signal d'émergence valide selon les méthodes de pharmacovigilance (hachuré) et les séquences potentiellement émergentes détectées avec les Fonctions Expertes (quadrillé).
- 14 Séquences associées aux Tumeurs Malignes de Vessie susceptibles d'émerger dont les séquences potentiellement émergentes (T = (1,1,1) quadrillé).
- 15 Séquences susceptibles d'émerger associées aux Lymphomes diffus Non hodgkiniens parmi lesquelles des séquences potentiellement émergentes (quadrillé).
- 16 Matrice **B** des groupes d'exposition simples et séquences potentiellement émergentes.
- 17 Dynamique temporelle des métriques relatives aux séquences potentiellement émergentes associées aux Sclérodermies Systémiques (code M34).

- 18 Dynamique temporelle des tailles et fréquences des groupes d'exposition simples incluant des nœuds potentiellement émergents associés aux Sclérodermies Systémiques.
- 19 Tableau de contingence 2x2 servant de base pour le calcul de la sensibilité.
- 20 Classement des séquences potentiellement nouvelles, selon la Fonction itérative.

Glossaire

Les termes définis ci-dessous sont spécifiques au travail présenté dans ce document.

Activité professionnelle	désigne au sens large le métier et le secteur d'activité suspectés d'être en relation avec l'apparition de la maladie
Exposition professionnelle	désigne les nuisances et l'activité professionnelle codifiées par le médecin à la suite de l'interrogatoire professionnel et suspectées d'être potentiellement en cause dans la survenue de la maladie
Maladie	Diagnostic retenu par le médecin à la suite d'investigations cliniques et d'examens médicaux
Nuisance professionnelle	Composé d'origine chimique, toxique ou industrielle, présents dans l'environnement du travail pouvant nuire à la santé du travailleur et suspectée par le médecin d'être potentiellement en relation avec la survenue de la maladie
Nœud / Séquence	Ensemble de PST enregistrés exactement à l'identique dans la base RNV3P, dont le poids est égal au nombre de copies de PST
PST	Problème de Santé au Travail désignant l'association entre une maladie et une exposition professionnelle composite (une à cinq nuisances, un métier et un secteur d'activité)

Contexte

Depuis les deux dernières décennies, l'industrialisation, l'urbanisation, le développement rapide des transports et de l'agriculture intensive ont entraîné des modifications de notre environnement à l'échelle individuelle et collective. Ces modifications pourraient être partiellement responsables de l'augmentation constatée de l'incidence de certains cancers [Peto 2001] du fait de l'émergence de facteurs environnementaux susceptibles de représenter une menace potentielle pour notre santé. Ces évolutions ont notamment marqué notre histoire avec l'apparition de plusieurs crises sanitaires de grande ampleur, induisant des coûts sociaux, économiques et politiques considérables.

› **Retour sur les crises sanitaires de grande ampleur**

Le risque sanitaire est la probabilité de survenue d'une maladie face à l'identification d'un danger. La notion de crise sanitaire qualifie ce risque comme réel ou supposé pour l'état de santé d'une population, rassemblant à la fois le caractère urgent et inédit de l'événement.

SRAS (Syndrome Respiratoire Aigu Sévère), vache folle, sang contaminé, canicule, amiante, pandémie grippale, les crises sanitaires de la fin des années 1990 à nos jours ont introduit une réelle inquiétude du fait de leur rapide émergence et de l'ampleur des conséquences. Largement médiatisées, ces crises suscitent l'indignation parce que l'opinion publique estime qu'elles pourraient être évitées.

› **Une demande sociale d'anticipation des risques**

Le seuil de tolérance à l'égard des risques environnementaux et sanitaires se transforme à partir de cette époque, induisant une demande sociale d'anticipation et de gestion des risques plus forte. Ces crises ont également mis en évidence des moyens insuffisants pour conduire des investigations épidémiologiques du fait d'un manque de collecte et de centralisation des connaissances relatives à la santé et à la multiplicité des expositions. Il est également apparu le besoin de développer une expertise pluridisciplinaire en alliant des compétences issues de plusieurs domaines scientifiques (médecine, pharmacie, biologie, épidémiologie). Enfin, ces crises ont souligné la nécessité de séparer les processus de décision publique des investigations de terrain.

› Les premiers pas de la veille sanitaire

Pour mieux gérer et anticiper ces situations de crises, le paysage de la santé publique a profondément évolué et le concept de "surveillance sanitaire", ou "veille sanitaire", s'est développé, initialement pour la sécurité des soins et des produits de santé. Ce concept de veille sanitaire se traduit par une surveillance continue de l'état de santé d'une population en vue de détecter et prévenir les menaces épidémiques. La veille sanitaire s'est plus largement appliquée aux dangers, d'origine anthropique et naturelle, susceptibles d'impacter la santé des populations. Le concept de sécurité sanitaire remet ainsi les questions de santé publique au cœur des missions de l'Etat, donnant naissance au principe de précaution à l'occasion des débats relatifs aux problèmes internationaux d'environnement (comme la crise de la vache folle), et reçoit une consécration publique en 1992 lors de la conférence de Rio.

En France, comme dans de nombreux pays, les dispositifs contribuant à la sécurité sanitaire se sont développés de manière successive au travers de divers organismes et agences nationales de sécurité. En parallèle, d'autres initiatives souvent peu coordonnées, provenant de réseaux de cliniciens ou de laboratoires, ont aussi contribué à dessiner le paysage de la surveillance et de l'alerte sanitaire tel qu'il est à ce jour. Ainsi, la création des agences nationales, au début des années 1990, peut se lire comme une réponse à une demande de sécurité et de prévention exprimée par les professionnels de santé, et plus largement par l'opinion publique ; comme les Agences Françaises de Sécurité Sanitaire de l'Environnement et du Travail (Afsset) et des Aliments (Afssa) réunies depuis le 1^{er} Juillet 2010 sous l'acronyme ANSES (Agence Nationale de Sécurité Sanitaire de l'Alimentation, de l'Environnement et du Travail).

› Le concept d'émergence

Le concept d'émergence est apparu avec la volonté des professionnels de santé d'anticiper et de prévenir le plus tôt possible les menaces sanitaires. Cependant, ce concept fait appel à un ensemble de notions interdépendantes, aux limites parfois floues et pas toujours consensuelles. Le concept d'émergence fait l'objet de nombreuses définitions correspondant à une réalité complexe. Les scientifiques, la presse et le grand public utilisent ce concept pour désigner traditionnellement les maladies infectieuses émergentes, dont une des définitions est la suivante :

"infections that are newly appeared in a population but are rapidly increasing in incidence or geographic range" [Morse 1995 ; Morse 2004].

Dans le contexte de la santé au travail qu'est le nôtre, une des premières questions sera de définir ce qu'est le concept d'émergence dans le cadre de notre travail.

› Organisation du manuscrit

Ce travail de thèse s'inscrit dans le contexte de surveillance sanitaire et de mise en évidence des maladies professionnelles potentiellement émergentes, dont l'objectif est de contribuer à une réflexion méthodologique autour de la surveillance et la détection des expositions professionnelles multiples présentant un risque potentiel d'émergence. Pour y répondre, des applications aux données du Réseau National de Vigilance et de Prévention des Pathologies Professionnelles (RNV3P) alimentent les aspects théoriques. Ce travail comprend quatre parties divisées en 10 chapitres :

- (1) La première partie est une introduction approfondie qui vise dans un premier temps à investiguer les méthodologies et outils statistiques existants pour la surveillance et la détection des événements sanitaires émergents (chapitre 1). La base de données du RNV3P est ensuite décrite, en positionnant son activité de recherche et les travaux déjà réalisés autour des problématiques de mise en évidence des Problèmes de Santé au Travail potentiellement émergents (chapitre 2). Enfin, nous avons défini le concept d'émergence au sens des données RNV3P (chapitre 3) et présenté les objectifs de ce travail ainsi que les défis méthodologiques auxquels nous devons faire face (chapitre 4).
- (2) La deuxième partie présente une approche basée sur les réseaux relationnels, que nous avons appelé "Exposomes". Nous avons présenté de manière générale le concept des réseaux (chapitre 5) et formalisé l'approche par Exposomes (chapitre 6). Cette approche permet ainsi de caractériser le spectre des expositions professionnelles communes à des individus atteints de la même maladie. L'approche par Exposome est illustrée à partir d'un échantillon extrait de la base RNV3P ; les Sclérodermies Systémiques (chapitre 7).
- (3) La troisième partie présente les travaux de modélisation menés sur la base de données RNV3P, que nous avons appelé "Fonctions Expertes" (chapitre 8). Ces fonctions formalisent le processus d'enregistrement des expositions professionnelles associées à une maladie. Nous avons ainsi développé une stratégie pour détecter des Problèmes de Santé au Travail potentiellement émergents (chapitre 9). Cette approche est illustrée à partir de trois échantillons extraits de la base RNV3P : les Sclérodermies Systémiques, les Lymphomes diffus Non Hodgkiniens et les Tumeurs Malignes de la Vessie. Enfin, nous avons exploré les Problèmes de Santé au Travail identifiés comme potentiellement émergents au sein de l'Exposome (chapitre 10).
- (4) La quatrième partie présente une discussion des approches développées dans ce travail ; Exposomes et Fonctions Expertes. Nous avons proposé des perspectives de travail afin d'améliorer la surveillance et la détection des Problèmes de Santé au Travail potentiellement émergents, ou potentiellement nouveaux.

Partie 1 : Introduction

"What drives epidemiologic quest for risk factors is the strong circumstantial evidence that what we eat, drink, breathe, and so on, are major factors in many devastating illnesses."

Gary TAUBES, 1995.
Journaliste scientifique, article dans SCIENCE.

Avec l'intensité des expositions sur le lieu de travail et l'importance de la population concernée, l'environnement professionnel tient une place particulièrement importante en matière de santé publique, notamment dans le guidage des actions de prévention et de réparation. Les maladies d'origine professionnelle présentent des caractéristiques spécifiques et les temps de latence entre l'exposition et la survenue de la maladie nécessitent de prendre en compte l'historique des activités professionnelles. Ces maladies peuvent être la conséquence de facteurs d'exposition professionnelle cumulés ou résultant d'interaction entre plusieurs expositions, on parle alors de multi-exposition, induisant des interactions ou synergies possibles.

La surveillance sanitaire en milieu professionnel [Imbernon 2006] permet d'évaluer l'impact du travail sur la santé, de repérer les secteurs d'activité, métiers et conditions de travail présentant des risques avérés ou probables sur la santé. Un des principaux enjeux de la surveillance est de suivre leurs évolutions et de repérer le plus tôt possible des problèmes de santé potentiellement émergents. Pour mener de telles études épidémiologiques, les méthodes sont diverses, allant de simples analyses statistiques descriptives à l'utilisation de méthodologies sophistiquées nécessitant des moyens et des ressources matérielles importantes. Cependant, la détection des risques professionnels potentiellement émergents constitue un défi tant pour les épidémiologistes que pour les décideurs de santé publique. En effet, la plupart des études dans le domaine de la santé environnementale modélisent la dispersion dans le temps et dans l'espace de maladies transmissibles, telles que la grippe. Ces modèles sont peu adaptés à la surveillance et la détection des risques professionnels.

Cette partie a pour objectif d'investiguer les méthodes, outils et algorithmes pour la surveillance temporelle des événements sanitaires, pour la détection des tendances, comportements ou événements atypiques. Nous avons ensuite décrit la base de données du RNV3P et présenté les travaux de recherche déjà réalisés. Nous proposons enfin une définition du concept d'émergence au sens des données du RNV3P et présentons les objectifs et les défis méthodologiques de ce travail.

1. Méthodes et outils pour la recherche en émergence

La surveillance sanitaire s'organise traditionnellement en réseau. Ces réseaux sont constitués de groupes d'experts qui mutualisent leurs connaissances et leurs compétences dans le but de collecter des données de santé sur une population spécifique et d'explorer ces informations à des fins de vigilance et de prévention des risques sanitaires.

1.1. Les réseaux de surveillance

On différencie les réseaux de médecins qui, forts de leur expérience de terrain, procèdent à l'enregistrement des cas ; les réseaux de vigilance qui collectent les incidents ou les situations à risques liés à l'utilisation d'un médicament ; et les programmes de surveillance et de détection des alertes sanitaires.

1.1.1. Les réseaux de médecins

- Les Réseaux de Recherche Fondés sur les Pratiques (RRFP)

Les RRFP sont des groupes de cliniciens qui s'organisent pour collecter des données médicales dans le but de comprendre et d'améliorer leur pratique médicale quotidienne. Alors que la recherche conventionnelle travaille sur des données spécifiques et sur de larges échantillons en milieu clinique, les RRFP collectent des données obtenues dans le cadre pratique du terrain. Les principales spécialités médicales concernées sont la médecine générale, la psychiatrie, la pédiatrie et la psychologie. Cette organisation confère aux praticiens une place importante dans la mesure où les conditions d'investigation des cas sur le terrain apportent moins d'interférences, notamment lorsqu'il s'agit de troubles du comportement (Annexe A-1).

- Les réseaux sentinelles

Cette surveillance est basée sur les cas rapportés spontanément par des médecins sentinelles, médecins généralistes libéraux bénévoles et volontaires répartis sur le territoire. Les données issues de tels réseaux permettent d'assurer le suivi d'une maladie donnée à partir d'indicateurs de santé et de développer des modèles de détection des alertes sanitaires [Costagliola 1991] ou des modèles de prévision des épidémies à différentes échelles géographiques [Viboud 2003].

1.1.2. Les réseaux de vigilance

Tous les risques liés à l'utilisation d'un produit ne peuvent pas être mis en évidence par des études cliniques avant la mise sur le marché, c'est pourquoi les réseaux de vigilance sanitaire ont pour vocation de surveiller et d'évaluer les incidents ou les effets indésirables résultant de l'utilisation des produits de santé après leur mise sur le marché. Ces réseaux de vigilance fonctionnent sur le principe de signalement volontaire émanant des professionnels de santé, ou des sujets exposés, et surveillent la sécurité d'utilisation des produits de santé en vue de prévenir tout risque lié à leur utilisation et de développer des actions correctives ou préventives (Annexe A-2).

1.1.3. Les programmes de surveillance et de détection des alertes sanitaires

Les Annexes B centralisent une revue de la littérature des dispositifs européens et internationaux qui contribuent à la surveillance environnementale et à l'alerte sanitaire. Parmi ces dispositifs, le réseau Modernet (*Monitoring Trends in Occupational Diseases and New and Emerging Occupation Risks*) où les équipes suivantes mutualisent leurs connaissances et leurs compétences pour la surveillance et la détection des maladies professionnelles émergentes, à l'échelle de l'Europe notamment avec les réseaux suivants :

- Le système britannique THOR-Extra (*The Health and Occupational Reporting Network*) est une extension du système THOR [Walsh 2005 ; Chen 2005 ; Turner 2005] permettant aux médecins sentinelles de rapporter dans une base de données les cas nouveaux ou inhabituels.
- Le NCvB (*Netherlands Center for Occupational Diseases*) enregistre et communique les statistiques des maladies professionnelles via un enregistrement national et des projets spécifiques de surveillance [Pal 2009]. L'objectif principal de ces enregistrements est d'améliorer les connaissances et la visibilité dans l'apparition et la prévention des maladies professionnelles. Ce système est basé sur trois types de surveillance : celui des maladies professionnelles de la peau (ADS-project) dont les données sont collectées par des dermatologistes, des maladies professionnelles du poumon par des pneumologues (PAL-project) et des cas d'encéphalopathies toxique chronique par les deux *Solvent Teams* (Amsterdam and Enschede).

Le RNV3P se situe parmi les réseaux de cliniciens qui enregistrent, de manière spontanée, des cas pathologiques dans une base de données nationale, et qui dispose d'un programme de surveillance et de détection des associations potentiellement émergentes.

1.2. Investigations des outils et méthodes

Dans les années 1990, les dispositifs de surveillance et de détection des risques sanitaires reposaient essentiellement sur le suivi de maladies à déclaration obligatoire car elles représentaient une menace connue et réelle, du fait de leur gravité et de leur susceptibilité à se propager rapidement dans une population. Aujourd'hui, les récentes avancées en épidémiologie permettent d'utiliser des méthodologies et outils performants pour détecter un risque sanitaire afin de générer une alerte [Wagner 2001]. Ces outils et méthodologies sont, pour la plupart, basés sur le signalement et la détection d'événements anormaux, et se sont diversifiés - et complexifiés - donnant ainsi naissance à des algorithmes sophistiqués. Récemment, la surveillance et la détection des alertes bioterroristes ont pris une place importante et ont ainsi renforcés les recherches pour la détection d'événements présentant une menace susceptible d'impacter la santé.

Les méthodes usuelles se basent sur la comparaison entre des données passées et des données actuelles. En ce qui concerne les émergences, ces méthodes supposent généralement que la maladie soit ancienne, et qu'il s'agit de l'amplification soudaine de cette endémie. Développées initialement pour la surveillance temporelle d'événements, ces méthodes sont basées sur les notions suivantes :

- le dépassement de seuils,
- le changement de régime dans la tendance d'une série chronologique,
- un maximum de cas dans des intervalles de temps,
- un accroissement significatif de la somme cumulée.

Nous avons ainsi identifié de nombreux outils et méthodes pour la surveillance temporelle et la détection des événements sanitaires émergents [Gaudarta 2007]. Nous détaillons leur principe de fonctionnement dans cette section.

1.2.1. Les méthodes de seuillage

L'outil des seuillages donne une réponse positive de détection si, pour une observation, la valeur est supérieure au seuil de détection fixé. Il est possible d'ajuster ce seuil en fonction des anomalies recherchées. Ces méthodes de seuillage sont bien adaptées à la détection des valeurs supérieures aux limites d'exposition ; comme par exemple avec la concentration de dioxine dans les aliments, les seuils épidémiques lors des épidémies hivernales de gastroentérite ou encore le déclenchement des alertes canicules.

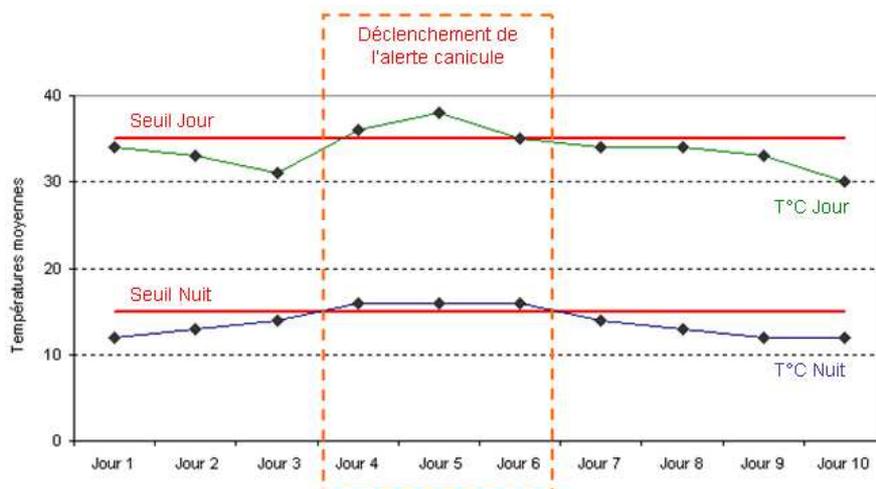


Figure 1 : Détection de l'alerte canicule avec la méthode de seuillage.

Par exemple, l'alerte canicule est déclenchée à Paris lorsque les températures moyennes minimales et maximales mesurées sur 3 jours consécutifs atteignent ou dépassent 21°C la nuit et 31°C le jour, contre 19°C et 32°C pour Brest, 24°C et 35°C à Marseille, 15°C et 35°C à Grenoble.

1.2.2. Les cartes de contrôle

Les cartes de contrôle sont des outils statistiques (MSP - Maîtrise Statistique des Procédés) qui permettent de déterminer le moment où une anomalie apparaît dans un processus. Ces cartes sont largement utilisées dans le domaine industriel, avec par exemple un système de détection des pièces présentant un défaut de fabrication [Kottner 2010]. Ce système de détection se présente sous la forme d'un graphique, où les points représentent le suivi dans le temps d'un indicateur, qualitatif ou quantitatif, dont la valeur centrale (souvent la moyenne) est encadrée de bornes de "conformité", appelées limites de contrôle inférieure et supérieure, bornes à l'intérieur desquelles le processus est dit sous contrôle. On distingue plusieurs cartes de contrôle, les plus utilisées sont les suivantes :

- CUSUM - CUmulative SUM control chart

La méthode CUSUM est une analyse séquentielle des données, initialement développée pour le contrôle de la qualité en milieu industriel. Comme son nom l'indique, cette méthode consiste à représenter la somme cumulée d'un indicateur quantitatif, et à construire les limites du processus. Cette méthode est souvent utilisée pour la surveillance de la mortalité ou de l'incidence d'une maladie [Sibanda 2007 ; Sonesson 2007 ; Fricker 2008 ; Watkins 2008 ; Boulkedid 2010].

- EWMA - Exponentially Weighted Moving Average

La méthode EWMA consiste à calculer les moyennes mobiles d'un indicateur quantitatif et à assigner un poids qui décroît exponentiellement du présent au passé. Cette méthode est utilisée par exemple pour la surveillance des infections chirurgicales ou encore la détection des excès de mortalité [Steiner 2010 ; Pilcher 2010].

- Carte de Shewhart

Les cartes de Shewhart représentent la valeur d'un indicateur qualitatif ou quantitatif, et utilisent classiquement des bornes pour le contrôle du processus. Un changement est détecté lorsque la valeur de l'indicateur dépasse la limite inférieure ou supérieure [Morton 2001].

Ces méthodes peuvent se combiner afin d'optimiser la surveillance et la détection des alertes comme par exemple avec l'utilisation des méthodes CUSUM et Shewhart [Westgard 1977 ; Blacksell 1994] dont l'utilisation simultanée permet de limiter les calculs de la méthode CUSUM et d'optimiser la détection des événements inhabituels.

1.2.3. Les méthodes de clustering

Les méthodes de clustering s'intéressent aux groupes atypiques d'observations en surveillant la distance temporelle entre les cas. Ces techniques incluent les *scan statistics*, l'indice de Tango, la technique des intervalles de temps et la statistique de Knox.

- *Scan Statistic*

Cette technique est basée sur le balayage d'une fenêtre temporelle sur un jeu de données. Le principe est de dénombrer le nombre d'événements observés dans les intervalles de temps [Naus 2006 ; Kulldorff 2007]. On calcule ensuite un rapport de vraisemblance afin de détecter des groupes d'observations atypiques.

- Statistique de Tango

La statistique de Tango est une généralisation de la statistique du χ^2 à l'échelle spatiale. Cette approche est développée par [Tango 1995] et conçu pour être fiable lorsque le nombre d'observations est petit [Rayens 1993].

- Technique des intervalles de temps

Cette technique, idéalement conçue pour les occurrences rares de maladies, détecte les augmentations des taux de maladies chroniques en mesurant l'intervalle de temps entre deux diagnostics [Chen 2003].

- Statistique de Knox

Le test de Knox est fréquemment utilisé pour tester les interactions entre des événements sanitaires occurrents dans un espace et une période maîtrisée. Ce test détermine si les événements sanitaires apparaissent avec une fréquence plus élevée que prévue dans une zone limitée. Pour appliquer le test de Knox, le moment de l'apparition d'une maladie doit être connu, ainsi que les coordonnées de l'endroit où la maladie a eu lieu. Par conséquent, une seule observation se définit par trois valeurs : le temps, la longitude et la latitude [Brooke Marshall 2007].

1.2.4. Les méthodes de régression

Il existe une grande diversité dans les méthodes de régression utilisées à des fins de détection d'événements atypiques. Nous détaillons dans cette section les méthodes de régression les plus couramment utilisées dans la littérature scientifique.

- KF - Kalman Filter (ou LQE - Linear Quadratic Estimation)

Le filtre de Kalman [Kalman 1960] utilise des mesures temporelles contenant du bruit (variations aléatoires) et autres inexactitudes. Il permet de produire des estimations qui ont tendance à être plus proches des valeurs vraies que des mesures originales. On trouve cette technique utilisée dans de nombreux travaux [Heath 1984 ; Gordon 1986 ; Hayton 2007 ; Ansermino 2009 ; Punithakumar 2009].

- Séries temporelles ou chronologiques

Une série temporelle, ou chronologique, est une suite d'observations quantitatives, recueillies à intervalles de temps réguliers. Une série est composée d'une tendance, qui traduit l'évolution à long terme des données d'observation, d'une saisonnalité ou phénomène périodique qui entraîne une variabilité des données ainsi que d'une composante irrégulière entraînant des variations de faible intensité mais de manière aléatoire (bruit). Les séries chronologiques permettent de caractériser et modéliser, à l'aide d'un modèle mathématique, l'évolution d'un processus au cours du temps, et ainsi de prédire son comportement. Il existe deux catégories de modèles pour étudier une série temporelle ; soit on considère que les données sont une fonction du temps permettant ainsi de réajuster les données avec des méthodes itératives comme la méthode des moindres carrés ; soit on cherche à déterminer les valeurs de la série temporelle en fonction des valeurs qui la précèdent, comme le modèle ARIMA (*AutoRegressive Integrated Moving Average*).

- Méthode des points de rupture

Cette approche consiste à détecter un point d'inflexion dans l'apparition d'une maladie ou à détecter un changement dans une distribution [Nobre 1994]. Cette approche est valable à condition que les indicateurs suivis soient représentatifs de la population d'où sont issues les observations.

- Méthode de Serfling

La méthode de Serfling, aussi appelée Serpent de Serfling [Serfling 1963] est basée sur la détection des excès de taux de mortalité ou de morbidité, basé sur un modèle de régression qui prédit une courbe en situation non épidémique et détecte les dépassements des taux réels comme anormaux.

- GLM - Generalized Linear Mixed Model

Lorsque les données présentent une tendance temporelle, un modèle GLM peut être complété d'un modèle de régression de type ANOVA – Analysis of VAriance. Les outils et méthodologies classiques de surveillance et de détection sont basés sur le suivi prospectif d'indicateurs de santé, tels que le taux d'incidence ou de mortalité, si l'échantillon est statistiquement représentatif de la population. Ces systèmes sont bien adaptés à l'évaluation des interventions de santé publique parce que l'élimination des causes a un impact immédiat sur les indicateurs. Dans ce contexte, cette surveillance produit des indicateurs à court terme pour évaluer l'efficacité de stratégies préventives et prédire des risques sanitaires susceptibles d'émerger.

1.3. Approche par modélisation

Face à des données recueillies de manière non conventionnelle, comme le RNV3P, plusieurs algorithmes sont utilisés en fonction du type et de la qualité des données recueillies. Les modèles détaillés ci-dessous font appel à des implémentations et des algorithmes complexes et sont pour la plupart appliqués au domaine de la surveillance syndromique, plus précisément celui des attaques bioterroristes. La surveillance syndromique a pour objectif d'identifier le plus tôt possible des grappes de malades, avant même que le diagnostic soit confirmé et rapporté aux organismes de santé publique [Henning 2004]. L'action de cette surveillance consiste à fournir une réponse rapide afin de réduire le taux de morbidité ou de mortalité.

Algorithme	Mission	Fonctionnement	Domaine d'application	Références
WSARE - <i>What's Strange About Recent Events?</i>	Détecter des anomalies pertinentes dans la surveillance d'une maladie donnée	Classer les dates d'apparition des événements Comparer les événements de ce jour avec les événements d'une distribution de références obtenues à partir des données empiriques Calcul de <i>p-value</i> pour classer les événements	Surveillance syndromique ; Epidémie influenza	Wong 2003 ; Wong 2005 ; Kaufman 2007
BCD - <i>Biosurveillance using a Change-point Detector</i>	Détecter une attaque terroriste le plus tôt possible	Inférer des probabilités sur des données récemment observées en posant des hypothèses sur l'apparition ou non de l'épidémie - Calculer un rapport de vraisemblance à partir de ces 2 probabilités Calcul de <i>p-value</i> à l'aide d'un modèle Monte Carlo	Surveillance syndromique	Buckeridge 2005
EARS - <i>Early Aberration Reporting System</i>	Fournir plusieurs alternatives de détection des aberrations à l'échelle du pays, de l'état et des services de santé locaux	Evaluation des changements rapides dans les fréquences et les taux Caractérisation des tendances ou des clusters inhabituels	Surveillance syndromique (Détection des attaques bio terroristes) ; Détection de l'épidémie influenza	Hutwagner 2003 ; Hutwagner 2005 ; Lawson 2005 ; Zhu 2005 ; McLeod 2009
BioSense	Identifier et détecter des anomalies au niveau local et ceux qui traversent les frontières juridictionnelles	Détecter des anomalies Estimer la provenance géographique et temporelle des anomalies Investiguer des faits complémentaires Comparer ces données avec des données provenant de plusieurs sources Développer des hypothèses Exclure ou valider l'existence d'un fait réel	Surveillance syndromique (Détection des épidémies de maladies infectieuses - intentionnelles et naturelles)	Sokolow 2005

Tableau 1a : Recensement des algorithmes de détection.

Algorithme	Mission	Fonctionnement	Domaine d'application	Références
ESSENCE - <i>Electronic Surveillance System for the Early Notification of Community-based Epidemics</i>	Détecter précocement des épidémies sur données pré-identifiées	Utilise SPC (Statistical Process Control) et MSPC (Multivariate SPC) pour normaliser les séries temporelles	Surveillance syndromique	Burkom 2004 ; Hurt-Mullen 2005 ; Najmi 2005 ; Burkom 2005
Bio-ALIRT – <i>Bio-Event Advanced Leading Indicator Recognition Technology</i>	Détecter des groupes d'épidémies dans un espace et un lieu donné	Détecter, en fonction d'un espace-temps donné, des groupes d'épidémies ayant développés des maladies respiratoires et gastro-intestinales Evaluer la performance de l'algorithme (sensibilité et spécificité - délais de détection)	Détection de la dissémination d'agents pathogènes terroristes	Siegrist 2004
ASAP - <i>Alternative Surveillance Alert Program</i>	Produire le plus rapidement possible des informations fiables sur la santé publique	Non renseigné	Surveillance syndromique des maladies intestinales	Edge 2003
PANDA - <i>Population-wide ANomaly Detection and Assessment</i>	Estimer le temps attendu pour que le clinicien détecte une épidémie de maladies	Approche bayésienne spatio-temporelle et multivariée pour modéliser la probabilité pour qu'une radiographie pulmonaire (ou une culture de sang) conduite à un diagnostic en cas d'épidémies sachant que le patient s'est présenté à telle heure au clinicien	Détection de maladies induites par une attaque bioterroriste (de type anthrax ou variole)	Shen 2008 ; Shen 2009 ; Shen 2010
BARD - <i>Bayesian Aerosol Release Detector</i>	Détecter les tendances dans les données compatibles avec la libération d'aérosols d'un agent pathogène	Modèle gaussien de dispersion d'agents pathogènes avec des données météorologiques récentes combinées avec des modèles bayésiens des symptômes de la maladie et les données de conditions normales	Bioterrorisme	Lawson 2006 ; Hogan 2007 ; Rolka 2007 ; Cami 2009
SSIC - <i>Syndromic Surveillance Information Collection</i>	Investiguer et détecter le plus précocement possible des syndromes cliniques	Non renseigné	Attaques bioterroristes et grandes flambées d'origine alimentaire et les infections émergentes et la grippe pandémique	Lober 2003

Tableau 1b : Recensement des algorithmes de détection.

Les algorithmes développés pour la surveillance syndromique constituent une alternative aux méthodologies conventionnelles car elles sont basées sur le développement d'un système adapté aux données de surveillance et contribuent à mettre en évidence des événements atypiques pour lesquels des hypothèses de risque sanitaire émergent sont sous-jacentes.

* * * * *

Pour répondre aux besoins de surveillance et de détection des événements de santé, les épidémiologistes utilisent des outils qui, malgré leurs limites, permettent aujourd'hui de satisfaire en partie aux besoins et aux exigences des décideurs. Ces méthodes atteignent leurs limites lorsque des décennies séparent les expositions et les premiers effets sur la santé, comme pour les cancers d'origine professionnelle.

Résumé

1. Méthodes et outils pour la recherche en émergence

La surveillance sanitaire s'organise traditionnellement en réseau où des cliniciens mutualisent leurs connaissances en enregistrant des événements de santé de manière standardisée. Les enregistrements des maladies à déclaration obligatoire, comme la tuberculose, se différencient des déclarations spontanées de médecins qui observent des cas de manière ponctuelle. Pour ce mode de recueil spontané, les épidémiologistes ont développé des méthodologies et outils statistiques, basés sur la détection d'anomalies, le plus souvent caractérisées par un changement statistiquement significatif concernant un ou plusieurs événements de santé par rapport à son historique. Ce procédé peut être résumé en deux étapes :

- 1/ La création d'une base de données qui stocke des descripteurs associés à une maladie,
- 2/ L'élaboration d'un seuil tel que la présence d'une anomalie soit détectée une fois le seuil dépassé.

Les méthodologies et outils statistiques utilisés pour la mise en évidence d'événements susceptibles d'émerger, développés le plus souvent pour la surveillance syndromique des attaques bioterroristes, sont utilisés selon les conditions suivantes :

- le délai entre l'exposition et la survenue de la maladie est court,
- la population de référence est connue (dénominateur),
- l'anomalie recherchée est connue (augmentation significative de l'incidence par exemple).

2. Présentation du RNV3P

L'intérêt croissant des pouvoirs publics pour l'amélioration des conditions de travail a ainsi favorisé la mise en place, en 2001, du Réseau National de Vigilance et de Prévention des Pathologies Professionnelles (RNV3P).

2.1. Le RNV3P

Le RNV3P est né du besoin de confronter et d'améliorer l'expertise des spécialistes par la mise en commun de leurs compétences et de leurs expériences. Initialement, les cas adressés par la médecine du travail étaient discutés au cours de réunions entre les membres du réseau. Après quelques années de ce type d'échanges et grâce à l'informatisation des données médicales, les connaissances et expériences des spécialistes sont mises en commun dans un Système d'Information (SI) qui rassemble des Problèmes de Santé au Travail (PST) dans une base de données à l'échelle nationale.

2.1.1. Historique et missions

La démarche est formalisée avec la création d'un réseau de Centres de Consultations des Pathologies Professionnelles (CCPP), avec l'aide des Caisses Régionales d'Assurance Maladie (CRAM), donnant naissance en premier lieu au Réseau National de Vigilance des Pathologies Professionnelles (RNVPP). Ces CCPP sont des centres d'expertise qui canalisent des observations à la suite de consultations, pour former des bases de données à l'échelle locale. La mise en commun de leurs compétences et de leurs expériences leur confère une place particulièrement importante en termes de vigilance et de prévention des risques professionnels.

Le RNVPP a été associé à l'Afsset en 2006 et est ainsi devenu le Réseau National de Vigilance et de Prévention des Pathologies Professionnelles (RNV3P) dont les partenaires sont la Caisse Nationale d'Assurance Maladie des Travailleurs Salariés (CNAM-TS), la Société Française de Médecine du Travail (SFMT) et le Centre Hospitalier Universitaire (CHU) de Grenoble. Ainsi, d'un système de recueil standardisé des consultations, le réseau devient un système coordonné de production de connaissances pour tous ses partenaires.



Figure 2 : Carte géographique des Centres de Consultations des Pathologies Professionnelles.

La mission du RNV3P est d'assurer la veille sanitaire à trois niveaux :

- (1) Une activité de veille épidémiologique programmée qui consiste à suivre au cours du temps l'évolution de couples pathologie - exposition prédéfinis.
- (2) Une activité de veille clinique qui permet aux partenaires du RNV3P, face à une nouvelle association, de rechercher la présence d'autres cas à l'échelle nationale et d'émettre un signalement pour une vigilance renforcée auprès des autres médecins du réseau.
- (3) Une activité de recherche pour élaborer des outils de mise en évidence d'associations potentiellement émergentes.

Les enjeux scientifiques sont de mettre en œuvre les outils nécessaires pour renforcer son rôle de vigie et d'alerte vis-à-vis de problèmes de santé potentiellement émergents, et pour améliorer sa capacité à décrire les situations professionnelles à risque.

2.1.2. Fonctionnement

Dès l'arrivée du patient dans le CCPP, un bordereau "Patient" est rempli de manière systématique à des fins administratives (Annexe C-1). Ensuite, le patient se présente à une ou plusieurs consultations qui aboutissent à la création d'un bordereau "Consultation" (Annexe C-2). Les CCPP reçoivent les patients pour diagnostic d'origine professionnelle dans 76% des cas (soit 62 268 / 81 132 PST enregistrés depuis 2001). La consultation consiste en un entretien oral individuel entre le patient et le médecin consultant qui se déroule dans les CCPP.

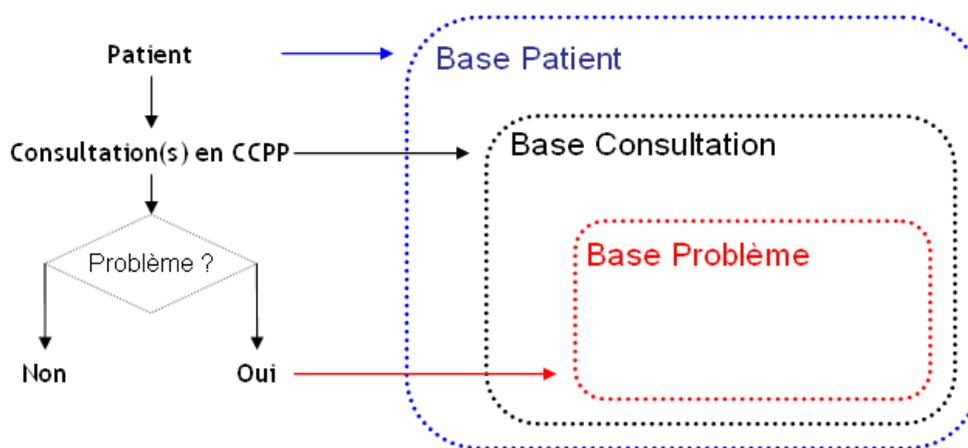


Figure 3 : Alimentation des bases de données locales.

2.1.2.1. Principales étapes de l'interrogatoire professionnel

L'objectif de l'interrogatoire professionnel est de rechercher et d'identifier une éventuelle étiologie professionnelle potentiellement responsable de la maladie diagnostiquée. Cet entretien n'est pas soumis à une méthodologie standardisée de recueil des informations, c'est pourquoi la description suivante est une réflexion sur les étapes clés dans le processus d'acquisition des informations.

L'interrogatoire professionnel constitue le point de départ de la création d'un PST.

ETAPE 1 : Reconstitution du parcours professionnel.

Le médecin consultant s'intéresse en premier lieu à toutes les entreprises dans lesquelles le patient a exercé son activité professionnelle. Il identifie et date de manière précise les postes de travail et les tâches réalisées au sein des entreprises. Il note également les changements de poste et des conditions de travail. La reconstitution du parcours professionnel est plus difficile si les faits sont anciens. L'entretien peut se compliquer notamment si la mémoire du patient est altérée à cause de son état de santé.

Par exemple, un patient atteint d'un glioblastome, présentant des pertes de mémoire et des troubles de la parole, a souhaité être accompagné de son épouse et d'un ancien salarié du site afin de mieux préciser certaines expositions.

ETAPE 2 : Recherche des nuisances dans le cadre de l'activité professionnelle.

A partir du parcours professionnel détaillé, le médecin procède à l'inventaire des nuisances professionnelles et identifie les processus de travail, comme les postures et gestes effectués, ainsi que les agents chimiques, physiques, biologiques et psychosociaux potentiellement en cause dans la survenue de la maladie.

Parallèlement à la recherche des nuisances, le médecin identifie les conditions d'hygiène et sécurité (moyens de protection collective et individuelle) mis en place. Ces informations visent à estimer l'exposition professionnelle et donc les niveaux de risques probablement encourus. Le médecin s'intéresse également aux activités de voisinage dans les mêmes ateliers qui peuvent générer des nuisances ; par exemple la diffusion de poussières ou de gaz sur un poste peut constituer une nuisance pour les postes de travail à proximité.

ETAPE 3 : Recherche de facteurs extra-professionnels.

La présence d'un facteur étiologique non professionnel n'élimine pas la recherche des expositions professionnelles potentiellement étiologiques, d'autant qu'il peut y avoir potentialisation des effets. Par exemple, les expositions au tabac et à l'amiante potentialisent le risque de survenue d'un cancer broncho-pulmonaire. Le médecin consultant interroge donc le patient sur ses activités extra-professionnelles (par exemple travaux de rénovation de son habitat) et loisirs (sports, jardinage) afin de n'écarter aucune piste de recherche.

ETAPE 4 : Chronologie des signes par l'activité professionnelle.

Le médecin consultant compare ensuite la chronologie des expositions aux nuisances professionnelles avec la chronologie de l'apparition ou de la disparition des signes de la maladie. L'établissement de l'imputabilité est plus facile s'il existe un délai court entre l'exposition à la nuisance et la survenue de la maladie. Il est possible de mener une étude plus approfondie du cas, en analysant par exemple les effets liés à la suppression (week ends et vacances) et la réadmission du patient au poste de travail pour des pathologies irritatives ou allergiques. L'asthme et les dermites se prêtent tout à fait à cette approche car on note une amélioration ou une disparition des signes cliniques après la cessation d'activité pendant plusieurs jours et la réapparition des symptômes lors de la nouvelle exposition professionnelle.

ETAPE 5 : Recherche des co-morbidités.

La recherche des co-morbidités constitue des arguments complémentaires sans pour autant prouver la relation causale. Il faut alors compléter l'interrogatoire professionnel individuel par des données provenant du médecin du travail au sein des entreprises dans lesquelles le patient a travaillé. Le médecin peut également demander la réalisation d'exams complémentaires comme des tests d'allergies ou encore un examen respiratoire fonctionnel.

L'interrogatoire professionnel est un mélange complexe entre objectivité et subjectivité. En effet, on peut parler de subjectivité pour les capacités d'observation et d'écoute du médecin vis-à-vis du patient, et d'objectivité pour les examens cliniques effectués lors de la consultation ou les examens complémentaires. Ainsi, plusieurs médecins de plusieurs centres de consultation n'enregistrent pas le PST de la même manière, tant au niveau des conclusions de l'interrogatoire qu'au niveau du codage des expositions retenues. L'interrogatoire professionnel constitue donc une étape fondamentale contenant une grande variabilité dans le recueil des données.

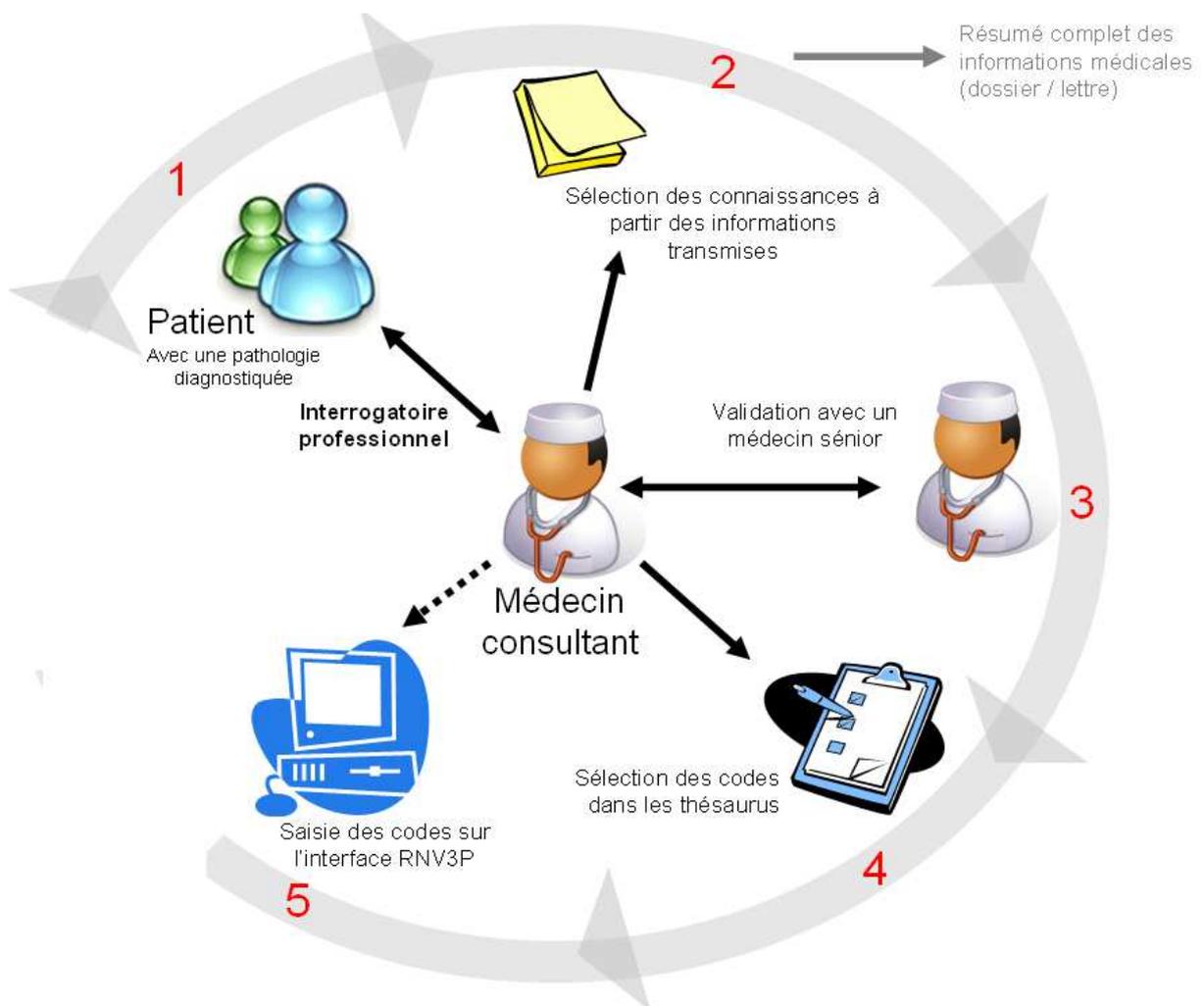


Figure 4 : Schéma de l'acquisition des informations à la production des connaissances.

A l'issue de la consultation, le médecin consultant remplit un bordereau "Problème" (Annexe C-3) et valide ses conclusions avec un médecin sénior. Puis, les codes sont sélectionnés dans les thésaurus et le PST est saisi dans l'interface du RNV3P.

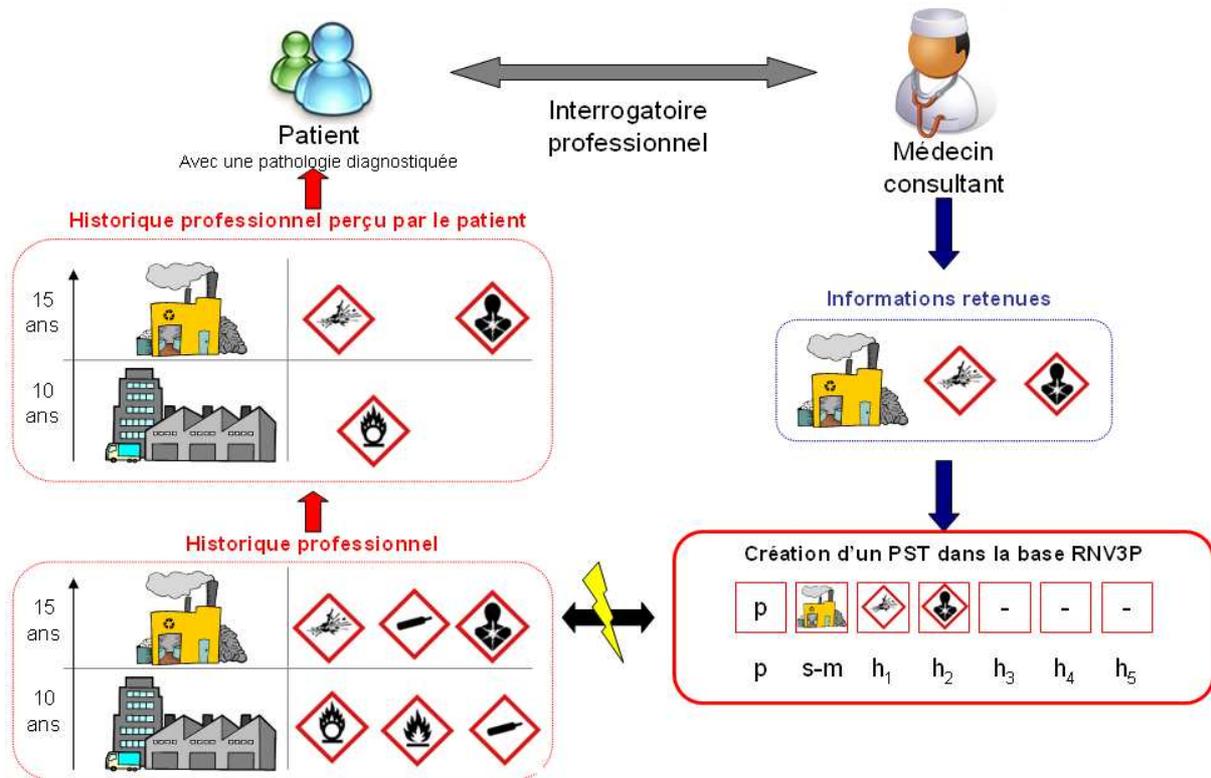


Figure 5 : Schéma illustratif des informations retenues à la suite de l'interrogatoire professionnel.

2.1.2.2. Codage et exportation des données locales au niveau national

Les informations recueillies lors des consultations de pathologies professionnelles sont enregistrées dans des bases de données locales sous la forme de Problèmes de Santé au Travail, puis anonymisées et exportées une fois par an dans la base de données nationale. La base "Problème" stocke ainsi des Problèmes de Santé au Travail (PST), définis comme association entre une pathologie et une exposition professionnelle composite dont une ou plusieurs composantes ont un rôle potentiellement étiologique.

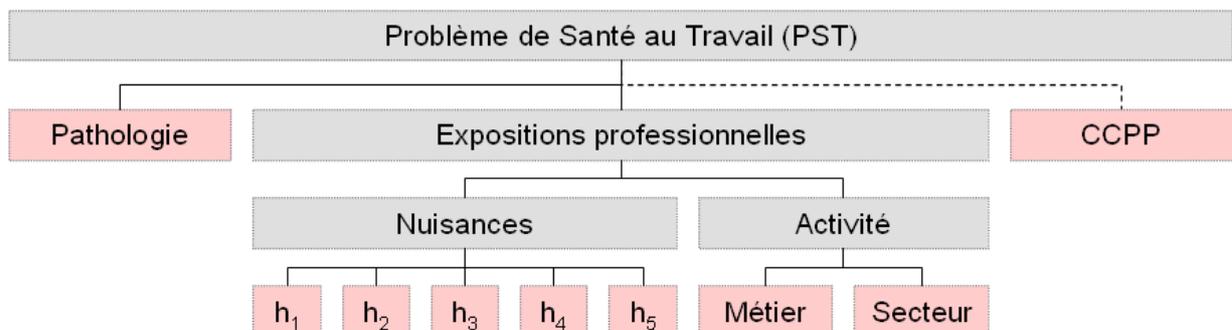


Figure 6 : Structure d'un Problème de Santé au Travail.

2.1.2.3. Procédures qualité

Des procédures qualité sont appliquées à la base de données afin de vérifier la cohérence des codages et des fiches saisies pour assurer la qualité des données extraites. Le volume total de la base RNV3P compte 136 766 PST de nature différente, rapportés de manière systématique et standardisée, au cours de la période 2001-2009.

La nature du PST est apportée par le médecin à la suite de l'interrogatoire professionnel et définit le cadre dans lequel ce problème sera archivé et exploité sur le plan statistique. Notre étude porte sur la base de données RNV3P des émergences, c'est-à-dire sur les données hors Service de Santé au Travail (SST), comprenant les PST de nature pathologie professionnelle, pathologie environnementale, pathologie ni professionnelle ni environnementale et les interrogatoires professionnels, soit un sous-ensemble du RNV3P, composé de 81 132 PST dont la répartition des natures est détaillée Figure 7. Un PST de nature pathologie professionnelle est une association dont le lien entre la pathologie et l'environnement de travail du patient est suspecté. Un PST de nature interrogatoire professionnel est une association dont la pathologie est connue mais pour laquelle un inventaire des expositions professionnelles est réalisé dans une démarche systématique. Un PST de nature pathologie environnementale est une association dont l'origine de la pathologie est suspectée être une nuisance environnementale, hors des expositions sur le lieu de travail. Un PST de nature pathologie ni professionnelle ni environnementale est une association pour laquelle le médecin n'a pas mis en évidence une relation causale entre la pathologie et les nuisances professionnelles ou environnementales connues.

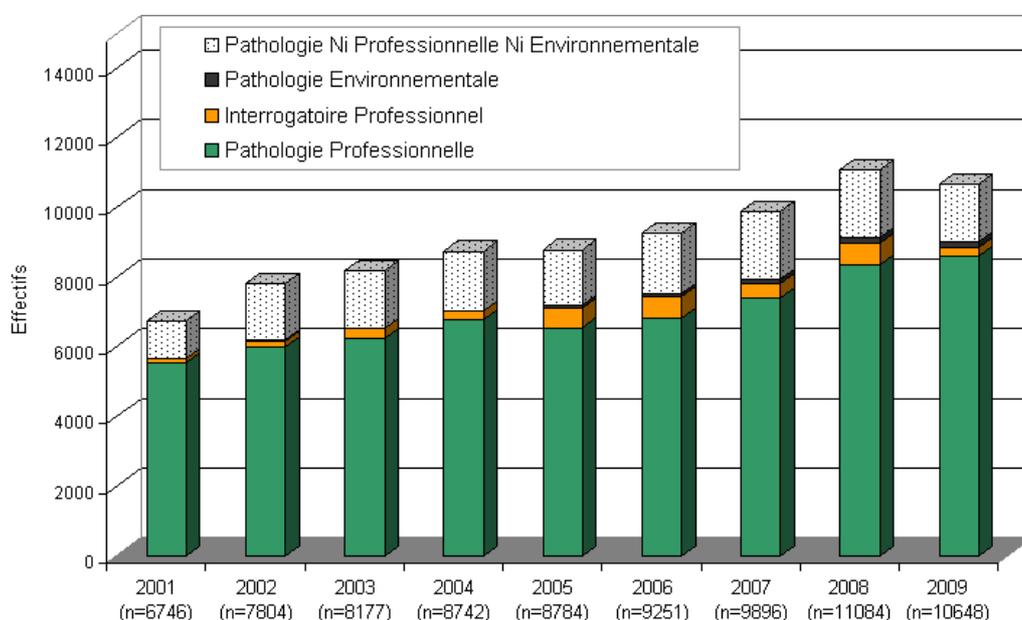


Figure 7 : Répartition des natures des Problèmes de Santé au Travail
(Données émergence RNV3P 2001-2009 ; N = 81 132 PST).

2.2. Travaux de recherche en émergence

Dans un contexte de veille sanitaire et avec l'enrichissement croissant de la base de données depuis 2001, les données font l'objet d'une veille statistique annuelle qui consiste à caractériser la provenance des PST enregistrés et à décrire les principales variables d'intérêt (pathologie, nuisances, métier et secteur d'activité) en fonction du temps, tant au niveau local que national.

Ces connaissances représentent un réservoir riche d'informations relatives à des pathologies potentiellement d'origine professionnelle et à l'environnement du travail composé d'expositions multiples. Des problématiques se sont alors posées pour acquérir un rôle de vigilance et d'alerte vis-à-vis des expositions professionnelles susceptibles de représenter un risque pour la santé.

2.2.1. Application des méthodes de pharmacovigilance

Ce travail est le premier volet des travaux de recherche appliqués sur les données du RNV3P dans le cadre d'une thèse de science [Bonneterre 2006-2010].

2.2.1.1. Objectif

L'objectif de ce travail était d'explorer et de tester les méthodes de pharmacovigilance qui semblaient pertinentes pour la surveillance temporelle des PST, et de contribuer à la mise en place de méthodes de veille sanitaire. La finalité de ces travaux était de soumettre au groupe de travail, au comité de pilotage et au comité scientifique du RNV3P, des propositions visant à renforcer la performance du système de détection des couples pathologie - exposition potentiellement émergents, pour lesquels la littérature est silencieuse, ou non concluante, et qui semblent justifier d'une exploration complémentaire grâce aux données de la base RNV3P.

2.2.1.2. Méthodologie

Couramment utilisées pour l'analyse des bases des enregistrements spontanés d'effets indésirables, les méthodes de pharmacovigilance (Annexe A-2) ont été appliquées aux données du RNV3P. Ces méthodes reposent sur des mesures de disproportions. Après comparaison de la structure des bases de données de pharmacovigilance et du RNV3P, sept méthodes ont été testées:

- Méthodes classiques et fréquentistes (Chi2, PRR, ROR, Yules),
- méthodes basées sur la loi de Poisson (Poisson et SPRT2),
- Méthodes bayésiennes décrites à ce jour (BCPNN)

La méthode PRR - Proportional Reporting Ratio [Wilson 2004] avait été retenue à la suite de comparaison avec les autres méthodes, et la démarche proposée était la suivante :

- 1) Détection de signaux selon la méthode PRR,
- 2) Tri automatique éliminant parmi les signaux générés ceux ayant droit à une reconnaissance en maladie professionnelle, traduisant ainsi des couples connus,
- 3) Investigation de chaque couple maladie - exposition non connu et générant un signal par investigation des données internes à la base (nombre de cas, nombre de centres distincts dont proviennent les cas, évolution temporelle, distribution des imputabilités),
- 4) Analyse des données de la bibliographie en distinguant les données humaines d'une part (données étiologiques et rapports de cas), et les données expérimentales d'autre part (toxicologie).

2.2.1.3. Résultats obtenus

La méthode PRR avait été appliquée aux données du RNV3P, pour les couples rapportés au moins 2 fois dans la base de données, et notamment à l'étude des Sclérodermies Systémiques (code M34) enregistrés dans la base RNV3P au cours de la période 2002-2007.

2.2.1.4. Conclusion générale

D'une part, les méthodes de pharmacovigilance sont appliquées à l'étude de la mono-exposition des données du RNV3P, c'est-à-dire aux couples pathologie - nuisance, pathologie - métier et pathologie - secteur d'activité. D'autre part, la génération d'un signal avec les méthodes de pharmacovigilance repose successivement sur le recrutement de plusieurs cas identiques, le codage homogène, l'aptitude des couples à générer un signal, et enfin un niveau informatif résiduel suffisamment explicite des enregistrements pour que ce signal puisse retenir l'intérêt des experts en charge de son investigation. Si l'automatisation de ces méthodes est possible, un accompagnement humain est toutefois nécessaire tant en amont (sur le codage) qu'en aval de ces méthodes (interprétation des signaux).

2.2.2. Veille des émergences cliniques

Face aux pertes d'informations engendrées par le codage des variables constituant un PST, un système complémentaire de mise en évidence de pathologies professionnelles potentiellement émergentes a été mis en place. Ce travail de veille clinique constitue le deuxième volet des travaux de recherche appliqués aux données du RNV3P et complète la première approche car elle permet de traiter spécifiquement les cas observés uniques (fréquence supérieure ou égale à 1).

2.2.2.1. Objectif

L'objectif de ce travail est de développer une procédure rapide et structurée, à l'échelle nationale, destinée aux cliniciens spécialistes de la médecine du travail. Cette procédure vise à prendre en compte les cas cliniques émergents directement identifiés au cours des consultations auprès des CCPP, en exploitant la richesse de l'information avant qu'elle ne soit résumée, voire même altérée par le codage.

2.2.2.2. Méthodologie

Cette procédure s'appuie sur l'expertise clinique des médecins du RNV3P. Or, les données d'exposition collectées et disponibles dans les dossiers de consultation, ne sont à ce jour pas directement accessibles dans les enregistrements du RNV3P (durées, fréquences et intensité d'exposition) et les codes nuisances, métier et secteur d'activité ont leurs propres limites.

Pour ces différentes raisons, il a donc semblé utile qu'un clinicien identifiant au moins un cas nouveau, a priori jamais décrit, puisse rapidement remonter cette information afin qu'elle soit gérée selon une procédure structurée, visant à conduire des actions d'alerte sanitaire puis de prévention appropriées. Pour ce faire, un groupe de travail dédié à l'émergence clinique a été initié mettant en place une procédure d'investigation des cas aujourd'hui validée par le comité de pilotage du RNV3P.

2.2.2.3. Résultats obtenus

La veille des émergences cliniques est illustrée avec l'investigation de pneumopathie d'hypersensibilité chez une prothésiste ongulaire soumise à la Cellule de Traitement des Alertes (CTA) et exposée aux méthacrylates, en particulier au méthacrylate d'éthyle.

2.2.2.4. Conclusion générale

La veille des émergences cliniques est bien adaptée à un réseau d'experts. Elle contourne l'ensemble des limites inhérentes au codage et offre la possibilité d'utiliser directement la richesse d'un dossier patient. En revanche, cette procédure présente une faible sensibilité puisque qu'elle est basée sur une hypothèse formulée par un seul médecin.

* * * * *

Ces travaux de recherche ont mis en évidence la capacité du RNV3P à décrire des situations professionnelles à risque grâce aux expertises médicales fournies par les médecins à l'issue des interrogatoires professionnels, et à utiliser les connaissances à disposition pour la surveillance et la détection de couples mono-exposés potentiellement émergents.

Résumé

2. Présentation du RNV3P

Le Réseau National de Vigilance et de Prévention des Pathologies Professionnelles (RNV3P) est un réseau d'experts hospitalo-universitaires qui enregistrent, dans une base de données nationale, des Problèmes de Santé au Travail (PST). Un PST est enregistré à la suite d'un interrogatoire professionnel entre le patient et le médecin consultant, et est défini comme association entre une pathologie et une exposition professionnelle composite, dont une ou plusieurs composantes ont un rôle potentiellement étiologique (une à cinq nuisances, un métier et un secteur d'activité).

Les missions du RNV3P s'articulent autour de la surveillance et de la détection des PST potentiellement émergents dans le but de générer des alertes sanitaires. Des travaux de recherche ont déjà été menés sur cette base de données dans le but de surveiller et de détecter des couples pathologie - nuisance, pathologie - métier et pathologie - secteur d'activité potentiellement émergents :

- La méthode PRR issue des méthodes de pharmacovigilance,
- La veille des émergences cliniques qui fait appel à l'expertise clinique des médecins du RNV3P.

3. Définition du concept d'émergence au sens du RNV3P

L'utilisation de méthodes statistiques pour la détection des risques émergents exige des définitions précises de ce qui doit être détecté. L'épidémiologiste doit donc veiller à l'adéquation entre la question posée et la méthode statistique proposée. Ce concept d'émergence se décline sous une infinité de questions liées aux échelles spatiale, temporelle ou spatio-temporelle, le plus souvent utilisé pour désigner des maladies infectieuses:

"Emerging infectious diseases are those due to newly identified and previously unknown infections which cause public health problems either locally or internationally" (OMS 1997)

"Infection nouvelle causée par l'évolution ou la modification d'un agent pathogène ou d'un parasite existant" (OIE 2006)

3.1. Le concept d'émergence dans la littérature

Une préoccupation majeure en santé publique est d'anticiper et de prévenir les pathologies. Dans cet objectif, le concept d'émergence tient une place centrale, dont la position ne cesse de prendre de l'importance avec l'augmentation des maladies nouvelles de notre siècle. La littérature scientifique fournit ainsi de nombreuses définitions de l'émergence désignant en général les maladies infectieuses émergentes. Dans certains cas, on parle même de "flambées épidémiques" pour caractériser l'émergence et la propagation d'une infection transmissible. Une maladie infectieuse émergente est traditionnellement définie comme :

"Infections that have newly appeared in a population or have existed but are rapidly increasing in incidence or geographic range" [Morse 1995 ; Morse 2004].

Mais le concept d'émergence ne se limite pas seulement aux maladies infectieuses. Plus généralement, Toma et al. définissent une maladie émergente comme:

"Une maladie dont l'incidence réelle augmente de manière significative dans une population donnée, d'une région donnée et durant une période donnée, par rapport à la situation épidémiologique habituelle de cette maladie" [Toma 2003].

Ces définitions incluent les notions de survenue nouvelle d'une pathologie déjà connue par le passé, ainsi que d'une augmentation de l'incidence dans une période et un espace géographique donné.

L'historique de la pathologie représente ainsi une composante essentielle dans la définition d'une maladie émergente. Les auteurs commentent et nuancent cette définition avec les remarques suivantes :

- Les maladies émergentes ne se limitent pas aux maladies transmissibles et infectieuses,
- Donner le qualificatif d'émergent à une maladie nouvelle dont l'incidence n'augmenterait pas significativement n'est pas pertinent,
- Le "risque émergentiel" (ou potentielle émergence) ne doit pas être considéré comme une véritable émergence du fait de l'incertitude de son évolution future, mais plutôt comme potentiellement émergent.

Aujourd'hui, des pathologies multifactorielles se développent et sont qualifiées de pathologies émergentes, généralement connues pour être en relation avec un risque, en rapport ou non avec l'environnement. Ne peut-on pas alors parler de risque émergent au même titre que de pathologies émergentes ?

3.1.1. Un risque sanitaire nouveau

Un risque sanitaire se définit comme l'exposition à un danger susceptible d'entraîner des conséquences néfastes sur la santé. Un risque sanitaire est dit nouveau si les conséquences d'une exposition, plus ou moins prolongée, entraînent des effets néfastes sur la santé, non connus historiquement. Un risque sanitaire nouveau peut aussi représenter un danger connu, considéré comme nouveau en fonction de l'évolution des connaissances et des preuves scientifiques.

Dans le cadre de la santé au travail, un risque nouveau est généralement dû à de nouveaux lieux de travail, des changements socio-organisationnels ou provoqué par l'introduction de nouvelles technologies ou procédés. Par exemple, l'inhalation des fibres d'amiante constitue, en 1906, un risque sanitaire nouveau, mis en évidence par une forte mortalité des ouvriers dans les filatures et dans les usines de tissage d'amiante près de Caen [Goldberg 2007].

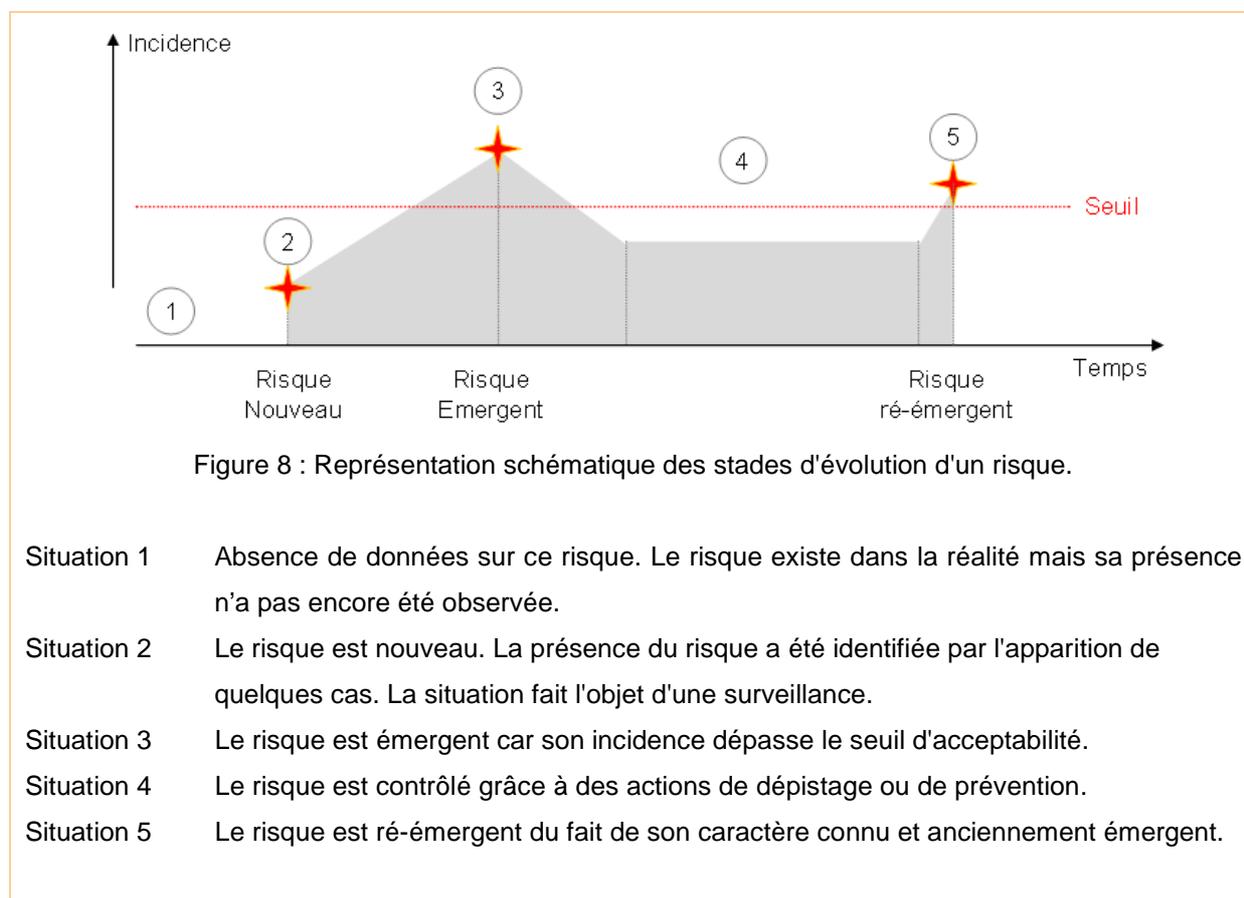
3.1.2. Un risque sanitaire émergent

Dans le domaine des maladies animales, on parle de "réussite émergentielle" pour définir l'aptitude d'une pathologie à se propager dans le temps et l'espace [Chastel 2000]. La notion de risque sanitaire émergent se distingue du risque nouveau du fait qu'il s'agit d'un risque déjà identifié dans le passé, et lorsqu'il dépasse un seuil d'acceptabilité [Jones 2008].

Dans l'exemple précédent, à partir du moment où l'augmentation des cas de mésothéliomes a dépassé un certain seuil de détection, le risque sanitaire lié à l'inhalation des fibres d'amiante, qualifié de nouveau à cette période, évolue en risque sanitaire émergent du fait du grand nombre de personnes exposées et présentant des signes cliniques de la pathologie.

3.1.3. Risque sanitaire ré-émergent ou résurgent

Le terme ré-émergent est le plus souvent employé pour les épidémies infectieuses ; par exemple les épidémies de choléra. On peut cependant qualifier le risque sanitaire ré-émergent comme risque émergent historiquement mais dont l'incidence augmente à nouveau de manière significative [Campos 1997].



Les maladies ré-émergentes se distinguent des maladies récurrentes par le délai d'apparition entre deux épisodes. Par exemple, la tuberculose est une maladie ré-émergente depuis quelques années alors qu'on n'observait que quelques cas sporadiques. La dengue est une maladie récurrente du fait de son apparition annuelle et prévisible.

3.1.4. Une évolution incertaine du risque

Les dispositifs d'alerte sont capables de détecter le plus tôt possible des pathologies nouvelles, émergentes ou ré-émergentes présentant un risque potentiel et non négligeable pour la santé des populations. La finalité de tels dispositifs est de produire une réponse rapide, par exemple sous la forme de mesures correctives de protection ou de prévention, afin de réduire les indicateurs de mortalité et/ou morbidité. Cependant, de tels dispositifs fonctionnent généralement sur des estimations à partir de connaissances empiriques qui permettent de modéliser la tendance d'un événement sanitaire, induisant une évolution incertaine de l'évolution du contexte général de l'événement. Dans ce cas, on parle de risque potentiellement nouveau ou potentiellement émergent.

3.2. Le concept d'émergence au sens du RNV3P

Par construction, la base de données du RNV3P n'a pas pour mission de mettre en évidence de nouvelles pathologies ou de nouvelles expositions professionnelles. En effet, l'enregistrement de ces variables dépend des référentiels de codages (thésaurus). Par conséquent, les pathologies et expositions professionnelles de la base, forcément référencées dans les thésaurus, sont connues.

3.2.1. Récurrence dans l'utilisation des codes

Les choix des médecins consultants sont discutés puis validés avec un médecin sénior dans 70% des cas (soit 50 612 / 81 132 PST). Les informations retenues sont ensuite formalisées en termes de connaissances, et enregistrées dans la base de données RNV3P sous la forme de codes afin d'être statistiquement exploitables.

Par convention, les degrés de précision retenus sont au minimum 3 digits pour les codes pathologies (CIM-10), 5 digits pour les nuisances (CNAM-TS), 2 digits pour les codes métier (CITP) et 2 digits pour les codes secteur d'activité (NAF). Les codes disponibles sont répertoriés dans les thésaurus ; qu'on peut qualifier comme "ensemble des possibles", caractérisé par un espace à 4 dimensions définies par les variables pathologie, nuisances, métier et secteur d'activité.

Une simple description de l'utilisation des thésaurus (Figure 9) met en évidence les utilisations des codages. Dans l'ensemble, on observe une large utilisation des codes relatifs à l'activité professionnelle, avec des codes récurrents d'une année sur l'autre, ce qui induit une certaine stabilité dans la répartition des métiers et secteurs d'activité des patients.

On note par exemple que 51% des codes pathologies ont été utilisés au moins une fois, et près de 8% des codes pathologies disponibles dans le thésaurus sont des codes récurrents, c'est-à-dire qu'ils apparaissent chaque année au moins une fois. Cependant, le thésaurus des pathologies est une classification internationale et est donc peu représentative des pathologies d'origine professionnelle. Concernant les codes nuisances, 20% des codes sont utilisés au moins une fois depuis 2001 et seulement 4% des codes sont récurrents.

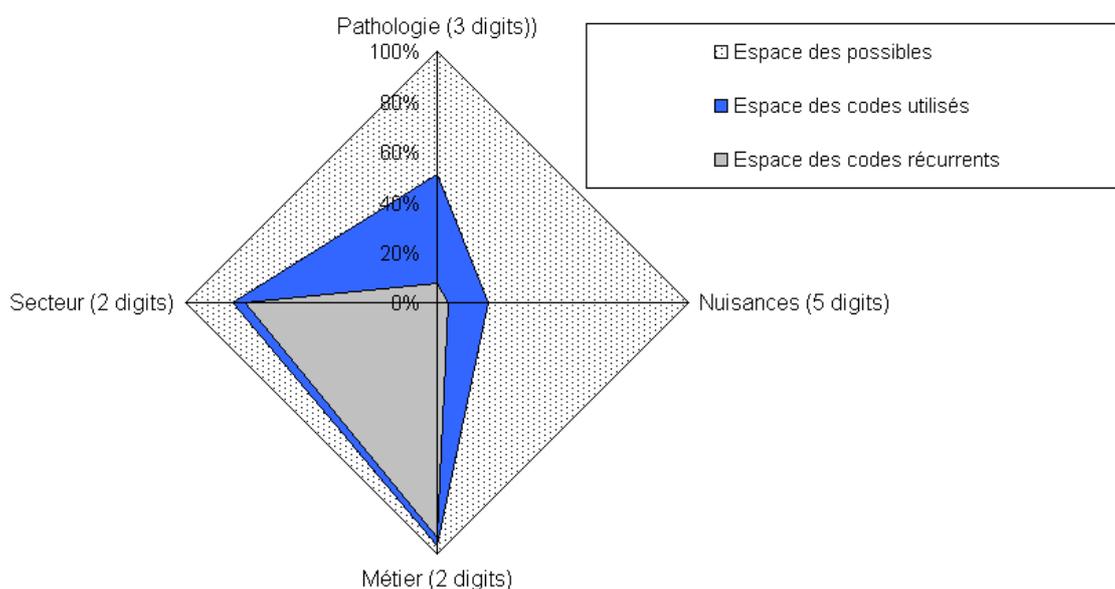


Figure 9 : Répartition de l'espace des codes utilisés et des codes par rapport à l'espace des possibles (Données émergence RNV3P 2001-2009 ; N = 81 132 PST).

Les codes récurrents tiennent une place importante parce qu'ils traduisent une habitude de codage, due à un certain nombre de paramètres, comme par exemple le recrutement local des patients ou encore la spécialité du centre. Par exemple, la récurrence d'un code métier peut signifier que les enregistrements des PST relatifs à ce métier sont régulièrement observés au niveau des CCPP, comparable à une situation endémique. A l'inverse, un code métier qui disparaît peut caractériser un changement du tissu socio-économique sur la zone locale autour d'un CCPP ou un changement de pratique de codage de la part du médecin.

3.2.1.1. Le thésaurus des pathologies

Les pathologies sont codées selon la CIM-10, thésaurus international publié par l'Organisation Mondiale de la Santé (OMS) et accessible gratuitement sur le net, pour l'enregistrement des causes de morbidité et de mortalité. La CIM permet le codage des maladies, des traumatismes et de

l'ensemble des motifs de recours aux services de santé. Cette classification est hiérarchisée en 21 chapitres, divisés en catégories et affectées d'un code alphanumérique à 3 digits.

Par exemple, le code J45 est attribué à l'asthme. La majorité des catégories propose un niveau de précision supplémentaire, précisé par un 4^{ème} digit. Par exemple, le code J450 est utilisé pour identifier l'asthme à prédominance allergique. Par convention, les codes pathologies retenus pour notre étude sont tronqués aux trois premiers digits.

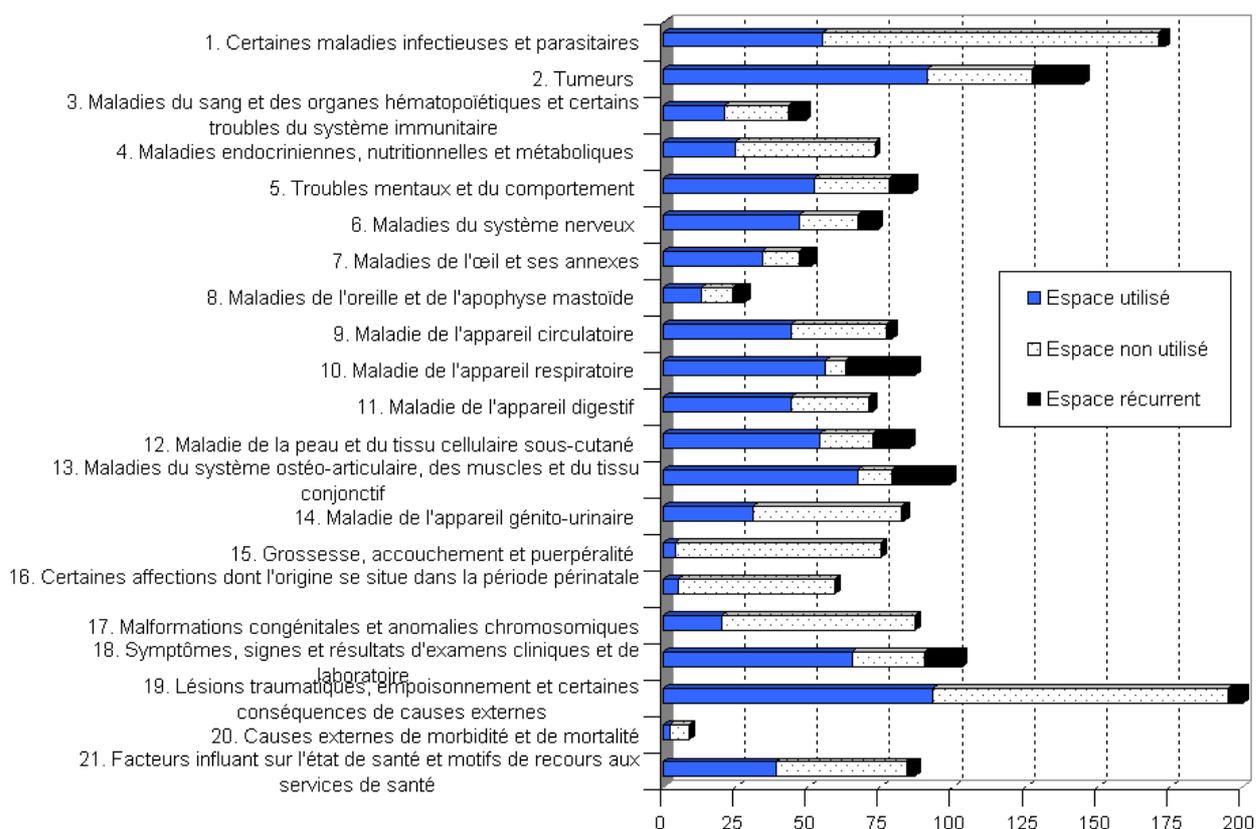


Figure 10 : Répartition de l'espace disponible, utilisé et récurrent des codes pathologie (3 digits)
(Données émergence RNV3P 2001-2009 ; N = 81 132 PST).

3.2.1.2. Le thésaurus des nuisances

Les nuisances sont codées selon le thésaurus établi par la CNAM-TS (Caisse Nationale d'Assurance Maladie des Travailleurs Salariés), accessible par les médecins consultants grâce à une interface access depuis les CCPP. Ce thésaurus est divisé en 14 catégories, classifiant les nuisances en fonction de leur nature et déterminant ainsi la racine du code à 5 digits.

Par exemple, le code 34110 indique que la nuisance est une substance industrielle (3****), faisant partie de la famille des fumées (34***), plus précisément des fumées de soudage (341**) et qu'il s'agit d'arcs électriques (34110).

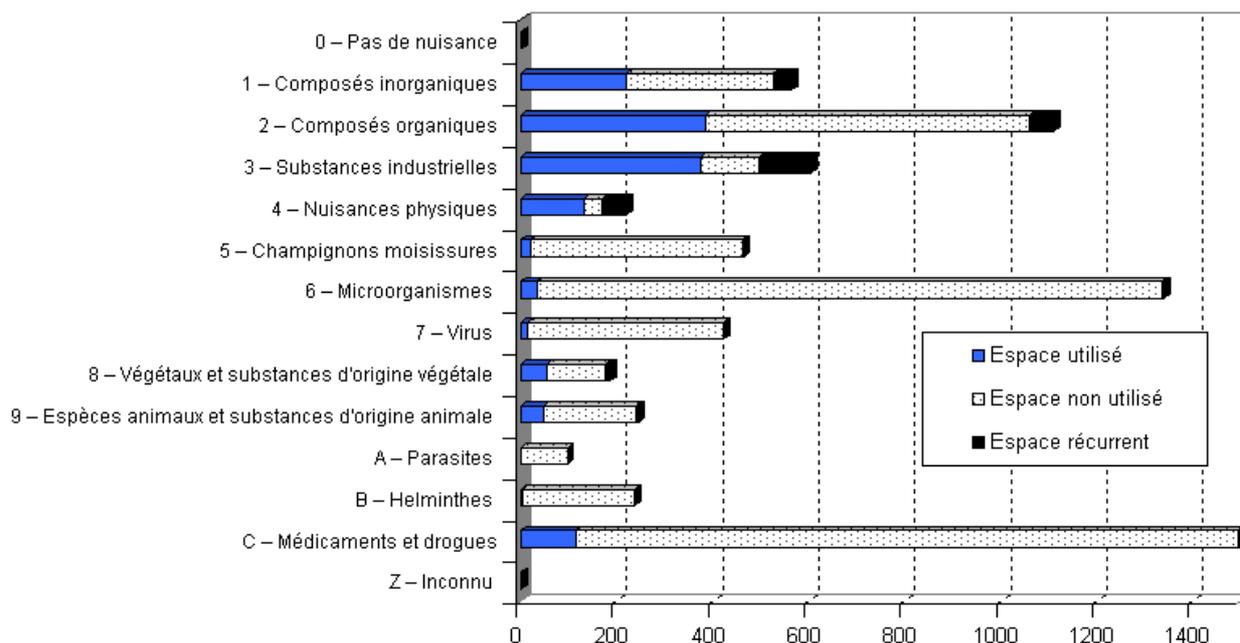


Figure 11 : Répartition de l'espace utilisé, non utilisé et récurrent des codes nuisances (5 digits)
(Données émergence RNV3P 2001-2009 ; N = 81 132 PST).

Certaines catégories de pathologie sont n'ont jamais été rapportées dans la base RNV3P (Figure 11). C'est le cas des catégories des parasites (exemple : leishmania) et des helminthes (exemple : nématodes), très peu présentes dans l'environnement professionnel.

3.2.1.3. Les thésaurus des métiers et secteurs d'activité

La classification CITP (Classification Internationale des Types de Professions) et la Nomenclature des Activités Françaises (NAF) fournissent les codes respectifs des métiers et des secteurs d'activité. Ces thésaurus sont accessibles gratuitement sur le net.

Dans la suite de ce travail, nous ferons référence au couple secteur - métier pour désigner l'activité professionnelle associée à la maladie. En effet, la liaison du métier et du secteur permet de mieux préciser l'activité exercée et donc le type d'expositions professionnelles susceptibles d'être en relation avec la maladie.

Par exemple, le poste de mécanicien et ajusteur de véhicules à moteur (code 7231) n'est pas sujet aux mêmes expositions selon qu'il est rattaché au secteur d'extraction de minerais d'uranium (code 120Z) ou au secteur de l'entretien et des réparations automobiles (502Z).

Alors que la Figure 9 illustre l'utilisation quasi-complète des codes métier et secteur (à 2 digits), la combinaison de ces codes (Figure 12) ne met pas en évidence de codages récurrents du fait de la diversité des métiers et des secteurs d'activité.

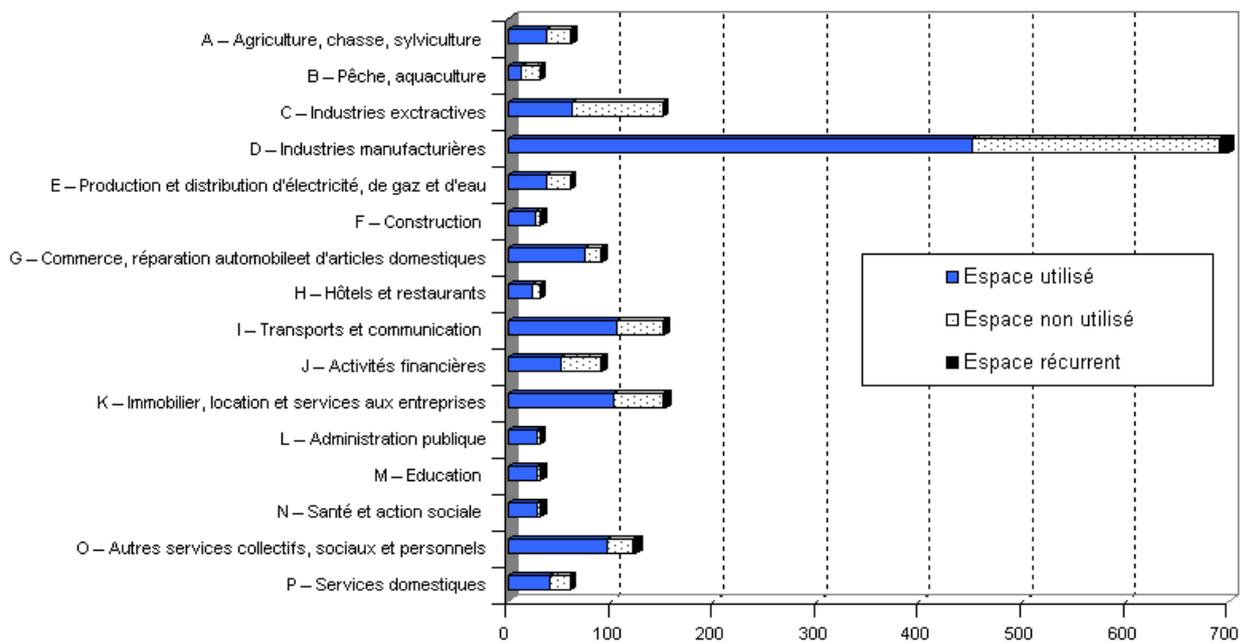


Figure 12 : Répartition de l'espace utilisé, non utilisé et récurrent des associations secteur-métier supposées responsables dans la survenue de la pathologie (2 x 2 digits) classés par catégorie de secteur d'activité (Données émergence RNV3P 2001-2009 ; N = 81 132 PST).

3.2.2. Le cumul des connaissances

La base RNV3P représente un entrepôt de données nationales qui, depuis 2001, s'enrichit des connaissances enregistrées par les médecins. L'application de méthode conventionnelle pour la surveillance épidémiologique consisterait à suivre l'évolution de la fréquence des pathologies enregistrées, et donc de l'utilisation des codes par les médecins consultants.

La Figure 13 illustre l'évolution de 4 catégories de pathologies, désignées par des codes couleurs, qui représentent 65% des PST enregistrés (soit 52 625 / 81 132 PST) depuis 2001. L'histogramme représente les effectifs annuels de PST enregistrés pour chacune de ces pathologies (échelle de gauche). Les droites représentent la taille du spectre des nuisances, ou nombre cumulé de nuisances distinctes associées à la pathologie (échelle de droite).

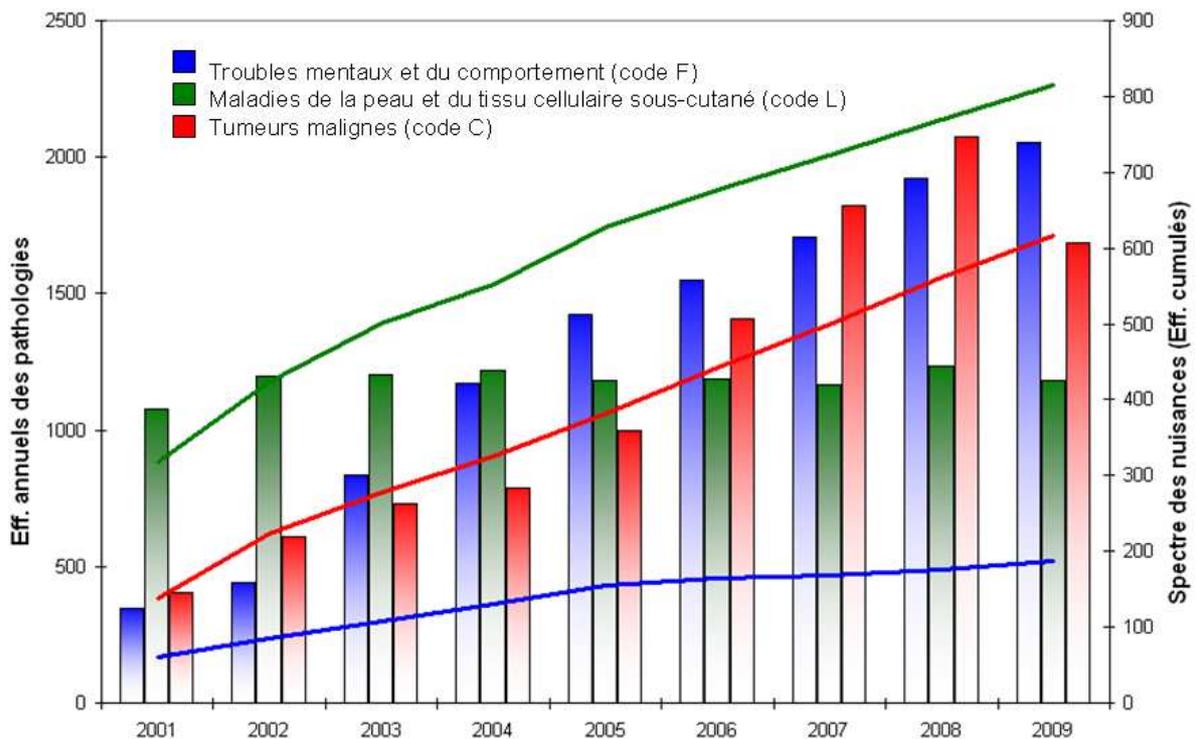


Figure 13 : Dynamique temporelle des effectifs des troubles mentaux et du comportement (code F), des maladies de la peau et du tissu cellulaire sous-cutané (code L) et des tumeurs malignes (code C) (Données émergence RNV3P 2001-2009 ; N = 81 132 PST).

Les histogrammes des troubles mentaux et du comportement (en bleu) et des tumeurs malignes (en rouge) sont caractérisés par une forme "épidémique" du fait de l'augmentation constante des cas enregistrés depuis 2001. Cependant, le nombre de nuisances associées à ces pathologies n'évolue pas de la même manière ; pour les troubles mentaux, la taille du spectre des nuisances associées stagne depuis 2005 alors que le nombre de nuisances associées aux tumeurs malignes est en évolution. Les effectifs annuels des maladies de la peau et du tissu cellulaire sous-cutané (en vert) sont relativement stables ; 1 200 PST enregistrés en moyenne chaque année. On peut qualifier son allure d' "endémique". Les connaissances associées à cette pathologie ne cessent de s'accroître depuis 2001.

Cette figure illustre trois situations observées dans le RNV3P et montre que l'augmentation des effectifs annuels de cas ne traduit pas forcément un apport de nouvelles connaissances sur l'exposition professionnelle associée à la pathologie. Le concept d'émergence ne peut donc pas se restreindre à la surveillance des fréquences d'enregistrements des cas mais doit prendre en compte la surveillance des connaissances nouvelles rapportées par les médecins.

3.3. Définition de l'émergence au sens du RNV3P

Dans le cadre des méthodes de pharmacovigilance appliquées au RNV3P, un couple pathologie - nuisance est potentiellement émergent si sa fréquence est *anormalement* élevée, c'est-à-dire si $LI_{95}(PRR) > 1$, et si cette association n'est pas mentionnée au titre des tableaux de maladies professionnelles indemnisables. En élargissant cette définition au cas de la multi-exposition, nous pouvons écrire que le terme d'association pathologie - expositions composites potentiellement nouvelle ou potentiellement émergente s'applique donc au contexte de la base, nous autorisant à émettre des hypothèses sur l'émergence probable d'une association entre une pathologie et des expositions professionnelles simples (mono-exposition) ou multiples (multi-exposition).

Sur la base de cette réflexion, nous formulons les définitions suivantes :

- ⇒ Une **association potentiellement émergente** est une association déjà existante dans la base RNV3P dont la fréquence d'apparition est *anormalement* élevée, c'est-à-dire supérieure à un critère défini.

- ⇒ Une **association potentiellement nouvelle** est une association jamais observée dans la base du RNV3P mais dont la probabilité d'apparaître est supérieure à un seuil défini.

Précisons toutefois que la définition que nous avons formulé mentionne une fréquence d'apparition anormale et non pas une augmentation anormale de la fréquence. Cette nuance tient compte des contraintes inhérentes à la base de données du RNV3P et implique également des conséquences quant au développement de la méthodologie.

Résumé

3. Définition du concept d'émergence au sens du RNV3P

La littérature scientifique offre de nombreuses définitions de l'émergence, très souvent employées pour désigner les maladies infectieuses émergentes, du fait de leur capacité à se propager rapidement dans le temps et dans l'espace. Pour les pathologies dont le délai de latence est long, comme pour les cancers d'origine professionnelle, où la pathologie apparaît plusieurs années après l'exposition, le concept d'émergence nécessite d'être clairement défini.

Pour ce faire, nous avons investigué dans un premier temps les différents concepts de l'émergence dans la littérature scientifique qui désigne, en règle générale, une augmentation significative d'une mesure de l'état de santé d'une population, comme l'incidence, dans une période et un espace définis.

Dans la base RNV3P, l'utilisation des codes décrivant les principales variables d'intérêt est facilement quantifiable en procédant à un inventaire systématique. L'apparition d'un code qui n'a jamais été rapporté auparavant ne signifie pas pour autant l'apparition d'une association potentiellement émergente. Nous avons ainsi apporté une définition statistique du concept d'émergence et de potentielle émergence au sens des données du RNV3P :

- Une association potentiellement émergente est une association déjà existante dans la base RNV3P dont la fréquence d'apparition est supérieure à un critère défini.
- Une association potentiellement nouvelle est une association jamais observée dans la base du RNV3P mais dont la probabilité d'apparaître est supérieure à un seuil défini.

4. De la mono à la multi-exposition

Dans le but de développer une approche pour l'étude de la multi-exposition, il serait instructif de commencer par comprendre comment l'analyse de la mono-exposition a été menée.

4.1. Introduction : la méthode PRR pour l'étude de la mono-exposition

Les méthodes de pharmacovigilance sont basées sur l'enregistrement des effets indésirables des médicaments après leur mise sur le marché, et les scientifiques exploitent les données issues d'enregistrements spontanés afin de générer des alertes sanitaires. Cette méthodologie repose sur le croisement des variables "Médicament" ou "Autres médicaments", et "Effet" ou "Autres effets", dans un tableau de contingence. A l'aide de ce tableau, des mesures de disproportions, comme la méthode PRR, sont calculées afin de détecter la fréquence anormalement élevée des effets provoqués par un médicament. Cette méthode a été appliquée aux données du RNV3P dans le but de détecter des PST potentiellement émergents [Bonneterre 2010], dont voici la méthodologie.

4.1.1. Découpage des associations composites

Comme décrit précédemment, les médecins associent 1 à 5 nuisances à la pathologie et forment ainsi une association composite. A la manière des couples médicament - effet, les séquences composites enregistrées dans la base RNV3P sont découpées en couples pathologie - nuisance, formés à partir des séquences composites.

Par exemple, une séquence composée de 5 nuisances est découpée en 5 couples pathologie - nuisance (Figure 14), formant ainsi 5 pseudo-PST mono-exposés à partir d'un seul PST multi-exposé.

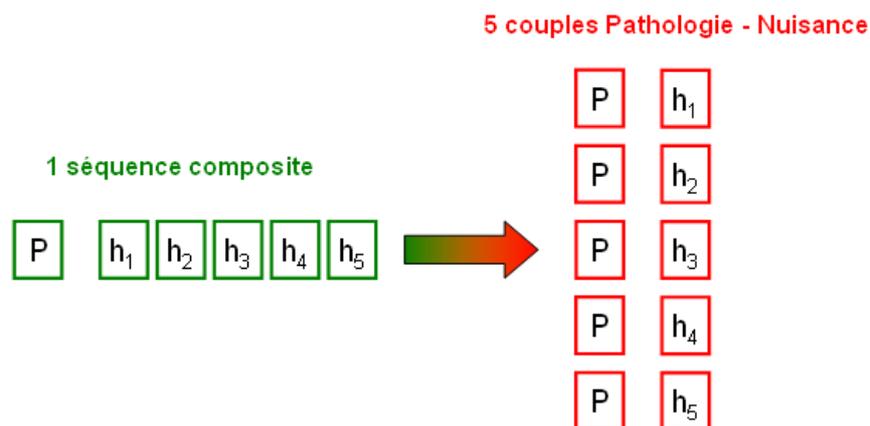


Figure 14 : Représentation schématique du découpage des séquences composites en couples pathologie - nuisance.

4.1.2. Mesure de disproportions

Les couples pathologie - nuisance sont classés dans le tableau de contingence suivant (Tableau 2) où a, b, c et d représentent les effectifs observés dans la base RNV3P, sur le modèle des tableaux de contingence médicament - effet.

	Pathologie	Autres Pathologies
Nuisance	a	b
Autres nuisances	c	d

Tableau 2 : Tableau de contingence 2x2 servant de base pour le calcul des mesures de disproportions utilisées en pharmacovigilance [Bonneterre 2006-2010].

Par exemple, si l'on s'intéresse au couple "sclérodémie systémique - silice", la fréquence d'observation du couple est noté en a, tous les couples "autre pathologie que sclérodémie systémique - silice" sont rapportées en b, toutes les couples "sclérodémie systémique - autre nuisance que silice" sont notées en c, et enfin tous les couples "autre pathologie que sclérodémie systémique - autre nuisance que silice" sont notées en d.

La méthode PRR repose sur le comptage des effectifs d'une nuisance associée à une pathologie par rapport aux autres nuisances. A partir du tableau de contingence, la méthode PRR consiste à comparer la probabilité pour qu'une nuisance soit associée à une pathologie par rapport à la probabilité que d'autres nuisances soient associées à cette pathologie.

$$\begin{cases} PRR = \frac{a/(a+b)}{c/(c+d)} \\ LI_{95\%} = e^{\ln(PRR) - 1.96 \times SE(\ln(PRR))}, SE(\ln(PRR)) = \sqrt{1/a - 1/(a+b) + 1/c + 1/(c+d)} \end{cases} \quad (4.1)$$

Un couple pathologie - nuisance est identifié comme "couple potentiellement émergent" si la borne inférieure de l'intervalle de confiance à 95% du PRR, notée $LI_{95\%}$, est supérieure à 1 et que ce couple n'est pas mentionné au titre des Tableaux MPI.

4.1.3. Résultats

Le Tableau 3 illustre les résultats concernant les couples pathologie - nuisance seulement (hors pathologie - métier et pathologie - secteur) qui émettent un signal valide, c'est-à-dire dont le PRR est significatif et le couple pathologie - nuisance n'est pas mentionné au titre des tableaux MPI.

Ces signaux valides sont ensuite remis aux experts pour évaluation, et sont confrontés à la littérature afin de voir s'il existe des données toxicologiques ou étiologiques en mesure d'appuyer l'hypothèse d'émergence. Un retour aux dossiers médicaux des patients est également nécessaire afin de permettre un retour vers la richesse des informations collectées lors de l'interrogatoire professionnel.

Exposure (INRS/CNAM code)	Systemic sclerosis cases		
	Nb	Nb centre	Years of signal
31160 Silica	38	18	2002-7
22224 Trichloroethylene	4	4	2003-7
31146 Hard metal dusts	3	2	2006-7
35110 Solvents	21	10	2003-7
36R00 Pesticides	3	3	2007
48000 Temperature conditions	5	5	2006-7
49000 Vibrations	3	3	2007

Tableau 3 : Tableau extrait de [Bonneterre 2010].

"Sclérose Systémique - exposition simple", associations générant un signal avec la méthode PRR de pharmacovigilance. Données issues des 30 centres hospitaliers.

Bien que les couples soient composés de nuisances distinctes, certains d'entre eux ont été rapportés pour un patient, décrivant ainsi un PST dans la base. Le principal inconvénient du découpage des séquences composites est la perte de l'information donnée sur les interactions et synergies potentielles entre les nuisances, du fait que celles-ci soient considérées comme des objets distincts.

Par exemple, la séquence "P-h₁-h₂" est découpée en 2 couples distincts "P-h₁" et "P-h₂". Si l'on étudie le couple "P-h₁" (ensemble a dans le tableau de contingence), le fait que le couple "P-h₂" (ensemble c) appartienne à la même séquence n'est pas pris en compte et apparaît même en dénominateur dans le calcul de la statistique.

4.2. Objectifs

Les premiers travaux de recherche appliqués à la base de données du RNV3P sont axés sur la surveillance de la mono-exposition et sur l'expertise des médecins consultants. Parce que **la survenue d'une pathologie ne résulte pas toujours de l'exposition à une seule nuisance**, il serait intéressant d'utiliser les informations existantes dans la base RNV3P, à savoir les facteurs d'exposition composites - 1 à 5 nuisances, métier et secteur d'activité - pour mettre en évidence des associations potentiellement émergentes.

Les objectifs de notre travail consistaient à développer une approche en prenant en compte la multi-exposition et leurs interactions dans la base de données RNV3P pour :

- (i) Surveiller des Problèmes de Santé au Travail
- (ii) Développer un système d'alerte pour les associations potentiellement émergentes.

La finalité de ce travail est d'une part de fournir aux acteurs de la santé publique des outils statistiques capables de suivre, année après année, l'évolution des pathologies et de pointer des associations pathologie-expositions nouvelles et/ou émergentes, en générant des alertes. D'autre part, ce travail doit permettre aux décideurs de s'approprier ces signaux d'émergences afin de mener des actions de prévention adaptées auprès des salariés et des entreprises.

4.3. Défis méthodologiques

Les travaux de recherche en pharmacovigilance et de veille clinique ont mis en évidence un certain nombre de limites inhérentes aux données qui complexifient le développement d'un système pour la surveillance temporelle et la détection d'alerte. Parmi ces limites, nous avons :

4.3.1. La multiplicité des expositions professionnelles

L'environnement professionnel représente une source multi-factorielle d'exposition dont les effets combinés sont peu connus ou peu documentés (effets additifs, antagonistes et synergiques). Sauf exception, comme le mésothéliome causé par l'amiante, les maladies liées à l'activité professionnelle ne sont pas spécifiques d'une exposition particulière, mais résultent de l'interaction entre plusieurs facteurs d'exposition.

C'est par exemple le cas de l'asthme qui, combinée à des facteurs d'exposition professionnelle, peut être révélé ou aggravé par la pollution de l'air. Il existe donc une multiplicité de facteurs pouvant intervenir dans l'apparition ou l'aggravation d'une maladie.

Les médecins ont la possibilité d'associer 1 à 5 nuisances professionnelles à une pathologie. Jusqu'à récemment, les consignes des écoles de codage étaient de simplifier l'information enregistrée dans la base RNV3P mais les recommandations actuelles incitent les médecins à enregistrer un maximum d'informations relatives aux données d'exposition. La multi-exposition professionnelle constitue donc une ressource importante en termes de connaissances et l'exploitation de ces données permettrait d'améliorer le processus de surveillance et de prévention des pathologies professionnelles.

4.3.2. Taille des échantillons

Contrairement aux études épidémiologiques conventionnelles qui exigent une puissance statistique suffisante pour pouvoir démontrer une liaison significative entre une pathologie et une exposition professionnelle, les systèmes de détection d'événements atypiques ne nécessitent pas forcément un nombre considérable de cas. En effet, ce type de méthodes est basé sur l'apparition, ou la réapparition de phénomènes potentiellement à risque. Les faibles effectifs de certaines pathologies de la base RNV3P ne présentent pas un problème, à condition que le système de détection soit adapté à cette particularité. Le volume des connaissances associé à une pathologie a donc plus d'importance que la fréquence des PST.

Par exemple, les cancers de la vessie (code C67) représentent moins de 1% des PST enregistrés depuis 2001 (soit 609 / 81 132 PST). Bien que peu de cas soient observés, on note un nombre conséquent d'expositions (182 nuisances). Au contraire, les plaques pleurales (code J92) représentent environ 10% des PST enregistrés (soit 8 071 / 81 132 PST) et on dénombre moins d'expositions professionnelles associées (124 nuisances dont les fibres d'amiante représentées dans 92% des cas).

4.3.3. Population-mère et population-cible

Les méthodologies de surveillance temporelle et de détection des événements émergents fonctionnent généralement à partir d'un échantillon représentatif d'une population. Une augmentation significative d'un indicateur au cours d'une période donnée caractérise l'apparition d'un événement potentiellement émergent.

Concernant la base RNV3P, les connaissances stockées représentent l'activité locale autour du centre. Comme les patients adressés auprès des CCPP ne représentent pas la population de référence d'où ils sont issus localement, toute étude épidémiologique conventionnelle à partir de ces données ne serait pas pertinente.

De plus, les patients reçus aux CCPP sont majoritairement adressés par les médecins du travail, ainsi que des médecins hospitaliers et médecins de ville, généralistes et spécialistes. Il arrive également que les patients eux mêmes prennent l'initiative de consulter. Ces médecins sentinelles, dont on ne connaît pas les motivations pour adresser un patient en CCPP, constituent un premier filtre difficilement caractérisable entre la population professionnelle et la population adressée en CCPP.

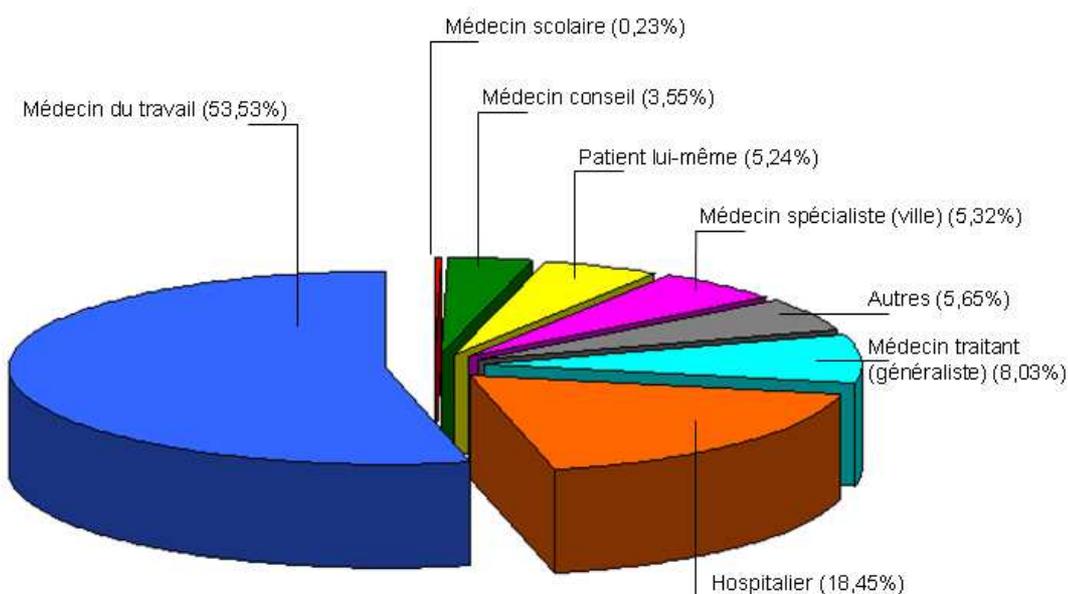


Figure 15 : Origine de la demande de consultation
(Données émergence RNV3P 2001-2009 ; N = 81 132 PST).

Par ailleurs, les CCPP ont d'autres spécificités qui influent sur le recrutement des patients : la taille du centre, la spécialité des consultations, la zone géographique et le tissu économique local. Ces biais entraîneraient des fluctuations peu prévisibles, dans le temps et dans l'espace, du nombre et de la nature des patients enregistrés dans le réseau.

* * * * *

Résumé

4. De la mono à la multi-exposition

L'application des méthodes de pharmacovigilance à la base de données du RNV3P, notamment la méthode PRR, met en évidence des couples pathologie - nuisance, pathologie - métier et pathologie - secteur d'activité potentiellement émergents, définis comme des situations pathologiques qui seraient la conséquence de risques professionnels non clairement identifiés.

Comme la survenue d'une pathologie ne résulte pas toujours de l'exposition à une seule nuisance, notre travail consistait à développer une approche pour :

- Surveiller les Problèmes de santé au Travail enregistrés dans la base de données RNV3P en prenant en compte l'ensemble des facteurs d'expositions professionnelles associés,
- Développer un système de détection des pathologies mono et multi-exposées potentiellement émergentes.

Pour développer une telle approche, les principaux défis méthodologiques auxquels nous devons faire face sont les suivants :

- Utiliser l'ensemble des connaissances relatives à la multi-exposition professionnelle, pouvant inclure des interactions ou synergies entre l'activité (métier et secteur) et l'ensemble des nuisances associées présentes sur le lieu de travail et potentiellement à l'origine de l'apparition de la maladie,
- Prendre en compte les échantillons présentant de faibles effectifs,
- Les populations mère et cible n'étant pas clairement connues, toute étude épidémiologique ne serait pas pertinente.

Partie 2 : Exposomes

"For a really long time, we've thought about individuals as though they were islands - a Robinson Crusoe model of social science. Being able to integrate information - not just about people, but about their relationships - is something that's completely new".

James FOWLER, 2009.

Professeur à l'École de médecine dans la Division des sciences sociales,
lors d'un entretien dans Seed Magazine avec Albert-László BARABASI,
Directeur du CCNR - Center for Complex Network Research.

La surveillance en santé-travail porte classiquement sur le suivi épidémiologique d'indicateurs en population générale (accidents et maladies du travail) et sur des études de cohortes ou cas-témoins spécifiques [Patel 2010] dans la mesure où les effectifs sont suffisants pour la puissance statistique attendue. Dans le cadre du RNV3P, la surveillance doit faire face à plusieurs défis méthodologiques : prendre en compte tous les éléments de l'exposition professionnelle (une à cinq nuisances, un métier et un secteur d'activité) et couvrir des échantillons dont les effectifs peuvent être peu nombreux et insuffisants pour être abordés par les méthodes épidémiologiques classiques. Basées sur les observations issues d'un réseau d'experts, les données du RNV3P ne permettent pas de faire des études épidémiologiques conventionnelles, comme par exemple l'estimation de l'incidence ou de la prévalence de certaines pathologies professionnelles dans des groupes de populations, du fait que la population de référence ainsi que la population des patients adressés dans les CCPP ne sont pas maîtrisées.

Face aux facteurs individuels et la complexité des situations professionnelles où des expositions d'origines diverses et variables en intensité et dans le temps se combinent dans les effets sur la santé, il pourrait se révéler très instructif d'investiguer sur les caractéristiques qui rapprochent ou éloignent les PST. En effet, de nombreuses maladies sont d'origines multifactorielles du fait de l'exposition prolongée et régulière à certains produits. De ce fait, plusieurs PST sont susceptibles de partager des expositions professionnelles identiques. Par conséquent, il conviendrait de développer une approche capable d'analyser l'ensemble des PST de manière globale pour l'étude des relations pathologie - expositions professionnelles multiples.

Dans cette perspective, nous nous sommes intéressés au concept des réseaux relationnels, et développés l'approche par Exposomes, comme un réseau de PST qui partagent un ou plusieurs éléments de l'exposition professionnelle. L'objectif de cette partie est de présenter dans un premier temps les bases des réseaux relationnels pour la compréhension des systèmes complexes. Dans un deuxième temps, nous définissons l'approche par Exposomes pour la surveillance programmée des PST et présentons une application à l'échantillon des Sclérodermies Systémiques issu de la base RNV3P.

5. Introduction au concept des réseaux

Désignant un ensemble d'éléments interconnectés, la notion de réseau s'est popularisée, notamment avec les réseaux sociaux de type Facebook, Twitter, LinkedIn ou encore BiomedExperts. Le concept des réseaux renvoie à la théorie des graphes faisant référence à la théorie des réseaux qui s'intéresse aux réseaux du monde réel plutôt qu'au concept mathématique. La théorie des réseaux est ainsi régie par des concept-clefs, dont la compréhension est un point fondamental pour l'exploitation de ces systèmes complexes.

5.1. Le concept des réseaux

Le point de départ de l'étude des réseaux consiste à structurer les individus de manière différente des bases de données conventionnelles.

5.1.1. Structure des données

Les bases de données conventionnelles sont construites sur un modèle classique où les individus, en lignes, sont décrits par des attributs en colonnes. L'analyse conventionnelle de ces informations consiste à examiner les distributions et les corrélations de ces variables parmi les individus.

Les données relatives aux réseaux se présentent sous la forme d'une matrice carrée symétrique (Figure 16). Les lignes et les colonnes de cette matrice décrivent exactement les mêmes individus. Chaque cellule de la matrice décrit une relation réciproque entre deux individus. A la différence des données conventionnelles qui se focalisent sur la distribution d'attributs parmi un ensemble d'individus, les données de réseaux s'intéressent aux relations entre les individus. La différence est conséquente pour les choix qu'un chercheur peut être amené à faire en décidant de la structure des données.

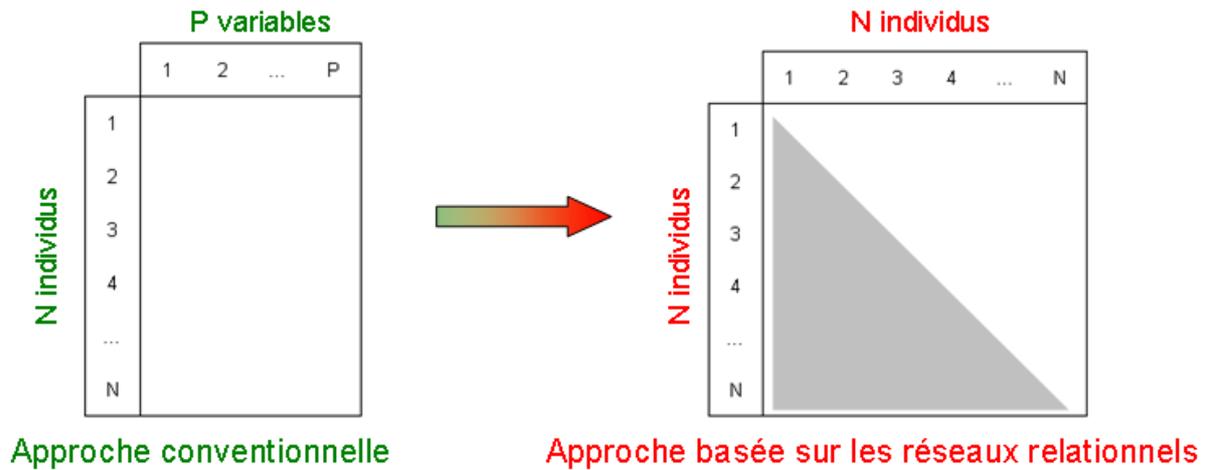


Figure 16 : Schéma illustratif de la structure des données.

Bien qu'il soit possible de décrire les données matricielles comme une forme spéciale de données conventionnelles, l'analyse des réseaux s'intéresse aux données de manière différente. Plutôt que de penser comment un individu se lie aux autres, on observe une structure dans laquelle les individus s'intègrent en fonction de leurs ressemblances. L'organisation des données sous la forme de réseaux nous amène à observer les individus sous un angle nouveau.

5.1.2. Composantes des réseaux

Un réseau, également appelé "graphe", est formalisé par deux composantes essentielles :

- Les nœuds

Les nœuds caractérisent les acteurs du réseau auxquels il est possible d'attribuer des descripteurs de nature qualitative et/ou quantitative. La théorie des graphes fait référence aux nœuds par des *vertices* (sommets), représentés le plus souvent par des cercles [Wasserman]. Par exemple, dans un réseau social, un nœud pourrait formaliser un individu de sexe masculin de poids supérieur à 70Kg.

- Les liens

Les relations entre ces nœuds sont conceptualisées par des liens, caractérisables par des descripteurs de type qualitatif ou quantitatif. Par exemple, un lien renseigne la nature de la relation entre deux nœuds ou individus, marital ou amical par exemple. Mais un lien peut aussi être de nature quantitative en renseignant par exemple le nombre de points communs entre deux individus.

Ces liens peuvent être "orientés", c'est-à-dire qu'un lien entre deux nœuds i et j relie soit i vers j , soit j vers i . Dans ce cas, le graphe se dit orienté. Les liens d'un graphe sont dits "symétriques" lorsqu'il n'existe pas de sens dans la relation, le graphe se dit alors non-orienté.

La théorie des réseaux a inspiré de nombreux travaux de recherche, favorisant l'interdisciplinarité des domaines ; comme la sociologie, la biologie moléculaire ou encore le domaine de la santé.

5.2. Les réseaux dans la littérature scientifique

Aujourd'hui, les réseaux sociaux connaissent une véritable explosion sur le net ; 77% des internautes déclarent être membres d'au moins un réseau social (96% chez les 18-24 ans) (source : ifop), nous donnant un aperçu de leur efficacité et de leur pouvoir communautaire. Mais les réseaux au sens large sont plus que des "communautés de contacts" : il s'agit d'un cadre conceptuel qui attire particulièrement l'attention des scientifiques pour résoudre des problèmes complexes [Oltvai 2002].

5.2.1. Les réseaux sociaux

Pour les sociologues, la théorie des graphes a permis de comprendre les phénomènes sociaux dans une grande variété de disciplines ; de la psychologie à l'économie [Borgatti 2009]. Un réseau social est défini comme un ensemble d'individus, représentés par des nœuds, connectés entre eux par des interactions sociales, représentées par des liens. La science des réseaux sociaux, ou *social networking*, étudie principalement les comportements entre des individus ou des communautés d'individus [Wasserman 1994].

A titre illustratif, Fowler et al. ont étudié la cohorte "*Framingham Heart Study*", composée de 12 067 individus de 1971 à 2003 [Fowler 2008]. Parmi les études réalisées sur ces données, leurs objectifs étaient d'évaluer la propagation du bonheur au sein d'un groupe d'individus et d'identifier les sources de bonheur. Ce travail a montré que le bonheur est un phénomène de réseau, pouvant se propager entre les individus en fonction de la nature des liens sociaux (époux, parents, amis).

5.2.2. Les réseaux en biologie moléculaire

Les biologistes appellent "interactome" un ensemble d'interactions moléculaires dans les cellules, et représentent ce système sous la forme d'un réseau. Ce terme a été créé en 1999 par un groupe de chercheurs français [Sanchez 1999]. L'interactome se réfère à des réseaux de type *protein-protein interaction network* (PPI) ou *protein interaction network* (PIN). Par exemple, des chercheurs ont démontré qu'il existait un faible nombre de protéines qui jouent un rôle essentiel dans la levure de bière car elles interagissent avec un grand nombre de protéines [Zotenko 2008]. Cette étude a notamment révélé que de ces protéines essentielles dépendait la survie de la cellule.

5.2.3. Les réseaux en santé

Dans les réseaux sociaux, les liens entre les individus déterminent la force de leur relation dans la survenue d'une maladie. Comme exemple, nous pouvons citer une étude sur la propagation de l'obésité où Christakis a utilisé le concept des réseaux relationnels en liant les individus qui partagent des liens sociaux de même nature pour suivre l'évolution de leur poids au cours du temps [Christakis 2007]. Partant de l'hypothèse que la prise de poids d'un individu peut être influencée par son entourage, cette approche a permis de soulever l'hypothèse que le réseau social pouvait être un des facteurs de l'obésité.

En médecine, le nombre de points communs entre les individus atteints de la même maladie déterminent l'intensité de leur relation. Sur cette idée, Barabási a conceptualisé ce réseau sous le terme de "diseasome" [Barabási 2007 a]. Ainsi, le réseau social des obèses, par exemple, peut être couplé à un réseau d'expositions plus large permettant ainsi d'étudier l'ensemble de ces facteurs génétiques, environnementaux ou comportementaux de manière collective pour prendre en compte la combinaison des composantes potentiellement étiologiques dans les effets sur la santé.

D'autres auteurs raisonnent en termes de "Human Disease Network", un réseau de pathologies humaines partageant des mutations génétiques similaires alors que les malades ne développent pas forcément la même pathologie. Cette approche a pour but d'émettre des hypothèses quant aux rôles des mutations génétiques dans l'apparition des maladies [Goh 2007 ; Barabási 2011].

Des approches prédictives ont également été développées sur le concept des réseaux pour l'étude de la propagation des épidémies dans un réseau social, comme par exemple la transmission des maladies infectieuses [Keeling 2005], où l'intensité de la relation est utilisée pour prédire la force de l'infection.

Ces exemples ont en commun la recherche de caractéristiques similaires pour la compréhension des mécanismes et l'identification des facteurs-clés dans le développement et la manifestation des pathologies. Ainsi, la recherche des similitudes entre des situations apparemment différentes est l'élément structurant des analyses qui peuvent en découler. Différents domaines de recherche utilisent les réseaux relationnels pour représenter et comprendre des systèmes complexes. Il s'agit d'une science commune dont les propriétés mécanistiques et structurelles sont universelles [Strogatz 2001].

5.3. Topologies des réseaux

Décrire la forme des réseaux constitue un travail essentiel pour comprendre son organisation. La structure des réseaux contribue ainsi à caractériser ses propriétés intrinsèques. On distingue classiquement trois formes de réseaux qui fonctionnent selon leurs propres propriétés (Figure 17).

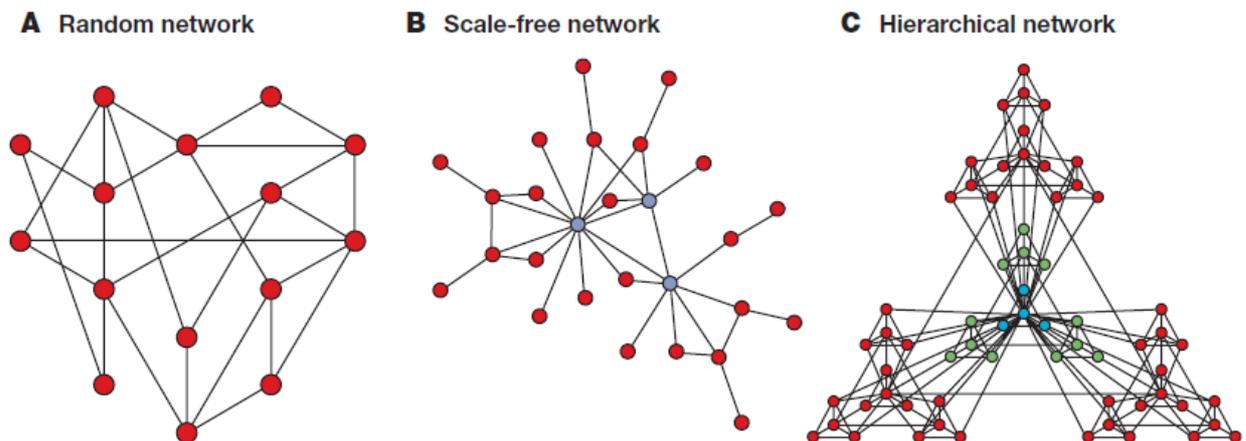


Figure 17 : Trois types de réseau (Illustration extraite de [Barabási 2004]).

A. *Random network*

Les *random networks* sont basés sur le modèle de Erdős - Rényi [Barabási 1999 ; Derényi 2005] conçus sur le principe que chaque paire de nœud a la même probabilité d'être connectée au sein du réseau. Cette caractéristique permet de créer un graphe dans lequel les nœuds ont approximativement le même nombre de liens, ou voisins directs.

B. *Scale-free network*

Les *scale-free networks* sont basés sur le modèle de Barabási - Albert [Barabási 1999] qui ont découvert leurs propriétés en 1999. Supposant que les sites internet étaient reliés entre eux de manière aléatoire, un logiciel a été utilisé pour analyser les liens entre les sites. Selon la théorie des graphes, la plupart des sites auraient du présenter le même nombre de liens. Mais en réalité, la plupart des sites étaient connectés à un très faible nombre d'autres sites, et seulement un faible nombre de sites possédaient un très grand nombre de liens. Cette découverte a transformé la perception des scientifiques dans différents domaines, par exemple en sociologie et en biologie moléculaire. Les *scale-free networks* permettent aujourd'hui de développer des modèles pour la lutte des épidémies ou encore l'analyse des faiblesses sur la Toile.

Ces réseaux se présentent donc sous la forme d'une organisation dépendante d'un faible nombre de nœuds fortement connectés et d'un nombre important de nœuds faiblement connectés [Sanchirico 2008 ; Zhou 2005 ; Boguna 2003 ; Barabási 2001]. Ces systèmes

nécessitent néanmoins un nombre suffisant de nœuds hautement connectés pour fonctionner. En l'absence de ces nœuds, la structure du réseau s'effondre.

C. Hierarchical network

Les *hierarchical networks* sont également représentés sous la forme d'arbre hiérarchique. Ces systèmes sont structurés de manière hiérarchique par de petits groupes de nœuds ; les nœuds sont reliés à un nœud central, appelés "root". Chaque nœud est relié à un certain nombre de nœuds de niveau inférieur dans la hiérarchie. Ce nombre est appelé facteur de branchement ("*branching factor*").

* * * * *

Résumé

5. Introduction au concept des réseaux

Un réseau est un ensemble d'éléments, appelés nœuds, et reliés entre eux par des liens. Un réseau est formalisé par un "graphe". L'étude des réseaux est basée sur des concepts fondamentaux, applicables et appliqués dans de nombreux domaines scientifiques.

Dans la littérature, le domaine de la sociologie utilise largement ce concept, mais ce système tend à se généraliser dans d'autres domaines comme en biologie moléculaire ou en santé. Dans ces domaines d'application, une des principales exploitations de ces systèmes est de caractériser la structure globale du réseau ("clustering") afin d'identifier des regroupements de nœuds en sous-réseaux. Nous présentons ainsi les formes les plus courantes de réseaux répertoriés dans la littérature ; les *scale-free networks*, les *random networks* et les *hierarchical networks*.

6. Formalisation des Exposomes

Le terme "Exposome" apparaît dans la littérature en 2005, où Wild définit l'Exposome comme une base de données qui rassemble l'ensemble des expositions environnementales qu'un individu peut subir au cours de sa vie [Wild 2005], parallèlement au "génom" comme ensemble du matériel génétique de l'individu. Cette approche nécessite des informations relatives à la traçabilité des expositions, mais pose néanmoins les bases d'une réflexion cohérente orientée vers le développement d'un système capable de prendre en compte l'ensemble des expositions à des fins exploratoires ou analytiques.

En 2010, Rappaport définit le concept de l'Exposome [Rappaport 2010] comme la totalité des expositions reçues par un individu au cours de sa vie, englobant toutes les sources d'expositions à des substances toxiques. L'auteur met en évidence les bénéfices d'une telle approche pour les scientifiques afin d'investiguer les causes environnementales de certaines maladies chroniques. Cependant, développer des mesures et moyens pour un historique complet des expositions auxquelles un individu est soumis tout au long de sa vie est une tâche difficile et coûteuse.

Notre définition d'un "Exposome professionnel" reprend le concept général de l'Exposome façon Wild mais se restreint à l'étude de l'environnement professionnel. L'Exposome professionnel que nous définissons dans ce chapitre ne représente pas seulement une idée, mais développe également une méthodologie basée sur le concept des réseaux pour prendre en compte l'ensemble des expositions - professionnelles en l'occurrence - pour comprendre et surveiller des problèmes de santé. Dans la suite de ce travail, nous emploierons le mot usuel "exposome" pour désigner l'exposome professionnel.

6.1. Définition des Exposomes

Au niveau de l'environnement professionnel, la base RNV3P représente une source d'informations précieuse. En effet, ces données recueillies par des médecins spécialistes permettent l'identification de facteurs d'expositions multiples associés à une pathologie. Il pourrait ainsi se révéler instructif d'étudier les points communs entre ces malades pour identifier un ou des facteurs-clefs potentiellement responsables dans la survenue de la pathologie.

Nous avons ainsi développé l'Approche par "Exposome" qui consiste à étudier les expositions professionnelles communes à des individus atteints de la même pathologie. Cette approche peut se conceptualiser sous la forme d'un réseau d'expositions multiples (Figure 18) ou des individus qui partagent la même pathologie (Plan Pathologie) sont susceptibles d'avoir en commun des nuisances professionnelles identiques (Plan Nuisances) et/ou exercés une activité professionnelle similaire (Plan Activité Professionnelle). Sur la base de cette réflexion, nous avons développé le concept d' "Exposome" et formulons la définition suivante :

L'Exposome professionnel est défini comme un réseau de PST connectés entre eux s'ils partagent un ou plusieurs éléments de l'exposition professionnelle identique(s) (nuisances, métier et/ou secteur).

Au sens de l'Exposome, un nœud est défini comme une séquence, ou pathologie associée à un ensemble d'expositions professionnelles multiples. Le poids de cette séquence, ou poids du nœud, est donné par le nombre de copies de PST identiques. Nous avons ainsi obtenu une structuration de la base de données RNV3P en plusieurs groupes de PST partageant des expositions professionnelles communes. De tels Exposomes résultent de la projection du réseau tripartite pathologie - expositions professionnelles multiples sur la dimension pathologie.

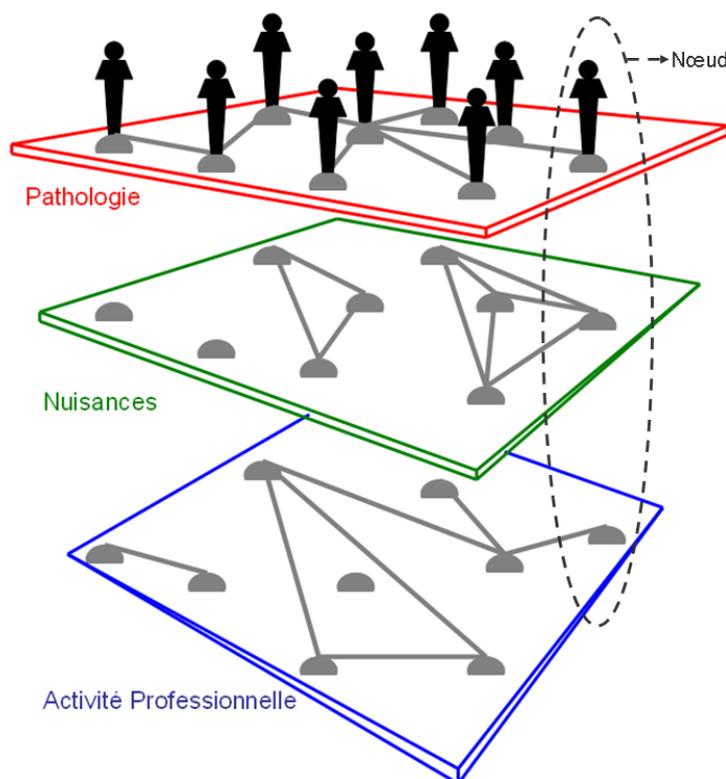


Figure 18 : Approche conceptuelle des Exposomes - Superposition des plans pathologie, nuisances et activité professionnelle.

La Figure 19 illustre la structure emboîtée du réseau avec des exposomes à différents niveaux d'organisation témoignant ainsi de la complexité relationnelle entre PST. Chaque nœud de l'exposome représente un PST, ou séquence, dont la taille est proportionnelle au nombre de PST identiques. Deux nœuds sont connectés lorsque ceux-ci ont au moins un élément de l'exposition professionnelle en commun. Le nombre de connexions entre nœuds indique à la fois la multi-exposition (ou nombre d'expositions différentes) et la diversité de connexion du nœud de par le nombre de PST distincts auxquels le nœud est relié.

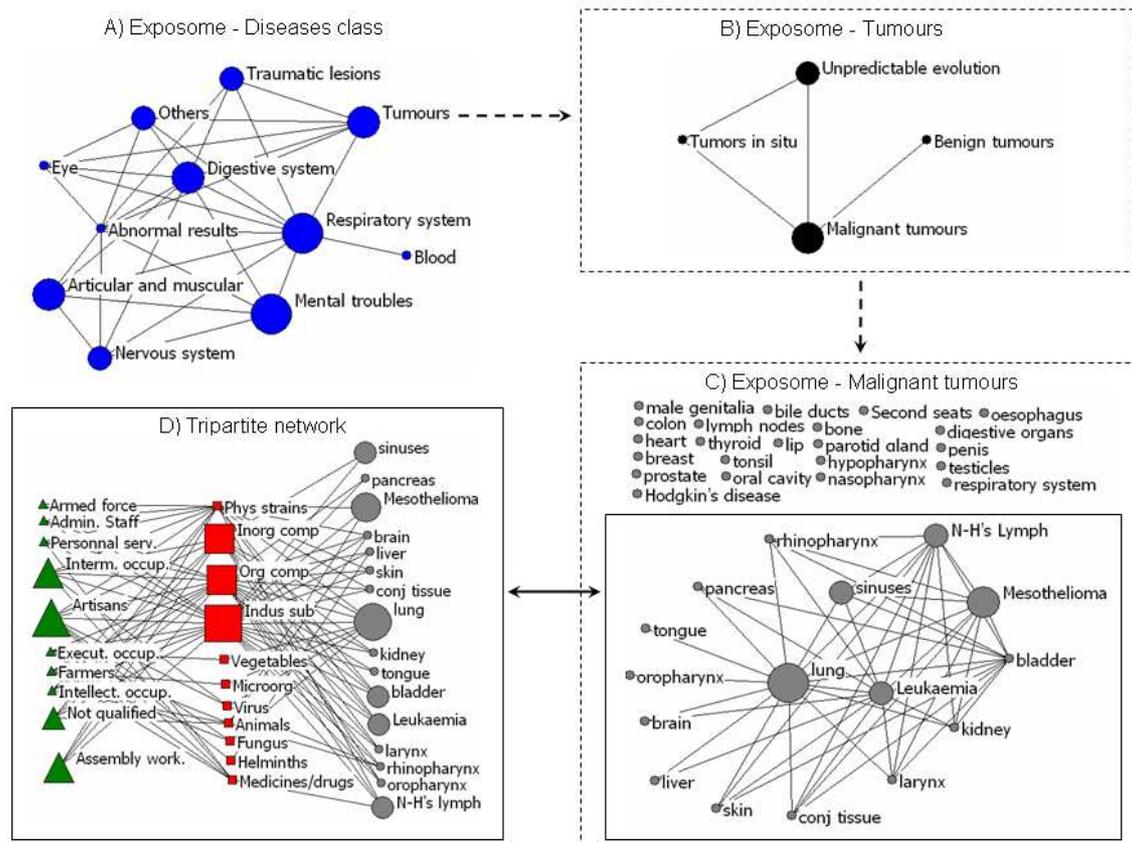


Figure 19 : Structure emboîtée du réseau avec des exposomes à différentes échelles [Faisandier 2009].

La Figure 19-A illustre l'Exposome des catégories de pathologies, dont celle des tumeurs, rapportées dans la base RNV3P. La Figure 19-B se focalise sur l'Exposome des tumeurs. Cette catégorie de pathologies rassemble 4 sous-ensembles que sont les tumeurs malignes, les tumeurs à évolution imprévisible, les tumeurs in situ et les tumeurs bénignes. La Figure 19-C représente les tumeurs malignes où 38 maladies sont distinctes dont 22 nœuds isolés n'ayant aucune nuisance commune avec les autres, et 16 nœuds connectés formant un réseau comme sur la Figure 19-D. Les cercles, carrés et triangles représentent respectivement les tumeurs malignes, les nuisances et les métiers relatifs à chaque observation. Les liens entre carrés et triangles indiquent la liaison nuisances - métier, et les liaisons entre carrés et cercles les liaisons nuisances - pathologie.

6.2. Construction des Exposomes

Dans le cadre de ce travail, nous avons rédigé un article soumis à *Journal of Biomedical Informatics*, qui pose les bases de cette approche et formalise notamment les différentes étapes de la construction de l'Exposome.

Occupational Exposome: a network-based approach for characterizing occupational health problems

**Soumis après corrections le 17 Janvier 2011 à *Journal of Biomedical Informatics*
Supplementary Files (Annexe E)**

Laurie Faisandier^a, Vincent Bonnetterre^{a,b}, Régis De Gaudemaris^{a,b}, Dominique J Bicot^c

^a Laboratoire Environnement et Prédiction de la Santé des Populations, UMR CNRS 5525 Université Joseph Fourier, Domaine de la Merci, 38 706 La Tronche, France

^b Service de Médecine et Santé au Travail, Centre Hospitalier Universitaire Grenoble, BP217, 38 043 Grenoble, France

^c Biomathématiques et Epidémiologie, Laboratoire Environnement et Prédiction de la Santé des Populations-TIMC, UMR CNRS 5525 Université Joseph Fourier, VetAgro Sup, Campus Vétérinaire de Lyon, 1 avenue Bourgelat, 69280 Marcy l'Etoile, France

Abstract.

Surveillance of work-related diseases and associated exposures is a major issue of public health, in particular for identifying and preventing new threats for health. In the occupational health context, the French national occupational disease surveillance and prevention network (RNV3P) have constructed a growing database that records every year all occupational health problems (OHPs) diagnosed by a network of physician specialists. The network aims to provide and develop an expertise on the disease-exposure relationships, and uses the RNV3P database for developing the surveillance of OHPs and for the detection of emerging associations between diseases and occupational exposures. In this paper, we have developed the theoretical framework of the occupational exposome, defined as a network of OHPs linked by similar occupational exposures, as a novel approach which allows to characterize and to analyze the disease - exposure associations reported in the RNV3P database in the form of a relational network. Next, the occupational exposome is structured in terms of occupational exposure groups which constitute informative subsets of hazards considered as the backbone tree spectrum of the occupational exposures potentially related to a disease.

To illustrate the wide possibilities of this method, the exposome approach is applied to the RNV3P database's sample of non-Hodgkin lymphomas (NHLs). As a result, we found that

the NHL occupational exposome could be described in terms of 86 embedded exposure groups, defined as a set of OHPs sharing at least one component of the occupational multi-exposure. For example, "organic solvents and thinners" is the most represented hazards related to NHLs, but is also co-associated to "benzene", "ionizing radiations" or "agricultural products". From the knowledge stored in the database by physician experts, the occupational exposome constitutes a decisive step towards the evolving monitoring of multi-exposure associated to a given disease.

Keywords: Occupational disease; multi-exposure; network

1. Introduction

Studies on health at work are designed primarily to identify occupational hazards, validate new atmospheric or biological indicators of exposure and effect, assess the impact of the occupational environment on health, gain information on the etiology of diseases and improve means of preventions. Surveillance of work-related diseases and exposures is a major issue of public health, in particular for identifying and preventing new threats for health. In such a context, surveillance conventionally involves in specific cohort and epidemiological follow-up of indicators (accidents and occupational diseases), often developed from surveillance networks which feed databases. An illustrative example of such a network is The Health and Occupational Reporting network (THOR) in the United Kingdom [1-3].

In France, the national occupational disease surveillance and prevention network (Réseau National de Vigilance et de Prévention des Pathologies Professionnelles, RNV3P) was set up in 2001 as a nation-wide network of experts from occupational health consultation centres in University Hospitals of metropolitan France that records in a structured and standardised RNV3P database all patient cases diagnosed with an Occupational Health Problem (OHP), i.e. patients with diseases potentially related to occupational exposures. The RNV3P network originated from the need of developing a common expertise on health problems by bringing together scientists and researchers working in multidisciplinary areas of health. Among others objectives, the network aims to provide and develop an expertise on the disease-exposure relationships, and uses the RNV3P database for developing the surveillance of OHPs and for the detection of emerging associations between diseases and occupational exposures. To address these issues, a systematic data-mining approach based on statistical tests of disproportionate as used in pharmacovigilance [4,5] was applied to RNV3P database to

generate pre-alerts of potentially emergent disease-exposure couples. Although very sensitive, this method appears to be less specific and fails to simultaneously handle all components of the exposure associated with a disease.

Following these first studies on the RNV3P database, we began to develop the occupational exposome approach [6,7], as an alternative and complementary approach to pharmacovigilance methods involving disease-single exposure couples, that incorporates all dimensions of the composite occupational exposures in analyses of the RNV3P database. Our aim is to develop an approach allowing to investigate characteristics or traits which gather or separate OHPs as many individual factors and complex occupational situations, where exposures of diverse origins and variable intensity over time, are combined in effects on health. In this framework, we conceptualized the occupational exposome as a network of OHPs sharing components of the set of occupational exposures. Such an occupational exposome will make sense only when dealing with diseases associated with multi-exposures, i.e., associated with more than one hazard.

Similar conceptual approaches can be found in literature including the wording exposome which had been already introduced by Wild [8]. In contrast to the occupational exposome, the Wild's exposome represents the collection and succession of individual and environmental exposures encountered during an individual's lifetime. The author proposes reconstructing for a given individual, an exposure network in order to better understanding the role of each exposure and thus generate research hypotheses for the disease etiology. Although this requires a certain amount of information on traceability of exposures, this exposome view lays the groundwork for a coherent debate geared towards the development of relational networks for monitoring of exposures of various origins. Beyond the idea of collecting multi-exposures, Barabasi [9] has constructed a "diseasome" to illustrate, under a network, the environmental and social factors which might have a potential role in the origins of obesity; and Goh et al. [10] have explored a network of human diseases that implicates similar genetic mutations. In a similar way, Christakis et al. [11] have used the concept of relational networks that link individuals sharing social ties to monitor their weight over time. Assuming that the weight of an individual may be influenced by his or her surroundings, this approach has highlighted the possibility that social networking could be a factor in the spread of obesity. What all these analyses and various views have in common is the search of similar characteristics for understanding underlying mechanisms and identifying key factors in the onset and development of diseases.

Our main goal is to develop a framework allowing analysing the RNV3P database in terms of an evolving and growing complex network. For this purpose, objectives of this paper are twofold: first, we outline the framework of the occupational exposome approach by providing definitions and showing how to construct exposomes, and second we show how to characterized and structured the topology of the constructed exposome by means of occupational exposure groups (sets of nodes or OHPs in a network sharing similar characteristics) which represent motifs carrying information on occupational exposures potentially related to the disease. To illustrate how this approach can be implemented, we considered the sample of Non-Hodgkin Lymphomas from the RNV3P database and, subsequently, we show the potential of this approach for monitoring and studying associations and relationships between diseases and occupational multi-exposures.

2. Material

Each diagnosed patient is recorded in the RNV3P database as an Occupational Health Problem (OHP) represented by a four codes item (Fig. 1) that includes a disease or pathology code (ICD-10 classification) plus a three-item exposure code holding information on the kind of industrial products, chemicals, psychological or organizational constraints (characterized by hazard codes) to which the patient was exposed during his or her occupational activity (characterized by occupation and activity sector codes). The hazards are coded using a hierarchical code owned by the national social security organism (Caisse Nationale d'Assurance Maladie, CNAM), the patient's professional activity is coded according to the international classification of occupational type (Classification Internationale Type des Professions, CITP-88), the French equivalent of the International Standard Classification of Occupations (ISCO), and the activity sector is coded according to the French occupational classification (Nomenclature des Activités Professionnelles, NAP-03). Two OHPs may differ either by their pathology and/or by at least one component of the exposure, i.e., hazards, occupation and activity sector. Likewise, several patients can present an identical OHP as well. From 2002 to 2007, the RNV3P database was full of 90,335 OHPs of which 75% were reported associated with one hazard, 17% with 2 hazards, 5% with 3 hazards, 2% with 4 hazards and less than 1% with 5 hazards.

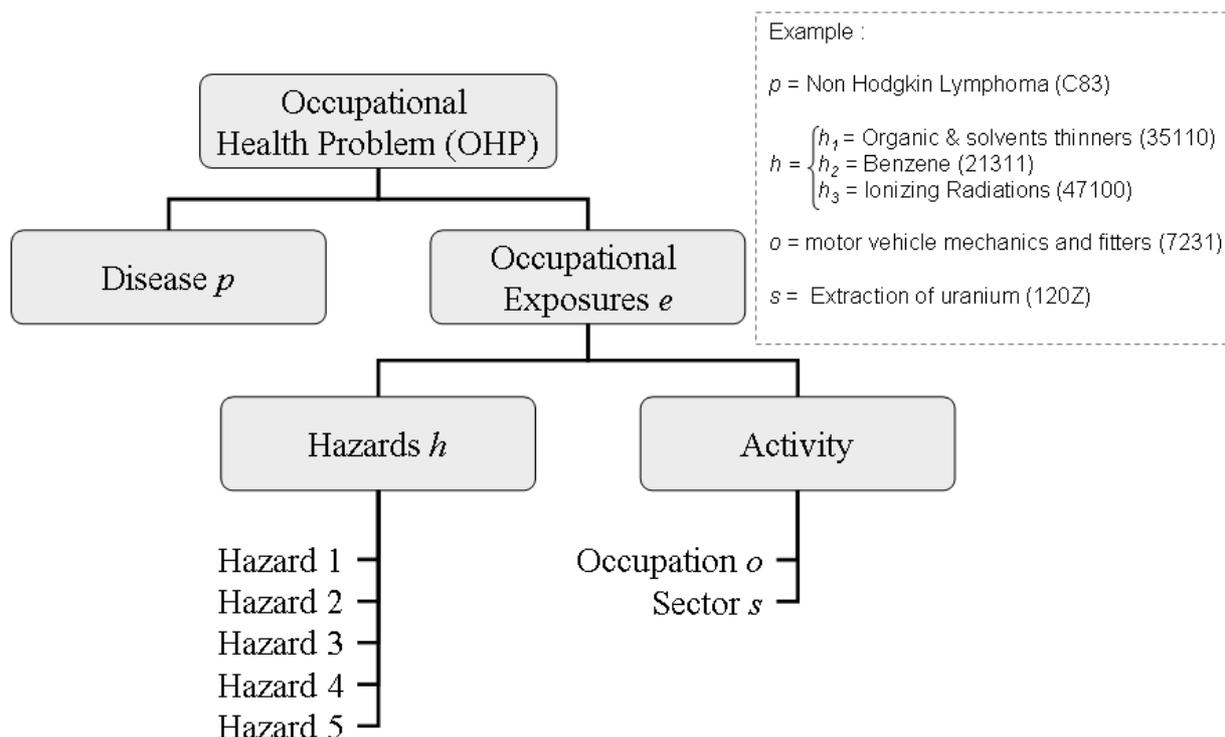


Fig. 1. Structure of an Occupational Health Problem (OHP).

Inset: example of an OHP with item codes given between parentheses.

3. Theory

3.1. Construction of the Occupational Exposome

To study the complexity of the relationships between diseases and composite occupational exposures and to seek similarities in associated exposures that potentially lead to identical diseases, we developed the concept of the occupational exposome. To process the RNV3P data in the form of a network of relationships, each OHP in the RNV3P database is represented by a node (vertex) $v = (p,e)T$, which is a unique combination of a disease (pathology) "p" and its associated three dimensional composite occupational exposure $e = (h,o,s)T$, characterized by a set of hazards "h", an occupation "o" and an activity sector "s" (Fig. 1). By convention, the hazard vector $h = (h_1,h_2,h_3,h_4,h_5)T$ may comprise from 1 up to 5 distinct hazards that were present in the patient's occupational environment as characterized by "o" and "s", and are suspected or confirmed to be related or cause the disease. Each node (i.e., OHP) is weighted by the total number "w" of identical OHP copies in the database (See Supplementary Information Fig.S3). In this way, the initial RNV3P database of OHPs is

mapped into an ensemble of V distinct nodes (each of which being an ensemble of identical OHPs) from which network analyses can now be conducted.

From now, the exposome graph is constructed as follows. Let $C_{ij,\eta}$ (with $0 \leq C_{ij,\eta} < 5$) denotes the number of hazards shared by two nodes v_i and v_j ($i, j = 1, 2, \dots, V$) appearing each

at least in η copies in the database. We have,
$$C_{ij,\eta} = \theta(w_i - \eta) \times \theta(w_j - \eta) \times \sum_{k,l=1}^5 \delta_{h_{ik}, h_{jl}}$$
, where

$\delta_{h_{ik}, h_{jl}}$ is the Kronecker symbol, which compares the components h_i and h_j of exposures e_i and e_j , with $\delta_{h_{ik}, h_{jl}} = 1$ if $h_{ik} = h_{jl}$, and $\delta_{h_{ik}, h_{jl}} = 0$ if $h_{ik} \neq h_{jl}$, and the product

$\theta(w_i - \eta) \times \theta(w_j - \eta)$ ensures that the numbers of OHP copies in nodes v_i and v_j are each greater than or equal to a threshold η (with $\eta \geq 1$), where $\theta(z) = 1$ if $z \geq 0$ and $\theta(z) = 0$ for $z < 0$. Now, we define the connection rule such that two nodes v_i and v_j are connected if their

associated $C_{ij,\eta} \geq D$, with $1 \leq D \leq 4$, and not connected otherwise. The D value is comprised between 1 and 4 because any node may be associated with up to 5 hazards and, therefore, two connected nodes may share D hazards. Thus, for a given D , we define the $D\eta$ – occupational

exposome as an undirected network characterized by the $V \times V$ adjacency matrix $\mathbf{A} = (A_{ij})$

between nodes v_i and v_j with the elements $A_{ij} = 1$ if $C_{ij,\eta} \geq D$ and $A_{ij} = 0$ otherwise. The choice of using the above connection rule for the network construction rather than similarity degree between nodes or other descriptors is dictated by the interest in identifying common or shared elements of the occupational exposure.

To summarize, the $D\eta$ – occupational exposome is an undirected (weighted) network of OHPs having in common at least D elements of the exposure and where each OHP appears at least η times in the database (See Supplementary Information). Such a network is represented by the graph $G = \{W, V, L, D, \eta\}$ made up of a set of W OHPs described by V distinct nodes connected between them by L heterogeneous links and characterised by the adjacency matrix defined by $A_{ij} = 1$ if $C_{ij,\eta} \geq D$ and by $A_{ij} = 0$ otherwise, with $i, j \in \{1, \dots, V\}$. Links in the exposome are heterogeneous because in the same network a node may both share more than one given hazards with a node and share other hazards with another node. The η and D can be regarded respectively as quantitative and qualitative control parameters of the exposome architecture as η controls both the node weight and degree and D , impacting the node degree,

allows zooming in or zooming out the network. The probability that two nodes of the network taken at random are connected is given by the density, $d = 2L/V(V-1)$.

3.2. Identification of occupational exposure groups

To structure and organize the occupational exposome in terms of informative motifs or subsets carrying information on occupational exposures related to a disease, we introduce the notion of the $D\eta$ – occupational exposure group defined as a cluster of nodes (OHPs), in the $D\eta$ - occupational exposome, sharing exactly D identical hazards ($1 \leq D \leq 4$, by definition). Let $\{h_\alpha, \alpha = 1, 2, \dots, H\}$ be the ensemble or list, of size H , of distinct hazards “ h_α ” in all connected nodes of the $D\eta$ – occupational exposome, and $\mathbf{B} = (\mathbf{B}_{\alpha i})$ the $H \times V$ matrix defined

as $\mathbf{B}_{\alpha i} = 1$ if $\sum_{k=1}^5 \delta_{h_{ik}, h_\alpha} = 1$, i.e. if one component of the hazard h_i , in the node v_i , is equal to h_α ,

and $\mathbf{B}_{\alpha i} = 0$ if $\sum_{k=1}^5 \delta_{h_{ik}, h_\alpha} = 0$. The 1η - occupational exposure group (or simple exposure group), coined “ h_η ” of the name of the hazard under consideration, is given by the ensemble $g_1(h_\alpha)$ of nodes all satisfying $\mathbf{B}_{\alpha i} = 1$, i.e. $g_1(h_\alpha) = \{\cup v_i \mid \mathbf{B}_{\alpha i} = 1, i = 1, 2, \dots, V\}$. The size or

number of nodes in $g_1(h_\alpha)$ is, $n_\alpha = \sum_{i=1}^V \mathbf{B}_{\alpha i}$, with frequency $q_\alpha = \sum_{i=1}^V w_i \mathbf{B}_{\alpha i} / \sum_{i=1}^V w_i$ normalized to one. High order ($D > 1$) occupational exposure groups are constructed from the intersection

of D simple exposure groups as, $g_D(h_{\alpha_1}, h_{\alpha_2}, \dots, h_{\alpha_D}) = g_1(h_{\alpha_1}) \cap g_1(h_{\alpha_2}) \cap \dots \cap g_1(h_{\alpha_D})$, or

equivalently, $g_D(h_{\alpha_1}, h_{\alpha_2}, \dots, h_{\alpha_D}) = \{\cup v_i \mid \mathbf{B}_{\alpha_1 i} \times \mathbf{B}_{\alpha_2 i} \times \dots \times \mathbf{B}_{\alpha_D i} = 1, i = 1, 2, \dots, V\}$, with

normalized frequencies $q_{\alpha_1, \alpha_2, \dots, \alpha_D} = \sum_{i=1}^V w_i [\mathbf{B}_{\alpha_1 i} \times \mathbf{B}_{\alpha_2 i} \times \dots \times \mathbf{B}_{\alpha_D i}] / \sum_{i=1}^V w_i$. Any node with D

hazards long is likely to belong to at most equivalently, $2^D - 1 - \delta_{D,5}$, distinct occupation exposure groups. As a consequence, the standard disease – exposure association (or simply, disease-hazard association) is now replaced by the disease-exposure group association where the identified occupational exposure group is considered as the most likely related candidate in the relationships between the disease and a set of exposure situations. Accordingly, the occupational exposome is described as an expansion in terms of occupational exposure groups

D_{η} – occupational exposome = $\sum_{j=D-1}^4 m_j g_j(\{h\}) \equiv (m_{D-1}, m_D, \dots, m_4)^T$
 as , where m_j is the number of $j\eta$ – occupational exposure groups. Beyond the nodes, what main matters in exposure groups are hazards which make up situations of occupational exposure. For $j = 0$, m_0 is the number of unconnected nodes and $g_0(\{h\})$ represents a single node with hazards not belonging to the list of shared hazards. As both nodes and links never disappear in the growing RNV3P data network, the ensemble of non-shared hazards in the 0_1 – occupational exposure groups required specific attention as newly incorporated nodes will be likely to create new connections between any nodes of the network and, therefore, create new exposure groups.

4. Results

To illustrate what an exposome looks like and how this approach can be used to gain information on OHPs, we consider the case of Non-Hodgkin Lymphomas (NHLs), a cancer whose incidence has been increasing since 1970s and whose risk factors are not yet well known [12]. In the present state of knowledge, NHLs are hematological malignancies involving a proliferation of lymphoid cells which tend to infiltrate all organs (gastrointestinal tract, skin, bones, etc) in an organism. Identified causes of NHLs mainly include immune system abnormalities and factors of infectious origin, although environmental factors, especially occupational exposures, are suspected to increase the risk of NHLs. Indeed, the literature suggests that exposure to substances such as dioxins, pesticides, benzene or organic solvents increase the risk of NHL [13-15]. As the level of proof issue is still very low, this disease is not yet considered in France as a compensated disease.

This knowledge can be extracted from the RNV3P database which, to date, comprises a total of 77 NHL patients or OHPs with a male/female ratio of 4.5 (63/14), mean age of 52.3 years (52 and 53.7 years for male and female, respectively) and median age of 54 years for both genders. This NHL sample involves 72 distinct hazard codes (or 54 aggregated codes), 55 different occupation codes and 28 different codes of activity sectors (Supplementary Information Table S3-5) thus showing the wide spectrum of occupational exposures that can be associated with NHLs.

4.1. Space of Shared hazards between Non-Hodgkin Lymphomas

Prior to the occupational exposome, Fig. 2 displays as a polar graph the information content of each node in terms of total number of hazards per node and number of hazards shared with other nodes. Each node is represented by a radius of length equal to the number of associated hazards (1 to 5). The entire surface of the graph (dashed area) represents the available space if all the 51 nodes of OHPs were associated with 5 hazards each. Our sample of NHLs occupies 60% (dark area) of the available space of which 32% of nodes (16/51) are of length one (associated to one hazard) and the remaining 68% (35/51) of nodes are multi-exposed (length > 1), thus indicating the likelihood that NHLs are associated with more than one hazard. Of the 68% (35/51) multi-exposed nodes, more than 80% (29/35) shares at least one reported hazard with other nodes (totalizing 48% (gray area) of the available space).

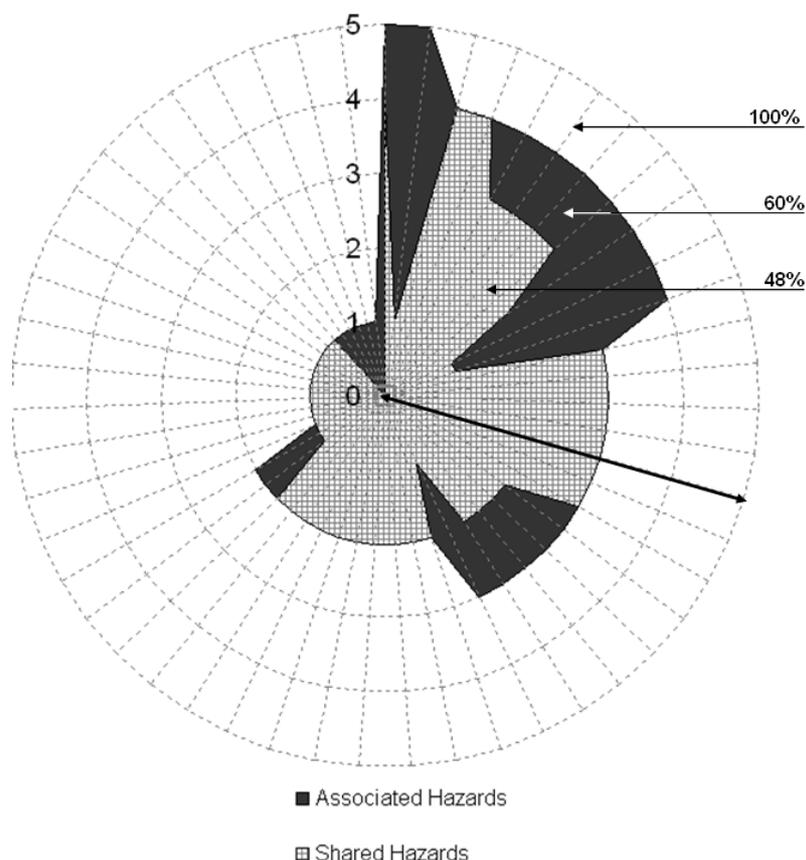


Fig. 2. Distribution of the occupied and shared space among NHLs nodes.

The polar graph displays 5 concentric circles, which represent the 5 possible associated hazards, with 51 radii each of which corresponding to a node of the NHL sample from the RNV3P database 2002-2007. Shaded gray area represents the space of shared hazards and dark area that of associated hazards. From the 77 NHL OHPs of which nearly half (38/77) are

found associated with more than one hazard, 51 distinct nodes are formed; each node representing w OHPs.

4.2. Presentation of Non-Hodgkin Lymphomas Exposome

As malignant tumors, NHLs are contained in a node that belongs within the hierarchy of the larger occupational exposome of malignant tumors (Supplementary Information Fig. S2 showing the 11 and 210 – exposomes). Zooming in that node, Fig. 3 shows the 11 - exposome of NHLs generated using NetDraw software [16]. Of the 51 nodes, 5 are singletons (not connected) as they do not share any hazards with any other nodes, and 46 nodes form a network showing the structure of the spectrum of exposures reported by physicians related to the NHLs. Other underlying dimensions of the exposome structure, the occupation, activity sector and timing of the record entry in the database, are also superimposed for a few nodes in Fig. 3. Using different values of η and D will lead to a different exposome profile as the number of nodes and the exposome density both decrease with η and D (Supplementary Information Fig.S2). As an example of variety in situational exposures, the occupation "mechanical engineering technicians" (solid line and open squares in Fig. 3) provided one case every year from 2004 to 2007 from different consultation centres, while the occupation "motor vehicle mechanics and fitters" (solid line and open circles in Fig. 3) reported one case in 2006 and 3 in 2007 (Supplementary Information Table.S4) from 3 different consultation centres. Motor vehicle mechanics belonging to different activity sectors, share similar hazards, in particular benzene as well as other solvents, whereas experts do not report the same hazards for mechanical engineering technicians.

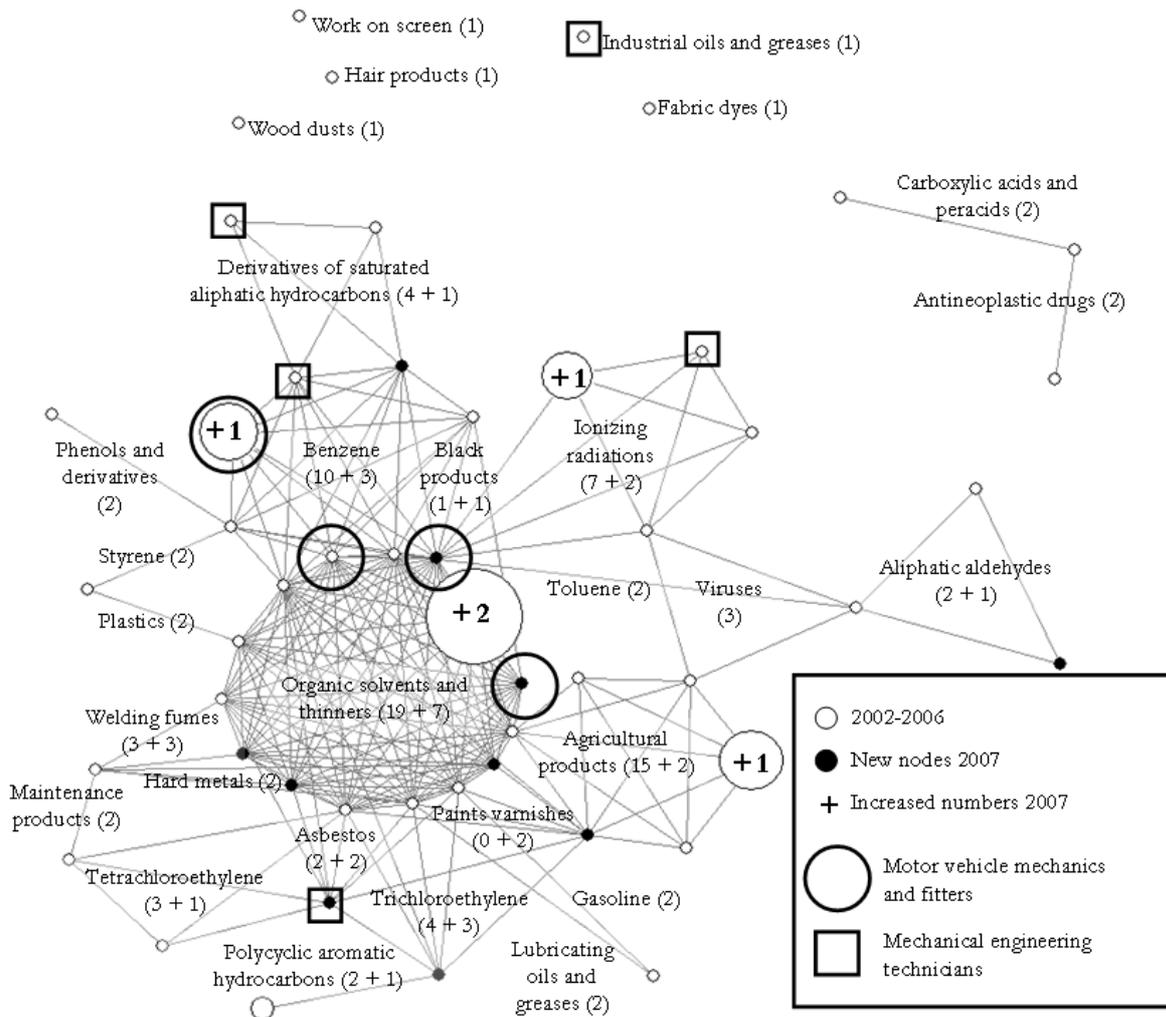


Fig. 3. The 1_1 - exposome of Non-Hodgkin Lymphomas.

RNV3P database 2002-2007, with $GNHL = \{W=77, V=51, L=210, D=1, \eta=1\}$; W: total number of OHPs, V: number of nodes and L: number of links. A node represents a set of identical NHL OHPs associated with a cortege of 1 to 5 hazards and contains at least $\eta = 1$ OHP. Nodes are connected if they share at least $D = 1$ hazard. The size of the nodes is proportional to their weight in number of OHPs. Quoted hazard names represent the 1_1 - occupational exposure groups with associated number of OHP content given in parentheses (total number of OHPs in period 2002-2006 + number of OHPs in 2007). There are 5 unconnected nodes at the top corresponding to the 01 - occupational exposure groups.

4.3. Topology of the Non-Hodgkin Lymphomas Exposome

Two indicators of interest can be used to characterize the NHL - exposome: the degree k_i of a node v_i , defined as the number of nodes in its nearest neighborhood, and the local clustering coefficient c_i for a node v_i , given by the proportion of links between the nodes within its

nearest neighborhood divided by the number of links that could possibly exist between them [17,18]. The degree k and clustering coefficient c for each node and their corresponding distributions for the NHL exposome in Fig. 3 are shown in Fig. 4.

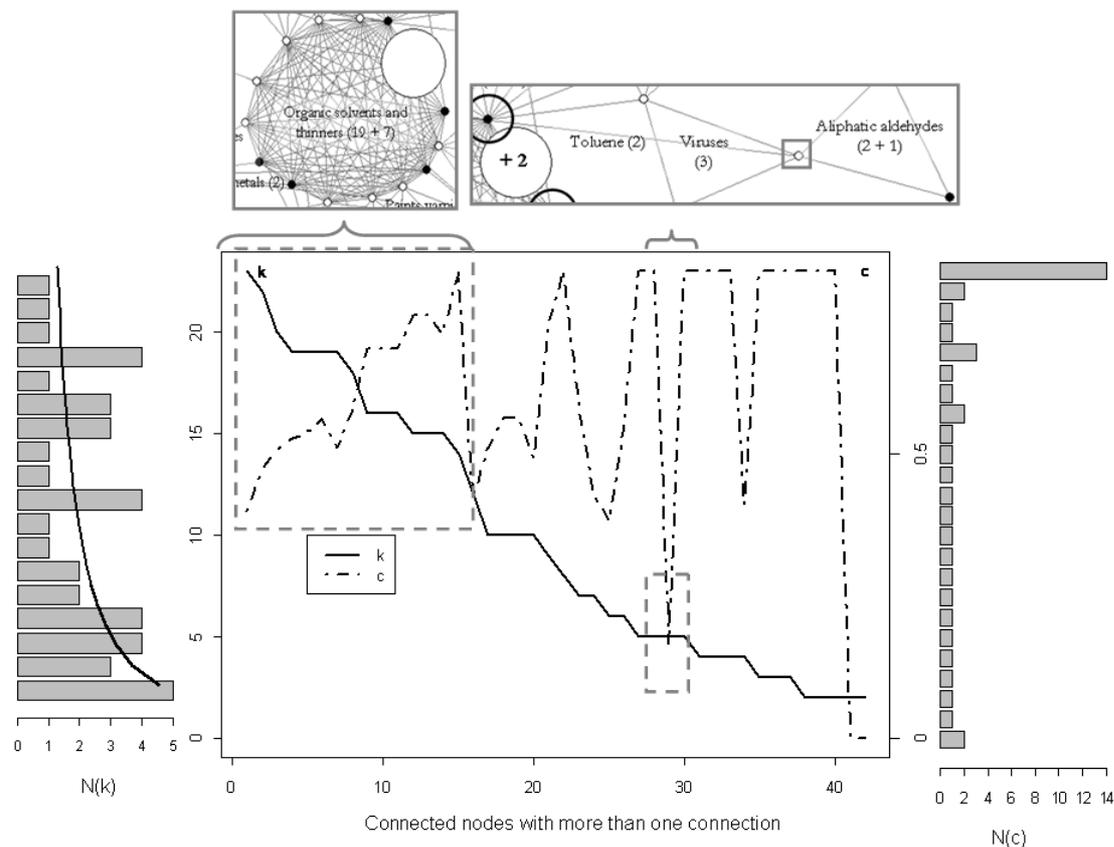


Fig. 4. Distribution of connectivity degree and clustering coefficient

The main frame shows the number of links per node or degree k (left y - axis) and the corresponding clustering coefficient c (right y - axis) versus the nodes with more than one connection in the 11 - exposome of NHLs. The insets in the top zoom on the corresponding parts of the NHL exposome in Fig. 3. The histograms on the left and right represent the distributions $N(k)$ of degree k and $N(c)$ of clustering coefficient c , respectively. The solid line through the histogram represents the best fit with the power law, $N(k) \sim k^{-0.52}$.

The NHL exposome architecture does not seem to fit with any of the well-known networks such as exponential, Erdos-Rényi, Barabasi-Albert and small world networks [19-22]. However, as in many empirically observed networks including the protein networks [23] and some social networks, the NHL exposome exhibits some features of scale-free networks where both the degree and clustering coefficient distributions follow a power law [24].

The first feature of scale-free networks is the commonness of nodes with a degree that greatly exceeds the average. Fig. 4 clearly shows more than half of nodes with a degree greater than the mean $\langle k \rangle = 8.24$ (median = 6, standard deviation = 6.98), and a number of nodes with a clustering coefficient greater than the mean $\langle c \rangle = 0.74$ (median = 0.77, standard deviation = 0.26). Note also that the distribution of degree fits with inverse of square root as $N(k) \sim 1/\sqrt{k}$, a decreasing law different from that found in the literature on scale-free networks. However, one has to keep in mind the yet small size of the NHL network.

The second feature in scale-free networks is that the local clustering coefficient decreases as the node degree increases. This means that the low-degree nodes belong to very dense sub-graphs and these sub-graphs are connected to each other through hubs. As illustrated in Fig. 4, nodes with low k and c bridge clusters of nodes like, for instance, the node "NHL x toluene x viruses x aliphatic aldehydes" which belongs to 3 clusters (dotted gray box in Fig. 4). Such bridging nodes have an important role in the cohesion and robustness of a cluster structured network such as the NHL exposome. Indeed, removing such hub nodes may lead to disintegration of the network [25,26]. The highest-degree nodes, associated with 3 hazards on average, are found within the largest cluster of nodes where k , ranging from 10 to 23 links (dotted gray box in Supplementary Information Fig.S3), increases as the clustering coefficient decreases. Inspection of the top ten highest-degree nodes (Supplementary Information Table S2) already suggests a tentative list of most frequently shared hazards, i.e., of leading occupational exposure groups. In addition, as the clustering coefficient of a node quantifies how close its neighbors are to forming a complete graph or a clique, the highest-degree nodes with a moderate clustering coefficient can be regarded as hubs between occupational exposure groups while those with high clustering coefficient belong to a single clique. Interestingly, we found that the NHL – occupational exposome is an assortative mixing network with an assortative index of $r = 0.321$, indicating that well connected nodes tend to connect one another [27]. As connected nodes in the exposome have in common at least an occupational hazard, a positive assortativeness suggests the occurrence of very popular hazards.

4.4. Occupational Exposure groups of Non-Hodgkin Lymphomas

The occupational exposome allows structuring expert knowledge in the form of occupational exposure groups. This supervised learning of individual characteristics enables to focus on a collective scale, as reported in Fig. 3. Indeed, we found that the complexity of the NHL

exposome is structured as $1_{\eta} - \text{NHL exposome} = (5, 24, 48, 11, 3)^T$, i.e., 5 unconnected nodes, 24 simple exposure groups (compared to the 54 distinct hazard codes), 48, 11 and 3 lists of 21, 31 and 41 – occupational exposure groups, totalizing 86 occupational exposure groups. These exposure groups constitute, within the framework of the RNV3P, the backbone tree spectrum of the most likely occupational exposures that may be related to NHLs.

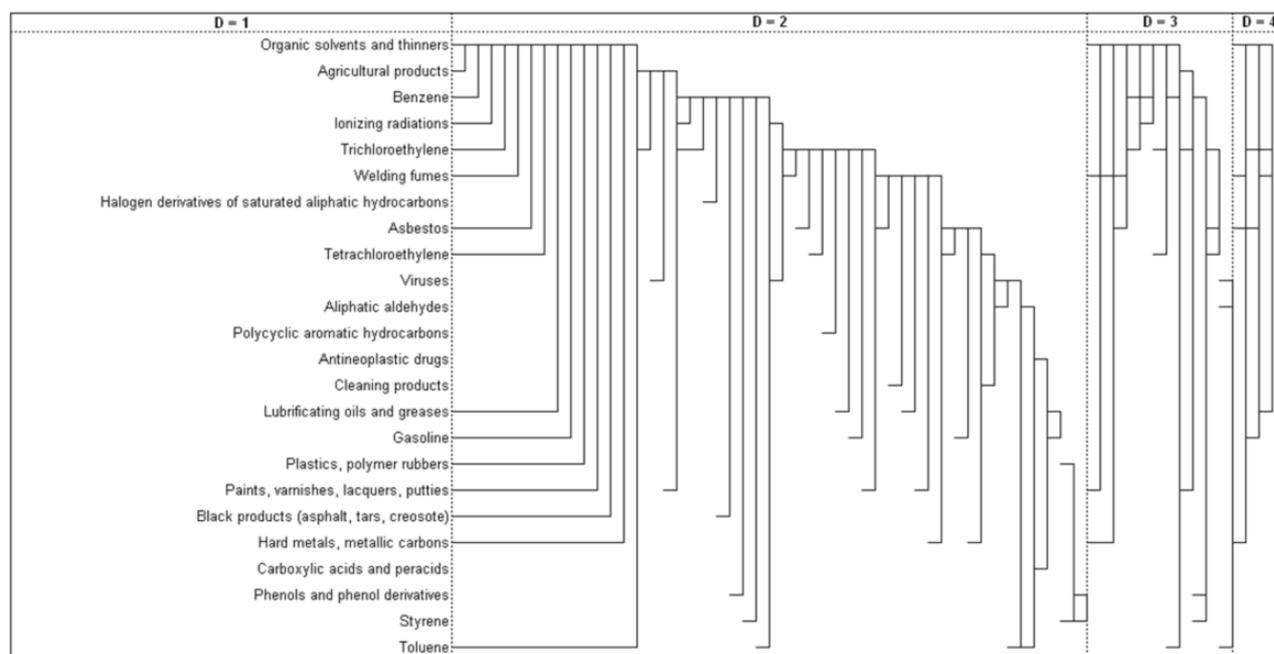


Fig. 5. Distribution of 24 shared hazards according to different multi-exposure levels (i.e. number of shared hazards)

List of the 86 exposure groups of NHLs comprising 24 simple exposure groups (D=1), and 48, 11 and 3 high order exposure groups 21, 31 and 41, respectively, RNV3P database 2002-2007. For $D = 1$, the exposure groups are sorted in decreasing order of their frequency (see Fig. 6) from the top to the bottom. For $D > 1$, each exposure group (constructed from D simple exposure groups) is represented by a vertical solid line with D branches each of which pointing on an associated simple exposure group.

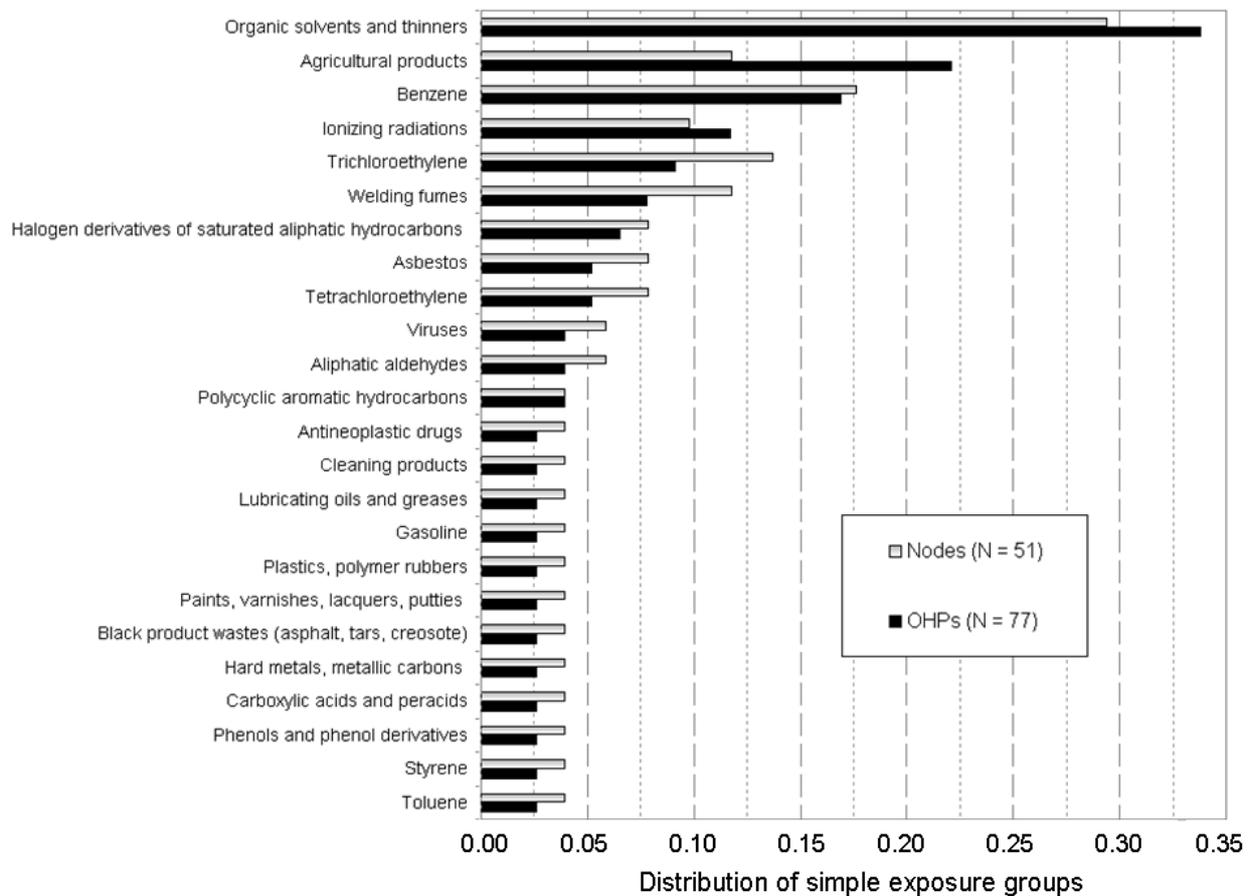


Fig. 6. Frequency of simple occupational exposure groups

Frequency q_{α} (dark histogram) of the h_{α} simple exposure groups of the 11 – exposome of NHLs, RNV3P database 2002-2007. Light histogram represents proportion of nodes in each exposure group.

The spectrum of simple occupational exposure groups (i.e., the most frequently shared hazards) are quoted in Fig. 5 and listed in Fig. 6 with corresponding size and frequency distributions. Both, the size and frequency distributions are similar with the largest (Fig. 5) and most frequent (Fig. 6) group being the "organic solvents and thinners". All top ten highest-degree nodes belong to the most frequent exposure groups although some of their associated hazards are not shared with other nodes (Supplementary Information Fig.S1). Fig. 5 also shows the overlaps between exposure groups and that high order exposure groups are often formed from larger and more frequent simple exposure groups, thus implying a high probability of occupational multi-exposure in NHLs. Overlaps between occupational exposure groups indicate that low order exposure groups are embedded in high order ones and/or sharing hazards for exposure groups of the same order > 1 .

Note that simple exposure groups can be considered as homogeneous cliques since a clique is formally defined as a sub-set of n nodes connected by $n-1$ links (regardless the nature of links) to other nodes in that clique [28,29]. Thus, because of the possible heterogeneity in the nature of links, the cliques generate "hybrid" exposure groups where links between nodes in the clique may be of different kinds (i.e., the shared hazards are different between nodes in the same clique). As an example of a hybrid exposure group, the "petrol x lubricating oils and greases x organic solvents and thinners" group is made up of 3 fully connected (through 2 links) nodes where each of 2 nodes share a single hazard "petrol", "lubricating oils and greases" and "organic solvents and thinners". Using algorithms to compute a census of all cliques, we found 22 hybrid exposure groups in the 11 – NHL occupational exposome.

5. Discussion

We have outlined the theoretical framework of the occupational exposome, defined as a network of OHPs linked by similar occupational exposures, which allows to characterize and to analyze the disease - exposure associations reported in the RNV3P database in the form of a relational network. The main underlying idea of this approach is to investigate characteristics or traits which gather or separate OHPs and therefore occupational exposure. Accordingly, the occupational exposome allows processing the RNV3P data in the form of an OHP network which, in turn, can be studied and characterized as any network and extract interesting and useful properties. In addition, we have structured the occupational exposome in terms of occupational exposure groups which, within the RNV3P framework and in the occupational health context, are informative subsets of hazards considered as the backbone tree spectrum of the occupational exposures which are potentially related to the disease under consideration. With this approach OHP data can be analysed at the scale of a single OHP or at that of an ensemble of OHPs as well.

For the sake of simplicity and of arguments, we have presented implementation of the exposome approach only at the scale of a single OHP, although the occupational exposome for several OHPs (malignant tumors) is given in Supplementary Information (Fig.S2). As illustrated with the NHLs sample (see Figs.3 – 6), the 11 – occupational exposome of the NHL is a “scale-free” like assortative mixing low dense ($d = 0.165$) network made up of 51 nodes with mean degree and mean clustering coefficient of 8.24 and 0.74, respectively, and structured in 86 occupational exposure groups plus 5 unconnected nodes corresponding to the 01 – occupational exposure groups (see Figs.5 and 6). These non-shared hazards in the 01 –

occupational exposure groups required specific attention as they are likely to create new occupational exposure groups when incorporated new nodes during the growing of the network. In addition, beside the small size of the NHL's sample, it would be informative to understand the underlying process leading to the inverse square root power law for the degree distribution in the NHL occupational network.

Although the small size of the NHL sample does not allow reliable fitting, the results found from the NHL occupational exposome can be considered as the state of the knowledge at the date of available data in the RNV3P. These findings turn out to be in coherence with the state of the knowledge on occupational and environmental risk factors of NHL. Indeed, recalling that the RNV3P database is a growing database implies that both nodes and links between nodes never disappear in the occupational exposome. The possible modifications in such a growing network will be to increase the weight of existing nodes, create new connections between already existing nodes, add new nodes and create connections between already present and new nodes. These changes occurring during the growing process of the network may affect some properties (like distributions of degree and clustering) of the occupational exposome but not destroy the structure in occupational exposure groups and rather increase the total number of occupational exposure groups with changes in their relative frequencies.

The approach that we are currently developing finds already immediate applications in addressing issues of the RNV3P. Indeed, organizing the ensemble of OHPs by mean of exposome serves as the starting point for further analyzes of the data. And, in contrast to analytic epidemiology that traditionally aims to test hypotheses advanced beforehand, the occupational exposome approach outlined above is intended to be used for 1) - scheduled surveillance of identified and targeted OHP groups (or disease – occupational exposure associations) and for 2) – prospective surveillance of the evolving network for detecting emergent events (like growth of already existing and/or appearance of new nodes and/or occupational exposure groups) leading to generation of new hypotheses in relationships between disease and occupational exposure that can be investigated by further epidemiological studies.

Finally, in the new version of the RNV3P information system, the surveillance approach will include three complementary analysis methods allowing generation of pre-alert signals: automatic data mining procedure based on statistical tests of disproportionate as used in pharmacovigilance which singles out potentially emergent disease-exposure couples, occupational exposome approach applied both on pre-selected single diseases and classes of diseases and, clinical alerts procedure. It is expected that the combined strength of the three

approaches to be altogether synergic for surveillance and detection of emerging OHPs: a pre-alert signal generated by one method may be also investigated by the others.

6. Conclusion

The occupational exposome approach opens up new horizons and we have hardly begun to explore the wide possibilities of explanation that it offers. The occupational exposomes could be analysed at several levels: at the level of a single disease (as illustrated above for the NHLs) or at the level of several diseases taking into account one, two or three dimensions of the occupational exposure (i.e., hazards, occupation and activity). Each scale of analysis has a corresponding occupational exposome, providing knowledge on the multi-exposure potentially related to the disease. Analysis at the level of several different diseases would result in an interlocking, hierarchical structure of exposomes. An additional interesting feature is that as the RNV3P is a growing database the exposome evolves over time as it has been illustrated in the exposome of NHLs without further analysis. In the framework of monitoring and surveillance over time, one may thus study trends in time of the exposomes as the dynamic of evolving and growing networks.

In prospective, for instance, one may make use of the underlying dynamic like in preferential attachment model (in which node links can be weighted depending on the number of nodes connected to it, i.e., the node degree) for generating expected occupational exposomes, and thus expected OHPs. Finally, this approach could be further applied to a global analysis of health problems that would include in addition to the RNV3P database other types of data, variables and descriptors such as, for instance, those used in the diseasome and the Wild like exposome, or approach employed by Patel et al. [30]. Such a step forward greatly exceeds the already complex framework of the RNV3P and necessitates going into and/or combining additional information and other databases.

GLOSSARY and ABBREVIATIONS

RNV3P: (Réseau National de Vigilance et de Prévention des Pathologies Professionnelles) the National Occupational Disease Surveillance and Prevention Network is a network of physician specialists from 30 occupational pathology centers of university hospitals in metropolitan France. Any patient diagnosed in these centers as having a disease potentially related to occupational exposures is recorded as an occupational health problem (OHP) in the national database.

OHP: corresponding to a patient case in the RNV3P database, an occupational health problem is an association of a disease (or pathology) and a composite occupational exposure (hazards, occupation and/or activity sector) of which one or several hazards are potentially causative.

D_{η} – occupational exposome: undirected network of OHPs having at least D components of the occupational exposure in common, and in which each OHP appears at least η times in the database.

D_{η} – occupational exposure group: cluster of OHPs, in the D_{η} – occupational exposome, sharing exactly D identical hazards of the occupational exposure.

NHL : Non-Hodgkin Lymphoma

Acknowledgements.

The authors thank the French Agency for Food, Environmental and Occupational Health and Safety (Agence nationale de sécurité sanitaire, ANSES) for supporting this work, the national health insurance organization (Caisse Nationale d'Assurance Maladie, CNAM) for funding of the consultations and participation in funding of the RNV3P, the engineers of the regional branches of the national health insurance organization (Caisses Régionales d'Assurance Maladie, CRAM), Sylvette Liaudy for her help with bibliography, Lynda Larabi for her contribution to the processing and quality control of data, the staff of the occupational health consultation centers who supply the RNV3P with data (C. Doutrelot-Philippon (Amiens), D. Penneau-Fontbonne Y. Roquelaure (Angers), I. Tahon (Besançon), P. Brochard, C. Verdun-Esquer (Bordeaux), J.D. Dewitte (Brest), M. Letourneux (Caen), M.F. Marquignon

(Cherbourg), A. Chamoux, L. Fontana (Clermont-Ferrand), J.C. Pairon (Créteil), H.J. Smolik (Dijon), J. Ameille, A. d'Escatha (Garches), A. Maitre, E. Michel (Grenoble), A. Gislard (Le Havre), P. Frimat, C. Nisse (Lille), D. Dumont (Limoges), A. Bergeret, J.C. Normand (Lyon), M.P. Le Hucher-Michel (Marseille), C. Paris (Nancy), D. Dupas, C. Geraut (Nantes), D. Choudat (Paris – Cochin), R. Garnier (Paris – Fernand Widal), D. Leger (Paris – Hotel-Dieu), E. Ben-Brik (Poitiers), F. Deschamps (Reims), A. Caubet, C. Verger (Rennes), J.F. Caillard, J.G. Gehanno (Rouen), D. Faucon (Saint-Etienne), A. Cantineau, (Strasbourg), J.M. Soulat (Toulouse), G. Lasfargues (Tours), the sentinel physicians of the occupational health services for recording and transmitting incident occupational health reports, and Nina Crowte for assistance in the translation from French of the manuscript.

References.

1. Chen Y, Turner S, Hussey L, Agius R. A study of work-related musculoskeletal case reports to the Health and Occupation Reporting network (THOR) from 2002 to 2003. *Occup Med (Lond)* 2005, 55:268-274.
2. Turner S, Lines S, Chen Y, Hussey L, Agius R. Work-related infectious disease to the Occupational Disease Intelligence Network and THE Health and Occupation Reporting network in the UK (2000-2003). *Occup Med (Lond)* 2005,55:275-281.
3. Walsh L, Turner S, Lines S, Hussey L, Chen Y, Agius R. The incidence of work-related illness in the UK health and social work sector : The Health and Occupation Reporting network 2002-2003. *Occup Med (Lond)*, 55:262-267.
4. Bonneterre V, Bicout DJ, Larabi L, Bernardet C, Maitre A, Tubert-Bitter P, De Gaudemaris R. Detection of emerging diseases in occupational health : usefulness and limitations of the application of pharmaco-vigilance methods to the database of the French National Occupational Disease Surveillance and Prevention network (RNV3P). *Occup Environ Med* 2008, 65:32-37.
5. Bonneterre V, Faisandier L, Bicout D, Bernardet C, Piollat J, Ameille J, De Clavière C, Aptel M, Lasfargues G, De Gaudemaris R, for RNV3P. Programmed health surveillance and detection of emerging diseases in occupational health: contribution of the French national occupational disease surveillance and prevention network (RNV3P). *Occup Environ Med* 2009, 67: 178-186.
6. Faisandier L, Bonneterre V, De Gaudemaris R, Bicout DJ. Development of a statistical method to detect emerging events : application to the French national

- occupational disease surveillance and prevention network (Réseau National de Vigilance et de Prévention des Pathologies Professionnelles, RNV3P) (Translated from French). *Epidémiol et Santé Anim* 2007, 51: 111-118.
7. Faisandier L, De Gaudemaris R, Bicout DJ. Occupational Health Problem Network : The Exposome. [<http://arxiv.org/abs/0907.3410>] Unpublished results
 8. Wild CP. Complementing the genome with an “exposome”: the outstanding challenge of environmental exposure measurement in molecular epidemiology. *Cancer Epidemiol Biomarkers Prev* 2005, 14: 1847-1850.
 9. Barabási AL. Network medicine – from obesity to the “diseasome”. *N Engl J Med* 2007, 357(Suppl 4), 404-407.
 10. Goh KI, Cusick ME, Valle D, Childs B, Vidal M, Barabasi AL. The human disease network. *Proc Natl Acad Sci U S A* 2007, 104: 8685-8690.
 11. Christakis NA, Fowler JH. The spread of obesity in a large social network over 32 years. *N Engl J Med* 2007, 357: 370-379.
 12. Müller AM, Ihorst G, Mertelsmann R, Engelhardt M. Epidemiology of non-Hodgkin’s lymphoma (NHL) : trends, geographic distribution, and etiology. *Ann Hematol* 2005, 84:1-12.
 13. Alexander DD, Mink PJ, Adami HO, Chang ET, Cole P, Mandel JS, Tricholopoulos D. The non-Hodgkin lymphomas: a review of the epidemiologic literature. *Int J Cancer* 2007, 120(Suppl 12):1-39.
 14. Fabro-Peray P, Daures JP, Rossi JF. Environmental risks factors for non-Hodgkin’s lymphoma : a population-based case-control study in Languedoc-Roussillon, France. *Cancer Causes Control* 2001, 12(Suppl 3): 201-212.
 15. Viel JF, Arveux P, Bavarel J, Cahn JY. Soft-tissue sarcoma and non-Hodgkin’s lymphoma clusters around a municipal solid waste incinerator with high dioxin emission levels. *Am J Epidemiol* 2000, 152(Suppl 1):13-19.
 16. Borgatti SP, Everett MG, Freeman LC *Ucinet for Windows: Software for Social Network Analysis*. Harvard, MA 2002: Analytic Technologies.
 17. Newman ME, Watts DJ, Strogatz SH. Random graph models of social networks. *Proc Natl Acad Sci USA* 2002, 99(Suppl 1):2566-2572.
 18. Watts DJ, Strogatz SH. Collective dynamics of 'small-world' networks. *Nature* 1998, 393(6684):440-2.
 19. Barabási AL, Albert R. Emergence of scaling in random networks. *Science* 1999, 286:509-512.

20. Barabási AL. The architecture of complexity. *IEEE Control Systems Magazine* 2000, 27(4), 33-42.
21. Palla G, Derenyi I, Farkas I, Vicsek T. Uncovering the overlapping community structure of complex networks in nature and society. *Nature* 2005, 435:814-818.
22. Park J, Barabási AL. Distribution of node characteristics in complex networks. *Proc Natl Acad Sci U S A* 2007, 104:17916-17920.
23. Yook, S.H Oltvai, Z.N. Barabasi, AL. Functional and topological characterization of protein. *Interaction networks* 2004, 4:928-942.
24. Albert R, Barabási AL. Statistical mechanics of complex networks. *Rev. Mod. Phys* 2002, 74:47-97.
25. Vallabhajosyula RR, Chakravarti D, Lutfeali S, Ray A, Raval A. Identifying hubs in protein interaction networks. *PLoS one* 2009;4(4):e5344.
26. He X, Zhang J. Why do hubs tend to be essential in protein networks? *PLoS Genet* 2006, 2(6):e88.
27. Newman MEJ. Assortative mixing in networks. *Phys. Rev. Lett.* 2002, 89:208701.
28. Adamcsek B, Palla G. CFinder : Locating cliques and overlapping in modules in biological networks. *Bioinformatics* 2006, 22(8):1021-1023.
29. Onnela J-P, Saramäki J, Hyvönen J, Szabo G, Lazer K, Kertesz J, Barabási AL. Structure and tie strengths in mobile communication networks. *Proc Natl Acad Sci U S A* 2007, 104(18):7332-7336.
30. Patel CJ, Bhattacharya J, Butte AJ. An Environment-Wide Association Study (EWAS) on Type 2 Diabetes Mellitus. *PLoS One* 2010, 5(5): e10746.

Résumé

6. Formalisation des Exposomes

A partir du concept des réseaux relationnels, nous avons développé une approche, nommée Exposome, qui consiste à étudier les ressemblances entre les PST.

L'Exposome professionnel est défini comme un réseau de PST connectés entre eux s'ils partagent au moins un élément de l'exposition professionnelle.

Pour une pathologie donnée, l'exposome décrit le spectre d'exposition auxquels les individus ont été exposés. Une des premières exploitations de l'approche par Exposome est l'identification de groupes d'exposition, définis comme un ensemble de PST qui partagent un ou plusieurs éléments de l'exposition professionnelle.

7. Caractérisation des Exposomes

Caractériser un réseau consiste à caractériser les particularités structurelles de son organisation et fait appel à des métriques, ou indicateurs, largement utilisées dans la science des réseaux sociaux. Ces métriques permettent de décrire les propriétés des nœuds, mais également des groupes d'exposition au sein de l'Exposome. L'identification de ces caractéristiques, à l'aide de métriques, permet de comprendre les propriétés topologiques des nœuds qui structurent un tel système [Barabási 2007 b ; Park 2007]. L'approche par Exposomes constitue une nouvelle manière d'exploiter les données de la base RNV3P et offre de larges possibilités d'analyse [The PLoS Medicine Editors 2008].

7.1. A l'échelle des nœuds

7.1.1. Degré de connectivité

Une des métriques les plus utilisées est le degré de connectivité d'un nœud i , noté k_i . Il s'agit d'une mesure individuelle qui consiste à compter le nombre de liens L_i du nœud i . En pratique, on somme chacune des lignes de la matrice adjacente, notée C_i . On note $P(k)$ la distribution des degrés de connectivité k dans un graphe, telle que :

$$\sum_{i=1}^N P(k_i) = 1 \quad (7.1)$$

Par exemple...

La Figure 20 illustre deux cas possibles. Dans le cas A, le nœud est dit isolé, ou indépendant, dans la mesure où il ne comporte pas de liens avec d'autres nœuds ($k_i = 0$). Dans le cas B, le nœud v_i est connecté à trois autres nœuds-voisins, son degré de connectivité est donc $k_i = 3$.

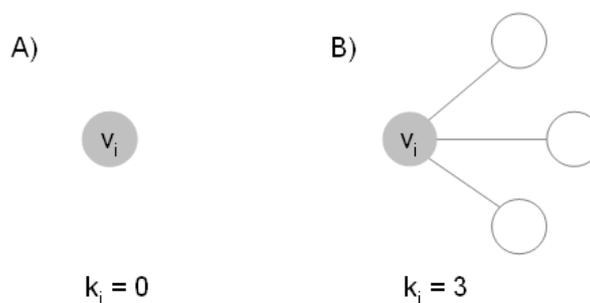


Figure 20 : Illustration du degré de connectivité.

Le degré de connectivité permet également de quantifier la popularité du nœud, définie comme le nombre de liens avec les autres nœuds. Sur la base de cette métrique, la densité du réseau mesure la cohésion des nœuds au sein d'un graphe, ou la probabilité pour que deux nœuds soient connectés. Dans un réseau non dirigé, comme celui de l'Exposome, la mesure de densité s'écrit :

$$d = \frac{L}{2N(N-1)} \quad (7.2)$$

Avec L le nombre total de liens observés et N le nombre total de nœuds.

7.1.2. Coefficient d'agglomération

Le coefficient d'agglomération, aussi appelé coefficient de clustering dans le langage de la théorie des réseaux, apporte plus de précision au degré de connectivité puisqu'il mesure la vraisemblance que les voisins d'un nœud soient également associés entre eux. Le coefficient d'agglomération C_i du nœud i est noté :

$$c_i = \frac{L_i(m)}{\text{Max}(L_i(m))} \quad (7.3)$$

Avec $L_i(m)$ le nombre de liens observés L_i entre les m voisins du nœud i et $\text{Max}(L_i(m))$ le nombre de liens possibles entre les voisins du nœud i . Le coefficient de clustering mesure la probabilité que deux nœuds connectés à un même nœud soient également connectés entre eux [Watts 1998].

Par exemple...

La Figure 21 illustre trois calculs possibles du coefficient d'agglomération. Si les voisins d'un nœud v_i ne sont pas connectés entre eux, alors le coefficient d'agglomération est nul (cas A). Si un nœud est relié à 3 voisins dont deux sont en relation, alors le coefficient de clustering est égal à 1/3 (cas B). Si tous les voisins d'un nœud sont reliés entre eux, alors le coefficient de clustering est égal à 1 (cas C).

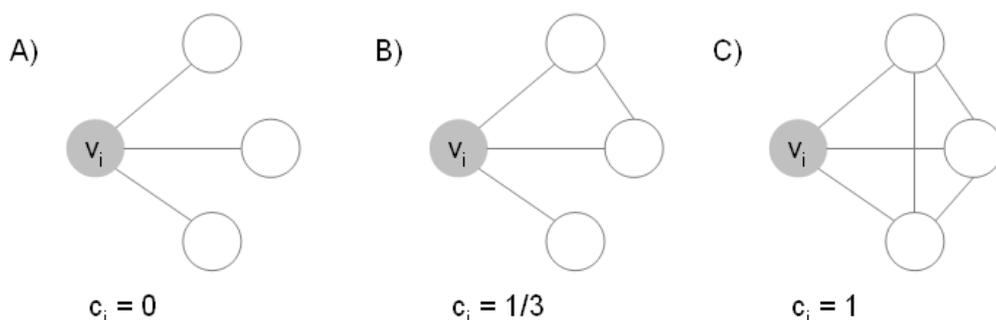


Figure 21 : Illustration du coefficient d'agglomération.

Cette métrique caractérise la place d'un nœud par rapport à ses voisins. Dans le cas C, le coefficient d'agglomération nous apporte une information essentielle : ce nœud comporte 3 voisins, tous connectés entre eux.

7.1.3. Similarité

Deux nœuds i et j sont susceptibles de partager les mêmes expositions professionnelles. Nous avons ainsi développé un indicateur qui mesure la similarité des nœuds entre eux, notée s_{ij} , en rapportant le nombre de nuisances communes aux nœuds i et j ($v_i \cap v_j$) par rapport au nombre de nuisances distinctes associées à ces deux nœuds ($v_i \cup v_j$), tel que :

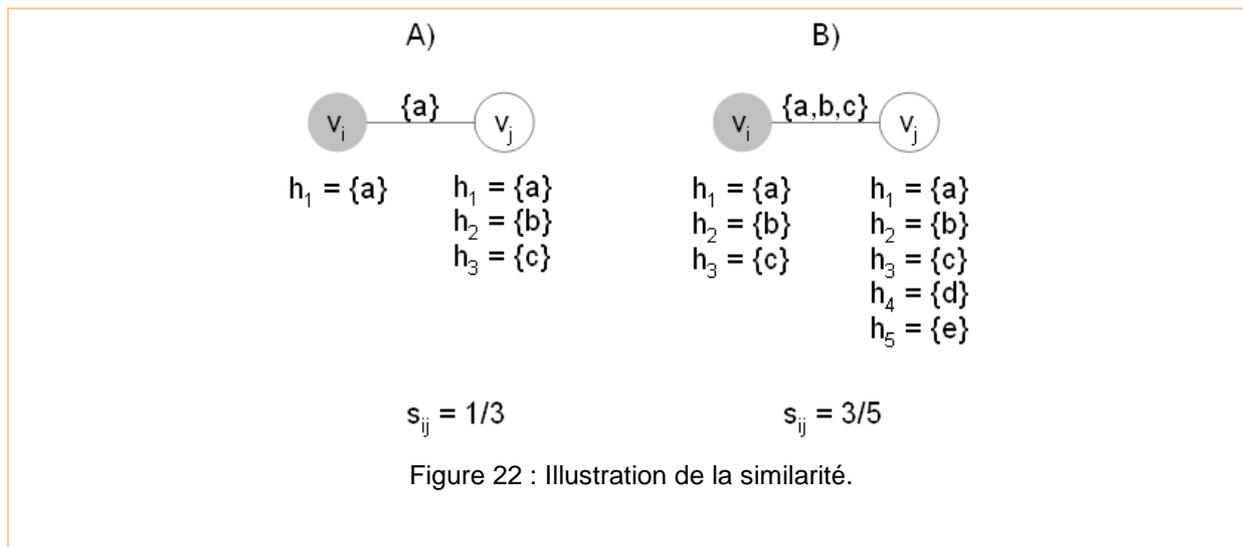
$$s_{ij} = \frac{v_i \cap v_j}{v_i \cup v_j} \quad (7.4)$$

Ainsi, le degré de similarité d'un nœud i avec les autres nœuds du réseau est donné par $\text{Max}(S_i)$.

NB : Le nombre de nuisances partagées entre deux nœuds détermine le nombre de groupes d'exposition auxquels ils appartiennent conjointement.

Par exemple...

Rappelons que chaque nœud est unique et, par conséquence, 2 nœuds peuvent partager au maximum 4 nuisances. La Figure 22 illustre la mesure de la similarité entre 2 nœuds. Dans le cas A, le nœud v_i est associé à une seule nuisance a , et le nœud v_j est associé à 3 nuisances qui sont a , b et c . Ces deux nœuds ont en commun la nuisance a . Le coefficient de similarité est donc égal à $1/3$. Dans le cas B, le nœud v_i est associé à 3 nuisances, et le nœud v_j à 5 nuisances. Ces deux nœuds partagent 3 nuisances. Le coefficient de similarité est donc égal à $3/5$.



7.2. Identification de groupes d'exposition

Une des premières exploitations de l'approche par Exposome est l'identification des groupes d'exposition, définis comme un ensemble de nœuds qui partagent un ou plusieurs éléments de l'exposition professionnelle. Ces groupes d'exposition permettent de structurer les connaissances acquises dans la base RNV3P pour une pathologie donnée. A partir de ces groupes, l'enjeu est de comprendre comment les individus s'intègrent dans la structure globale de l'Exposome. En d'autres termes, l'identification des règles de connectivité entre les nœuds nous permettrait d'une part de comprendre comment les PST sont structurés, et d'autre part de suivre leurs évolutions au cours du temps.

7.2.1. Groupe d'exposition simple

On parle de groupe d'exposition simple lorsque les nœuds partagent exactement la même exposition. Ce groupe porte alors le nom de l'élément partagé entre les nœuds. Le poids du groupe d'exposition simple est défini par le nombre de nœuds inclus dans le groupe, et pas le nombre total de PST décrits.

Notre définition du groupe d'exposition s'apparente à celle d'un *clique* dans le langage de la théorie des graphes. Au sens strict, un *clique* est un ensemble de n nœuds connectés entre eux par exactement $n-1$ liens. Cette définition d'un clique n'impose aucune restriction sur la nature des liens.

7.2.2. Groupe d'exposition multiple

7.2.2.1. Groupe d'exposition hybride

Dans le contexte de l'Exposome, les cliques peuvent former des groupes d'exposition "hybrides", définis comme un ensemble de n nœuds reliés entre eux par exactement $n-1$ liens pouvant être de natures différentes (plusieurs éléments de l'exposition sont partagés).

La Figure 23 illustre la formation des groupes d'exposition ainsi définis. Dans le cas A, les nœuds forment un groupe d'exposition simple parce qu'ils sont reliés par la ou les mêmes expositions. Dans le cas B, le groupe d'exposition A est formé de nœuds reliés par exactement la ou les mêmes éléments de l'exposition professionnelle, dont deux nœuds partagent un élément supplémentaire de l'exposition. Les groupes C et D ont en communs deux nœuds qui partagent les mêmes expositions, à des niveaux différents.

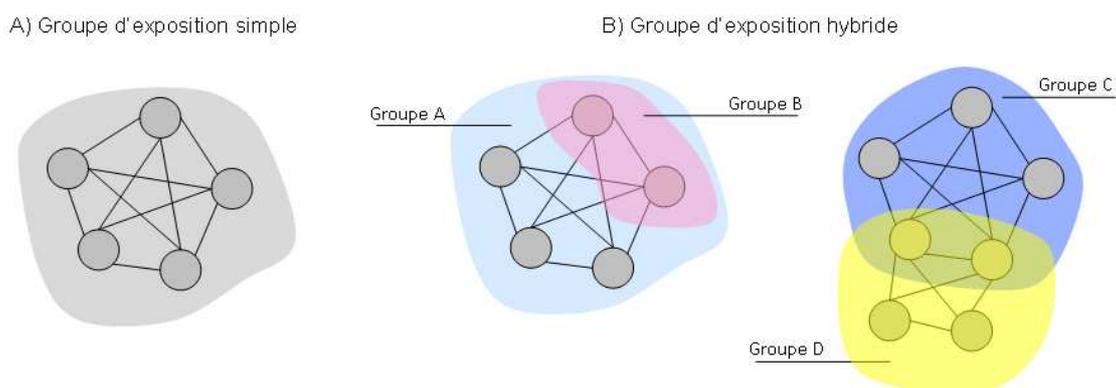


Figure 23 : Illustration des groupes d'exposition. A) Groupe d'exposition simple : Les nœuds sont reliés entre eux par le même élément de l'exposition. B) Groupes d'exposition hybride : Les nœuds partagent une ou plusieurs expositions de nature différente.

7.2.2.2. Recouvrement des groupes d'exposition

Les nœuds multi-exposés peuvent appartenir à plusieurs groupes d'exposition, et ainsi faire le lien entre des expositions professionnelles. Nous avons développé un indicateur qui permet de mesurer le recouvrement des groupes d'exposition. Soit r_{ij} le degré de recouvrement des groupes i et j , noté :

$$r_{ij} = \frac{N_i \cap N_j}{N_{ij}} \quad (7.5)$$

Avec $N_i \cap N_j$ le nombre de nœuds communs aux groupes d'exposition i et j , et N_{ij} le nombre de nœuds appartenant à la fois aux groupes i et j .

Par exemple...

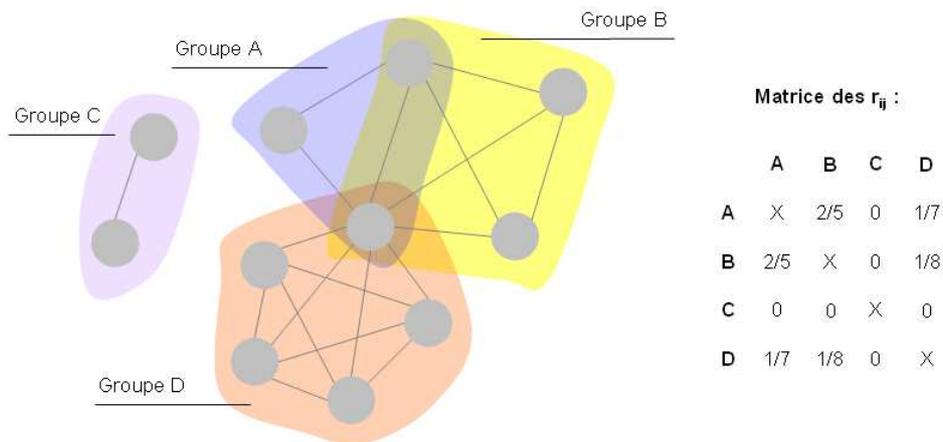


Figure 24 : Illustration du recouvrement des groupes d'exposition.

7.2.2.3. Coefficient d'agglomération des groupes d'exposition

Dans la littérature scientifique, Yip et al. ont développé une méthodologie appelée "Generalized Topological Overlap Measure" qui permet d'étudier, à l'aide d'indicateurs quantitatifs, le chevauchement des groupes de nœuds dans un réseau [Yip 2007]. Cette approche a été reprise par Onnela et al. pour étudier la structure d'un réseau de communication [Onnela 2007]. Ces auteurs mesurent le recouvrement relatif au voisinage de deux nœuds, représenté par la proportion d'amis communs entre deux nœuds i et j , notée O_{ij} telle que :

$$O_{ij} = \frac{n_{ij}}{(k_i - 1) + (k_j - 1) - n_{ij}} \quad (7.6)$$

Avec n_{ij} le nombre de voisins communs aux nœuds i et j , k_i et k_j décrivent les degrés de connectivité associés aux nœuds i et j . Lorsque $O_{ij} = 0$, cela signifie que les nœuds i et j ne présentent pas de points communs. A l'inverse, si deux nœuds font partis du même voisinage, alors $O_{ij} = 1$.

Par exemple...

La Figure 25 illustre quatre valeurs possibles de O_{ij} . Lorsque $O_{ij} = 0$, alors les nœuds i et j n'ont pas de voisins communs.

Dans la seconde situation, les nœuds i et j ont 4 voisins chacun et en partagent 2. La valeur de O_{ij} est donc de $2/6$ ou $1/3$.

Lorsque $O_{ij} = 2/3$, le nœud i partage ses 4 voisins avec le nœud j le nœud j partage 2 voisins sur 4, donc $O_{ij} = 4/6$ ou $2/3$.

Lorsque O_{ij} est égal à 1 (maximum) alors les voisins du nœud i sont également les voisins du nœud j .

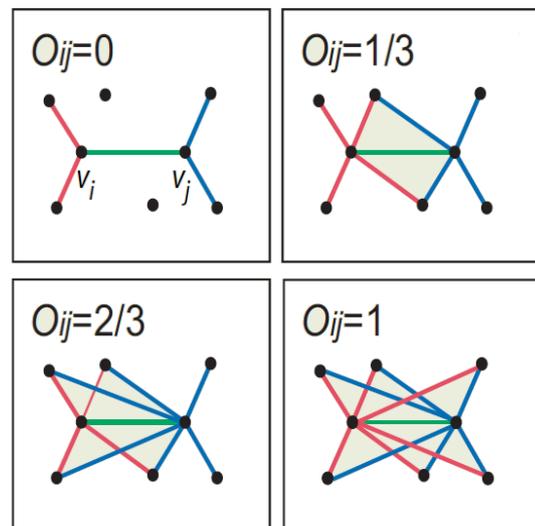


Figure 25 : Illustration du recouvrement "élaboré" des groupes d'exposition. Illustration extraite de [Onnella 2007].

7.3. Application aux données RNV3P

L'Exposome fournit une cartographie structurée des expositions professionnelles potentiellement à l'origine de la pathologie, et partagées par les PST. Nous explorons dans un premier temps la topologie de l'Exposome à partir de l'échantillon des 171 PST (90 nœuds) associés aux Sclérodermies Systémiques au cours de la période 2001-2009 dans la base RNV3P.

7.3.1. Topologie des nœuds

L'Exposome des Sclérodermies Systémiques présente les 171 PST, caractérisés en 90 séquences ou nœuds, de poids supérieur ou égal à 1, enregistrés au cours de la période 2001-2009 (Figure 26). Chaque nœud décrit un PST, défini comme association entre la pathologie étudiée et l'ensemble des nuisances professionnelles associées. La taille d'un nœud est proportionnelle au nombre de PST identiques ainsi décrits.

Le 1_1 -Exposome (obtenu en prenant les nœuds comportant au moins $\eta = 1$ PST et reliés par au moins $D = 1$ nuisance) est un réseau peu dense (densité = 0,051) où 31 nuisances distinctes sont partagées entre les nœuds. Les nœuds décrivant un PST unique représentent plus de 85% des nœuds de l'Exposome (soit 78 / 90 nœuds). On dénombre ainsi 14 nœuds isolés (non connectés) et 76 nœuds connectés, partageant au moins un élément de l'exposition professionnelle.

Au regard du 1₁-Exposome, trois ensembles de nœuds hautement connectés occupent une position centrale, il s'agit des groupes d'exposition des Solvants et diluants organiques (43 PST), Silice sable (37 PST) et Silice cristallisée (32 PST).

En excluant les nœuds qui ne décrivent qu'un seul PST, le 1₂-Exposome présente une autre dimension incluse dans la structure du 1₁-Exposome et illustre également l'une des conditions du test 0 de susceptibilité émergentielle, précisément le test des fréquences minimales. Le 1₂-Exposome (obtenu en prenant les nœuds comportant au moins $\eta = 2$ PST et reliés par au moins $D = 1$ nuisance) représente seulement 13% des nœuds (soit 12 / 90 séquences) mais concentrent toutefois 53% des PST enregistrés (soit 91 / 171 PST). Les groupes d'exposition restants sont reconnus au titre des Tableaux MPI et la plupart des nœuds sont isolés.

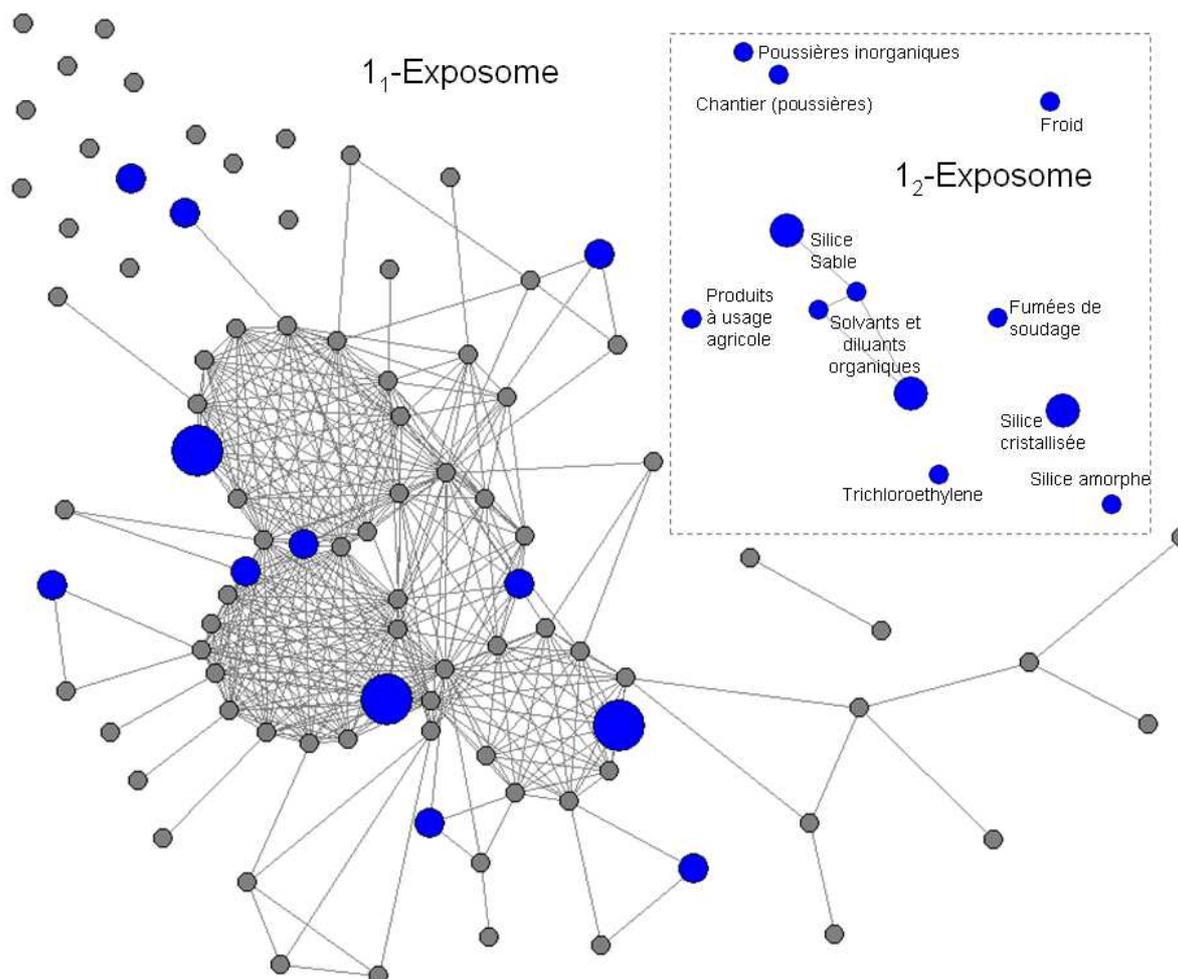


Figure 26 : 1₁-Exposome et 1₂-Exposome des Sclérodermies Systémiques (Données émergence RNV3P 2001-2009, N = 77 913 PST hors "99999 - Pas de Nuisance Identifiée").

Sur la base de ces observations, nous nous intéressons à l'échelle des groupes d'exposition, plutôt qu'à l'échelle individuelle des nœuds, pour mettre en évidence les connaissances acquises dans la base RNV3P.

7.3.2. Topologie des groupes d'exposition

En structurant l'Exposome en groupes d'expositions, les similitudes entre les nœuds associées à une pathologie sont mises en évidence. Ainsi, la Figure 27 illustre les Exposomes des groupes d'exposition. Chaque carré représente un groupe d'exposition simple (ou nuisance partagée), composé d'au moins 2 nœuds ($\eta \geq 2$). La présence d'au moins 1 nœud appartenant à deux groupes d'exposition est marquée par un lien entre ces deux groupes. Pour optimiser la lisibilité sur cette figure, nous avons inscrit les codes des expositions partagées plutôt que les libellés. La Figure 28 fait référence aux libellés de ces groupes d'exposition.

Nous orientons notre analyse sur les similitudes entre les nœuds plutôt que sur leurs caractéristiques individuelles. De cette manière, les connaissances sont structurées sous la forme de groupes de nœuds partageant des similitudes en termes d'expositions professionnelles (1₃-Exposome Figure 27). Le recouvrement des groupes d'exposition est un marqueur de multi-exposition. En effet, lorsque deux groupes d'exposition se chevauchent, cela signifie qu'un nœud appartient à la fois à ces deux groupes, et est donc composé d'au minimum deux éléments de l'exposition professionnelle. Les nœuds peuvent ainsi partager jusqu'à 4 nuisances entre eux.

Sur ce principe, les groupes d'exposition simples sont appelés groupes d'exposition d'ordre 1, les groupes d'exposition qui se chevauchent deux à deux sont des groupes d'ordre 2, etc. Les Sclérodermies Systémiques sont associées à 83 nuisances distinctes au cours de la période 2001-2009. Parmi ces nuisances, seulement 31 sont partagées entre les 90 nœuds de l'Exposome et représentent des groupes d'exposition simple (ordre 1). On dénombre 55 groupes d'exposition d'ordre 2, 18 groupes d'ordre 3, 2 groupes d'ordre 4 et 1 groupe d'ordre 5. La Figure 28 indique la présence ou l'absence des nuisances partagées pour chaque niveau de recouvrement des groupes d'exposition (D).

Sur la base de cette topologie, il serait instructif de connaître la probabilité pour qu'un nœud, pris au hasard, appartienne à un ou plusieurs groupes d'exposition. Cette expérience fournirait des connaissances non seulement sur la multi-exposition associée à un nœud, mais aussi sur les similitudes avec les autres nœuds en termes d'expositions professionnelles partagées.

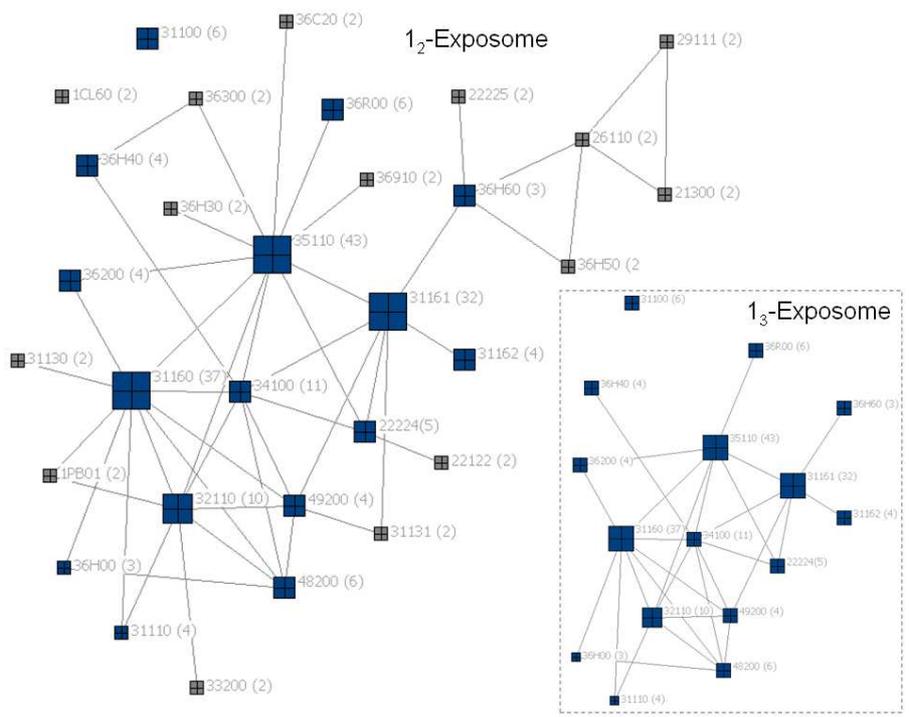


Figure 27 : 1₂- Exposome et 1₃-Exposome des groupes d'expositions associés aux Sclérodermies Systémiques. Entre parenthèses, le nombre de PST dans chaque groupe d'exposition.

	D = 1	D = 2	D = 3	D = 4	D = 5
35110 - Solvants et diluants organiques					
31160 - Silice sable					
31161 - Silice cristallisée					
34100 - Fumées de soudage					
32110 - amiante (Fibres)					
31100 - Composés inorganiques					
48200 - Froid					
36R00 - Produits à usage agricole					
22224 - Trichloroethylene					
31110 - Chantier (Poussières)					
31162 - Silice amorphe					
36200 - Peintures Vernis Laques Mastics					
49200 - Vibrations					
36H40 - Dégraissant					
36H00 - Produits d'entretien					
36H60 - Lessives détergent					
21300 - Hydrocarbures aromatiques					
22122 - Dichlorométhane					
22225 - Tetrachloroethylene					
26110 - Ethers					
29111 - Formaldéhyde méthanal					
31130 - Ciment (Poussières)					
31131 - Béton (Poussières)					
33200 - Gaz d'échappement					
36300 - Adhésifs colles					
36910 - Encres					
1CL60 - Hypochlorites inorganiques					
1PB01 - Plomb					
36C20 - Huiles et graisses lubrifiantes					
36H30 - Décapant					
36H50 - Désinfectants					

Figure 28 : Présence / absence des nuisances dans les groupes d'exposition d'ordre 1 à 5. En vert, les nuisances reconnues au titre des tableaux MPI, en rouge les nuisances non reconnues.

* * * * *

L'idée principale de l'approche par exposome est d'investiguer l'exposition professionnelle commune aux PST en organisant les connaissances de la base RNV3P sous la forme d'un réseau, caractérisable par des métriques spécifiques. Nous avons ainsi structuré l'exposome en groupes d'exposition professionnelle, définis comme sous-ensembles d'information ou spectre des expositions professionnelles potentiellement liés à la maladie. Les données du RNV3P peuvent ainsi être analysées à la fois à l'échelle individuelle (PST) et à l'échelle collective (groupes d'exposition).

Si les exposomes permettent de comprendre et structurer la multi-exposition associée à une pathologie et de suivre ces connaissances en cours du temps, cette approche ne permet pas de pointer des profils atypiques d'exposition professionnelles, mais de générer des hypothèses de recherche suite à la présence d'un ou plusieurs facteurs d'expositions professionnelles associés à la pathologie. La détection des PST potentiellement émergents est développée et mise en œuvre dans la partie suivante.

Résumé

7. Caractérisation des Exposomes

L'approche par Exposomes fournit une cartographie des expositions, ou spectre des expositions professionnelles, potentiellement responsable de la survenue de la pathologie. Sur la base de cette organisation, un groupe d'exposition est défini comme un ensemble de nœuds partageant au moins "D" éléments de l'exposition professionnelle. On distingue les groupes d'exposition simples où les nœuds partagent exactement la même nuisance, des groupes d'expositions hybrides où la nature et le nombre d'éléments partagés entre les nœuds peuvent être multiples.

La caractérisation de l'Exposome passe donc par l'identification des groupes d'exposition, mais également par la mesure des métriques propres aux réseaux relationnels à l'échelle des nœuds, type degré de connectivité, coefficient d'agglomération ou encore similarité, et à l'échelle des groupes d'exposition, type degré de recouvrement entre plusieurs groupes.

Cette approche fournit aux décideurs une description des expositions professionnelles associées à une pathologie et permet également de générer des hypothèses quant à la relation entre la pathologie étudiée et les expositions professionnelles apportées, simples ou multiples.

Partie 3 : Fonctions Expertes

"With epidemiology you can tell a little thing from a big thing. What's very hard to do is to tell a little thing from nothing at all."

Michael THUN, 1995.

Director of analytic epidemiology for the American Cancer Society.

Les disparités dans le fonctionnement des CCPP et dans le recrutement des patients, et donc la diversité d'analyse dans le processus de l'interrogatoire professionnel conduisant à la création d'un PST, justifient la construction d'une fonction qui reflète les connaissances et pratiques des médecins. Une manière de construire une telle fonction aurait été de travailler en amont (*top-down*) des données RNV3P en réalisant, par exemple, une enquête sur questionnaire auprès des médecins. Nous avons plutôt choisi de travailler en aval (*bottom-up*) des données pour construire un modèle qui décrit au mieux ces enregistrements. Nous avons ainsi introduit l'approche par "Fonction Experte" qui reproduit les choix des médecins dans l'enregistrement de la multi-exposition professionnelle associée à une pathologie donnée.

On appelle "**Fonction Experte**", un modèle probabiliste qui prend comme données d'entrée une maladie et une activité professionnelle, et fournit en sortie une séquence de une à cinq nuisances. Cette fonction génère ainsi des PST reflétant les pratiques de codage et les règles d'associativité entre une pathologie et ses expositions professionnelles multiples.

Comme son nom l'indique, cette fonction représente une modélisation des enregistrements faits par les médecins qui génèrent des PST à la suite des interrogatoires professionnels. En d'autres termes, une Fonction Experte représente la distribution potentiellement observable des PST pour une pathologie donnée, caractérisant ainsi les contributions des médecins à la base des connaissances du RNV3P. Pour construire une telle fonction, nous développons plusieurs scénarios, selon qu'on s'intéresse en premier lieu à la nuisance principale et ses co-nuisances, ou aux descripteurs d'un PST.

Les objectifs de cette partie sont de formaliser les composantes de la base RNV3P et de classer les associations pathologie - expositions observées, observables et permises. Nous construisons ainsi un cadre théorique et élaborons différentes stratégies de construction des PST selon plusieurs hypothèses, que nous avons détaillées dans cette partie. A partir de ces Fonctions Expertes, nous avons élaboré une stratégie pour détecter des PST potentiellement émergents et illustré cette approche sur trois échantillons extraits de la base RNV3P.

8. Formalisation des Fonctions Expertes

Nous formalisons les objets de la base RNV3P et construisons des descripteurs qui caractérisent les informations structurant un PST. A partir de ces notations, nous estimons la diversité des configurations possibles des variables par une simple analyse combinatoire, puis les comparons avec les configurations observées dans la base RNV3P.

8.1. Formalisation du système

La formalisation passe par la structuration du système en définissant les variables qui décrivent un PST, ses descripteurs de type qualitatif et quantitatif ainsi que les paramètres.

8.1.1. Variables

8.1.1.1. La pathologie

La pathologie, notée p , renseigne le diagnostic retenu aux termes d'investigations cliniques et para-cliniques. Cette variable est codée selon la nomenclature de la CIM-10, dont l'espace disponible, utilisé et récurrent est présenté dans la partie 1.

8.1.1.2. Les nuisances

Les nuisances retenues sont celles suspectées ou confirmées à l'origine de la pathologie et présentes dans l'environnement professionnel du patient. Cet ensemble, de une à cinq nuisances, se présente sous la forme d'un vecteur, que l'on notera $\mathbf{h} = (h_1, h_2, h_3, h_4, h_5)$, contenant les codes sélectionnés à partir du thésaurus CNAM, et enregistrés dans la base RNV3P à la suite de l'interrogatoire professionnel.

8.1.1.3. L'activité professionnelle

L'activité professionnelle potentiellement responsable de la survenue de la maladie est codifiée par le médecin à la suite de l'interrogatoire professionnel. Cette activité comprend un métier, noté m, dont la nomenclature est le CITP-88 fournie par le BIT, et un secteur d'activité, noté s, codé selon la nomenclature de la NAF-93.

8.1.1.4. Formalisation d'une séquence

Un PST résulte de la combinaison d'éléments médicaux (pathologie) et de données qualitatives d'exposition (nuisances). Les variables qui composent un PST se décrivent ainsi :

$$\text{PST} = \boxed{p} \boxed{h_1 h_2 h_3 h_4 h_5} \boxed{m} \boxed{s} \quad (8.1)$$

Où p renseigne la pathologie, h_1 la nuisance principale et h_2 à h_5 les co-nuisances associées, m le métier et s le secteur d'activité.

On appelle "**séquence**", une ou plusieurs associations pathologie - expositions professionnelles enregistrées exactement à l'identique. Le poids d'une séquence est défini comme le nombre de copies de ces PST.

Sur la base de ces variables, nous décrivons les informations contenues dans une séquence, en fournissant des descripteurs qualitatifs et quantitatifs.

8.1.2. Descripteurs

Une fois défini, un PST peut être décrit par les descripteurs suivants :

8.1.2.1. Longueur des séquences

On note L_i la longueur d'une séquence i définie comme le nombre de nuisances associées à la pathologie. Par définition, L est compris entre 1 et 5. Une séquence est dite "multi-exposée" si sa longueur est supérieure à 1, et "mono-exposée" si une seule nuisance est associée à la pathologie.

8.1.2.2. Imputabilité des couples pathologie - nuisance

Au cours de l'interrogatoire professionnel, le médecin juge de l'intensité de la relation entre la pathologie et les nuisances retenues en fournissant un indicateur subjectif d'imputabilité. Par

convention, pour les séquences dont la longueur est supérieure à 1, la nuisance principale est enregistrée avec l'imputabilité la plus forte, et les co-nuisances associées ont des degrés d'imputabilité inférieurs ou égaux à celui de la nuisance principale. La Figure 29 fournit l'échelle ordinaire des imputabilités.

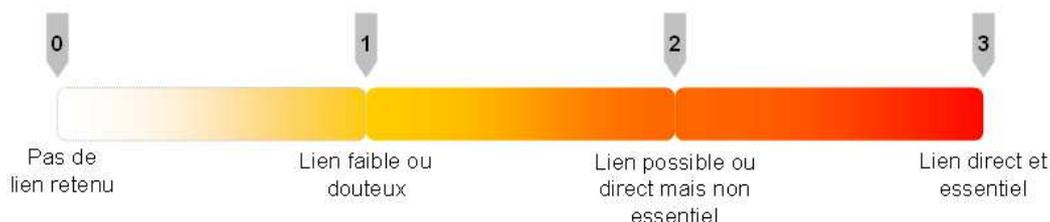


Figure 29 : Echelle ordinaire du degré d'imputabilité entre une pathologie et une nuisance.

Parmi les données émergence RNV3P 2001-2009, les degrés d'imputabilité à 4 correspondent à des avis pour aptitude formulé par l'expert au sujet de l'adéquation entre l'état de santé du patient et son activité professionnelle (51 / 81 132 pour la nuisance principale, 38 / 19 397 pour la co-nuisance 1, 11 / 5 899 pour la co-nuisance 2, 5 / 1 682 pour la co-nuisance 3, 0 / 455 pour la co-nuisance 4).

L'évaluation du niveau d'imputabilité entre une maladie et une nuisance est une tâche difficile lorsque le délai entre le début de l'exposition et la survenue de la maladie est long, pouvant aller jusqu'à plusieurs dizaines d'années pour les maladies chroniques ou les cancers. La vision subjective du médecin consultant induit une variabilité inter-opérateur dans l'enregistrement des PST. C'est pourquoi nous ne prenons pas en compte le degré d'imputabilité dans la construction des Fonctions Expertes.

8.1.2.3. Reconnaissance des couples pathologie – nuisance

L'indemnisation d'un couple pathologie-nuisance est une donnée administrative et juridique. Les maladies professionnelles indemnifiables (MPI) sont celles inscrites dans des tableaux MPI, ou celles reconnues par le Comité Régional de Reconnaissance des Maladies Professionnelles (CRRMP) :

"Art. L. 461-1, 2e al. Est présumée d'origine professionnelle toute maladie désignée dans un tableau de maladies professionnelles et contractée dans les conditions mentionnées à ce tableau."

En France, les maladies professionnelles sont essentiellement répertoriées dans des tableaux et annexes de décrets ministériels. A ce jour, 117 tableaux MPI existent, mais seulement 12 représentent 98% des déclarations.

La reconnaissance des couples pathologie - nuisance en tant que couple reconnu au titre des Tableaux MPI atteste des connaissances acquises sur la pathologie. Un couple reconnu ne peut donc pas être considéré comme potentiellement émergent.

Tableau n°30 bis du Régime Général Cancer broncho-pulmonaire provoqué par l'inhalation de poussières d'amiante Date de création : Décret du 22 Mai 1996 Dernière mise à jour : Décret du 14 Avril 2000		
Désignation de la maladie	Délai de prise en charge	Liste limitative des travaux susceptibles de provoquer cette maladie
Cancer broncho-pulmonaire primitif	40 ans (sous réserve d'une durée d'exposition de 10 ans)	Travaux directement associés à la production des matières contenant de l'amiante Travaux nécessitant l'utilisation d'amiante en vrac Travaux d'isolation utilisant des matériaux contenant de l'amiante Travaux de retrait d'amiante Travaux de pose et de dépose de matériaux isolants à base d'amiante Travaux de construction et de réparation navale Travaux d'usinage, de découpe et de ponçage de matériaux contenant de l'amiante Fabrication d'entretien ou de maintenance effectués sur des équipements contenant des matériaux à base d'amiante Travaux d'entretien ou de maintenance effectués sur des équipements contenant des matériaux à base d'amiante

Tableau 4 : Illustration d'un Tableau de Maladie Professionnelle Indemnisable.

Pour une séquence i de longueur L_i , nous noterons l_i , $0 \leq l_i \leq L_i$, le nombre de couples pathologie - nuisances reconnus au titre des Tableaux MPI.

	$l_i = 0$	$0 < l_i < L_i$	$l_i = L_i$	TOTAL
$L_i = 1$	29,3%	-	46,1%	75,4%
$L_i = 2$	4,6%	4,2%	8,3%	17,1%
$L_i = 3$	1,0%	2,3%	2,0%	5,4%
$L_i = 4$	0,3%	0,9%	0,4%	1,6%
$L_i = 5$	0,1%	0,4%	0,1%	0,6%
TOTAL	35,3%	7,8%	56,9%	100%

Tableau 5 : Répartition des proportions des séquences en fonction de leurs longueurs et du nombre de nuisances reconnues indemnissables au titre des Tableaux des Maladies Professionnelles.

Avec plus de 75% des PST de longueur 1 (soit 61 137 / 81 132 PST) dont 46% sont reconnus au titre des Tableaux MPI (soit 37 385 / 81 132 PST), l'espace des PST potentiellement émergents et multi-exposés est petit.

8.1.3. Paramètres

Les enregistrements des PST dépendent de deux principaux paramètres identifiables à partir de la base de données RNV3P : les Centre de Consultation des Pathologies Professionnelles (CCPP) et l'activité professionnelle (secteur - métier) retenue par le médecin consultant. D'autres paramètres, comme le tissu socio-économique de la zone locale ou le *turnover* des médecins consultants dans les CCPP, influent potentiellement sur les enregistrements des PST, mais sont difficilement appréciables du fait de leur dynamique temporelle peu prévisible.

8.1.3.1. Le Centre de Consultation

La provenance des PST constitue le principal déterminant dans le recueil des données. En effet, des paramètres tels que la spécialité du centre ou sa localisation géographique peuvent induire un biais dans le recrutement des patients ou dans le choix des expositions professionnelles, incluant les médecins consultants et leur spécialité médicale. Par exemple, le CCPP de l'Hôtel Dieu enregistre essentiellement des troubles du sommeil (code G47), qui représentent 37% des PST enregistrés pour ce centre au cours de la période 2001-2009 (soit 136 / 362 PST).

8.1.3.2. L'activité professionnelle

Le secteur d'activité et le métier supposés responsables de la survenue de la pathologie structurent les nuisances professionnelles associées à une pathologie. En effet, pour un secteur-métier donné, correspond un spectre d'exposition spécifique. Par exemple, le métier de coiffeur (code 5141, 2 438 PST) est caractérisé par une ensemble de nuisances récurrentes (utilisées chaque année), telles que teinture (cosmétiques) (code 36J12), oxydants et décolorants (cosmétiques) (code 36J11), phénylénédiamine (code 2Q319), etc.

Pour illustrer l'impact de ces deux paramètres, nous avons choisi l'exemple du CCPP de Brest, dont les enregistrements de PST sont caractérisés par 22% (soit 895 / 3 895 PST) de patients exerçant le métier d'artisans et ouvriers des métiers de la métallurgie (code 35) dans le secteur de la construction des matériels de transports (code 72), et principalement celui de la construction navale. Parmi ces patients, 99% (soit 888 / 895 PST) sont exposés aux fibres d'amiante (code 32110). Ce CCPP présente un processus d'enregistrement des PST spécifique à son environnement géographique et socio-économique. Il serait donc intéressant de paramétrer les Fonctions Expertes selon la provenance des CCPP pour l'étude de cette activité professionnelle.

Les variables, paramètres et descripteurs d'un PST définis, nous sommes en mesure d'estimer les configurations possibles des PST pour les comparer aux configurations observées.

8.2. Configuration des séquences

Une simple analyse descriptive de la base RNV3P montre que les pathologies sont associées de manière combinatoire à plusieurs expositions professionnelles (nuisances, métier et secteur d'activité). Cette observation nous amène alors à poser les questions suivantes. A partir d'une pathologie donnée,

- 1) Quelle est la diversité des séquences possibles ?
- 2) Quelle est la diversité des séquences observées dans la base RNV3P ?
- 3) Toutes les séquences sont-elles observables et permises ?

Les réponses à ces questions suggèrent d'aborder le problème sous un angle probabiliste. Si la réponse à la question 1) est un problème simple d'analyse combinatoire, et la question 2) un dénombrement des cortèges enregistrés dans la base RNV3P, la question 3) nécessite d'introduire une fonction probabiliste.

8.2.1. Estimation des séquences possibles

A partir de la formalisation des variables du système, nous estimons le nombre de séquences possibles. Soit H le nombre total de nuisances associables à une pathologie et L la longueur des séquences. Le nombre total d'associations entre une pathologie et une nuisance est donc donné par :

$$N_{pst} = \sum_{L=1}^5 \binom{H}{L} \theta(H-L) = \sum_{L=1}^5 \frac{H!}{L!(H-L)!} \theta(H-L) \quad (8.2)$$

Où chaque terme $\binom{H}{L}$ représente le nombre de PST de longueur L , et $\theta(x) = 1$ si $x \geq 0$ et $\theta(x) = 0$

si $x < 0$. Par exemple, pour $H = 5$, on a $N_{pst} = 5 + 10 + 10 + 5 + 1 = 31$ combinaisons possibles et pour $H = 10$, on dénombre $N_{pst} = 10 + 45 + 120 + 210 + 252 = 637$ combinaisons possibles.

Le nombre de PST possibles croît exponentiellement avec H . Ce calcul suppose que les longueurs des séquences soient de même fréquence. Or, la Figure 30 met en évidence les probabilités d'obtenir une séquence de longueur L , notée P_L . Par exemple, 75% des PST enregistrés dans la base RNV3P sont mono-exposés, c'est-à-dire associés à une seule nuisance (soit 61 137 / 81 132 PST).

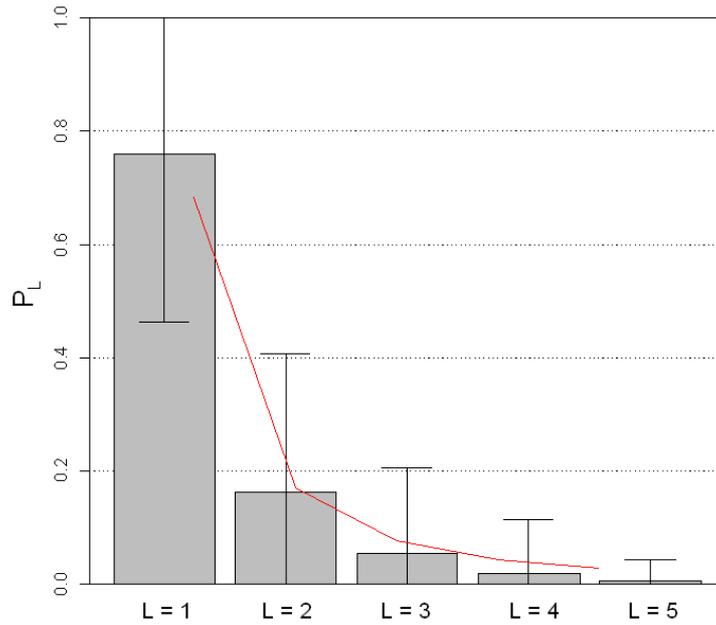


Figure 30 : Probabilité moyenne des longueurs des PST, notée P_L (moyenne +/- écart type). La répartition est décrite par la loi : $P(L) \sim 0.6832L^{-2}$ (Pearson = 0.99) (Données émergence RNV3P 2001-2009 ; N = 81 132 PST).

Pour affiner le nombre de configurations possibles, nous ajustons l'estimation des configurations sur les fréquences observées des longueurs des séquences. Ainsi, Le nombre ajusté d'associations pathologie - nuisance possibles devient :

$$Np_{st_{ajusté}} = \sum_{n=1}^5 p_L \frac{H!}{L!(H-L)!} \theta(H-L) \quad (8.3)$$

Avec p_L la probabilité pour que la séquence soit de longueur L (Figure 30). Par exemple, pour $H = 5$, on a $Np_{st_{ajusté}} = 4 + 2 + 0 + 0 + 0 = 6$ au lieu de 31, et pour $H = 10$, $Np_{st_{ajusté}} = 8 + 7 + 6 + 8 + 25 = 54$ au lieu de 637.

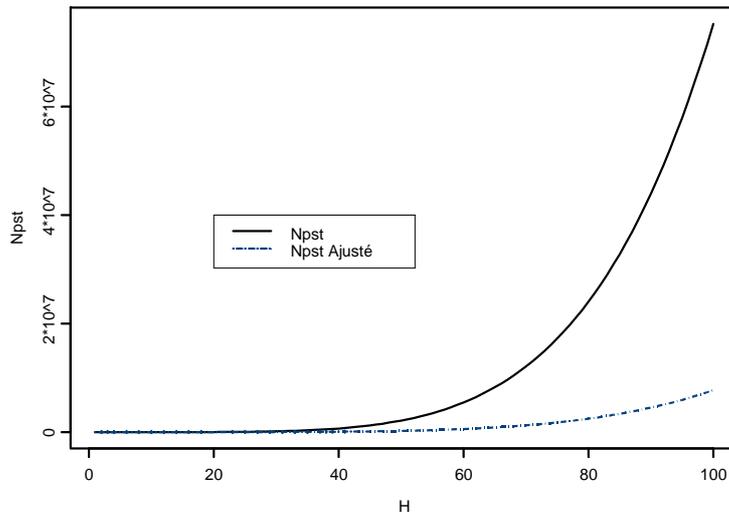


Figure 31 : Estimations des configurations de nuisances possibles.

Malgré l'ajustement, le nombre de configurations possibles croît énormément et rapidement en fonction de la taille de H.

8.2.2. Propriétés des séquences observées

Les séquences observées dans la base RNV3P ne correspondent pas à nos estimations. Par exemple, pour un spectre de 5 codes nuisances ($H = 5$), 26 codes pathologies distincts sont associées avec en moyenne $3,6 \pm 1,6$ séquences associées alors qu'on retrouve entre 6 et 31 combinaisons possibles lors des estimations. La Figure 32 illustre, pour chaque pathologie (point) enregistrées dans la base RNV3P, la taille du spectre des nuisances (en abscisses) par rapport au nombre de séquences observées (en ordonnées).

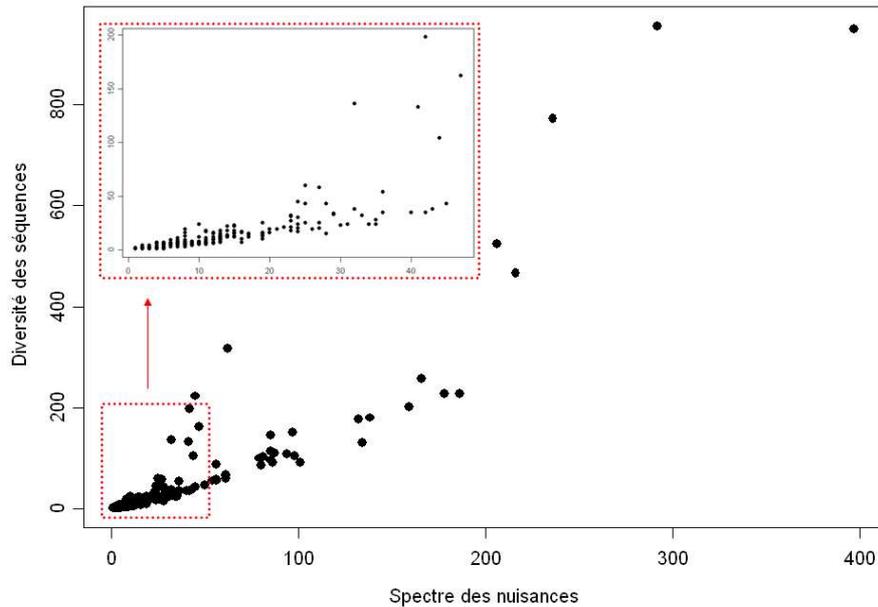


Figure 32 : Diversité des séquences observées par pathologie en fonction du spectre des nuisances (Données émergence RNV3P 2001-2009 ; N = 81 132 PST).

Il semble que le processus d'associations des expositions professionnelles à une pathologie soit régi par des règles, structurant ainsi les informations enregistrées pour un PST. A partir des variables, descripteurs et paramètres définis, nous construisons un générateur de PST par modélisation.

8.3. Modélisation du processus

Ces règles "cachées" peuvent être modélisées en se basant sur les connaissances empiriques contenues dans la base RNV3P. L'exercice consiste à comprendre et modéliser les pratiques de codage afin de construire des séquences.

Dans la littérature scientifique, on retrouve des travaux de modélisation qui ont pour objectif de détecter des comportements ou des observations atypiques. Par exemple, Fawcett et al. ont construit et modélisé des comportements sociaux [Fawcett 1999] afin de détecter des anomalies dans le comportement des individus, comme la détection des fraudes téléphoniques. Dans un autre contexte plus proche du notre, l'algorithme de détection, nommé PANDA, est basé sur une approche bayésienne qui, à partir de probabilités *a priori*, modélise des pathologies afin des détecter les causes de leur survenue [Shen 2008 ; Shen 2009 ; Shen 2010].

Les points communs entre ces systèmes sont de comprendre et de décrire les mécanismes intrinsèques d'association des variables dans une base de données, de construire un ou plusieurs modèles afin de caractériser ces mécanismes et confronter ces résultats aux données empiriques.

Nous avons ainsi élaboré un algorithme pour modéliser le processus d'enregistrement des PST dans la base RNV3P, dont les principales étapes sont décrites Figure 33 et détaillées dans cette section.

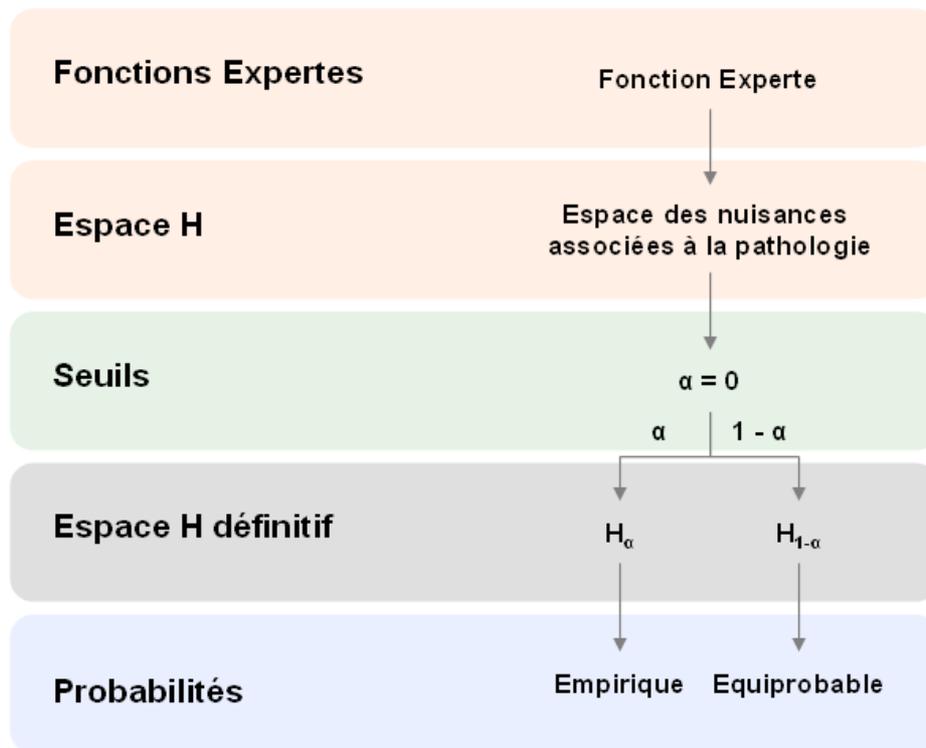


Figure 33 : Paramétrage des Fonctions Expertes.

8.3.1. Construction de l'espace des nuisances associables H

Le spectre des expositions professionnelles associées à une pathologie est construit à partir de l'historique des nuisances et des activités professionnelles dans la base RNV3P. Ce spectre s'enrichit chaque année des connaissances apportées par les médecins consultants. Ainsi, à l'issue de l'interrogatoire professionnel, on distingue trois possibilités dans le processus d'enregistrement d'un PST :

- (1) Soit le médecin enregistre un PST déjà existant dans la base de données RNV3P. Il n'apporte pas de nouvelles connaissances, mais augmente la fréquence de ce PST.
- (2) Soit le médecin crée un PST à partir de nuisances déjà associées à la pathologie dans la base RNV3P mais jamais observées dans une même séquence. Un nouveau PST est créé à partir du spectre des expositions professionnelles déjà rapportées.

- (3) Soit le médecin crée un nouveau PST avec une ou plusieurs nuisances jamais associées à la pathologie auparavant. La base des connaissances sur les expositions associées à cette pathologie augmente.

A la différence des possibilités (1) et (2), la possibilité (3) nécessite l'introduction de nouvelles nuisances dans le processus d'enregistrement d'un PST. Ainsi, pour permettre aux Fonctions Expertes de créer de nouveaux PST, nous délimitons un espace de nuisances susceptibles d'être associées à une pathologie, que l'on notera H, composé des nuisances déjà associées à la pathologie dans la base RNV3P, noté H_c , et de nuisances que nous définissons à partir d'une hypothèse.

Nous supposons que l'activité professionnelle conditionne l'ensemble des nuisances associables à une pathologie et définissons ainsi un ensemble de nuisances, non identifiées dans les connaissances empiriques associée à une pathologie donnée, mais déduite des activités professionnelles associées cette pathologie.

La Figure 34 illustre à l'aide d'un exemple, la construction de l'espace des nuisances associables pour une pathologie donnée.

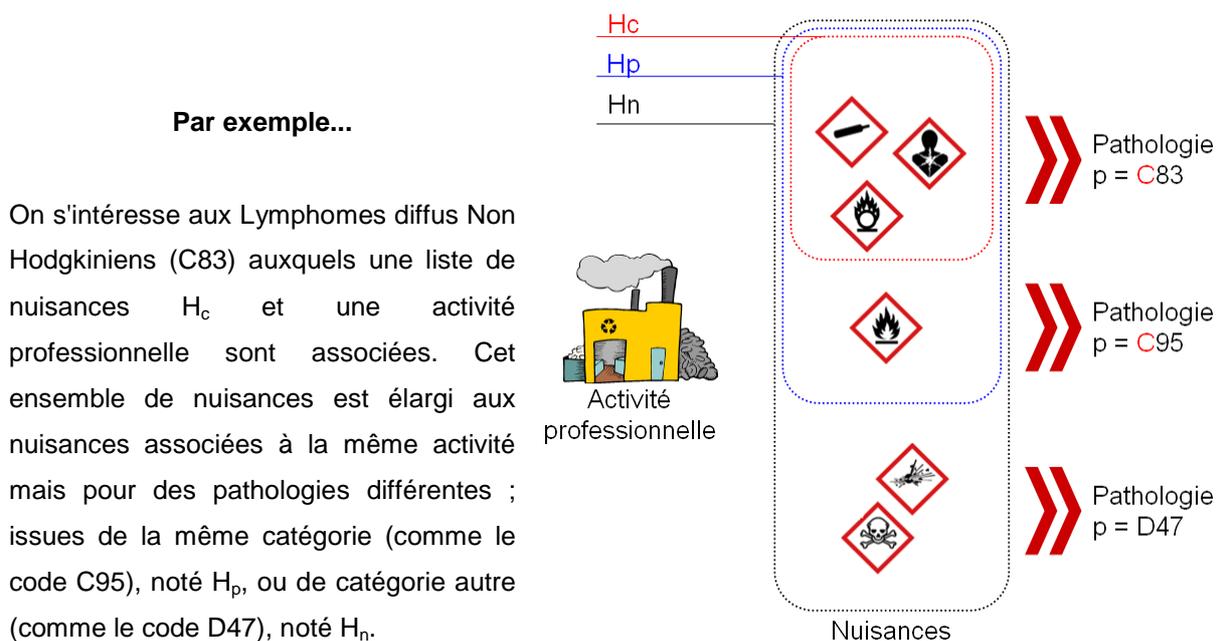


Figure 34 : Espace des nuisances associables à une pathologie et activité données.

La Figure 35 illustre l'espace des nuisances associables à une pathologie (par définition, $H_c \leq H_p \leq H_n$), composé des ensembles :

- H_c : espace des nuisances associées à la pathologie dans l'historique de la base RNV3P,
- H_p : espace des nuisances associées à l'activité professionnelle de la pathologie et dont le 1^{er} digit de la pathologie est identique au 1^{er} digit de la pathologie en question,
- H_n : espace des nuisances associées aux secteur-métier d'une pathologie donnée, quelque soit la pathologie.

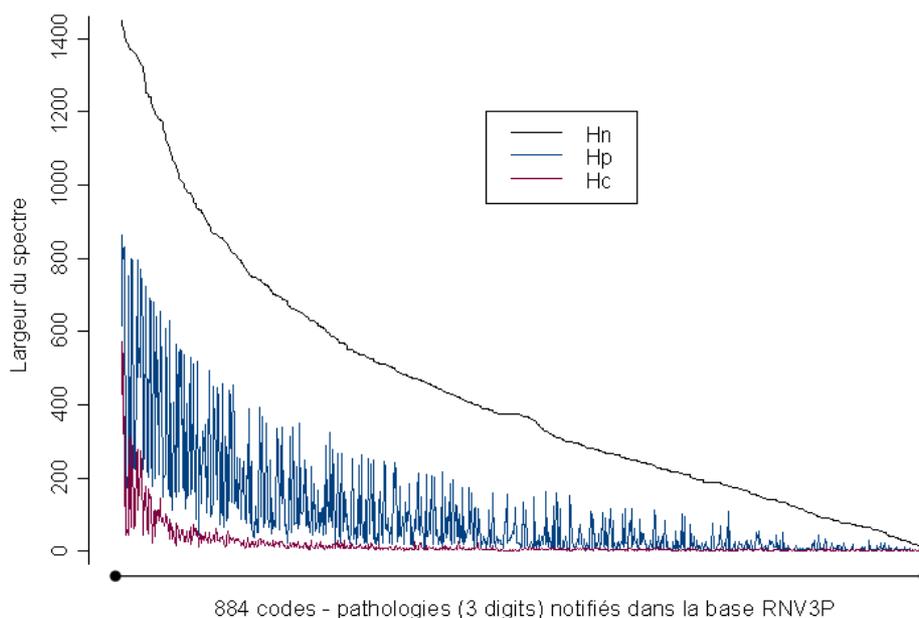


Figure 35 : Espace des nuisances associables aux codes-pathologies enregistrés dans la base RNV3P (Données émergence RNV3P 2001-2009 ; N = 81 132 PST).

8.3.2. Construction des seuils

Soit α la proportion de "nouvelles" nuisances introduites dans les étapes décisionnelles qui aboutissent à la construction d'un PST. La Figure 36 illustre 3 valeurs possibles de α , comprises entre 0 et 1.

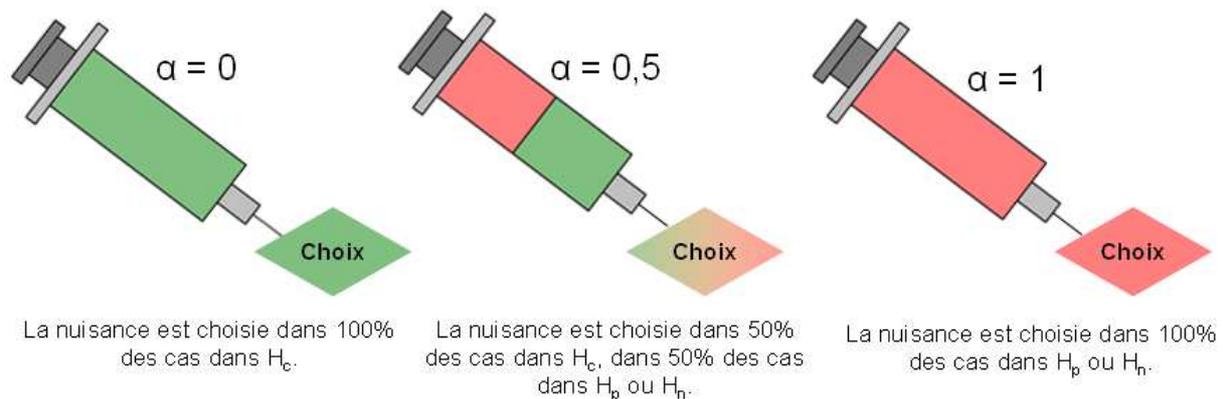


Figure 36 : Schéma illustratif du seuil α .

Pour un α donné, la proportion des nouvelles nuisances est proportionnelle à la taille des ensembles H_p ou H_n , de tailles respectives N_p et N_n :

$$\begin{cases} \alpha_p + \alpha_n + (1 - \alpha) = 1 \\ \alpha_p = \frac{\alpha \times N_p}{N_p + N_c} ; \alpha_n = \frac{\alpha \times N_n}{N_p + N_c} \end{cases} \quad (8.4)$$

Avec α_p et α_n les proportions des nuisances issues respectivement des espaces H_p et H_n .

8.3.3. Fonctions Expertes

Un PST provient d'une série de choix successifs réalisés par le médecin consultant à l'issue de l'interrogatoire professionnel avec le patient. Les Fonctions Expertes permettent de construire les étapes du processus décisionnel, conduisant à la création d'un PST. A partir de l'espace des nuisances associables dans lequel le médecin serait amené à choisir les nuisances à la suite de l'interrogatoire professionnel, nous avons développé des algorithmes de construction d'un PST modélisant comment un médecin mènerait un interrogatoire professionnel.

Selon la chronologie du bordereau de saisie d'un PST (Annexe C-3), le choix initial qu'un médecin serait amené à effectuer est celui de l'activité professionnelle potentiellement responsable de la survenue de la pathologie. Ainsi, pour les différents modèles que nous avons développés, les codes secteur d'activité et métier sont sélectionnés en amont selon une loi de probabilité empirique, notée $P(s-m)$, se référant aux connaissances stockées dans la base de données RNV3P.

Par exemple...

La loi de probabilité empirique des activités professionnelles associée à la pathologie sclérodémie systémique (code M34, N = 171 PST hors "99999 - Pas de nuisance identifiée") est définie comme la liste des fréquences associées aux codes secteur-métier dans la base des connaissances du RNV3P (Figure 37).

Le code 45-71 correspond au secteur de la construction et au métier ouvriers et artisans des métiers de l'extraction et du bâtiment, qui est l'activité professionnelle la plus représentée pour cette pathologie, avec 38 / 171 PST.

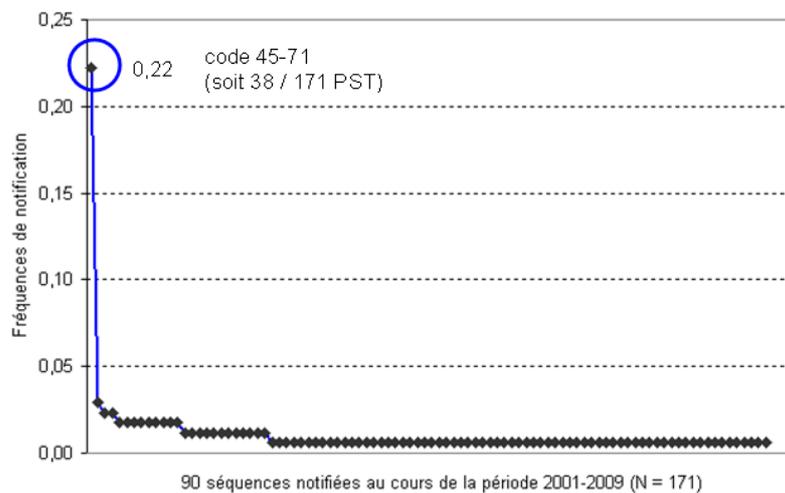


Figure 37 : Distribution de la loi de probabilité empirique des activités professionnelles associées aux sclérodémies systémiques (code M34, N = 171 PST hors "99999 - Pas de nuisance identifiée").

8.3.3.1. Fonction itérative

A partir de la pathologie diagnostiquée et de l'activité professionnelle, le modèle sélectionne le choix des nuisances auxquelles le patient a été exposé. La Fonction itérative suppose que les nuisances sont choisies l'une après l'autre dans l'espace des nuisances associables H selon une loi de probabilités empiriques.

Le choix de l'activité professionnelle influe sur le choix de la nuisance principale, qui influe sur le choix de la première co-nuisance, qui influe sur le choix de la deuxième co-nuisance, et ainsi de suite.

Ce modèle s'écrit mathématiquement :

$$F_{\text{itérative}} = P(s-m) \times P_1(h_1|s-m) \times P_2(h_2|s-m, h_1) \times P_3(h_3|s-m, h_1, h_2) \times P_4(h_4|s-m, h_1, h_2, h_3) \times P_5(h_5|s-m, h_1, h_2, h_3, h_4) \quad (8.5)$$

Où $P(s-m)$ est la probabilité de sélectionner l'activité professionnelle. Comme on suppose que ce choix influe sur la sélection des co-nuisances ; P_1 est la probabilité pour que le médecin choisisse une première nuisance sachant l'activité professionnelle. Les probabilités suivantes tiennent compte des informations apportées aux étapes précédentes. Ainsi, P_2 désigne la probabilité d'associer la nuisance h_2 sachant que h_1 et $s-m$ ont déjà été choisis auparavant. On renouvelle le processus pour les choix des 2^{ème}, 3^{ème} et 4^{ème} co-nuisances, selon les probabilités P_3 , P_4 et P_5 . La Figure 38 illustre ce processus décisionnel.

Remarque :

Les descripteurs, tels que la longueur des séquences et le degré d'indemnisation, ne constituent pas des étapes décisionnelles mais sont calculés une fois le PST construit. Les degrés d'imputabilité et d'indemnisation peuvent faire l'objet d'une étape décisionnelle à incorporer une fois la nuisance h_i choisie.

Par exemple...

La liste des nuisances associées aux sclérodermies systémiques (code M34) et à l'activité ouvriers et artisans des métiers de l'extraction et du bâtiment dans le secteur de la construction (code 45-71) est composée de 20 nuisances. La Figure 39 illustre les 3 nuisances les plus représentées des 20 nuisances associées au niveau h_1 .

Au niveau h_2 , 7 nuisances sont associables à la nuisance "Silice sable" dont l'amiante qui est enregistrée 2 fois (3 des 7 nuisances sont représentées). Au niveau h_3 , la nuisance "Fibres d'amiante" a la même probabilité d'être associée au plomb ou aux poussières de chantier.

Notons que, pour produire M échantillons comparables à l'échantillon de la base RNV3P, les probabilités sur les longueurs des séquences sont introduites à chaque itération du modèle.

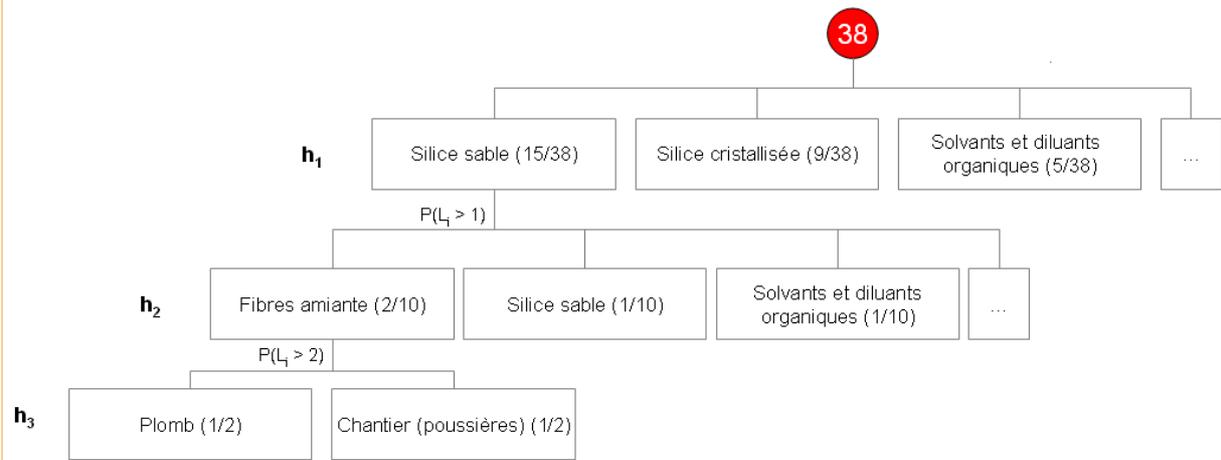


Figure 39 : Illustration des lois de probabilité empirique associée aux sclérodermies systémiques (code M34) et à l'activité ouvriers et artisans des métiers de l'extraction et du bâtiment dans la construction (code 45-71).

8.3.3.2. Fonction par blocs

Partant du fait que les descripteurs d'un PST sont fixés par le médecin à l'issue de l'interrogatoire professionnel, nous faisons l'hypothèse que le médecin pourrait choisir les nuisances en blocs en fonction des descripteurs ; comme la longueur des séquences et le degré d'indemnité de l'association pathologie - nuisance. Les descripteurs sont définis par la probabilité $P(s-m)$ et $P(l,L)$ où le médecin choisit, à l'issue de l'interrogatoire professionnel, l'activité professionnelle et le nombre de nuisances qui, selon lui, seraient potentiellement à l'origine de la pathologie. Le médecin sélectionne ensuite deux blocs de nuisances selon les connaissances empiriques de la base RNV3P ; un bloc de nuisances reconnues au titre des Tableaux MPI et un bloc de nuisances non reconnues.

Soit P_l la probabilité de choisir l nuisances reconnues et P_{L-l} la probabilité de choisir $L-l$ nuisances non reconnues. Ce modèle s'écrit mathématiquement :

$$F_{\text{bloc}} = P(s-m) \times P(L|s-m) \times P(l|s-m, L) \times P_l(h_1, \dots, h_l) \times P_{L-l}(h_{l+1}, \dots, h_L) \quad (8.6)$$

Où $P(l|s-m, L)$ est la probabilité de choisir l nuisances reconnues ayant déjà choisi L la longueur de la séquence et $s-m$ l'activité professionnelle associée. La Figure 40 illustre ce processus décisionnel.

Remarque : Les descripteurs font partie intégrante du processus de décision de cette fonction. On pourrait également incorporer le choix du niveau d'imputabilité, avant de sélectionner les nuisances en blocs.

Par exemple...

Les sclérodermies systémiques (N = 171 PST) associées à l'activité ouvriers et artisans des métiers de l'extraction et du bâtiment dans le secteur de la construction représentent 38 / 171 PST (Figure 41). Parmi ces PST, 60% sont associés à une seule nuisance (soit 23/38 PST), 26% des cas à deux nuisances (soit 10/38 PST) et 14% sont associés à 3 nuisances (soit 5/38 PST). Pour les 23 PST associés à une seule nuisance, 21 sont associés à une nuisance reconnue (soit une probabilité de $21/23 = 0,91$) et 3 sont associés à une nuisance non reconnue (soit une probabilité de $2/23 = 0,09$). Parmi les 10 PST associés à 2 nuisances, 2 sont associés à 2 nuisances non reconnues, 5 à une nuisance reconnue et une nuisance non reconnue, et 3 à 2 nuisances reconnues. A partir de ces descripteurs, la fonction par blocs choisit les nuisances reconnues dans la liste des nuisances reconnues et les nuisances non reconnues dans la liste des nuisances non reconnues selon les probabilités empiriques (entre parenthèses).

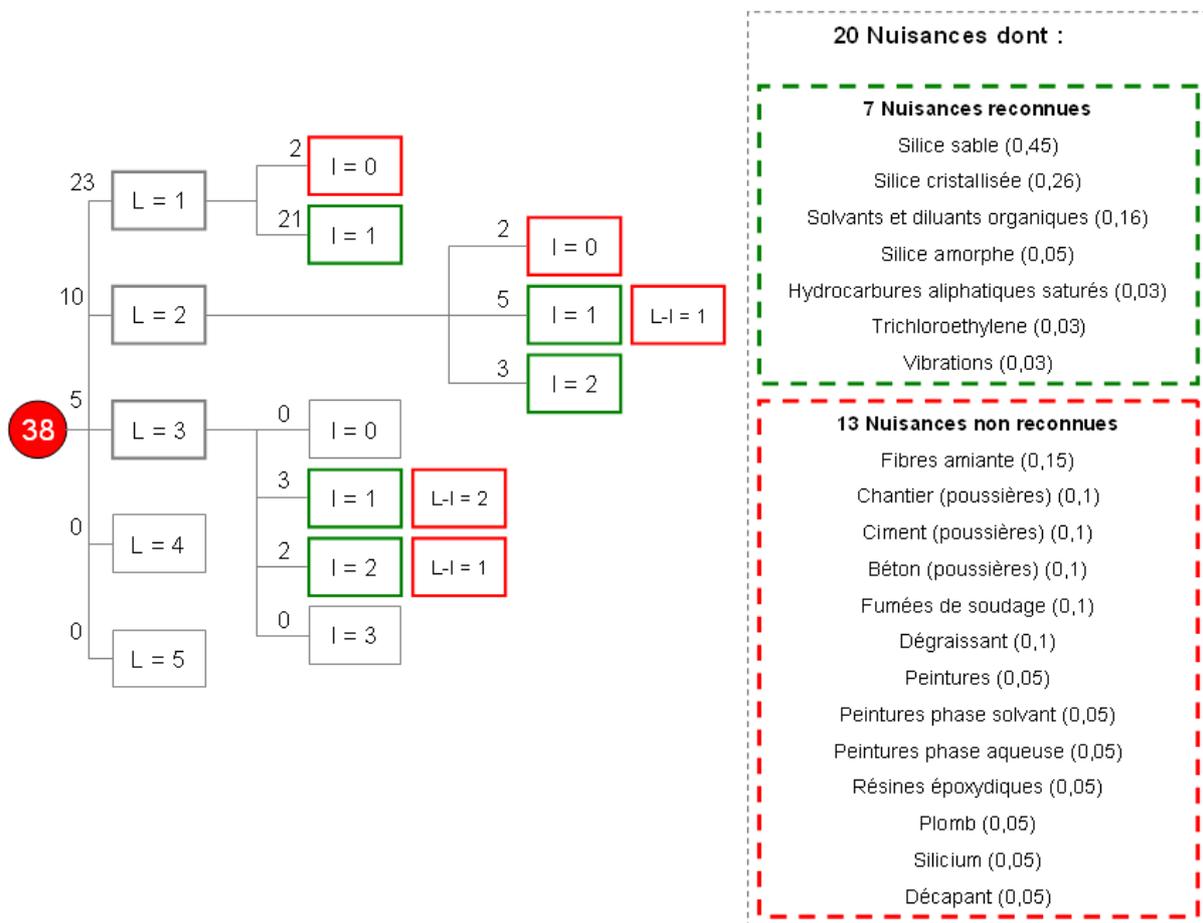


Figure 41 : Organigramme de la Fonction par blocs correspondant aux sclérodermies systémiques associées au métier d'ouvriers et artisans des métiers de l'extraction et du bâtiment dans le secteur de la construction.

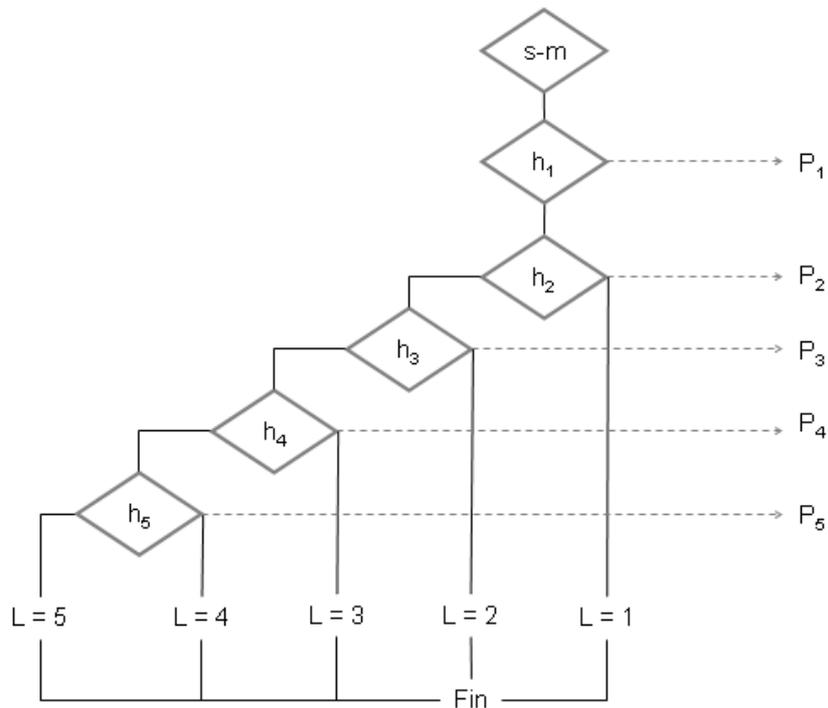


Figure 38 : Organigramme de la fonction itérative, avec s-m l'activité professionnelle, h_1 à h_5 les nuisances, L les longueurs des séquences, et P les probabilités d'association à chaque niveau de L.

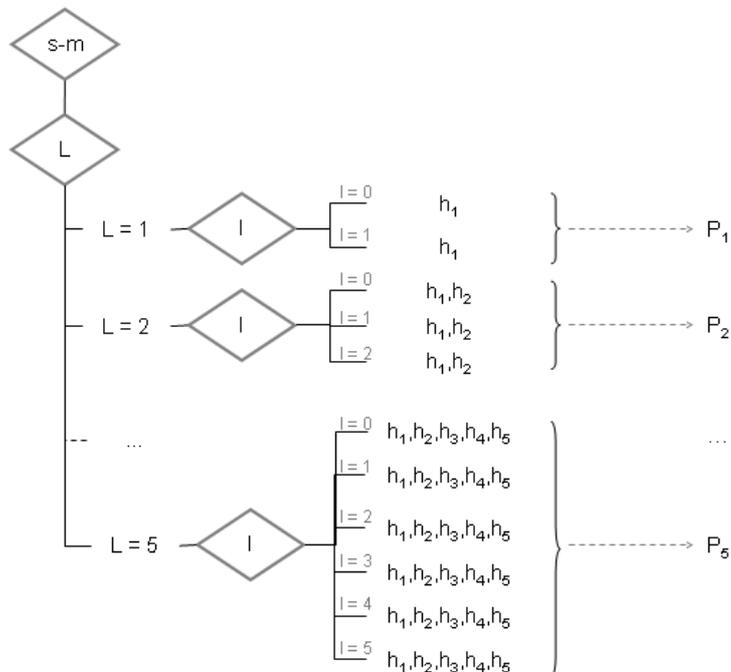


Figure 40 : Organigramme de la fonction par bloc, avec s-m l'activité professionnelle, L les longueurs des séquences, I le nombre de nuisances reconnues, h_1 à h_5 les nuisances associées, et P les probabilités d'association à chaque niveau de L.

Résumé

8. Formalisation des Fonctions Expertes

Les médecins du RNV3P sont des créateurs de PST. De la précision des enregistrements dépend la détection des associations. Afin de détecter des PST potentiellement émergents, nous avons choisi de travailler en amont des enregistrements afin de conceptualiser le processus de recueil des données en répondant à la question suivante :

Quel est le processus ou mécanisme qui guide le médecin consultant dans l'enregistrement d'une ou plusieurs expositions professionnelles associées à un diagnostic ?

Nous avons ainsi introduit des "Fonctions Expertes" pour traiter ce problème, comme modèle probabiliste qui, pour une pathologie donnée, modélisent des PST en se basant sur les connaissances cumulées de la base RNV3P.

9. Détection des associations potentiellement émergentes

Rappelons la définition d'une association potentiellement émergente au sens du RNV3P, définie dans le chapitre 3.

Une séquence est potentiellement émergente si cette association était déjà existante dans la base RNV3P dont la fréquence d'apparition devient anormalement élevée selon un critère défini.

Nous avons ainsi développé les trois tests suivants, basés sur des critères statistiques et présentés de manière indépendante:

- Test 0** Teste la susceptibilité d'une séquence d'être émergente.
- Test 1** Elimine les séquences dues au hasard et donc à un bruit statistique.
- Test 2** Identifie les séquences potentiellement émergentes ou nécessitant un suivi.

9.1. Critères statistiques

Le test 0 consiste à vérifier deux conditions : la présence de connaissances acquises dans la séquence au titre des Tableaux MPI (Tableau 4) et la fréquence minimale de la séquence. Le principe des tests 1 et 2 est conforme aux règles de tout raisonnement scientifique. La première étape consiste à formuler une hypothèse nulle, et à observer les valeurs si cette hypothèse était vraie. Il s'agit ensuite de vérifier si les observations sont effectivement identiques à celles obtenues sous l'hypothèse nulle.

9.1.1. Test 0 : Susceptibilité émergentielle

Le test 0 est une sélection qualitative des séquences présentant une "susceptibilité émergentielle", basé sur les conditions suivantes :

9.1.1.1. Condition sur les connaissances acquises

A l'échelle de la multi-exposition, nous avons défini une séquence i comme "reconnue" si le nombre de nuisances mentionnées au titre des Tableaux MPI, noté l_i , est égal à la longueur de la séquence i , notée L_i . Par définition, $l_i \leq L_i$. L'une des conditions du test 0 consiste donc à exclure les séquences reconnues, au profit des séquences comportant au moins une nuisance non reconnue, avec $l_i < L_i$.

9.1.1.2. Condition des fréquences minimales

Pour une pathologie p donnée, l'échantillon E_p décrit N séquences, dont le nombre de copies (ou poids) d'une séquence i est noté n_i . Les séquences rapportées une seule fois ($n_i = 1$) sont considérées comme du bruit statistique.

En conclusion, le test 0 consiste à vérifier :

$$t_0 = \begin{cases} 1 & ; \text{ si } l_i < L_i \text{ et } n_i > 1 \\ 0 & ; \text{ sinon} \end{cases} \quad (9.1)$$

Où n_i est le poids d'une séquence, L_i sa longueur et l_i le nombre de nuisances reconnues.

Le Tableau 6 illustre, à l'aide d'un exemple illustratif la règle de décision du test 0 :

Seq _i	L _i	l _i	n _i	l _i < L _i	n _i > 1	Test 0
a	5	2	1	oui	non	0
b	1	0	5	oui	oui	1
c	2	2	2	non	oui	0

Tableau 6 : Exemple illustratif de la règle de décision du test 0.

9.1.2. Test 1 : Equiprobabilité des fréquences

9.1.2.1. Hypothèse nulle

Le test 1 consiste à vérifier si la fréquence d'apparition d'une séquence i , notée $f_i = n_i / N$, est due au hasard. Pour ce faire, nous utilisons l'hypothèse nulle :

H_0 : " Pour une pathologie donnée, les fréquences des séquences sont dues au hasard "

Pour vérifier cette hypothèse, nous calculons les fréquences théoriques obtenues sous l'hypothèse nulle pour une pathologie donnée.

9.1.2.2. Expérience

Soit le vecteur \mathbf{F} de taille N , contenant les fréquences observées pour chacune des séquences associées à une pathologie donnée, tel que :

$$\left\{ \begin{array}{l} \mathbf{F} = [f_1 \dots f_i \dots f_N] \\ \sum_{i=1}^N f_i = 1 \end{array} \right. \quad (9.2)$$

Pour chaque séquence i , sa fréquence théorique notée $\langle g_i \rangle$ – obtenue sous l'hypothèse nulle – correspond à la moyenne des fréquences contenues dans le vecteur \mathbf{F} permutées M fois de manière aléatoire, telles que:

$$\langle g_i \rangle = \frac{1}{M} \sum_{k=1}^M g_{ik} \quad (9.3)$$

Avec g_{ik} la fréquence théorique obtenu pour la séquence i de l'échantillon k .

9.1.2.3. Statistique du test

Soit RR_i le risque relatif associé à la séquence i . Son intervalle de confiance à 95%, noté $[RR_i]$, tels que :

$$\left\{ \begin{array}{l} RR_i = \frac{f_i}{\langle g_i \rangle} \\ [RR_i] = RR_i \pm t(\alpha) \times \Delta_{RR_i} \end{array} \right. \quad \text{avec } \Delta_{RR_i} = \frac{\sigma_{RR_i}}{\sqrt{M}} \quad \text{et } \sigma_{RR_i} = \frac{1}{M-1} \sum_{k=1}^M (g_{ik} - \langle g_i \rangle)^2 \quad (9.4)$$

Avec $t(\alpha)$ la probabilité de rejeter l'hypothèse nulle à tort (par défaut 5%; $t(\alpha) = 1,96$).

9.1.2.4. Règle de décision

On rejette l'hypothèse nulle lorsque $RR_i \geq 1$; ce qui signifie que la fréquence observée f_i est supérieure à la moyenne des fréquences théoriques $\langle g_i \rangle$. L'observation de cette séquence n'est pas due au hasard.

NB : Une étude épidémiologique ayant pour objectif de quantifier la relation entre la pathologie et une exposition utilise la borne inférieure du risque relatif pour préciser la significativité de la relation. La règle de décision présentée ci-dessus n'a pas un tel objectif ; RR est le ratio entre les fréquences observées et la moyenne des fréquences théoriques. Ainsi, l'intervalle de confiance est présenté à titre indicatif.

En conclusion, le test 1 consiste à vérifier pour chaque séquence :

$$t_1 = \begin{cases} 1 & ; \text{ si } RR_i \geq 1 \\ 0 & ; \text{ si } RR_i < 1 \end{cases} \quad (9.5)$$

Où RR_i est le rapport entre la fréquence observée f_i et la moyenne des fréquences théoriques $\langle g_i \rangle$ obtenues sous l'hypothèse nulle.

Le Tableau 7 illustre, à l'aide d'un exemple illustratif la règle de décision du test 1 :

Seq _i	n _i	f _i		g ₁	g ₂	...	g _M	$\langle g \rangle$	RR _i	Test 1
a	1	0,09		0,37	0,18	...	0,09	0,22	0,45	0
b	1	0,09		0,09	0,37	...	0,27	0,21	0,45	0
c	3	0,27		0,18	0,09	...	0,18	0,21	1,36	1
d	2	0,18		0,09	0,27	...	0,09	0,22	0,90	0
e	4	0,37		0,09	0,09	...	0,37	0,22	1,81	1

$M = 10\ 000$

Tableau 7 : Exemple illustratif de la règle de décision du test 1.

9.1.3. Test 2 : Réussite émergentielle

9.1.3.1. Hypothèse nulle

Si le médecin consultant constituait une séquence en piochant les nuisances de manière équiprobable (au hasard), alors chaque nuisance aurait la même probabilité d'apparaître. Nous utilisons l'hypothèse nulle suivante :

H'_0 : "Toutes les nuisances sont équiprobables".

Pour décider si l'hypothèse formulée est supportée ou non par les séquences d'un échantillon donné, nous construisons une méthode pour conclure si l'écart observé entre la valeur de la statistique obtenue dans l'échantillon et celle paramétrée selon notre hypothèse est trop important pour être uniquement due au hasard.

9.1.3.2. Expérience

Ce test fait appel aux Fonctions Expertes. Nous utilisons le seuil des nuisances associables $\alpha = 0$ sous l'hypothèse d'équiprobabilité, avec l'ensemble des nuisances associées à la pathologie dans la base de données RNV3P (ensemble H_0).

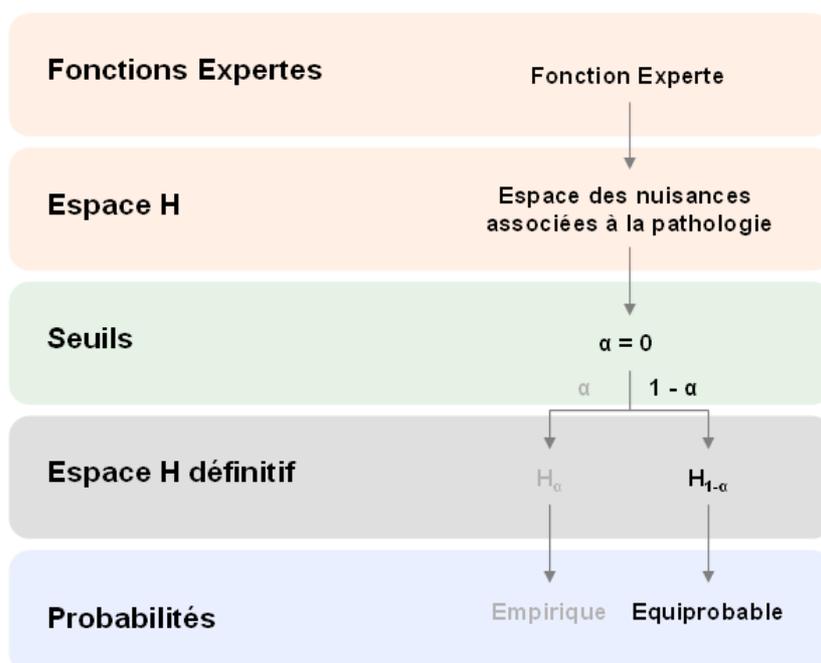


Figure 42 : Schéma d'utilisation de la Fonction Experte pour le test 2.

Nous générons ainsi M échantillons de taille N , comparables à l'échantillon des séquences empiriques et identifions les séquences reproduites exactement à l'identique des observées. Puis, nous associons le nombre de copies, ou poids, noté m_{ik} , à chaque séquence théorique i de l'échantillon k .

9.1.3.3. Statistique du test

Soit RR_i le risque relatif associé à la séquence i tel que :

$$RR_i = \frac{f_i}{\langle g_i \rangle} \quad \text{avec} \quad g_i = \frac{m_{ik}}{NM} \quad \text{et} \quad \sum_{i=1}^{NM} g_i = 1 \quad (9.7)$$

Où g_i est la fréquence moyenne d'apparition des cortèges générés par la fonction, $\langle g_i \rangle$ la fréquence moyennée des g_i sur les M échantillons, g_{ik} la fréquence de la séquence i dans l'échantillon k . On fournit également son intervalle de confiance à 95%, noté $[RR_i]$, tel que :

$$[RR_i] = RR_i \pm t(\alpha) \times \Delta_{RR_i} \quad \text{avec} \quad \begin{cases} \Delta_{RR_i} = \frac{\sigma_{RR_i}}{\sqrt{M}} \\ \sigma_{RR_i} = \frac{1}{M-1} \sum_{k=1}^M (g_{ik} - f_i)^2 \end{cases} \quad (9.8)$$

Avec $t(\alpha)$ la probabilité de rejeter l'hypothèse nulle à tort (par défaut 5%; $t(\alpha) = 1,96$).

9.1.3.4. Règle de décision

On rejette l'hypothèse nulle selon laquelle les nuisances sont distribuées de manière équiprobable dans une séquence lorsque la fréquence observée f_i est supérieure à la moyenne des fréquences théoriques $\langle g_i \rangle$. L'observation de cette séquence n'est pas due au hasard.

En conclusion, le test 2 consiste à vérifier :

$$t_2 = \begin{cases} 1 ; & \text{si } RR_i \geq 1 \\ 0 ; & \text{si } RR_i < 1 \end{cases} \quad (9.9)$$

Où RR_i est le rapport entre la fréquence observée f_i et la moyenne des fréquences théoriques $\langle g_i \rangle$ obtenues sous l'hypothèse nulle.

Le Tableau 8 illustre, à l'aide d'un exemple illustratif la règle de décision du test 2 :

Seq _i	n _i	f _i	Fonction Experte				< g _i >	RR _i	Test 2
			g ₁	g ₂	...	g _M			
a	1	0,09	0,09	0,09	...	0	0,15	0,6	0
b	1	0,09	0,27	0,09	...	0,37	0,08	1,25	1
c	3	0,27	0,27	0,27	...	0,18	0,31	3,85	1
d	2	0,18	0,27	0,18	...	0,09	0,25	0,8	0
e	4	0,37	0,27	0,18	...	0,37	0,36	1,11	1

Tableau 8 : Exemple illustratif de la règle de décision du test 2.

9.2. Stratégie de détection des PST potentiellement émergents

A partir de ces tests, nous définissons une stratégie pour détecter des PST potentiellement émergents. Nous avons ainsi élaboré un test statistique composite, noté T. Ce test se présente comme une suite de décisions obtenues à partir des tests t_0 , t_1 et t_2 , que nous détaillons en trois étapes indépendantes.

$$T = (t_0, t_1, t_2) \quad (9.10)$$

La Figure 43 illustre ces étapes décisionnelles, sous la forme d'un arbre de décisions composé des 3 étapes suivantes :

ETAPE 1 : Tester la susceptibilité émergentielle des séquences.

Les séquences observées une seule fois composées de nuisances reconnues ne sont pas susceptibles d'émerger ($t_0 = 0$), du fait des connaissances établies entre la pathologie et les nuisances professionnelles, et du caractère unique du cas.

En revanche, on considère une séquence susceptible d'émerger ($t_0 = 1$) lorsqu'elle est composée d'au moins une nuisance non reconnue au titre des Tableaux MPI et que sa fréquence d'enregistrement est au moins égale à 2.

ETAPE 2 : Vérifier la distribution non équiprobable des fréquences.

Les séquences présentant une susceptibilité émergentielle ($t_0 = 1$) sont soit assimilées à un bruit statistique ($t_0 = 1$ et $t_1 = 0$) lorsque les fréquences des séquences sont dues au hasard, soit présentent un risque émergentiel ($t_0 = 1$ et $t_1 = 1$), c'est-à-dire que leur fréquence d'enregistrement sont supérieures au hasard.

ETAPE 3 : Vérifier la distribution non équiprobable des nuisances.

Les séquences classées comme susceptibles d'émerger et dont leur fréquence d'enregistrement sont supérieures au hasard ($t_0 = 1$ et $t_1 = 1$) sont soumises au test 2. Celles qui présentent des nuisances non dues au hasard sont classées en séquences potentiellement émergentes ($t_0 = 1$ et $t_1 = 1$ et $t_2 = 1$). Les séquences qui ne vérifient pas cette condition ($t_0 = 1$ et $t_1 = 1$ et $t_2 = 0$) nécessitent un suivi du fait de leur susceptibilité à émerger et du risque émergentiel qu'elles représentent.

En pratique, ces trois étapes sont réalisées de manière indépendante, sur l'ensemble des séquences associées à une pathologie donnée.

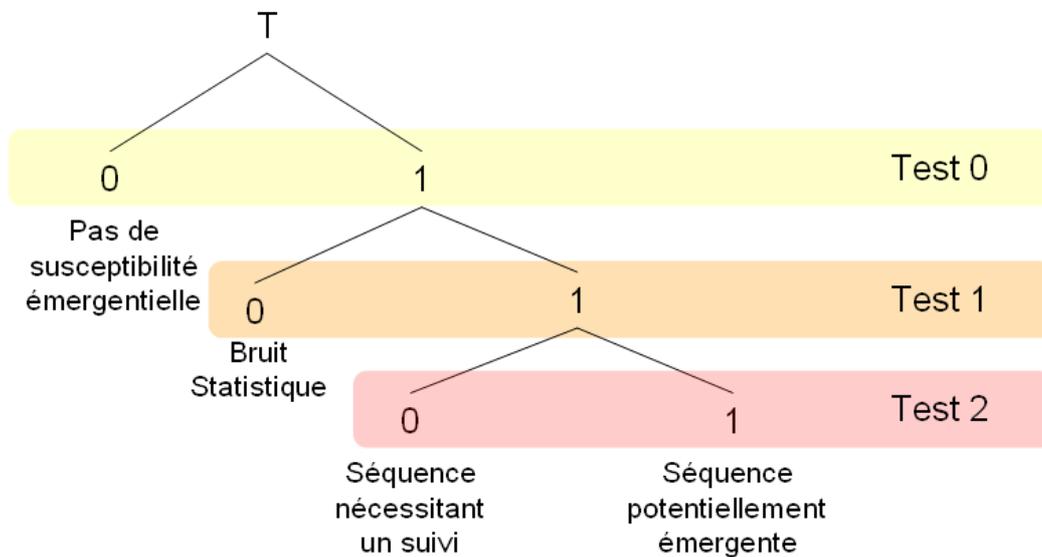


Figure 43 : Stratégie de détection des séquences potentiellement émergentes.

9.3. Application aux données RNV3P

Nous avons extrait de la base RNV3P les PST composés du code nuisance "99999 - Pas de nuisance identifiée", principalement issus des PST de nature interrogatoire professionnel.

Au cours de la période 2001-2009, la base RNV3P compte 77 913 PST qui décrivent 20 070 séquences distinctes. La base rassemble ainsi **moins de 10% des séquences susceptibles d'émerger** ($t_0 = 1$) (soit 1 874 / 20 070 séquences), dont **38% de ces séquences sont multi-exposées** (soit 716 / 1 874 séquences). Ainsi, la probabilité pour une séquence d'être multi-exposée et susceptible d'émerger est de 0,38.

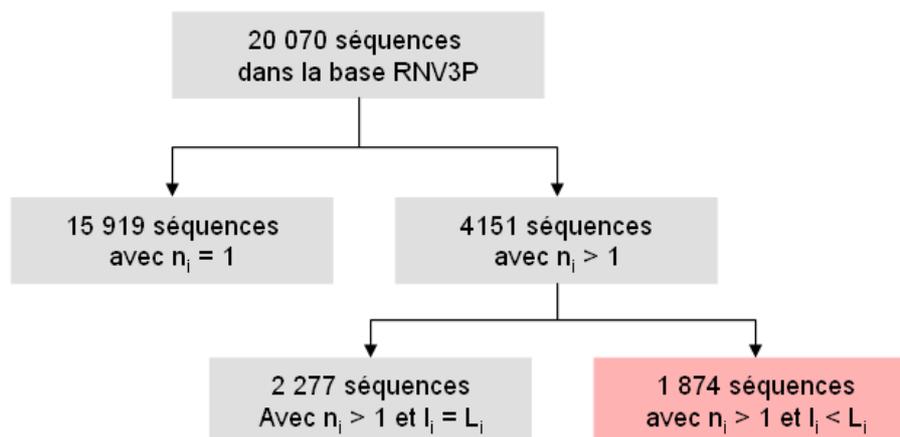


Figure 44 : Schéma détaillé de la taille de l'ensemble des PST susceptibles d'émerger au cours de la période 2001-2009.

Le Tableau 9 représente les pathologies de la base RNV3P, classées en fonction de l'importance des séquences susceptibles d'émerger. Nous avons choisis trois échantillons extraits de la base RNV3P : les Sclérodémies Systémiques (code M34), les Tumeurs Malignes de la Vessie (code C67) et les Lymphomes diffus Non Hodgkiniens (code C83).

Rang		Pathologie	Effectifs des PST	Effectifs des séquences	Effectifs des séquences susceptibles d'émerger ($t_0 = 1$) dont les séquences de longueur 1 ($L_i = 1$)
1	C34	Tumeur maligne des bronches et du poumon	6 016	1 315	209	76
2	F32	Episodes dépressifs	2 894	167	74	26
3	F43	Réaction à un facteur de stress sévère et troubles de l'adaptation	4 202	152	72	20
4	F41	Autres troubles anxieux	3 481	164	66	23
5	J45	Asthme	4 378	1 514	57	41
...
12	C67	Tumeur Maligne de la Vessie	495	286	30	17
...
81	M34	Sclérodémie Systémique	171	90	5	4
...
107	C83	Lymphome diffus Non Hodgkinien	31	26	3	3
...
862	Z98	Autres états post-chirurgicaux	1	1	0	0
TOTAL			77 913	20 070	1 874	1 158

Tableau 9 : Classement des pathologies selon leur susceptibilité émergentielle (Données émergence RNV3P 2001-2009 ; N = 77 913 PST hors "99999 - Pas de nuisance identifiée").

9.3.1. Echantillon des Sclérodermies Systémiques

Les Sclérodermies Systémiques sont des maladies relativement rares dont la prévalence varie entre 7 et 500 cas par million d'habitants [Chiffot 2008 ; Hinchcliff 2008]. Les causes et les mécanismes de cette maladie sont imparfaitement connus ; les déterminants familiaux et auto-immuns ne sont pas convaincants et des facteurs environnementaux sont fortement suspectés [Inachi 1996 ; Nietert 2000 ; Mora 2009]. Cette pathologie est associée à 35 codes nuisances mentionnés au titre des tableaux MPI. L'échantillon des Sclérodermies Systémiques a été testé avec les méthodes de pharmacovigilance sur la période 2001-2007, et a fait l'objet d'une publication [Bonnetterre 2010].

9.3.1.1. Résultats obtenus avec la méthode PRR

Les données relatives aux Sclérodermies Systémiques extraites de la base 2001-2007 comprennent 106 PST enregistrés en pathologie principale ou en co-morbidités. Après découpage des séquences, la méthode PRR prend en compte 68 couples pathologie - nuisance, parmi lesquels 7 génèrent un signal d'émergence, dont 4 un signal d'émergence valide (hors couples reconnus comme sclérodermie - silice, sclérodermie - solvants et diluants organiques, sclérodermie - trichloréthylène). Le Tableau 10 illustre l'historique des signaux d'émergence valides au cours de la période 2001-2007 selon la méthode PRR.

Couples sclérodermie systémique - nuisance		2001	2002	2003	2004	2005	2006	2007
31146	Métaux durs	0	0	0	1	2	3	3
36R00	Produit à usage agricole	0	0	0	0	1	2	3
48200	Froid	0	0	0	1	1	3	5
49200	Vibrations transmises aux membres (machines outils)	1	1	1	1	1	2	3

Tableau 10 : Couples pathologie - nuisance émettant un signal d'émergence valide (cellules grisées) selon la méthode PRR issues des méthodes de pharmacovigilance [Bonnetterre 2010].

Ces couples émettant un signal d'émergence sont composés de nuisances appartenant à des séquences multi-exposées, comme par exemple, la nuisance "Métaux durs" (code 31146) est enregistrée dans trois séquences multi-exposées :

- Silice - Carbone - Métaux durs - Solvants et diluants organiques (codes 31161-1C001-31146-35110 observé en 2004 auprès d'un ouvrier (code 7222),
- Métaux durs - Tungstène - Cobalt (codes 31146-1W001-1CO01) observé en 2005 auprès d'un conducteur d'installations de traitement thermique des métaux (code 8123),
- Silice - Métaux durs - Vibrations (codes 31161-31146-49200) observé en 2006 auprès d'un éboueur (code 9161).

9.3.1.2. Résultats obtenus avec les Fonctions Expertes

A partir de la même base de données, nous appliquons le test composite dans le but de comparer les résultats obtenus selon la méthode PRR et l'approche par Fonctions Expertes ; d'abord à partir de la base 2001-2007, puis en actualisant avec les données récentes de la base 2001-2009.

- Données RNV3P 2001-2007 (N = 106 PST)

A partir de cette base de données 2001-2007, nous appliquons le test composite et déterminons quelles sont les séquences potentiellement émergentes, pour $M = 10\ 000$. La Figure 45 illustre la répartition des séquences enregistrées dans la base RNV3P au cours de la période 2001-2007 (effectifs cumulés), et la part des séquences susceptibles d'émerger ($t_0 = 1$).

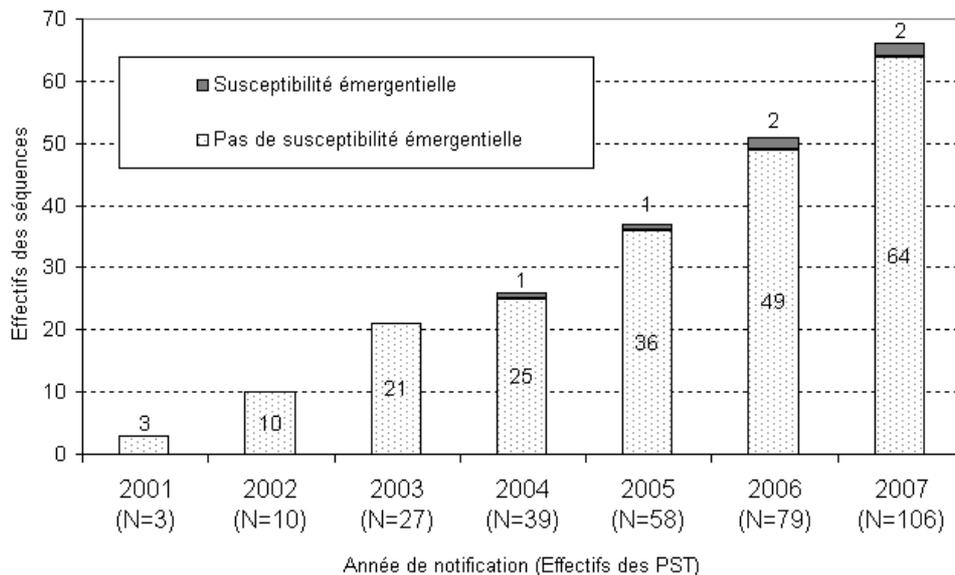


Figure 45 : Répartition des séquences associées aux sclérodermies systémiques et enregistrées dans la base RNV3P au cours de la période 2001-2007 (N = 106 PST) et résultats du test composite.

Les résultats des Fonctions Expertes itérative et par bloc sont identiques. La dynamique temporelle des séquences potentiellement émergentes (cellules hachurées) est détaillée dans le Tableau 11.

Séquences	2001	2002	2003	2004	2005	2006	2007
48200 Froid	0	0	0	1	1	2	2
31110 Chantier (Poussières)	0	0	1	2	2	2	2

Tableau 11 : Dynamique temporelle des séquences potentiellement émergentes ($T = (1, 1, 1)$: cellules hachurées) associées aux sclérodermies systémiques au cours de la période 2001-2007 (N = 106 PST).

- Données RNV3P 2001-2009 (N = 171 PST)

La base de données "Emergence" du RNV3P 2001-2009 rassemble 171 PST associés à la Sclérodémie Systémique (pathologie principale seulement), après exclusion du code "99999 - Pas de nuisance identifiée". La Figure 46 illustre la répartition des séquences enregistrées dans la base RNV3P au cours de cette période, et la part des séquences susceptibles d'émerger ($t_0 = 1$).

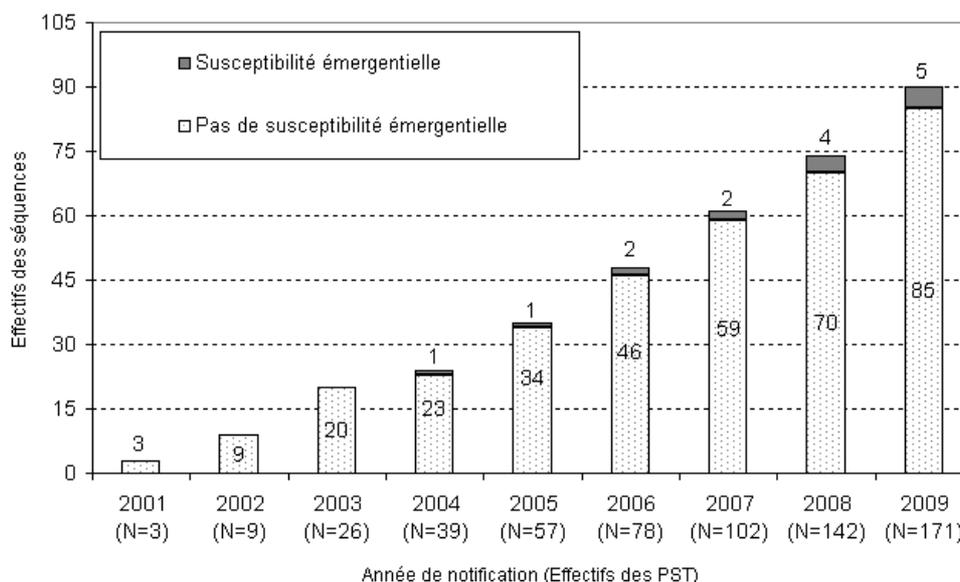


Figure 46 : Répartition des séquences associées aux sclérodémies systémiques et enregistrées dans la base RNV3P au cours de la période 2001-2009 (N = 171 PST) et résultats du test composite.

Les résultats des Fonctions Expertes itérative et par bloc sont identiques. La dynamique temporelle des séquences potentiellement émergentes (cellules hachurées) est détaillée dans le Tableau 12 (Résultats détaillés en Annexe D-1).

Séquences	2001	2002	2003	2004	2005	2006	2007	2008	2009
36R00 Produits à usage agricole	0	0	0	0	0	1	1	4	4
31110 Chantier (Poussières)	0	0	1	2	2	2	2	2	3
48200 Froid	0	0	0	1	1	2	2	3	3
34100 Fumées de soudage	0	0	0	0	0	0	1	1	2
36200- (35110) Peintures, Vernis, Laques, Mastics - Solvants et diluants organiques	0	0	0	0	1	1	1	2	2

Tableau 12 : Dynamique temporelle des séquences potentiellement émergentes (cellules hachurées) associées aux sclérodémies systémiques au cours de la période 2001-2009 (N = 171 PST). Entre parenthèses, les nuisances reconnues au titre des Tableaux MPI.

9.3.1.3. Bilan

Les discordances entre les deux approches s'expliquent en partie par les conditions basées sur les fréquences minimales. En effet, la méthode PRR s'applique aux couples dont la fréquence d'apparition est supérieure à 2, et une des conditions du test composite T consiste à prendre en compte les séquences qui apparaissent au moins 2 fois.

Mais la définition des seuils n'explique pas toutes les discordances observées entre ces deux approches, et relèvent d'avantage des stratégies méthodologiques employées. Sur la base de cette observation, nous avons investigué les couples ou séquences qui suscitent un intérêt en termes d'émergence. Le Tableau 13 synthétise les fréquences cumulées des couples émettant un signal d'émergence valide selon la méthode PRR (hachuré) et les séquences potentiellement émergentes détectées selon le test composite T (quadrillé).

Couples / Séquences		2001	2002	2003	2004	2005	2006	2007	2008	2009
48200	Froid	0	0	0	1	1	3	5	-	-
		0	0	0	1	1	2	2	3	3
36R00	Produits à usage agricole	0	0	0	0	1	2	3	-	-
		0	0	0	0	0	1	1	4	4
49200	Vibrations	1	1	1	1	1	2	3	-	-
31146	Métaux durs	0	0	0	1	2	3	3	-	-
31110	Chantier (Poussières)	0	0	1	2	2	2	2	2	3
34100	Fumées de soudage	0	0	0	0	0	0	1	1	2
36200- (35110)	Peintures, vernis, laques, mastics- Solvants et diluants organiques	0	0	0	0	1	1	1	2	2

Tableau 13 : Synthèse des couples émettant un signal d'émergence valide selon les méthodes de pharmacovigilance (cellules grisées) et les séquences potentiellement émergentes détectées avec les Fonctions Expertes (cellules hachurées). Entre parenthèses, les nuisances reconnues au titre des Tableaux MPI.

Ce tableau souligne des concordances au niveau des expositions professionnelles qui suscitent un intérêt en termes d'émergence, comme par exemple les expositions au froid ou aux produits à usage agricole. Mais les deux approches présentent des spécificités de détection propres à la méthodologie mise en œuvre. A titre illustratif, les sclérodermies systémiques associées aux vibrations ou aux métaux durs génèrent un signal d'émergence valide avec la méthode PRR, mais ne présentent pas de susceptibilité émergentielle avec l'approche par Fonctions Expertes du fait que ces nuisances sont co-exposées et appartiennent donc à une séquence multi-exposée.

L'approche par Fonctions Expertes a notamment permis de mettre en évidence une séquence multi-exposée potentiellement émergente composée d'une nuisance reconnue au titre des tableaux MPI, les solvants et diluants organiques, et d'une nuisance non reconnue les peintures, vernis, laques et mastics.

9.3.2. Echantillon des Tumeurs Malignes de la Vessie

Les Tumeurs Malignes de la Vessie (code C67) sont mentionnées au titre des tableaux MPI "Lésions prolifératives de la vessie provoquées par les amines aromatiques et leurs sels et la N-nitrosodibutylamine et ses sels" (Tableau n°15 TER du Régime Général) et "Affections cancéreuses provoquées par les goudrons de houille, les huiles de houille, les brais de houille et les suies de combustion du charbon" (Tableau n°16 BIS du Régime Général), soit 29 nuisances.

Cet échantillon représente 31 PST enregistrés dans la base RNV3P au cours de la période 2001-2009. La Figure 47 illustre la répartition des séquences enregistrées dont celles susceptibles d'émerger.

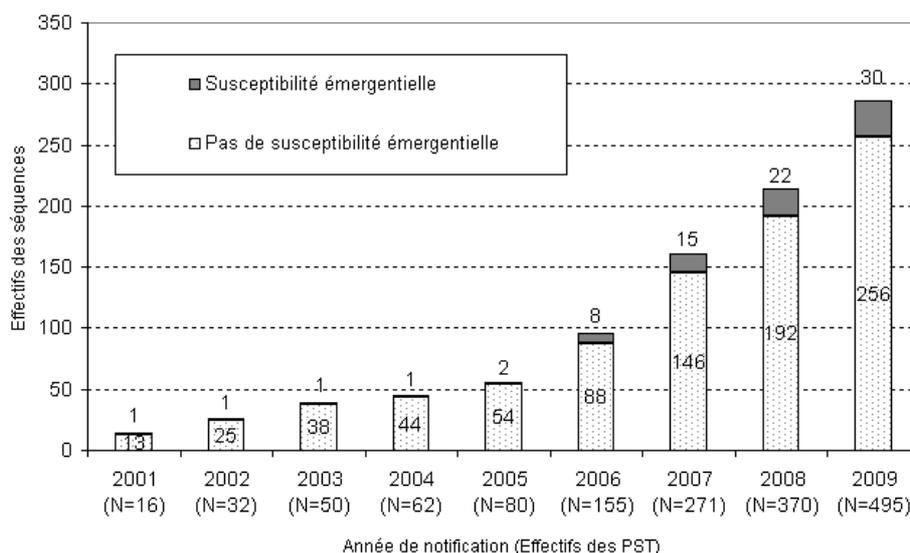


Figure 47 : Répartition des séquences associées aux Tumeurs Malignes de la Vessie et enregistrées dans la base RNV3P au cours de la période 2001-2009 (N = 495 PST) et résultats du test composite.

Le Tableau 14 illustre les détails des signaux d'émergence potentielle pour les 30 séquences (Résultats détaillés en Annexe D-2). Deux conclusions divergentes sont observées sur la séquence "36F00 - Produits noirs (asphalte, goudrons, créosote)" où la Fonction Experte itérative identifie la séquence comme potentiellement émergente et la Fonction par blocs identifie la séquence comme nécessitant un suivi.

Séquences		2001	2002	2003	2004	2005	2006	2007	2008	2009
36200	Peintures, vernis, laques, mastics	2	3	3	4	4	5	7	10	14
33210	Diesel (gaz)	0	0	0	0	0	10	14	17	19
36R00	Produits à usage agricole	0	0	0	0	0	0	2	5	8
36500- 36700	Matières plastiques caoutchoucs polymères - Matières colorantes	0	1	1	1	1	1	2	2	2
(21320)- 31146	HAP - Métaux durs	0	0	0	0	0	0	2	2	2
22224	Trichloroethylene	0	0	0	0	0	4	6	6	6
32300	Fibres végétales textile T66A	0	0	0	0	0	0	0	2	2
(21320)- 36C20- 334C0	HAP - Huiles et graisses lubrifiantes - Produits noirs (goudrons bitumes asphaltes)	0	0	0	0	0	0	0	0	2
36410	Résines époxydiques	0	0	0	0	0	0	1	1	2
2Q31Z	Autre amine aromatique	0	0	0	0	0	0	1	1	2
36210- (35110)	Peintures vernis laques mastics en phase solvants - Solvants et diluants organiques	0	0	0	0	1	2	2	3	3
(21320)- 33210	HAP - Diesel (gaz)	0	0	0	0	0	2	2	4	4
312B0- 22224	Suie - Trichloroethylene	0	0	0	0	0	2	2	2	2
36F00	Produits noirs (asphalte, goudrons, créosote)	0	0	0	0	0	0	0	1	2
47100	Radiations ionisantes	0	0	1	1	1	1	1	2	3
36B11	Carburants automobile diesel gas oil fioul gazole	0	0	0	0	0	0	1	2	2
(35110)- 36200	Solvants et diluants organiques - Peintures vernis laques mastics	0	1	1	1	1	1	1	2	3
36J10	Produits capillaires	0	1	1	1	1	1	1	1	3
(21320)- 2Q150- (2Q300)	HAP - Dérivés nitrosés des amines aromatiques aliphatiques - Amines aromatiques et dérivés	0	0	0	0	0	0	2	2	2
36H00	Produits d'entretien	0	0	0	0	0	0	0	1	2
31220	Bois (Poussières)	0	0	0	1	1	1	1	1	2
36J11	Oxydants et décolorants (cosmétiques)	0	0	0	0	0	0	0	1	2
(21320)- 32110	HAP - Amiante (Fibres)	0	0	0	0	0	0	0	2	3
(21320)- 34100	HAP - Fumées de soudage	0	0	0	0	1	1	1	2	3
(21320)- 32110- (35110)	HAP - Amiante (Fibres) - Solvants et diluants organiques	0	0	0	0	2	2	2	2	2
(35110)- 32110	Solvants et diluants organiques - Amiante (Fibres)	0	0	0	0	0	1	2	2	2
(36200)- 2Q300	Peintures vernis laques mastics - Amines aromatiques et dérivés	0	0	0	0	0	1	3	3	3
31330	Tabac (poussières) T66A	0	0	0	0	0	4	6	7	8
34100	Fumées de soudage	0	0	0	0	0	1	4	4	4
32110	Amiante (Fibres)	0	0	0	0	0	0	1	3	7

Tableau 14 : Séquences associées aux Tumeurs Malignes de Vessie susceptibles d'émerger dont les séquences présentant un intérêt en termes d'émergence (potentiellement émergentes : cellules hachurées). Entre parenthèses, les nuisances reconnues au titre des Tableaux MPI.

9.3.3. Echantillon des Lymphomes Non Hodgkiniens

Les Lymphomes diffus Non Hodgkiniens (LNH) sont les hémopathies malignes les plus fréquentes chez l'adulte avec plus de 100 000 nouveaux cas en France chaque année. Avec des facteurs de risque infectieux avérés ou probables, l'environnement professionnel constitue une des pistes d'investigations possibles [Alexander 2007 ; Boffetta 2007]. Les LNH ont fait l'objet de nombreuses études épidémiologiques et plusieurs facteurs d'exposition professionnelle déterminants ont été identifiés [Persson 1999]. Parmi ces expositions, des travaux scientifiques ont identifiés le benzène [O'Connor 1999 ; Smith 2007] ou encore les radiations ionisantes [Karipidis 2007] mais aucune liaison significative n'a pu être mise en évidence.

Dans la base RNV3P, l'échantillon des LNH (code C83) représente 31 PST enregistrés au cours de la période 2001-2009, pour 23 séquences. La taille du spectre des nuisances est de 36 codes. Cette pathologie ne fait pas partie des tableaux MPI, donc aucune nuisance reconnue n'est identifiée. La Figure 48 illustre la répartition des séquences potentiellement émergentes au cours de la période 2001-2009, dont celles susceptibles d'émerger. Le Tableau 15 détaille les séquences identifiées comme potentiellement émergentes (Résultats détaillés en Annexe D-3).

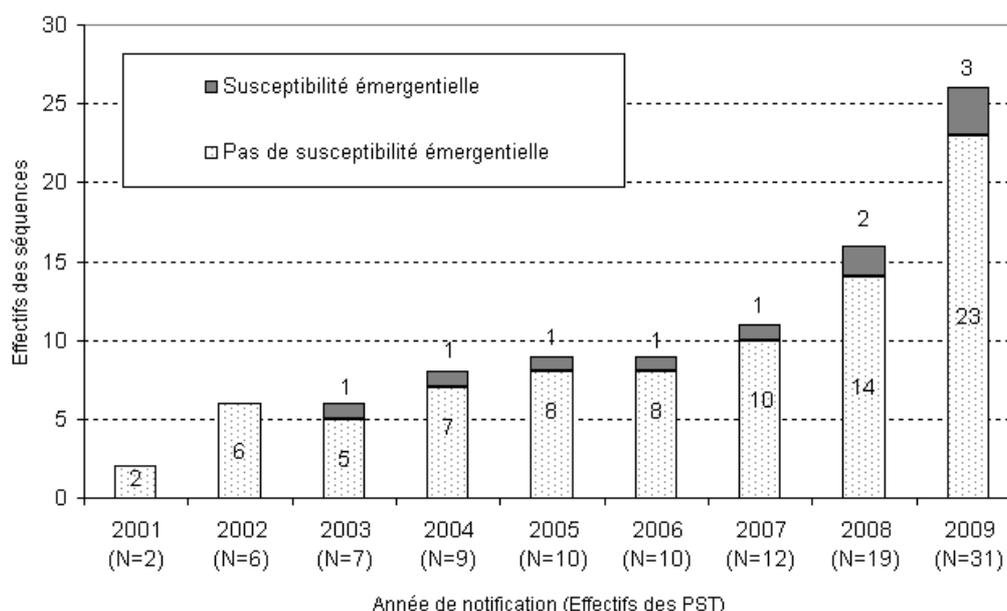


Figure 48 : Répartition des séquences associées aux Lymphomes Non Hodgkiniens enregistrés dans la base RNV3P au cours de la période 2001-2009 (N = 31 PST).

Séquences	2001	2002	2003	2004	2005	2006	2007	2008	2009
21311 Benzène	0	1	2	2	2	2	2	3	4
47100 Radiations ionisantes	0	0	0	0	1	1	1	2	2
36200 Peintures, vernis, laques, mastics	0	0	1	1	1	1	1	1	2

Tableau 15 : Séquences susceptibles d'émerger associées aux Lymphomes diffus Non hodgkiniens parmi lesquelles des séquences potentiellement émergentes (quadrillé).

La nuisance benzène (code 21311) est également associée aux LNH parmi des séquences multi-exposées :

- En 2004, Arsenic - Benzène - Toluène - Chlore (codes 1AS01 - 21311 - 21312 - 1CL01),
- En 2008, Solvants et diluants organiques - Virus oncogènes - Benzène - Herbicides - Trichloroethylene (codes 35110 – 70003 - 21311 - 36R40 - 22224),
- En 2008, Fibres amiante - Benzène - Arc électrique (fumées de soudage) (codes 32110 - 21311 - 34110),
- En 2009, Peintures, vernis, laques, mastics - Solvants et diluants organiques - Benzène (codes 36200 - 35110 - 21311).

La nuisance radiations ionisantes (code 47100) est trouvée dans deux autres séquences :

- En 2002, Virus des bactéries (virus phages) - Radiations ionisantes - Solvants et diluants organiques (codes 70001 - 47100 - 35110),
- En 2009, Radiations ionisantes - Fibres amiante - Solvants et diluants organiques (codes 47100 - 32110 - 35110).

La séquence peintures vernis laques mastics (code 36200) apparait également dans une séquence multi-exposée composée des nuisances Peintures, vernis, laques, mastics - Solvants et diluants organiques - Benzène (codes 36200 - 35110 - 21311) enregistrée en 2009.

* * * * *

D'une part, ces exemples illustratifs mettent en évidence l'apparition d'émergences potentielles en prenant en compte les informations cumulées dans la base RNV3P. Ainsi, une séquence identifiée comme potentiellement émergente l'année N aura probablement le même statut les années suivantes N+1, N+2, etc. L'avantage de cette approche est de conserver les informations enregistrées au cours du temps. Toutefois, avec l'augmentation des effectifs, de telles séquences méritent-elles encore le statut d'émergence potentielle ?

D'autre part, bien que les séquences identifiées comme potentiellement émergentes sont pour la plupart mono-exposées ($L_i=1$) et présentent en faibles effectifs, elles sont susceptibles de partager des expositions professionnelles communes. Il conviendrait alors de développer une approche pour explorer les similitudes des expositions de ces séquences potentiellement émergentes, quelque soit leur fréquence d'enregistrement.

Résumé

9. Détection des associations potentiellement émergentes

*Une **séquence est potentiellement émergente** si elle existe déjà dans la base RNV3P mais dont sa fréquence d'apparition devient anormalement élevée, selon un critère défini.*

Dans ce chapitre, nous avons défini ce critère qui permet de cibler l'apparition "anormale" d'une séquence. Il s'agit d'un critère statistique et composite qui permet de filtrer les séquences observées et d'identifier les séquences dont la fréquence d'apparition n'est pas due au hasard :

- Test 0 : Test de susceptibilité émergentielle (Teste la susceptibilité d'une séquence d'être émergente),
- Test 1 : Equiprobabilité des fréquences (Elimine les séquences dues au hasard et donc à un bruit statistique),
- Test 2 : Réussite émergentielle (Identifie les séquences potentiellement émergentes ou nécessitant un suivi).

Cette méthodologie est appliquée à trois échantillons extraits de la base RNV3P ; les Sclérodermies Systémiques, les Lymphomes diffus Non Hodgkiniens et les Tumeurs Malignes de la Vessie.

10. Exploration de l'émergence au sein de l'exposome

Nous avons développé une méthodologie basée sur des critères quantitatifs pour la détection des séquences présentant un intérêt en termes d'émergence. Trois observations justifient l'utilisation de l'approche par Exposomes pour exploiter et interpréter ces séquences :

- (1) Près de 80% des séquences enregistrées dans la base RNV3P sont des séquences qui n'apparaissent qu'une seule fois au cours de la période 2001-2009 (soit 15 919 / 20 070 séquences). Ce volume de données est exclu de la recherche des séquences potentiellement émergentes avec le test 0 de susceptibilité émergentielle.
- (2) Moins de 6% des séquences enregistrées dans la base RNV3P sont multi-exposées et observées plus d'une fois (soit 1 158 / 20 070 séquences). Cette observation explique le fait que les PST présentant un intérêt en termes d'émergence sont pour la plupart mono-exposés.
- (3) Les séquences présentant un intérêt en termes d'émergence sont susceptibles de partager des nuisances identiques avec les autres séquences associées à une pathologie. En effet, des patients diagnostiqués d'une même maladie peuvent avoir été exposés à diverses nuisances. Inversement, des individus exposés aux mêmes nuisances peuvent développer des pathologies distinctes.

Nous avons montré que l'Exposome est un réseau structuré des connaissances acquises associées à une pathologie donnée. Avec l'identification de séquences potentiellement émergentes, nous nous interrogeons ainsi sur le rôle de chacune dans la structure globale de l'Exposome :

- Les séquences identifiées comme potentiellement émergentes présentent-elles des similitudes ou bien sont-elles isolées des autres séquences ?
- De telles séquences appartiennent-elles à un ou plusieurs groupes d'exposition ?

Pour répondre à ces questions, nous projetons les séquences identifiées comme potentiellement émergentes sur l'Exposome global de la pathologie.

10.2. Caractérisation des PST potentiellement émergents

Si les nœuds potentiellement émergents appartiennent à des groupes d'exposition, quelles sont leurs similitudes en termes d'expositions professionnelles avec les autres nœuds ? Leur popularité est-elle plus élevée ? Nous investiguons alors les propriétés à l'échelle des nœuds, puis à l'échelle des groupes d'exposition.

10.2.1. A l'échelle des nœuds

La Figure 50 illustre la distribution des métriques sur chacun des 90 nœuds associés aux Sclérodermies Systémiques au cours de la période 2001-2009 :

A - 36R00 - Produits à usage agricole

B - 31110 - Chantier (Poussières)

C - 48200 - Froid

D - 34100 - Fumées de soudage

E - 36200-35110 - Peintures, Vernis, Laques, Mastics et Solvants et diluants organiques

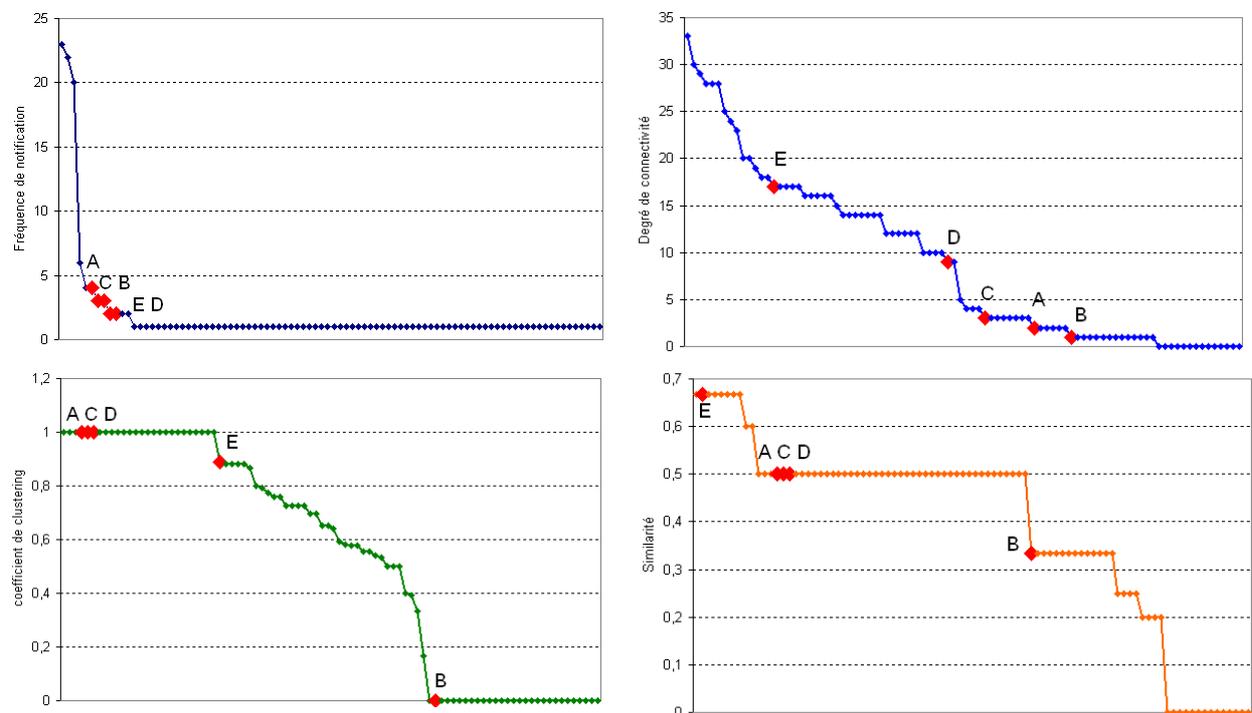


Figure 50 : Distributions des métriques relatives aux 90 séquences associées aux Sclérodermies Systémiques et identification des séquences potentiellement émergentes.

Au niveau des fréquences d'enregistrement, les nœuds identifiés comme potentiellement émergents sont des nœuds observés plus d'une fois du fait qu'ils sont sélectionnés selon les tests t_0 et t_1 (H_0 : "Les fréquences des nuisances sont équiprobables"). De tels nœuds ne sont pas les plus populaires, et présentent des degrés de connectivité moyens, voire faibles. Le nœud E multi-exposé a la plus forte popularité au sein de l'Exposome du fait qu'il partage deux nuisances, 35110 - Solvants et diluants organiques et 36200 - Peintures Vernis Laques Mastics. Les nœuds A, C et D présentent un coefficient de clustering maximum (égal à 1) parce que tous les nœuds auxquels ils sont connectés sont également connectés entre eux. Le nœud B est relié à des nœuds qui ne partagent pas de nuisances entre eux ($c = 0$).

A partir de ces trois métriques, on ne différencie pas de particularités dans leurs propriétés topologiques qui pourraient marquer leur susceptibilité à émerger, du fait qu'ils sont imbriqués dans la topologie de l'Exposome. Ces nœuds identifiés comme potentiellement émergents présentent des valeurs semblables à d'autres nœuds au sein de l'Exposome. Il semble que leur place dans le système ne les prédispose pas à une susceptibilité émergente. Les nœuds A, C, D et E présentent en revanche un indice de similarité élevé, une particularité témoignant du partage d'une ou de plusieurs expositions professionnelles avec les autres nœuds. Nous investiguons alors les métriques à l'échelle des groupes d'exposition.

10.2.2. A l'échelle des groupes d'exposition

Les nœuds mono-exposés ($L = 1$) ne peuvent appartenir qu'à un seul groupe d'exposition. C'est précisément le cas des nœuds A, B, C et D (Tableau 16). En revanche, les groupes d'exposition auxquels ils appartiennent peuvent être des groupes d'exposition hybrides, composés d'autres nuisances partagés ou reliés à d'autres groupes d'exposition. Le nœud multi-exposé E ($L > 1$) est composé de deux nuisances, partagées avec d'autres nœuds au sein de l'Exposome.

	1CL60	1PB01	21300	22122	22224	22225	26110	29111	31100	31110	31130	31131	31160	31161	31162	32110	33200	34100	35110	36200	36300	36910	36C20	36H00	36H30	36H40	36H50	36H60	36R00	48200	49200
A	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
B	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
C	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
D	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
E	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0

Tableau 16 : Matrice **B** des groupes d'exposition simples et séquences potentiellement émergentes.

La Figure 51 illustre le nœud E, multi-exposés dont les nuisances sont Solvants et diluants organiques et Peintures Vernis Laques Mastics. Ce nœud joue ainsi le rôle de "passerelle" entre les deux groupes d'exposition.

Dans la structure globale de l'Exposome, ce nœud présente également des proximités avec d'autres groupes d'exposition (Figure 51).

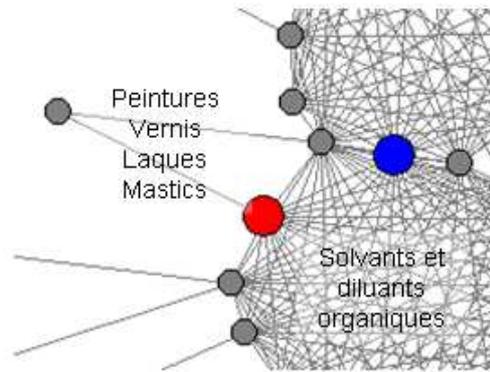


Figure 51 : Nœud E dans le 1^r-Exposome: 35110-36200 - Solvants et diluants organiques - Peintures Vernis Laques Mastics.

Ces nœuds n'appartiennent pas seulement à des groupes d'exposition simples mais sont imbriqués dans la structure globale du réseau. Nous investiguons alors la position des nœuds identifiés comme potentiellement émergents au sein de l'Exposome (Figure 52).

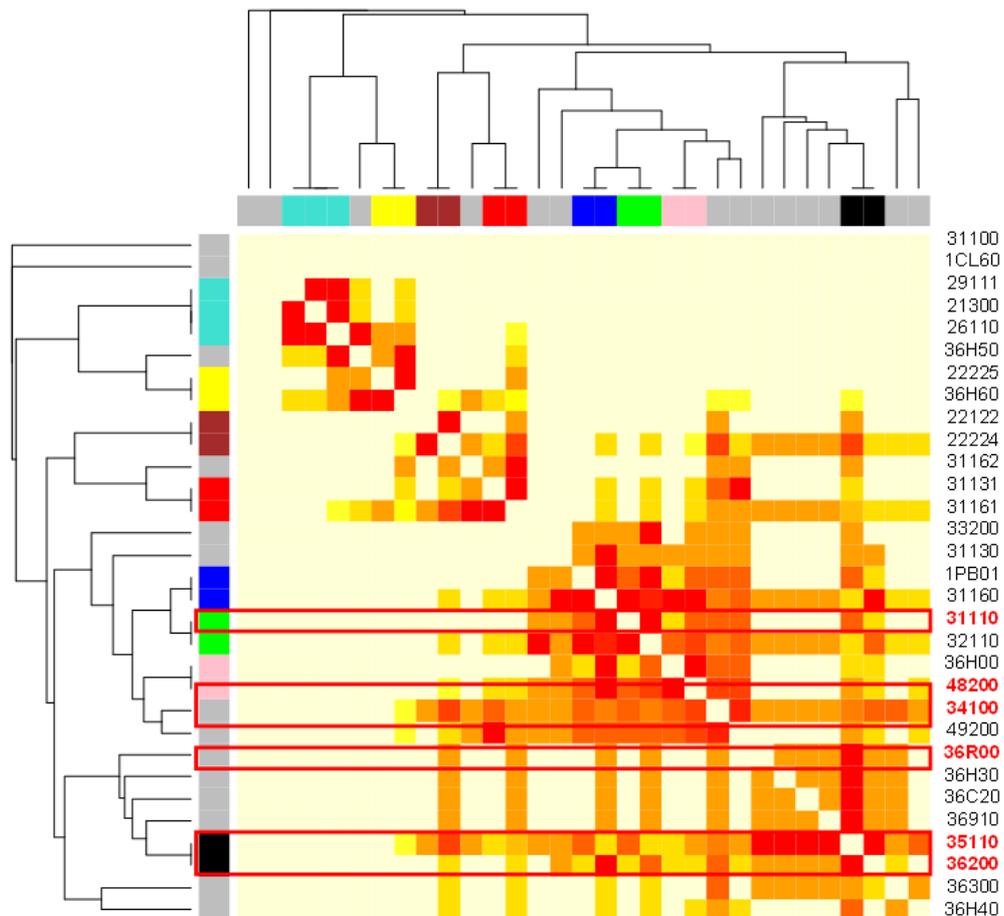


Figure 52 : "Generalized Topological Overlap Measure" des groupes d'exposition associés aux Sclérodermies Systémiques au cours de la période 2001-2009.

La caractérisation des nœuds identifiés comme potentiellement émergents ne nous permet pas d'identifier des marqueurs de l'émergence car les métriques associées aux nœuds ne présentent pas de valeurs "atypiques", pour cette pathologie. Nous avons cependant mis en évidence le ou les groupes d'exposition auxquels ces nœuds appartiennent dans la structure globale de l'Exposome.

10.3. Dynamique temporelle des PST potentiellement émergents

Sur la base de ces observations, peut-on identifier des marqueurs caractéristiques de l'apparition des signaux d'émergence ? Pour répondre à cette question, nous nous intéressons à la dynamique temporelle de l'Exposome.

10.3.1. A l'échelle des nœuds

Afin d'étudier la dynamique temporelle de ces séquences, et de repérer des éventuelles marqueurs d'émergences potentielles, nous avons synthétisé les métriques relatives aux nœuds potentiellement émergents au cours de la période cumulée 2001-2009 (Tableau 17).

	2001	2002	2003	2004	2005	2006	2007	2008	2009
A) 36R00 - Produits à usage agricole									
Connectivité k	-	-	-	-	-	1	2	2	2
Agglomération c	-	-	-	-	-	0	1	1	1
Similarité s	-	-	-	-	-	0,5	0,5	0,5	0,5
B) 31110 - Chantier (Poussières)									
Connectivité k	-	-	0	0	0	0	0	0	1
Agglomération c	-	-	0	0	0	0	0	0	0
Similarité s	-	-	0	0	0	0	0	0	0,33
C) 48200 - Froid									
Connectivité k	-	-	-	0	0	1	3	3	3
Agglomération c	-	-	-	0	0	0	1	1	1
Similarité s	-	-	-	0	0	0,5	0,5	0,5	0,5
D) 34100 - Fumées de soudage									
Connectivité k	-	-	-	-	-	-	4	8	9
Agglomération c	-	-	-	-	-	-	1	1	1
Similarité s	-	-	-	-	-	-	0,5	0,5	0,5
E) 36200-35110 - Peintures, Vernis, Laques, Mastics et Solvants et diluants organiques									
Connectivité k	-	-	-	-	8	11	11	16	17
Agglomération c	-	-	-	-	1	0,81	0,81	0,875	1
Similarité s	-	-	-	-	0,5	0,5	0,5	0,5	0,66

Tableau 17 : Dynamique temporelle des métriques relatives aux séquences potentiellement émergentes associées aux Sclérodermies Systémiques (code M34).

Sur la base de ce tableau, les nœuds connectés présentent des métriques non nulles, c'est-à-dire que le nœud appartient à un ou plusieurs groupes d'exposition. A l'inverse, un nœud isolé présente des métriques nulles, et ne partage donc pas d'exposition avec les autres nœuds.

Avec l'apparition des signaux d'émergence potentielle, on observe des nœuds actifs qui conservent leur dynamique en se déployant au sein de l'Exposome. C'est précisément le cas des nœuds A, D et E. Inversement, le nœud B, isolé, conserve une évolution lente, à faible pouvoir agrégatif, en dépit de son potentiel émergent. On observe également l'apparition d'une activité agrégative pour le nœud C, qui va de paire avec l'identification du nœud comme potentiellement émergent.

10.3.2. A l'échelle des groupes d'exposition

A l'échelle des groupes d'exposition, la question est de savoir si les groupes d'exposition incluant les nœuds potentiellement émergents existaient avant la détection ou si l'enregistrement d'un nouveau nœud (ou nouvelle séquence) marque la naissance d'un groupe d'exposition. Le Tableau 18 synthétise les tailles et fréquences des groupes d'exposition simples composés des nœuds potentiellement émergents.

	2001	2002	2003	2004	2005	2006	2007	2008	2009
Nœuds	3	9	20	25	36	50	63	78	90
PST	3	9	26	39	57	78	102	142	171
Groupes d'exposition	0	2	4	6	8	15	24	26	31
36R00 - Produits à usage agricole									
Taille	-	-	-	-	-	2	3	3	3
Fréquence	-	-	-	-	-	2	3	6	6
31110 - Chantier (Poussières)									
Taille	-	-	-	1	1	1	1	1	2
Fréquence	-	-	-	2	2	2	2	2	4
48200 - Froid									
Taille	-	-	-	-	-	2	4	4	4
Fréquence	-	-	-	-	-	3	5	6	6
34100 - Fumées de soudage									
Taille	-	-	-	-	2	2	5	9	10
Fréquence	-	-	-	-	2	2	5	9	11
36200 - Peintures, Vernis, Laques, Mastics									
Taille	-	-	-	-	-	2	2	2	3
Fréquence	-	-	-	-	-	2	2	3	4
35110 - Solvants et diluants organiques									
Taille	-	2	4	4	9	11	11	16	17
Fréquence	-	2	8	11	17	19	23	35	43

Tableau 18 : Dynamique temporelle des tailles et fréquences des groupes d'exposition simples incluant des nœuds potentiellement émergents associés aux Sclérodermies Systémiques.

Les métriques ne semblent pas indiquer la prédisposition d'un nœud ou d'un groupe d'exposition à devenir potentiellement émergent. En effet, des nœuds sont identifiés comme potentiellement émergents avant d'être connecté et inversement. Nous recommandons alors d'établir un suivi sur l'ensemble du groupe d'exposition auquel le ou les nœuds identifiés comme potentiellement émergents appartiennent.

Résumé

10. Exploration de l'émergence au sein de l'Exposome

L'Exposome est une approche conceptuelle qui permet d'étudier les similitudes entre les séquences associées à une pathologie donnée. Cette approche fournit une structuration des données RNV3P sous la forme d'un réseau relationnel. La caractérisation de ce réseau permet d'accéder à une nouvelle dimension de compréhension avec l'identification des groupes d'exposition, simples ou hybrides.

Ces groupes permettent de formaliser les connaissances acquises au cours de l'interrogatoire professionnel et permettent de cartographier le spectre des expositions potentiellement responsables dans la survenue de la pathologie, à différents niveaux de compréhension.

Nous illustrons l'approche par Exposomes avec un échantillon extrait de la base RNV3P, les Sclérodermies Systémiques, et explorons la place occupée par les PST identifiés comme potentiellement émergents et leur dynamique temporelle.

Partie 4 : Discussion - Perspectives

Synthèse - Discussion

Les objectifs de ce travail consistaient à développer une approche pour surveiller les PST enregistrés dans la base RNV3P en prenant en compte la multi-exposition associée, pour détecter des PST potentiellement émergents. Le principal défi méthodologique auquel nous devons faire face était de prendre en compte l'ensemble des éléments de l'exposition professionnelle potentiellement à l'origine de la pathologie.

Nous avons ainsi développé dans un premier temps l'approche par Exposomes, définie comme un réseau de PST partageant au moins "D" élément(s) de l'exposition professionnelle. Dans un deuxième temps, nous avons développé les "Fonctions Expertes" et construit une stratégie pour la détection des PST potentiellement émergents.

Surveillance programmée : Exposomes

La surveillance programmée au sein de la base RNV3P consiste à suivre au cours du temps des pathologies présélectionnées, enregistrées par les médecins à la suite des interrogatoires professionnels. Une simple inspection de la base permet de dénombrer les PST associés à une pathologie ou une exposition professionnelle spécifique.

Compte tenu des contraintes liées à la prise en compte de la multi-exposition et aux faibles effectifs des échantillons de pathologies, nous nous sommes intéressés aux similitudes entre les PST pour construire une cartographie des connaissances acquises par les médecins à la suite des interrogatoires professionnels, appelée approche par Exposomes. Nous exploitons le potentiel de cette approche qui s'inscrit dans la surveillance programmée à deux niveaux :

(1) Caractérisation et surveillance de l'Exposome

La caractérisation de l'Exposome se fait à l'échelle des nœuds définis comme ensembles de PST strictement identiques, et à l'échelle des groupes d'exposition définis comme ensembles de PST partageant au moins "D" éléments de l'exposition professionnelle. Nous rappelons à cette occasion que l'Exposome n'a de sens que si les PST sont multi-exposés, et donc susceptibles de partager des expositions professionnelles.

A l'échelle des nœuds, les métriques relatives aux réseaux relationnels, telles que le degré de connectivité, le coefficient d'agglomération ou encore la similarité, permettent de caractériser leurs propriétés intrinsèques au sein de l'Exposome. Par exemple dans l'échantillon des Sclérodermies Systémiques, les degrés de connectivité ont permis d'identifier les nœuds "populaires", c'est-à-dire des PST qui partagent le plus de similitudes avec les autres PST associés à la pathologie, composés pour la plupart de nuisances reconnues au titre des Tableaux MPI.

Les groupes d'exposition permettent de structurer les connaissances acquises par les médecins pour une pathologie donnée, indiquant la probabilité pour que la pathologie étudiée soit associée à une nuisance, à deux nuisances, etc. Avec en moyenne près de 8 000 PST chaque année, ces enregistrements apportent de nouvelles connaissances entraînant des modifications structurelles au sein de l'Exposome. Ainsi, le suivi de cette dynamique permet de repérer la naissance d'un nœud ou d'un groupe d'exposition, ou de suivre la croissance d'un groupe d'exposition. Par exemple, le groupe des Solvants et diluants organiques dans l'échantillon des Sclérodermies Systémiques est composé de 2 nœuds (pour 2 PST) en 2002. Sa taille augmente de manière linéaire jusqu'en 2009, avec 17 nœuds (pour 43 PST). Pour d'autres groupes, l'évolution est plus lente. C'est le cas par exemple des poussières de chantier. Cette nuisance est rapportée pour la première fois en 2004 avec 1 nœud (pour 2 PST). Il faut attendre l'année 2009 pour voir apparaître 1 nouveau nœud (pour 2 PST) associé à cette nuisance.

Nous avons ainsi structuré les connaissances stockées dans la base de données RNV3P sous la forme de groupe d'exposition. Ces groupes nous apportent ainsi des informations sur une pathologie donnée à savoir le spectre des expositions professionnelles potentiellement à l'origine d'une pathologie sélectionnée et les interactions entre ces éléments à plusieurs niveaux.

(2) Dimension de l'activité professionnelle

L'approche par Exposomes est illustrée dans ce travail à partir de l'échantillon des Sclérodermies Systémiques. Les nœuds formalisent une pathologie associée à un ensemble de nuisances, toute activité professionnelle confondue. Cependant, l'Exposome ne se restreint pas seulement à l'étude d'une pathologie et de ses nuisances, et peut prendre en compte l'activité professionnelle (association secteur - métier).

Par exemple, un nœud peut se formaliser comme une association entre une pathologie, une à cinq nuisances professionnelle, un métier et un secteur d'activité. Ainsi, deux nœuds peuvent être reliés par des nuisances identiques, un métier et/ou un secteur d'activité identique. Un tel Exposome ne fournirait pas seulement le spectre des nuisances associées à une pathologie, mais apporterait une dimension supplémentaire à l'exploration de l'exposition professionnelle multiple. L'intérêt de cette approche serait de construire des groupes d'activité professionnelle, en parallèle des groupes d'exposition.

La Figure 53 synthétise l'utilisation de l'approche par Exposome. Nous avons présenté dans ce travail une application basée sur un échantillon extrait de la base RNV3P, les Sclérodermies Systémiques, afin de garder une bonne visibilité dans l'exploration de l'Exposome. Cependant, les données d'entrée pourraient également englober l'ensemble des données enregistrées dans la base RNV3P à condition de se référer à l'identification des groupes d'exposition compte tenu du manque de lisibilité des sorties graphiques.

A partir de la définition d'une règle de connexion, modulable selon le nombre d'expositions partagées et du poids des nœuds, l'exposome constitue une méthode de data-mining pour identifier des groupes d'exposition. Cette étape aboutit enfin à la surveillance de ces groupes, en termes de suivi structurel et temporel avec le repérage de nouveaux groupes ou d'augmentation des effectifs de nœuds dans un groupe en fonction des connaissances cumulées dans la base RNV3P.

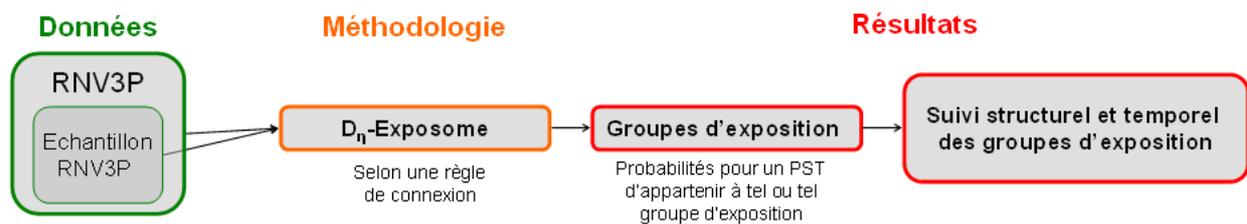


Figure 53 : Synthèse de l'utilisation de l'approche par Exposome.

L'approche par Exposome ne permet pas de répondre aux objectifs de détection des PST potentiellement émergents mais pose les bases d'une surveillance ciblée sur les expositions partagées entre les PST pour une pathologie. La dimension de l'activité professionnelle n'a pas été prise en compte dans ce travail mais est implicitement utilisée dans les Fonctions Expertes.

Détection des PST potentiellement émergents : Fonctions Expertes

Compte tenu des disparités dans le fonctionnement des CCPP et dans le recrutement des patients, nous nous sommes intéressés au processus de construction des PST, notamment aux principales étapes de l'interrogatoire professionnel. Sous l'hypothèse que les nuisances soient choisies en fonction de l'activité professionnelle potentiellement à l'origine de la pathologie, nous avons modélisé le processus de construction des PST en développant des Fonctions Expertes. Ces fonctions reproduisent les choix que les médecins seraient susceptibles d'effectuer pour enregistrer les PST dans la base RNV3P, et reflètent ainsi le cumul de leurs connaissances et pratiques acquises lors de l'interrogatoire professionnel.

- **Fonctions Expertes**

La fonction itérative repose sur l'hypothèse suivante : au fil des étapes de l'interrogatoire professionnel, le médecin construit le PST en choisissant une première nuisance, puis une seconde, une troisième, etc. en fonction du métier et du secteur d'activité, qui selon lui, seraient potentiellement à l'origine de la pathologie.

La fonction par blocs repose sur l'hypothèse suivante : le médecin enregistre le PST une fois que l'interrogatoire professionnel a eu lieu, en paramétrant dans un premier temps les descripteurs des PST (activité professionnelle, nombre de nuisances retenues et leur caractère reconnu au titre des Tableaux MPI), puis en choisissant les nuisances dans le thésaurus selon ces paramètres.

A partir de ces Fonctions Expertes, nous avons élaboré une stratégie, applicable à l'ensemble des pathologies enregistrées dans la base de données RNV3P, pour filtrer les PST et détecter de potentielles émergences. Cette stratégie repose sur les décisions obtenues à la suite d'une série de trois tests successifs.

- **Stratégie de détection des PST potentiellement émergents**

Les applications aux données du RNV3P sont basées sur l'échantillon d'une pathologie. D'une part, pour qualifier une séquence potentiellement émergente ou nécessitant un suivi, la série de tests doit être réalisée de manière indépendante sur l'échantillon. D'autre part, à la différence de la surveillance programmée, cette série de tests peut être appliquée à l'ensemble de la base de données RNV3P, à condition que les paramètres d'entrée englobent également la pathologie.

Le test 0 consiste à exclure de la liste des séquences susceptibles d'émerger les séquences enregistrées seulement une fois au cours de la période étudiée et composées exclusivement de nuisances reconnues au titre des Tableaux MPI.

La condition des fréquences minimales élimine près de 80% des séquences enregistrées dans la base RNV3P (soit 15 919 / 20 070 séquences). On peut cependant introduire une flexibilité à ce niveau pour retenir les cas détectés par la Cellule de Traitement des Alertes (CTA) lors de la veille clinique des émergences, et les classer directement comme susceptibles d'émerger ($t_0 = 1$).

A titre d'exemple, le premier cas soumis à la CTA est une patiente atteinte d'une pneumopathie d'hypersensibilité (J679), exerçant le métier de prothésiste ongulaire (code secteur 930E - code métier 5141) et exposée aux méthacrylates (2F17Z), en particulier aux méthacrylates d'éthyle. Il s'agit d'un cas unique repéré lors d'une consultation pour interrogatoire professionnel et suspecté d'être émergent par le médecin consultant en charge du dossier. Si l'on appliquait les deux conditions du test 0, ce cas ne serait pas identifié comme susceptible d'émerger parce qu'il est enregistré une seule fois.

Le test 1 consiste à vérifier que la fréquence d'apparition d'une séquence n'est pas due au hasard. Pour tester cette hypothèse, l'expérience consiste à distribuer de manière équiprobable les fréquences associées aux séquences. Si une séquence apparaît en quantité moindre que le hasard, alors les connaissances contenues dans les séquences sont assimilées à du "bruit" ($t_1 = 0$). Au contraire, une séquence qui apparaît plus que dans la situation d'équiprobabilité est une séquence présentant un intérêt en termes d'émergence potentielle ($t_1 = 1$).

Nous avons défini les séquences susceptibles d'émerger ($t_0 = 1$) mais dont la fréquence d'apparition est due au hasard ($t_1 = 0$) comme bruit. Cependant, le risque émergentiel peut être visible à ce stade de l'analyse car on suppose que les séquences potentiellement émergentes apparaissent dans un premier temps en faible effectif, avant de présenter des fréquences d'enregistrement anormalement élevées. Ainsi, les séquences dont la série de tests est égal à $T = (1,0,1)$ peuvent être des séquences présentant un intérêt en termes d'émergence car susceptibles d'émerger ($t_0 = 1$) identifiées comme potentiellement émergentes selon une Fonction Experte ($t_2 = 1$). Ces séquences n'apparaissent pas en quantité suffisante ($t_1 = 0$) pour être considérées comme de "vraies émergences potentielles" mais peuvent toutefois être classées comme "nécessitant un suivi".

Le test 2 consiste à vérifier l'hypothèse que les nuisances seraient distribuées de manière équiprobable, faisant intervenir les Fonctions Expertes. Pour tester cette hypothèse, l'expérience consiste à générer des séquences de manière équiprobable, selon les Fonctions Expertes itérative et par blocs. A l'issue de ces simulations, les fréquences théoriques sont comparées aux fréquences observées. Si la fréquence théorique d'une séquence est inférieure à la fréquence observée, alors l'hypothèse de départ n'est pas vérifiée ; les nuisances ne sont pas distribuées de manière équiprobable et la séquence ne présente pas de potentiel émergent.

Pour les trois échantillons testés selon les Fonctions Expertes itérative et par blocs, pas ou peu de différence n'est observée dans les décisions. Cependant, les fréquences théoriques selon les deux Fonctions Expertes sont relativement proches voire identiques (Annexes D).

On peut toutefois se demander si les résultats obtenus ne dépendent pas des schémas décisionnels. Une manière possible de vérifier cette "fonction - dépendance" serait de mesurer la sensibilité et la spécificité des séquences générées selon les Fonctions Expertes. Cette approche nous permettrait de définir laquelle des Fonctions Expertes reproduirait avec le plus de précisions les séquences de la base RNV3P.

La Figure 54 synthétise l'utilisation de la série de tests pour la détection des séquences potentiellement émergentes. Comme pour l'approche par Exposome, les données initiales peuvent être focalisées sur l'échantillon d'une pathologie, comme présenté dans ce travail, ou bien englobées l'ensemble des PST enregistrés dans la base de données RNV3P.

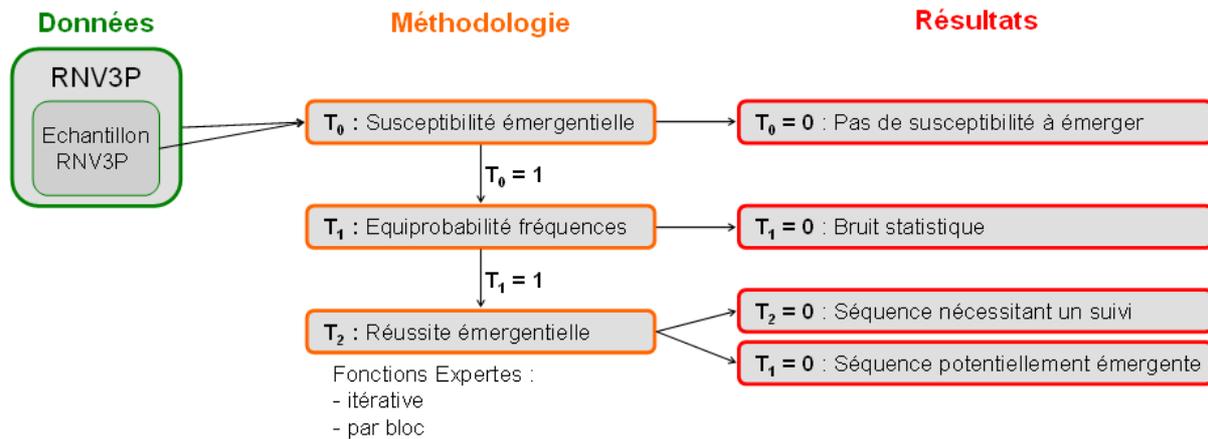


Figure 54 : Synthèse de détection des séquences potentiellement émergentes.

La série de tests, avec l'implémentation des Fonctions Expertes, a ainsi permis d'identifier les séquences potentiellement émergentes à partir des connaissances stockées dans la base RNV3P. La projection de ces séquences sur l'Exposome structure ces informations par rapport à l'ensemble des PST enregistrés pour une pathologie donnée. Une perspective de travail serait de prendre en compte l'ensemble des métriques relatives aux Exposomes dans les schémas décisionnels des Fonctions Expertes. Cela se traduirait par exemple par une étape décisionnelle supplémentaire qui consisterait à modéliser les similitudes en termes d'activités professionnelles ou de nuisances professionnelles.

Les degrés d'imputabilité n'ont pas été pris en compte dans le développement des Fonctions Expertes. Ce choix méthodologique se justifie par plusieurs points :

1/ Variabilité inter-médecins. Les degrés d'imputabilité sont codés dans la base RNV3P comme variables ordinales. Le médecin consultant fournit une appréciation qualitative de l'intensité de la relation entre la maladie et une nuisance retenue à la suite de l'interrogatoire professionnel. Cette donnée est donc soumise à la subjectivité des médecins et implique une variabilité dans la codification, et donc dans son interprétation.

2/ Influence du système de reconnaissance. Le code retenu par le médecin ne nous permet pas de préciser si l'intensité de la relation entre la pathologie et la nuisance est réelle, ou s'il s'agit de pratique de codage liée au système d'indemnisation où la pathologie doit être associée à une nuisance reconnue pour aboutir à une reconnaissance.

3/ Codage peu adapté à la multi-exposition. Un PST multi-exposé est composé de n nuisances ($n > 1$), et forme ainsi n couples pathologie-nuisance. Les n couples sont décrits par n degrés d'imputabilité. Comment évaluer la contribution de chaque nuisance dans l'apparition de la maladie ? D'autant plus que les co-nuisances peuvent être associées à des degrés d'imputabilité égaux.

La prise en compte des degrés d'imputabilité dans le paramétrage des Fonctions Expertes pourrait cependant apporter des informations quant à l'intensité de la relation entre la maladie et une nuisance et ainsi permettre un meilleur classement des séquences potentiellement émergentes. Ce codage nécessiterait l'introduction de variables factuelles et mesurables ; telles que la durée d'exposition.

Les conclusions des applications extraites de la base RNV3P dépendent essentiellement du processus d'enregistrement des PST. C'est pourquoi nous recommandons une grande vigilance lors des phases de sélection et de codification des connaissances enregistrées dans la base RNV3P lors de l'interrogatoire professionnel.

Nous avons précédemment abordé le fait que 75% des PST enregistrés dans la base RNV3P sont mono-exposés. Rappelons que l'approche par exposome n'a de sens que si les PST sont multi-exposés. Un enregistrement systématique des nuisances, comme l'utilisation de matrice emploi-exposition, pourrait-il permettre d'encourager le codage de la multi-exposition ?

Les matrices emploi-exposition s'inscrivent dans le cadre d'étude épidémiologique et ont fait leur preuve pour évaluer rétrospectivement les expositions professionnelles [Bouyer 1993]. Ces matrices sont présentées sous la forme d'un tableau à double entrées croisant les emplois et les nuisances. L'intersection des cellules peut comprendre un ou plusieurs indices d'exposition. Ainsi, pour identifier des expositions rencontrées sur un emploi spécifique, les épidémiologistes se réfèrent à cet outil et les nuisances sont attribuées automatiquement aux malades en fonction de leur activité professionnelle.

Dans le contexte de la base RNV3P, l'utilisation de matrices emplois-expositions présenterait certes un avantage pour mieux caractériser les nuisances professionnelles présentes dans l'environnement professionnel du malade, mais homogénéiserait les informations relatives à l'exposition et constituerait également un frein à la recherche en émergence pour l'enregistrement d'expositions rares, atypiques ou nouvelles.

Perspectives

Les Fonctions Expertes que nous avons développé constituent deux modélisations possibles du processus d'enregistrement des PST dans la base de données. Comme décrit Annexes D, les Fonctions Expertes itérative et par blocs ne produisent pas forcément les mêmes métriques statistiques. Il est tout à fait envisageable de développer d'autres modèles et présentons dans cette section l'implémentation d'une Fonction Experte hybride.

Implémentation d'une Fonction Experte hybride

Un autre modèle possible de Fonctions Expertes combinerait les algorithmes développés pour les Fonctions Expertes itérative et par bloc. La première étape, commune aux autres fonctions que nous avons développées, consisterait à choisir l'activité professionnelle potentiellement responsable de la survenue de la pathologie. Ensuite, le médecin serait amené à rapporter la nuisance principale, conditionnellement à l'activité retenue. Ensuite, le médecin choisit de rapporter L-1 co-nuisances en blocs ; L-1-l co-nuisances reconnues au titre des Tableaux MPI et l le nombre de co-nuisances non reconnues. Cette Fonction Experte hybride s'écrirait alors :

$$F_{\text{hybride}} = P(s-m) \times P(h_1 | s-m) \times P(L | s-m, h_1) \times P(l | s-m, h_1, L) \times P_{L-l}(h_2, \dots, h_5 | s-m, h_1, L, l) \times P_l(h_2, \dots, h_5 | s-m, h_1, L, l)$$

Avec $P(s-m)$ la probabilité d'associer le secteur s et le métier m à la pathologie diagnostiquée, P_{L-l} la probabilité d'associer L-l co-nuisances non reconnues et P_l la probabilité d'associer l co-nuisances reconnues. Ce schéma décisionnel est illustré Figure 55.

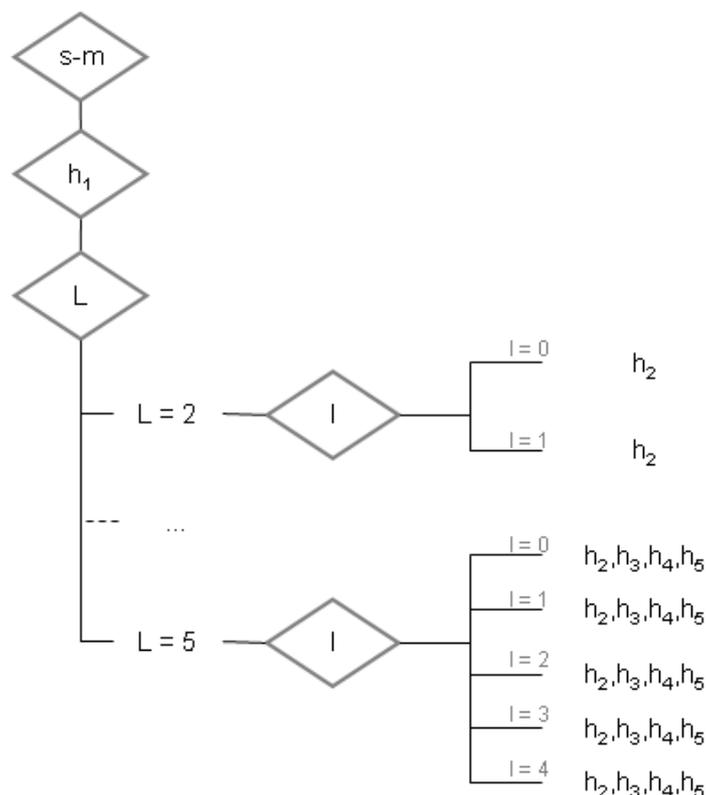


Figure 55 : Organigramme de la Fonction Experte hybride, où la variable s-m définit l'activité professionnelle, h_1 la nuisance principale, L la longueur de la séquence, I le nombre de nuisance reconnues au titre des Tableaux MPI et h_2 à h_5 les co-nuisances associées.

Vers une approche prospective de surveillance et d'émergence

Les Fonctions Expertes sont basées sur l'apprentissage des connaissances accumulées dans la base RNV3P, permettant d'identifier des PST potentiellement émergents. Cette approche ne nécessite pas l'introduction de "nouvelles" nuisances dans les étapes décisionnelles des Fonctions Expertes ($\alpha = 0$). Une perspective de ce travail serait d'utiliser ces connaissances en vue de prédire des PST potentiellement nouveaux. Ainsi, nous avons inclus la possibilité d'introduire des connaissances nouvelles, jamais associées à la pathologie dans la base RNV3P, dans le processus décisionnel ($\alpha > 0$). Rappelons la définition d'une séquence potentiellement nouvelle au sens du RNV3P :

*Une **séquence potentiellement nouvelle** est une séquence jamais observée dans la base RNV3P mais dont la probabilité d'apparaître est supérieure à un seuil défini.*

Pour générer de telles séquences, de nouvelles nuisances jamais associées à la pathologie doivent être combinées pour former de nouvelles séquences. Le travail consiste ensuite à calculer les probabilités d'apparition des séquences jamais observées dans la base, mais susceptible d'apparaître.

- **Génération des séquences selon un modèle empirique**

Par défaut, on fixe le seuil $\alpha = 0,05$ comme la proportion de nouvelles nuisances introduites dans la liste H des nuisances associables à une pathologie donnée. A la manière des tests diagnostiques qui classent les "malades" et les "non malades" en fonction du résultat du test "positif" ou "négatif", les séquences obtenues à la suite des simulations sont classées dans le Tableau 19.

	Observés	Non Observés
Simulés	a	b
Non Simulés	c	∞

Tableau 19 : Tableau de contingence 2x2 servant de base pour le calcul de la sensibilité.

L'ensemble a représente les séquences simulées strictement à l'identiques des séquences observées dans la base RNV3P, l'ensemble b représente les séquences simulées jamais observées dans la base RNV3P, l'ensemble c représente les cortèges observés non simulés.

Du fait que les cortèges soient observés ou simulés, l'ensemble d pourrait rassembler toutes les séquences possibles mais jamais observés ou simulés. Nous avons déjà démontré que le nombre de configurations possibles était exponentiel. L'ensemble d n'est donc pas renseigné.

- **Classement des séquences potentiellement nouvelles**

Dans les tests diagnostiques, la sensibilité d'un diagnostic est sa capacité à donner un résultat positif lorsque la maladie est présente. Cet indicateur mesure la probabilité pour que le test identifie un vrai malade avec un test positif. Dans le cadre des Fonctions Expertes, cette mesure représente la capacité des algorithmes à reproduire à l'identique des séquences observées. Soit Se la sensibilité d'une Fonction Experte, ou la proportion des séquences simulées "bien classées", telle que :

$$Se = \frac{a}{a+b} \quad ; \quad [Se] = Se \pm t(\alpha) \times \sqrt{(1-Se)/(a+b)}$$

Où a est l'ensemble des séquences simulées reproduites à l'identique des séquences observées par rapport à l'ensemble des séquences générées par l'algorithme (a + b), avec $t(\alpha)$ égal à 1,96 pour un risque d'erreur de 5%.

Les séquences potentiellement nouvelles se situent dans l'ensemble des séquences simulées mais non observées dans l'échantillon de la base RNV3P (ensemble b). Nous classons ainsi les séquences en fonction de leur fréquence d'apparition dans les simulations et en fonction de leur degré de reconnaissance au titre des Tableaux MPI.

Note : Comme il s'agit d'une approche prospective, nous devrions estimer le nombre de nouvelles séquences enregistrées dans la base de données pour l'année N+1 à partir des informations dont nous disposons sur la période cumulée disponible dans la base. L'application suivante ne considère pas ce paramètre et génère des séquences

- **Application illustrative**

Nous choisissons l'échantillon des Sclérodermies Systémiques. Les séquences sont générées selon M = 10 000 échantillons de taille N = 171 PST. La Fonction itérative présente la meilleure sensibilité (0,98 [0,95 ; 0,99]) par rapport à la Fonction par blocs (0,72 [0,65 ; 0,78]). Ainsi, nous présentons un extrait du classement des 33 séquences identifiées comme potentiellement nouvelles selon la Fonction itérative.

Parmi ces 33 séquences associées aux Sclérodermies Systémiques et identifiées comme potentiellement nouvelles, 72% sont multi-exposées (soit 24 / 33 séquences) et 90% sont composées d'au moins une nuisance non reconnue au titre des Tableaux MPI (soit 30 / 33 séquences). Nous présentons seulement les 10 séquences dont les fréquences d'apparition sont les plus élevées (Tableau 20).

Rang	Séquences potentiellement nouvelles	Fréquences d'apparition [Intervalle de confiance à 95%]	L	I
1	(31160)- 32110 - Silice sable - Amiante(Fibres)	0,74 [0,57 ; 0,92]	2	1
2	(1SI01) - Silicium	0,64 [0,49 ; 0,79]	1	1
3	(31160)- 31110 - Silice sable - Chantier (Poussières)	0,63 [0,48 ; 0,78]	2	1
4	31131 - Béton (Poussières)	0,61 [0,47 ; 0,76]	1	0
5	36C10 - Huiles de coupe Huiles d'usinage	0,60 [0,44 ; 0,76]	1	0
6	(31160)- 1PB01 - Silice sable - Plomb	0,58 [0,43 ; 0,73]	2	1
7	36210 - Peintures Vernis Laques Mastics en phase solvant	0,54 [0,41 ; 0,68]	1	0
8	22110 - Dérivés fluorés des hydrocarbures aliphatiques saturés	0,53 [0,38 ; 0,69]	1	0
9	(22220) - Dérivés fluorés des hydrocarbures aliphatiques insaturés	0,50 [0,36 ; 0,64]	1	1
10	(49200) - Vibrations transmises aux membres	0,37 [0,26 ; 0,47]	1	1

Tableau 20 : Classement des séquences potentiellement nouvelles, selon la Fonction itérative. Entre parenthèses, les nuisances reconnues au titre des Tableaux MPI.

Certaines séquences sont composées de nuisances non reconnues au titre des Tableaux MPI (I < L). Il peut également s'agir de recombinaisons de nuisances déjà existantes dans la base RNV3P, ou de nuisances jamais associées à la pathologie.

Par exemple, la séquence Silice sable - Fibres d'amiante est composée d'une nuisance reconnue au titre des Tableaux MPI (Silice sable) et d'une nuisance non reconnues (Fibres d'amiante). Ces deux expositions étaient déjà associées à la pathologie dans l'historique des connaissances de la base, mais jamais associées dans la même séquence.

L'approche par Exposome permettrait ainsi d'étudier la relation de ces séquences potentiellement nouvelles avec les séquences existantes dans la base RNV3P. De telles séquences entraîneraient la naissance de nouveau nœud, et potentiellement de nouveaux groupes d'exposition.

Conclusion

Le travail présenté dans cette thèse a abouti au développement d'approches pour la surveillance et la détection des PST potentiellement émergents. La question est maintenant de savoir comment les décideurs peuvent s'approprier ces méthodologies et satisfaire aux besoins de vigie et d'alerte du réseau RNV3P.

L'approche par Exposome et les Fonctions Expertes ont été développée sous des logiciels spécifiques (SPlus pour la programmation statistique et des logiciels spécifiques aux réseaux relationnels). L'utilisation de ces méthodologies n'est pas facilement accessible. L'automatisation de ces méthodologies semble être un point important à développer pouvant faire l'objet d'un logiciel ou d'une application web.

Les PST identifiés comme potentiellement émergents ou potentiellement nouveaux pourraient être transmis à la Cellule de Traitement des Alertes (CTA) pour investigation scientifique. Les alertes émanant d'un seul centre de consultation pourraient conduire à la diffusion de l'information auprès des entreprises locales par le biais des médecins référents du centre. Les alertes émanant de plusieurs centres de consultation pourraient faire l'objet d'investigations scientifiques, pour aboutir le cas échéant sur la mise en place d'études épidémiologiques analytiques, de type cohorte ou cas-témoins, afin de déterminer le rôle potentiellement étiologique d'une exposition professionnelle dans la survenue de la maladie.

Références

Alexander DD, Mink PJ, Adami HO, Chang ET, Cole P, Mandel JS, Trichopoulos D.

The non-Hodgkin lymphomas: a review of the epidemiologic literature
Int J Cancer. 2007 ; 120 Suppl 12 : 1-39.

Ansermino JM, Daniels JP, Hewgill RT, Lim J, Yang P, Brouse CJ, Dumont GA, Bowering JB.

An evaluation of a novel software tool for detecting changes in physiological monitoring
Anesth Analg. 2009 ; 108 (3) : 873-80

Barabási AL, Albert R.

Emergence of scaling in random networks
Science. 1999 ; 286 (5439) : 509-512.

Barabási AL, Ravasz E, Vicsek, T.

Deterministic Scale-Free Networks
Physica A : Statistical Mechanics and its Applications. 2001 ; 299 (3-4) : 559-564.

Barabási AL, Oltvai ZN.

Network biology : understanding the cell's functional organization
Nat Rev Genet. 2004 ; 5 (2) : 101-113.

Barabási AL. [a]

Network medicine--from obesity to the "diseasome"
N Engl J Med. 2007 ; 357 (4) : 404-407.

Barabási AL. [b]

The Architecture of Complexity
IEEE Control Systems Magazine. 2007 ; 27 (4) : 33-42.

Barabási AL, Gulbahce N, Loscalzo J.

Network medicine: a network-based approach to human disease.
Nat Rev Genet. 2011 ; 12(1) : 56-68.

Blacksell SD, Gleeson LJ, Lunt RA, Chamnanpood C.

Use of combined Shewhart-CUSUM control charts in internal quality control of enzyme-linked immunosorbent assays for the typing of foot and mouth disease virus antigen
Rev Sci Tech. 1994 ; 13 (3) : 687-699

Boffetta P, de Vocht F.

Occupation and the risk of non-Hodgkin lymphoma
Cancer Epidemiol Biomarkers Prev. 2007 ; 16 (3) : 369-372.

Boguñá M, Pastor-Satorras R, Vespignani A.

Absence of epidemic threshold in scale-free networks with degree correlations
Phys Rev Lett. 2003 ; 90 (2) : 028701.

Bonneterre V.

Détection et investigation de maladies professionnelles potentiellement émergentes à partir du Réseau National de Vigilance et de Prévention des Pathologies Professionnelles (RNV3P)
Thèse de doctorat : Modèles, méthodes et algorithmes pour la biologie, la santé et l'environnement : Université de Grenoble ; 2010

Bonneterre V, Faisandier L, Bicout D, Bernardet C, Piollat J, Ameille J, de Clavière C, Aptel M, Lasfargues G, de Gaudemaris R, RNV3P

Programmed health surveillance and detection of emerging diseases in occupational health: contribution of the French national occupational disease surveillance and prevention network (RNV3P)
Occup Environ Med. 2010 ; 67 (3) : 178-186

Borgatti SP, Mehra A, Brass DJ, Labianca G.

Network Analysis in the Social Sciences
Science. 2009 ; 323 (5916) : 892-895.

Boulkedid R, Sibony O, Bossu-Salvador C, Oury JF, Alberti C.

Monitoring healthcare quality in an obstetrics and gynaecology department using a CUSUM chart
BJOG. 2010 ; 117 (10) : 1225-1235

Bouyer J, Hémon D.

Studying the performance of a job exposure matrix
Int J Epidemiol. 1993 ; 22 Suppl 2 : S65-71.

Brooke Marshall J, Spitzner DJ, Woodall WH.

Use of the local Knox statistic for the prospective monitoring of disease occurrences in space and time
Stat Med. 2007 ; 26 (7) : 1579-93

Buckeridge DL, Burkom H, Campbell M, Hogan WR, Moore AW.

Algorithms for rapid outbreak detection: a research synthesis
J Biomed Inform. 2005 ; 38 (2) : 99-113

Burkom HS, Elbert Y, Feldman A, Lin J.

Role of data aggregation in biosurveillance detection strategies with applications from ESSENCE
MMWR Morb Mortal Wkly Rep. 2004 ; 53 Suppl : 67-73

Burkom HS, Murphy S, Coberly J, Hurt-Mullen K.

Public health monitoring tools for multiple data streams
MMWR Morb Mortal Wkly Rep. 2005 ; 54 Suppl : 55-62

Cami A, Wallstrom GL, Hogan WR.

Measuring the effect of commuting on the performance of the Bayesian Aerosol Release Detector
BMC Med Inform Decis Mak. 2009 ; 9 Suppl 1 : S7

Campos J.

Enfermedades infecciosas nuevas y reemergentes: su importancia para la salud pública [New and re-emergent infectious diseases: their importance for public health]
Enferm Infecc Microbiol Clin. 1997 ; 15 (5) : 235-236. Article en espagnol.

Chastel C.

Emergential success: a new concept for a better appraisal of viral emergences and reemergences
Acta Virol. 2000 ; 44 (6) : 375-376

Chen R, Froom P.

The CUSCORE test and the q-interval in cluster analyses of colon cancer and of lymphoma among asbestos workers
Stat Med. 2003 ; 22 (19) : 3101-9

Chen Y, Turner S, Hussey L, Agius R.

A study of work-related musculoskeletal case reports to The Health and Occupation Reporting network (THOR) from 2002 to 2003
Occup Med (Lond). 2005 ; 55 (4) : 268-74

Chiffrot H, Fautrel B, Sordet C, Chatelus E, Sibilia J.

Incidence and prevalence of systemic sclerosis: a systematic literature review
Semin Arthritis Rheum. 2008 ; 37 (4) : 223-35.

Christakis NA, Fowler JH.

The spread of obesity in a large social network over 32 years
N Engl J Med. 2007 ; 357 (4) : 370-379.

Costagliola D, Flahault A, Galinec D, Garnerin P, Menares J, Valleron AJ.

A routine tool for detection and assessment of epidemics of influenza-like syndromes in France
Am J Public Health. 1991 ; 81 (1) : 97-99

Derényi I, Palla G, Vicsek T.

Clique percolation in random networks
Phys Rev Lett. 2005 ; 94 (16) : 160202.

Edge VL, Lim GH, Aramini JJ, Sockett P, Pollari FL.

Development of an Alternative Surveillance Alert Program (ASAP): Syndromic Surveillance of Gastrointestinal Illness Using Pharmacy Over-the-Counter Sales
J Urban Health : Bull N Y Acad Med. 2003 ; 80 Suppl 1 : i138-i139

Faisandier L, Bonneterre V, De Gaudemaris R, Bicout D.

Elaboration d'une méthode statistique pour la détection d'événements émergents : application au Réseau National de Vigilance et de Prévention des Pathologies Professionnelles (RNV3P).
Epidémiol santé anim. 2007 ; 51 : 111-118.

Faisandier L, De Gaudemaris R, Bicout DJ.

Occupational Health Problem Network: the Exposome.
2009 July : 3 p. Article disponible sur : <http://fr.arxiv.org/abs/0907.3410>

Faisandier L, Bonneterre V, De Gaudemaris R, Bicout D.

Occupational Exposome: a network-based approach for characterizing occupational health problems.
J Biomed Inform. Soumis le 18 Avril 2010. Révisé le 17 Janvier 2011.

Fawcett T, Provost F.

Activity monitoring: Noticing interesting changes in behavior
In : Proceedings of the Fifth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, August 15-18, 1999, San Diego, USA

Fowler JH, Christakis NA.

Dynamic spread of happiness in a large social network: longitudinal analysis over 20 years in the Framingham Heart Study
BMJ. 2008 December 4 ; 337 : a2338.

Fricker RD Jr, Hegler BL, Dunfee DA.

Comparing syndromic surveillance detection methods: EARS' versus a CUSUM-based methodology
Stat Med. 2008 ; 27 (17) : 3407-3429

Gaudart J, Giorgi R, Poudiougou B, Touré O, Ranque S, Doumbo O, Demongeot J.

Détection de clusters spatiaux sans point source prédéfini : utilisation de cinq méthodes et comparaison de leurs résultats
Rev Epidémiol Santé Publique. 2007 ; 55 (4) : 297-306

Goh KI, Cusick ME, Valle D, Childs B, Vidal M, Barabási AL.

The human disease network
Proc Natl Acad Sci U S A. 2007 ; 104 (21) : 8685-8690.

Goldberg M.

Amiante : une surveillance épidémiologique à maintenir
Bulletin Epidémiologique Hebdomadaire. 2007 ; 41-42 : 345-346.

Gordon K.

The multi-state Kalman Filter in medical monitoring
Comput Methods Programs Biomed. 1986 ; 23 (2) : 147-154

Hayton P, Utete S, King D, King S, Anuzis P, Tarassenko L.

Static and dynamic novelty detection methods for jet engine health monitoring
Philos Transact A Math Phys Eng Sci. 2007 ; 365 (1851) : 493-514

He X, Zhang J.

Why do hubs tend to be essential in protein networks ?
PLoS Genetics. 2006 ; 2 (6) : 826-834.

Heath RA.

Detection of change in physiological measures using an adaptive Kalman filter algorithm
Psychol Bull. 1984 ; 96 (3) : 581-588

Henning KJ.

What is syndromic surveillance ?
MMWR Morb Mortal Wkly Rep. 2004 ; 53 Suppl : 5-11

Hinchcliff M, Varga J.

Systemic sclerosis/scleroderma: a treatable multisystem disease
Am Fam Physician. 2008 ; 78 (8) : 961-968.

Hogan WR, Cooper GF, Wallstrom GL, Wagner MM, Depinay JM.

The Bayesian aerosol release detector: an algorithm for detecting and characterizing outbreaks caused by an atmospheric release of Bacillus anthracis
Stat Med. 2007 ; 26 (29) : 5225-52

Hurt-Mullen KJ, Coberly J.

Syndromic surveillance on the epidemiologist's desktop: making sense of much data
MMWR Morb Mortal Wkly Rep. 2005 ; 54 Suppl : 141-6

Hutwagner L, Thompson W, Seeman, GM, Treadwell T.

The bioterrorism preparedness and response Early Aberration Reporting System (EARS)
J Urban Health. 2003 ; 80 (2 Suppl 1) : i89-i96

Hutwagner L, Barson JV.

Use of the early aberration reporting system (EARS) for detection of bioterrorism agent attacks
Aviat Space Environ Med. 2005 ; 76 (10) : 1001-2

Inachi S, Mizutani H, Ando Y, Shimizu M.

Progressive systemic sclerosis sine scleroderma which developed after exposure to epoxy resin polymerization
J Dermatol. 1996 ; 23 (5) : 344-346.

Imbernon E, Goldberg M.

Surveillance épidémiologique des risques professionnels : pourquoi et comment ?
Bulletin Epidémiologique Hebdomadaire. 2006 ; 46-47 : 356-358.

Jones KE, Patel NG, Levy MA, Storeygard A, Balk D, Gittleman JL, Daszak P.

Global trends in emerging infectious diseases
Nature. 2008 ; 451 (7181) : 990-993

Kalman RE.

A new approach to linear filtering and prediction problems
Trans ASME J Basic Eng. 1960 ; 82 (Series D) : 35-45

Karipidis KK, Benke G, Sim MR, Kauppinen T, Kricker A, Hugues AM, Grulich AE, Vajdic CM, Kaldor J, Armstrong B, Fritschi L.

Occupational exposure to ionizing and non-ionizing radiation and risk of non-Hodgkin lymphoma
Int Arch Occup Environ Health. 2007 ; 80 (8) : 663-670.

Kaufman Z, Wong WK, Peled-Leviatan T, Cohen E, Lavy C, Aharonowitz G, Dichtiar R, Bromberg M, Havkin O, Kokia E, Green MS.

Evaluation of a syndromic surveillance system using the WSARE algorithm for early detection of an unusual, localized summer outbreak of influenza B: implications for bioterrorism surveillance
Isr Med Assoc J. 2007 ; 9 (1) : 3-7

Keeling MJ, Eames KTD.

Networks and epidemic models
J R Soc Interface. 2005 ; 2 (4) : 295-307.

Kottner J, Halfens R.

Using Statistical Process Control for Monitoring the Prevalence of Hospital-acquired Pressure Ulcers
Ostomy Wound Manage. 2010 ; 56 (5) : 54-59

Kulldorff M, Mostashari F, Duczmal L, Katherine Yih W, Kleinman K, Platt R.

Multivariate scan statistics for disease surveillance
Stat Med. 2007 ; 26 (8) : 1824-33

Lawson BM, Fitzhugh EC, Hall SP, Franklin C, Hutwagner LC, Seeman GM, Craig AS.

Multifaceted syndromic surveillance in a public health department using the early aberration reporting system
J Public Health Manag Pract. 2005 ; 11 (4) : 274-81

Lawson AB.

Disease cluster detection: a critique and a Bayesian proposal
Stat Med. 2006 ; 25 (5) : 897-916

Lober WB, Trigg LJ, Karras BT, Bliss D, Ciliberti J, Stewart L, Duchin JS.

Syndromic surveillance using automated collection of computerized discharge diagnoses
J Urban Health. 2003 ; 80 (2 Suppl 1) : i97-i106

McLeod M, Mason K, White P, Read D.

The 2005 Wellington influenza outbreak: syndromic surveillance of Wellington Hospital Emergency Department activity may have provided early warning
Aust N Z J Public Health. 2009 ; 33 (3) : 289-94

Mora GF.

Systemic sclerosis: environmental factors
J Rheumatol. 2009 ; 36 (11) : 2383-2396.

Morse SS.

Factors in the emergence of infectious diseases
Emerg Infect Dis. 1995 ; 1 (1) : 7-15

Morse SS.

Factors and determinants of disease emergence
Rev Sci Tech. 2004 : 23 (2) : 443-451

Morton AP, Whitby M, McLaws ML, Dobson A, McElwain S, Looke D, Stackelroth J, Sartor A.

The application of statistical process control charts to the detection and monitoring of hospital-acquired infections
J Qual Clin Pract. 2001 ; 21 (4) : 112-117

Najmi AH, Magruder SF.

An adaptive prediction and detection algorithm for multistream syndromic surveillance
BMC Med Inform Decis Mak. 2005 ; 5 : 33

Naus J, Wallenstein S.

Temporal surveillance using scan statistics
Stat Med. 2006 ; 25 (2) : 311-324

Nietert PJ, Silver RM.

Systemic sclerosis: environmental and occupational risk factors
Curr Opin Rheumatol. 2000 ; 12 (6) : 520-526.

Nobre FF, Stroup DF.

A monitoring system to detect changes in public health surveillance data
Int J Epidemiol. 1994 ; 23 (2) : 408-18

O'Connor SR, Farmer PB, Lauder I.

Benzene and non-Hodgkin's lymphoma
J Pathol. 1999 ; 189 (4) : 448-453.

Oltvai ZN, Barabási AL.

Systems biology. Life's Complexity Pyramid
Science. 2002 ; 298 (5594) : 763-764.

Onnela JP, Saramäki J, Hyvönen J, Szabó G, Lazer D, Kaski K, Kertész J, Barabási AL.

Structure and tie strengths in mobile communication networks
Proc Natl Acad Sci U S A. 2007 ; 104 (18) : 7332-7336.

Pal TM, de Wilde NS, van Beurden MM, Coenraads PJ, Bruynzeel DP.

Notification of occupational skin diseases by dermatologists in The Netherlands
Occup Med (Lond). 2009 ; 59 (1) : 38-43.

Park J, Barabási AL.

Distribution of node characteristics in complex networks
Proc Natl Acad Sci U S A. 2007 ; 104 (46) : 17916-20.

Patel CJ, Bhattacharya J, Butte AT.

An Environment-Wide Association Study (EWAS) on type 2 diabetes Mellitus
PLoS ONE. 2010 ; 5 (5) : e10746.

Persson B, Fredrikson M.

Some risk factors for non-Hodgkin's lymphoma
Int J Occup Med Environ Health. 1999 ; 12 (2) : 135-142.

Peto J.

Cancer epidemiology in the last century and the next decades
Nature. 2001 ; 411 (6835) : 390-395

Pilcher DV, Hoffman T, Thomas C, Ernest D, Hart GK.

Risk-adjusted continuous outcome monitoring with an EWMA chart: could it have detected excess mortality among intensive care patients at Bundaberg Base Hospital?
Crit Care Resusc. 2010 ; 12 (1) : 36-41

Punithakumar K, Li S, Ben Ayed I, Ross I, Islam A, Chong J.

Heart motion abnormality detection via an information measure and Bayesian filtering
Med Image Comput Comput Assist Interv. 2009 ; 12 (Pt 2) : 373-380

Rappaport SM.

Implications of the exposome for exposure sciences
J Expo Sci Environ Epidemiol. 17 nov 2010 [Epub ahead of print]

Rayens MK, Kryscio RJ.

Properties of Tango's index for detecting clustering in time
Stat Med. 1993 ; 12 (19-20) : 1813-27

Rolka H, Burkom H, Cooper GF, Kulldorff M, Madigan D, Wong WK.

Issues in applied statistics for public health bioterrorism surveillance using multiple data streams: research needs
Stat Med. 2007 ; 26 (8) : 1834-1856

Sanchez C, Lachaize C, Janody F, Bellon B, Roder L, Euzenat J, Rechenmann F, Jacq B.

Grasping at molecular interactions and genetic networks in Drosophila melanogaster using FlyNets, an Internet database
Nucleic Acids Res. 1999 ; 27 (1) : 89-94.

Sanchirico A, Fiorentino M.

Scale-free networks as entropy competition
Phys Rev E Stat Nonlin Soft Matter Phys. 2008 ; 78 (4 Pt 2) : 046114.

Serfling RE.

Methods for current statistical analysis of excess pneumonia-influenza deaths
Public Health Rep. 1963 ; 78 (6) : 494-506

Shen Y, Adamou C, Dowling JN, Cooper GF.

Estimating the joint disease outbreak-detection time when an automated biosurveillance system is augmenting traditional clinical case finding
J Biomed Inform. 2008 ; 41 (2) : 224-31

Shen Y, Cooper GF.

Bayesian modeling of unknown diseases for biosurveillance
AMIA Annu Symp Proc. 2009 : 589-93

Shen Y, Cooper GF.

A new prior for bayesian anomaly detection: application to biosurveillance
Methods Inf Med. 2010 ; 49 (1) : 44-53

Sibanda T, Sibanda N.

The CUSUM chart method as a tool for continuous monitoring of clinical outcomes using routinely collected data
BMC Med Res Methodol. 2007 ; 7 : 46

Siegrist D, Pavlin J.

Bio-ALIRT biosurveillance detection algorithm evaluation
MMWR Morb Mortal Wkly Rep. 2004 ; 53 Suppl : 152-158

Smith MT, Jones RM, Smith AH.

Benzene exposure and risk of non-Hodgkin lymphoma
Cancer Epidemiol Biomarkers Prev. 2007 ; 16 (3) : 385-391.

Sokolow LZ, Grady N, Rolka H, Walker D, McMurray P, English-Bullard R, Loonsk J.

Deciphering data anomalies in BioSense
MMWR Morb Mortal Wkly Rep. 2005 ; 54 Suppl : 133-139

Sonesson C.

A CUSUM framework for detection of space-time disease clusters using scan statistics
Stat Med. 2007 ; 26 (26) : 4770-4789

Steiner SH, Jones M.

Risk-adjusted survival time monitoring with an updating exponentially weighted moving average (EWMA) control chart
Stat Med. 2010 ; 29 (4) : 444-454

Strogatz SH.

Exploring complex networks
Nature. 2001 ; 410 (6825) : 268-76.

Tango T.

A class of tests for detecting 'general' and 'focused' clustering of rare diseases
Stat Med. 1995 ; 14 (21-22) : 2323-2334

The PLoS Medicine Editors.

It's the Network, Stupid: Why Everything in Medicine Is Connected
PLoS Med. 2008 ; 5 (3) : 333-334.

Toma B, Thiry E.

Qu'est-ce qu'une maladie émergente ?
Epidémiologie et santé animale. 2003 ; 44 : 1-11

Turner S, Lines S, Chen Y, Hussey L, Agius R.

Work-related infectious disease reported to the Occupational Disease Intelligence Network and The Health and Occupation Reporting network in the UK (2000-2003)
Occup Med (Lond). 2005 ; 55 (4) : 275-81

Viboud C, Boëlle PY, Carrat F, Valleron AJ, Flahault A.

Prediction of the spread of influenza epidemics by the method of analogues
Am J Epidemiol. 2003 ; 158 (10) : 996-1006

Wagner MM, Tsui FC, Espino JU, Dato VM, Sittig DF, Caruana RA, McGinnis LF, Deerfield DW, Druzdzel MJ, Fridsma, DB.

The emerging science of very early detection of disease outbreaks
J Public Health Manag Pract. 2001 ; 7 (6) : 51-59

Wagner MM, Moore AW, Aryel RM.

Handbook of biosurveillance.
California USA : Elsevier Academic Press ; 2006

Walsh L, Turner S, Lines S, Hussey L, Chen Y, Agius R.

The incidence of work-related illness in the UK health and social work sector: The Health and Occupation Reporting network 2002-2003
Occup Med (Lond). 2005 ; 55 (4) : 262-267

Wasserman S, Faust K.

Social Network analysis - Methods and applications. Series: Structural Analysis in the Social Sciences n°8. Cambridge UK : Cambridge University Press ; 1994

Watkins RE, Eagleson S, Veenendaal B, Wright G, Plant AJ.

Applying cusum-based methods for the detection of outbreaks of Ross River virus disease in Western Australia
BMC Med Inform Decis Mak. 2008 ; 8 : 37

Watts DJ, Strogatz SH.

Collective dynamics of 'small-world' networks
Nature. 1998 ; 393 (6684) : 440-442.

Westgard JO, Groth T, Aronsson T, de Verdier CH,

Combined Shewhart-cusum control chart for improved quality control in clinical chemistry
Clin Chem. 1977 ; 23 (10) : 1881-1887

Wild CP.

Complementing the genome with an "exposome": the outstanding challenge of environmental exposure measurement in molecular epidemiology
Cancer Epidemiol Biomarkers Prev. 2005 ; 14 (8) : 1847-1850.

Wilson AM, Thabane L, Holbrook A.

Application of data mining techniques in pharmacovigilance
Br J Clin Pharmacol. 2004 ; 57(2):127-134.

Wong WK, Moore A, Cooper G, Wagner M.

WSARE: What's Strange About Recent Events ?
J Urban Health : Bull N Y Acad Med. 2003 ; 80 (2 Suppl 1) : i66-i75

Wong WK, Moore A, Cooper G, Wagner M.

What's Strange About Recent Events (WSARE) : An algorithm for the early detection of disease outbreaks
J Mach Learn Res. 2005 ; 6 : 1961-1998

Woolhouse ME, Gowtage-Sequeria S.

Host range and emerging and reemerging pathogens
Emerg Infect Dis. 2005 ; 11 (12) : 1842-1847

Yip AM, Horvath S.

Gene network interconnectedness and the generalized topological overlap measure
BMC Bioinformatics. 2007 ; 8 : 22.

Zhou H, Lipowsky R.

Dynamic pattern evolution on scale-free networks
Proc Natl Acad Sci U S A. 2005 ; 102 (29) : 10052-7.

Zhu Y, Wang W, Atrubin D, Wu Y.

Initial evaluation of the early aberration reporting system--Florida
MMWR Morb Mortal Wkly Rep. 2005 ; 54 Suppl : 123-30

Zotenko E, Mestre J, O'Leary DP, Przytycka TM.

Why do hubs in the yeast protein interaction network tend to be essential : reexamining the connection between the network topology and essentiality
PLoS Comput Biol. 2008 ; 4 (8) : e1000140.

Logiciels

PROGRAMMATION

S-PLUS® 6.1 for Windows PROFESSIONAL EDITION Release 1

S-PLUS : Copyright 1988, 2002 Insightful Corp.

S : Copyright Lucent Technologies, Inc.

R version 2.9.2

Copyright © 2009 The R Foundation for Statistical Computing

RESEAUX

Ucinet 6 for Windows Version 6.221

Copyright © 1999-2007 Analytic technologies

NetDraw 2.086

Graph Visualization Software 2002

Harvard: Analytic Technologies

CFinder

The cluster (Community) Finding Program

Copyright ©, Department of Biological Physics, Eötvös university

Valorisation des travaux

ARTICLES SCIENTIFIQUES

Faisandier L, Bonneterre V, De Gaudemaris R, Bicout D.

Occupational Exposome : a network-based approach for characterizing occupational health problems.

J Biomed Inform. Soumis le 18 Avril 2010. Révisé le 17 Janvier 2011.

Bonneterre V, Faisandier L, Bicout D, Bernardet C, Piollat J, Ameille J, de Clavière C, Aptel M, Lasfargues G, de Gaudemaris R, RNV3P.

Programmed health surveillance and detection of emerging diseases in occupational health: contribution of the French national occupational disease surveillance and prevention network (RNV3P).

Occup Environ Med. 2010 ; 67 (3) : 178-186.

Faisandier L, Bonneterre V, Bicout DJ, De Gaudemaris R.

Élaboration d'une méthode statistique pour le suivi des pathologies à expositions professionnelles multiples.

Revue d'Epidémiologie et de Santé Publique. 2008 ; 56 (5S) : 259-260.

Faisandier L, Bonneterre V, De Gaudemaris R, Bicout D.

Elaboration d'une méthode statistique pour la détection d'événements émergents : application au Réseau National de Vigilance et de Prévention des Pathologies Professionnelles (RNV3P).

Epidémiologie et santé animale. 2007 ; 51 : 111-118.

COMMUNICATIONS ORALES

Faisandier L, De Gaudemaris R, Bicout DJ.

Using Expert Functions for Identification of the Potentially Emerging Occupational Health Problems.

Meeting Modernet, Monitoring trends in Occupational Diseases and tracing new and Emerging Risks in a NETWORK, Milan, Italie, 18-19 novembre 2010.

Faisandier L, De Gaudemaris R, Bicout DJ.

Approche par modélisation pour la détection des Problèmes de Santé au Travail émergents. Journée de l'Atelier Modélisation du Cluster Environnement, Lyon, 7 octobre 2010.

Faisandier L, Bonneterre V, De Gaudemaris R, Bicout DJ.

New approaches to exploit health Surveillance - The Exposome.

Workshop Occupational Health Surveillance and Related Areas, Manchester, Grande-Bretagne, 1er juillet 2010.

Faisandier L, Bonneterre V, De Gaudemaris R, Bicout DJ.

De la surveillance des expositions professionnelles vers la détection des Problèmes de Santé émergents dans le RNV3P - Approche par exposome.

31ème congrès national de Médecine et Santé au Travail - Symposium Afsset, Toulouse, 31 mai - 4 juin 2010.

Faisandier L, Bonneterre V, De Gaudemaris R, Bicout DJ.

The Exposome, a network-based approach for surveillance of occupational health exposures.

Congrès international EPICOH (Epidemiology in Occupational Health)-MEDICHEM, Occupational Health under Globalization and New Technology, Taipei, Taïwan, 20 - 25 avril 2010.

Faisandier L, De Gaudemaris R, Bicout DJ.

The "Exposome", a network based approach for surveillance of occupational health problems.

Meeting Modernet, Monitoring trends in Occupational Diseases and tracing new and Emerging Risks in a NETWORK, Paris, 17 septembre 2009.

Faisandier L, Bonneterre V, Bicout DJ, De Gaudemaris R.

Caractérisation typologique du réseau national de vigilance et de prévention des maladies professionnelles.

Journées scientifiques de l'AEEMA, Association pour l'Étude de l'Épidémiologie des Maladies Animales, Paris, 30 mai - 1^{er} juin 2007.

Faisandier L, Bonneterre V, De Gaudemaris R, Bicout D.

Surveillance des données du Réseau national de vigilance et de prévention des pathologies professionnelles : élaboration d'une méthode statistique originale pour le suivi des expositions à des nuisances multiples.

11ème colloque ADEREST, Association pour le Développement des Études et Recherches Épidémiologiques en Santé Travail, Nantes, 22-23 novembre 2007.

Faisandier L, Bonneterre V, Bicout DJ, De Gaudemaris R.

Elaboration d'une approche statistique pour le suivi des pathologies à expositions professionnelles multiples.

Congrès international de l'ADEF-EPITER, Association des épidémiologistes de langue française et Association pour l'épidémiologie de terrain, Paris, 10 -12 septembre 2008.

COMMUNICATIONS AFFICHEES

Faisandier L, De Gaudemaris R, Bicout DJ.

Le concept des réseaux relationnels comme une approche pour caractériser et surveiller des Problèmes de Santé au Travail.

8ème journée scientifique de l'EDISCE, Ecole Doctorale Ingénierie pour la Santé, la Cognition et l'Environnement, Université de Grenoble, Grenoble, 7 mai 2009.

Bonneterre V, Faisandier L, De Gaudemaris R, Bicout DJ.

Réseau de vigilance des Pathologies Professionnelles. 1 – Identification et caractérisation des couples pathologie x nuisances

Journées de la Recherche Médicale, Université de Grenoble, UFR de Médecine, Grenoble, 4 avril 2008.

Faisandier L, Bonneterre V, De Gaudemaris R, Bicout DJ.

Réseau de vigilance des Pathologies Professionnelles. 2 – Prise en compte de la multiplicité des expositions.

Journées de la Recherche Médicale, Université de Grenoble, UFR de Médecine, Grenoble, 4 avril 2008.

Programmed health surveillance and detection of emerging diseases in occupational health: contribution of the French national occupational disease surveillance and prevention network (RNV3P)

Vincent Bonneterre

Equipe EPSP, Laboratoire TIMC, UFR de Médecine, Domaine de la Merci, 38706 La Tronche Cedex, France

E-mail vbonneterre@chu-grenoble.fr –

Tel. (33) 4 76 76 54 42

Fax (33) 4 76 76 89 10

Laurie Faisandier, Laboratoire Environnement et Prédiction de la Santé des Populations, TIMC, Université Joseph Fourier, Grenoble, France

Dominique Bicout, Laboratoire Environnement et Prédiction de la Santé des Populations-TIMC, Université Joseph Fourier, Grenoble, France

Cyril Bernardet, Service de Médecine et Santé au Travail, CHU Grenoble, France

Jacques Piollat, Service de Santé au Travail MT2I, Grenoble, France

Jacques Ameille, AP-HP, Unité de Pathologie Professionnelle, de Santé au Travail et d'Insertion, Hôpital Raymond Poincaré, Garches, France

Caroline de Clavière, Agence Française de Sécurité Sanitaire Environnement et Travail (AFSSET), Maisons-Alfort, France

Michel Aptel, Agence Française de Sécurité Sanitaire Environnement et Travail (AFSSET), Maisons-Alfort, France

Gérard Lasfargues, Agence Française de Sécurité Sanitaire Environnement et Travail (AFSSET), Maisons-Alfort, France

Régis de Gaudemaris, for RNV3P* Laboratoire Environnement et Prédiction de la Santé des Populations-TIMC, Université Joseph Fourier, and Service de Médecine et Santé au Travail, CHU Grenoble, France,

RNV3P* : C. Doutrelot-Philippon (Amiens), D. Penneau-Fontobonne, Y. Roquelaure (Angers), I. Tahon (Besançon), P. Brochard, C. Verdun-Esquer (Bordeaux), J.D. Dewitte, B. Lodde (Brest), M. Letourneux (Caen), M.F. Marquignon (Cherbourg), A. Chamoux, L. Fontana (Clermont-Ferrand), J.C. Paireon (Créteil), H.J. Smolik (Dijon), J. Ameille, A. D'Escatha (Garches), A. Maitre, E. Michel (Grenoble), A. Gislard (Le Havre), P. Frimat, A. Sobaszek, (Lille), D. Dumont (Limoges), A. Bergeret, J.C. Normand (Lyon), M.P. Le Hucher-Michel (Marseille), C. Paris (Nancy), D. Dupas, C. Geraut (Nantes), D. Choudat, L. Bensefa (Paris – Cochin), R. Garnier (Paris – Fernand Widal), D. Leger (Paris Hotel Dieu), E. Ben-Brik (Poitiers), F. Deschamps (Reims), C. Verger, A. Caubet (Rennes), J.F. Caillard, J.G. Gehanno (Rouen), D. Faucon (Saint-Etienne), A. Cantineau (Strasbourg), J.M. Soulat (Toulouse), G. Lasfargues (Tours)

* RNV3P is indexed in PubMed

Keywords: epidemiology, statistical models, occupational diseases, prevention and control, registries

Word count: 4698 (+116 for acknowledgments)

ABSTRACT

Objective: The French national occupational disease surveillance and prevention network (RNV3P) includes the 30 occupational disease consultation centres in university hospitals, where patients are referred for a potentially work-related disease, and an occupational health service (OHS). The aim of this work is to demonstrate the contribution of this network to national health surveillance.

Methods: Data from consultations are recorded in standardised occupational health reports and coded using international or national classifications. *Programmed health surveillance* is carried out through annual follow-up of annual referrals to experts for pre-selected disease-exposure associations, as well as incidence estimations for the well characterized working population followed by the OHS. *Hypotheses on new emerging diseases* are generated using statistical methods employed in pharmacosurveillance and by modelisation as an exposome to analyse multiple exposures. **Results:** 58 777 occupational health reports were collected and analysed from 2001 to 2007. Referrals to the 30 university hospital centres increased significantly for asbestos-related diseases, mood disorders and adjustment disorders related to psychological and organisational demands, and for elbow and shoulder disorders related to manual handling. Referrals significantly decreased for asthma, rhinitis related to exposure to organic dusts (vegetable or animal) or to chemicals, except for cosmetics and cleaning products. Estimation of incidences by the occupational health services showed a different pattern in different sectors of activity. The methods for detecting emerging diseases are presented and illustrated by the example of systemic sclerosis, identifying new exposures and new sectors of activity to be investigated. **Conclusion:** The RNV3P collects data from two complementary samples: 30 university hospital centres (workers or former workers) and an occupational health service (current workers). This dual approach is useful for surveillance and for hypothesis generation on new emerging disease-exposure associations.

INTRODUCTION

Health surveillance is an important issue in occupational health since working conditions and occupational exposures are rapidly changing. It is therefore important that national and international systems should be available which are capable of following changes in known occupational diseases and exposures, and above all capable of detecting emerging disease-exposure associations, in particular those related to new technologies and work organisations.

The European Agency for Safety and Health at Work in Bilbao has recently made an inventory of the various national occupational safety and health monitoring systems in Europe.(1) They cover in particular health-related events (health monitoring) such as occupational accidents, diseases and absenteeism, and monitor working conditions and professional exposure (risk monitoring). The European Agency now encourages the integration of data from these various sources. While official systems make it possible to produce European statistics on occupational accidents and diseases, the statistics provided by insurance organisations must be interpreted with great caution because of underestimation of the role attributed to occupation and frequent voluntary underdeclaration by individuals.(2-4)

Health surveillance in France, based on a statutory declaration system, encounters the same problem of underreporting (5), and is focused on already well known risks. A national surveillance system became necessary, based on the network of occupational disease clinical centres which covers all the 30 university hospitals in France and where expert physicians collect standardised reports of work-related diseases, independently of compensation considerations. The national occupational disease surveillance and prevention network (Réseau National de Vigilance et Prévention des Pathologies Professionnelles, RNV3P), created in 2001 with the participation of the national health insurance system (Caisse Nationale d'Assurance Maladie, CNAM) and also supported since 2006 by the French Agency for Environmental and Occupational Health Safety (Agence Française de Sécurité Sanitaire de l'Environnement et du travail, AFSSET), collects every year more than 8000 new occupational health reports throughout France, with the exception of the overseas territories.

The aim of the present study is to describe the practical methods the RNV3P uses for health surveillance and to present some of the findings from the 58 777 occupational health reports collected by the network between 2001 and 2007, with regard to both programmed health surveillance and to detection of emerging diseases, on which research is continuing.

METHODS

The RNV3P combines three different approaches: follow-up of predefined occupational diseases, estimation of annual incidences concerning the data provided by the OHS, and signal generation of potential emerging diseases.

- a) follow-up of predefined occupational diseases reported in the 30 university hospital centres:

The 30 university hospitals throughout the country have an occupational disease consultation centre to which patients are referred by their general practitioner, occupational physician or other specialist, for investigation of the possible occupational nature of their disease. The principal reasons for referring patients are resort to expert knowledge in characterizing the link between exposures and diseases, or need for technical support (diagnoses tests or strategy). Conversely, if the occupational physician himself/herself can handle the problem, patients are not referred to university hospital centres as this is the case for most of the musculoskeletal disorders or cases of hearing loss. Patients are not specifically referred to these centres for compensation procedures as these are not the main experts' scope. The catchment areas mainly cover most big cities, and urban areas, and to a less extend semi-rural or rural areas (1.10 % of our patients work in the agricultural sector –fishing including-

whereas this population represent 4.16% of the working population, according to the 1999 census).

For each patient referred, the experts produce a structured occupational health report whose main coded items are principal disease and comorbid diseases (ICD-10), principal exposure and four other possible exposures (code developed by M Falcy from INRS Institut National de Recherche et Sécurité for the Caisse Nationale d'Assurance Maladie), occupation title (ISCO-88, developed by the ILO) and sector of professional activity (Nomenclature des Activités Françaises, NAF) related to the principal exposure (6, 7). For each of the exposures recorded, the expert gives an opinion on the degree of attributability of the principal disease to professional exposures: "3" for a direct and essential link, "2" for a possible or direct but not essential link, "1" for a weak or doubtful link, and "0" when the expert thinks the disease is not work-related. The level of attributability is assessed by taking into account the patient's history of occupational exposures, estimated levels of exposures, and chronology of appearance of the disease. Demographic characteristics and administrative information on the patients are also recorded. Coding is done by expert physicians in each university hospital centre.

Standardisation of practices and quality improvement are ensured by logistic support provided by AFSSET. Every 18 months, all experts from the university hospital centres undergo quality training to improve coding and procedures are under way for better definition of levels of attributability. Many automated procedures are performed either during data entry on the computer in each centre (out of range data, test of missing data) or during importation into the national data base (data coherence tests according to the typology of occupational health report, age and employment status). An annual potential error report is generated for each centre and corrections are required. Programmed health surveillance is carried out on the basis of the occupational health reports, from 2002 to 2007 inclusive, which have been classified by the experts as work-related diseases (attributability > 0) using conventional statistical methods (χ^2 tests) in order to follow, year by year, the changes in predefined groups of disease-principal exposure associations which were selected for their value in terms of prevention and their high number of reports in the network (at least 20 cases annually). This analysis does not include the year 2001, which was the first year that occupational health reports were collected, so that trends could be studied only in data that were known to have been exhaustively recorded by the centres.

b) Programmed health surveillance based on data from the occupational health service:

This standardised data collection has been extended to an occupational health service near Grenoble university hospital since 2004, whose population affiliated consists of salaried workers of the urban and suburban population working in Grenoble and in nearby towns. The activity of the salaried workers followed by this OHS is presented at the top of the table 4. Some activity sectors are not represented as agricultural activities, construction and education that all depend on specific OHS.

In this service, eight sentinel occupational physicians record all incident work-related diseases (attributability > 0), in the employees whom they follow regularly (10 000 – 12 000). These employees have a medical interview and are examined by their occupational physician at least once every two years. During these clinical visits, the physicians note any newly diagnosed disease supposed to be work-related that has developed since the last visit and produce the corresponding occupational record that is then validated by the expert in the centre.

Their occupational health reports are coded in the same way as those of the university hospital network and the observations are validated with the expert from the associated university hospital centre.(8). Analyses are done separately on this dataset. The incidence of the disease-principal exposure pairs can be estimated, and this was done for the years 2004 to

2007 through experimental extension of the RNV3P network to the occupational health service mentioned above, as the population followed by the sentinel physicians was known (number, sectors of activity). This approach to surveillance is based essentially on the ICD-10 at 2-digit level, while more detailed codes are used to analyse separately musculoskeletal disorders (MSD) of the spine, MSD of the upper limb, carpal tunnel syndrome and other neurological disorders. Annual incidences are calculated as the ratio of the number of new cases of disease discovered by the sentinel physicians to the number of employees that they monitor each year. Incidences are expressed as number of cases/1000 employees per year.

c) Signal generation of potential emerging occupational diseases

The development of original methods of detecting emerging disease-exposure associations is one of the network's essential axes of research. Two ways are currently being explored: methods issued from pharmacosurveillance and a multi-exposure approach, the exposome. The first method applies methods of *signal generation* used by international pharmacosurveillance databases (9, 10), including proportional reporting ratios (PRR) (11) as employed by the British Medicines Control Agency to analyse the data of the pharmacosurveillance computer system Adverse Drug Reactions On-line Information Tracking (ADROIT). The terms "signal" in Pharmacovigilance is defined by WHO as "*reported information on a possible causal relationship between an adverse event and a drug, of which the relationship is unknown or incompletely documented previously*". In our context, we are interested in reported information on possible causal relationship between an adverse medical event and an occupational exposure. These safety Data Mining methods generate a statistical signal when there is a discrepancy between the true number of a disease-nuisance pair reported in the RNV3P database and the expected number for this same pair in the database. Disease-nuisance association giving entitlement to compensation are then excluded, because referring to an already well characterized occupational risk. This approach has been presented in a preliminary publication (12), where the use of this method was restricted to occupational health reports with attributability at least equal to 1, and taking only the principal exposure into account. In the present study, the same method is applied on 2001-2007 data (one more year), taking into account all exposures recorded, even when attributability is nil, in order to look for previously unknown associations (n = 58 777 occupational health reports). The same analysis is also done by sectors of activity and occupations, detecting disease-exposure, disease-occupation and disease-sector of activity pairs which generate a "signal".

The emerging signals are then complemented by the corresponding internal data from the occupational health reports: centres which sent in the reports, level of attributability, associated coexisting exposures, etc. The associations identified in this way are then submitted to national experts for further external investigation, in particular the search for what is known concerning these disease-exposure associations in toxicological and epidemiological publications. The present work takes as an example the results yielded by the PRR method for systemic sclerosis (ICD-10 code M34), a serious disease in which certain occupational factors are established and others suspected.

The second approach named 'exposome' is a network-based approach that we performed to present groups of common exposures shared by patients with multiple exposures and the same disease. The reports are displayed graphically as a network linking occupational situations which share at least one exposure. This approach, familiar to sociologists (13, 14), is known in molecular biology as an interactome and describes the network of interactions between proteins or amino acids within a cell.(15) Within the RNV3P, methodology using exposomes has previously been described for displaying occupational health reports in the network.(16) This was developed with Ucinet 6 (17) and

NetDraw software. (18). The exposome has many interests in terms of hypothesis generation. First of all, it allows identifying the main exposures shared by patients presenting the same disease, in particular the exposures that are not currently being known as risk factors. Secondly, taking the time (in a yearly base) into account, it identifies the appearance of new nodes (new exposures notified for patients presenting the disease), and the speed of growth of all the nodes. If one is growing faster than the other, this is a signal that should be analysed: does it testify better referral to expert and better recognition of a disease-nuisance association, or a diffusion of a new risk factor? An other point of interest of the exposome is the analyse of occupational health reports that are common to different groups of exposure. It may testify that some groups reported with a known risk factor of a disease, may also be exposed to an other under-reported exposure that could also be a risk factor. To illustrate the exposome, we keep the same example consisting of occupational health reports in which systemic sclerosis was the principal disease.

RESULTS

Programmed health surveillance based on patients referred to the university hospital centres

Between January 1, 2001, and December 31, 2007, the network collected 58 777 occupational or environmental health reports concerning potentially work-related diseases. The majority (79%) were potentially related to occupational exposures (attributability >0), 1% were attributed to environmental exposure and in 20% the experts were unable to attribute the disease to either type of exposure (attributability =0). Change from 2001 to 2007 in the number of occupational health reports, patients' demographic characteristics and type of referring physician are shown in table 1.

Table 1 Number of occupational health reports, patients' demographic characteristics and type of referring physician, 2001-2007, in the 30 university hospital centres.

	2001	2002	2003	2004	2005	2006	2007
Number of occupational health reports	6818	7873	8146	8786	8699	9054	9401
Mean age (yr)	46.2	46.7	46.8	47.5	47.6	48.6	49.1
Men (%)	69.3	68.3	66.7	66.2	62.6	62.4	61.7
Referring physician (%)							
Occupational physician	60.7	60.3	61.2	59.1	57.2	54.2	54.0
General practitioner	7.8	6.7	5.7	6.8	6.5	7.9	8.0
Specialist physician	24.0	20.3	22.0	23.3	25.3	27.4	28.1
Other	7.5	12.6	11.1	10.8	10.9	10.5	10.0
Status (%)							
Employed	80.6	82.0	81.8	80.1	77.8	73.8	74.5
Unemployed	3	2.3	2.7	2.0	2.6	2.6	2.5
Retired	13.5	13.4	13.2	15.9	16.7	19.2	19.7
Other	2.9	2.3	2.3	1.9	2.9	4.4	3.2

The number of occupational health reports showed a moderate increase over the years. The trend to an increase in patients' mean age, a higher proportion of referring physicians who were not occupational health specialists, and an increase in the proportion of retired patients confirmed that the network continued to be accessible both to a working population and to those no longer in employment.

Table 2 shows yearly change in the 20 preselected disease-exposure pairs, among the 40 819 health reports from 2002 to 2007 with attributability > 0.

Table 2 Annual variation in numbers of preselected disease-exposure pairs in the 30 university hospital centres

Diseases (ICD-10 codes)	Exposures	2002	2003	2004	2005	2006	2007	p
		6134	6438	6967	6799	7034	7447	
Primary malignant neoplasms of the respiratory tract (C34-C39)	Asbestos and other inorganic substances	229	289	289	299	453	549	0.00
Primary malignant neoplasms of the respiratory tract (C34-C39)	Other chemical products	13	23	20	17	26	45	0.00
Mesothelioma (C45)	Asbestos and other inorganic substances	63	68	87	76	99	94	0.06
Pleural plaque (J92)	Asbestos and other inorganic substances	865	801	1109	1004	985	804	0.00
Mood (affective) disorders (F30-F39) and Adjustment disorders, neurotic, stress-related and somatoform disorders (F40-F48, T74)	Mental exposures and organisational stress	631	881	1151	1306	1466	1614	0.00
Disorders of the cervical spine (M4712, M4822, M50, M500-M509, M530, M531, M5310, M5312, M5313, M5412, M542, M5422, M5423, M5429, M5482)	Physical workload, manual handling, manipulation and postural demands	22	37	44	48	29	35	0,97
Disorders of the thoracic, lumbar and sacral spine (M4316, M4327, M4795, M4806, M4855, M510, M511, M5324, M5338, M5385, M5395, M54, M5416, M543, M5430, M5436, M5437, M5439, M544, M5385, M5446, M5447, M5449, M545, M5450, M5456, M5457, M5459, M546, M5465, M5466, M5484, M5485, M5486, M5488)	Physical workload, manual handling, manipulation and postural demands	237	253	294	240	239	286	0.22
Synovitis, tenosynovitis and other disorders of shoulder or elbow tendons (M75, M70, M77)	Physical workload, manual handling, manipulation and postural demands	264	297	295	361	322	408	0.00
Carpal tunnel syndrome (G50-G59)	Physical workload, manual handling, manipulation and postural demands	174	169	168	198	189	202	0.93
Asthma (J45-J46)	Organic substances (vegetable and animal)	192	147	173	138	129	110	0.00
Asthma (J45-J46)	Cleaning products	20	19	22	27	23	24	0.80
Asthma (J45-J46)	Cosmetics	36	39	43	50	39	34	0.33
Asthma (J45-J46)	Other chemical products	163	145	121	138	107	68	0.00
Rhinitis, rhinopharyngitis (J30-J31)	Organic substances (animal and vegetable)	105	107	106	87	61	83	0.00
Rhinitis, rhinopharyngitis (J30-J31)	Cosmetics	23	28	23	30	22	23	0.32
Rhinitis, rhinopharyngitis (J30-J31)	Other chemical products	38	38	25	31	20	26	0.00
Dermatitis (allergic, irritant, other) (L20-L30)	Organic substances (animal and vegetable)	99	129	124	121	112	114	0.18
Dermatitis (allergic, irritant, other) (L20-L30)	Cleaning products	151	155	151	143	157	133	0.01
Dermatitis (allergic, irritant, other) (L20-L30)	Cosmetics	92	94	87	97	113	126	0.17
Dermatitis (allergic, irritant, other) (L20-L30)	Other chemical products	259	259	250	218	233	213	0.00
Primary or presumed primary malignant neoplasms of lymphoid, haematopoietic and related tissue (C81-C96)	Benzene and other chemical products	16	19	19	28	27	39	0.00

p column: χ^2 test for trend (Armitage)

The main significant increase in referrals to the expert network concern adjustment disorders attributed to mental exposures and organisational stress (about 2.3% increase per year), followed by primary lung cancer attributed to asbestos, and musculo skeletal disorders concerning shoulder and elbow related to manual handling. Referrals for asthma and rhinitis showed a significant annual decrease in exposure to organic dusts (vegetable or animal) as well as in exposure to chemical products other than cleaning products and cosmetics.

Programmed health surveillance based on data from the occupational health service

From 2004 to 2007, the eight physicians of the occupational health service followed a population of 10 725 to 12 781 employees, depending on the year, and collected 907 occupational health reports. Table 3 shows estimated annual incidences in the reports by type of disease (ICD-10).

Table 3 Annual incidence of diseases (ICD-10, 2-digit level) expressed as number of cases per 1000 employees followed in the Grenoble occupational health service

Diseases (ICD-10, 2-digit level) Nb of persons followed	2004 10796		2005 10725		2006 10439		2007 12781	
	n	I	n	I	n	I	n	I
01 Certain infectious and parasitic diseases	2	0.2	3	0.3	1	0.1	1	0.1
04 Endocrine, nutritional and metabolic diseases	1	0.1	1	0.1	1	0.1	2	0.2
05 Mental and behavioural disorders	29	2.7	41	3.8	50	4.8	87	6.8
06 Carpal tunnel syndrome	17	1.6	19	1.8	15	1.4	8	0.6
07 Diseases of the eye and adnexa					1	0.1	1	0.1
08 Diseases of the ear and mastoid process	5	0.5	1	0.1	8	0.8	2	0.2
09 Diseases of the circulatory system	10	0.9	7	0.7	7	0.7	7	0.5
10 Diseases of the respiratory system	9	0.8	11	1	5	0.5	8	0.6
11 Diseases of the digestive system					5	0.5	3	0.2
12 Diseases of the skin and subcutaneous tissue	5	0.5	10	0.9	6	0.6	7	0.5
13 Diseases of the musculoskeletal system, upper limb	43	4	67	6.2	62	5.9	50	3.9
13 Diseases of the musculoskeletal system, spine	64	5.9	81	7.6	31	3	43	3.4
13 Diseases of the musculoskeletal system, other	8	0.7	13	1.2	19	1.8	14	1.1
13 Other rheumatological disorders					1	0.1		
22 Other			2	0.2				
Total	197	18.3	260	24.3	214	20.6	236	18.4

The two disease categories with the highest annual incidence were diseases of the musculoskeletal system (upper limb) and mental and behavioural disorders. The latter increased regularly (from 2.5/1000 in 2004 to 6.8/1000 in 2007). Table 4 shows the mean annual estimated incidences (2004 to 2007) of diseases (ICD-10, 2-digit level), analysed by sector of professional activity.

Table 4 Mean annual incidences, 2004 to 2007, of diseases (ICD-10, 2-digit level) by sectors of activity in the Grenoble occupational health service.

Diseases (ICD-10, 2-digit level)	<i>D</i> Manufacturing industries		<i>G</i> Trade, vehicle and domestic appliance repairs		<i>H</i> Hotel and catering		<i>J</i> Transports, communication		<i>K</i> Property business, rentals, company services		<i>L</i> Public administrations		<i>N</i> Health care and social work		<i>O</i> Other community and social services	
	N° of employees followed		7336		1896		2090		12434		1508		7537		3354	
N° of cases, incidence	n	I	n	I	n	I	n	I	n	I	n	I	n	I	n	I
05 Mental and behavioural disorders	30	4.4	49	6.4	8	4	24	14.2	39	3.2	3	2	36	4.8	12	3.6
06 Carpal tunnel syndrome	8	1.2	10	1.4	2	1.2	1	1.4	8	0.7	5	3.4	19	2.4	4	1.1
06 Other than carpal tunnel syndrome	2	0.3	2	0.3	1	0.6	1	1.4	3	0.3	1	0.7	2	0.3	1	0.3
08 Diseases of the ear and mastoid	9	1.3	4	0.6												
09 Diseases of the circulatory system	7	1	6	0.9			1	0.2	5	0.4			10	1.4	1	0.4
10 Diseases of the respiratory system	13	2	2	0.3					8	0.7	3	2	3	0.4	3	0.9
12 Diseases of the skin	3	0.5	1	0.1	2	1			4	0.3			11	1.4	4	1.1
13 MSD, upper limb	36	5.4	42	6	17	9.3	1	1.4	40	3.2	10	6.7	56	7.2	13	3.8
13 MSD, spine	25	3.9	41	5.7	10	5.2	2	1.8	40	3.3	4	2.5	71	8.9	8	2.3
13 MSD, other	6	0.9	8	1.1	6	3.3			13	1	2	1.2	15	2	2	0.6
13 Other rheumatic diseases			1	0.2					1	0.1						

n : number of cases, I : incidence

For greater clarity, the columns contain only the sectors in which at least 1500 employees were followed during that period, and the rows give only the groups of disorders with at least 10 incident cases. Patterns of disease incidence differed according to sector of activity. Musculoskeletal disorders of the upper limb were most frequent in the hotel and catering business ($p=0.012$), mental and behavioural disorders in the transport and communication sector ($p<0.001$), and musculoskeletal disorders of the spine in health care and social work ($p<0.001$).

Detection of potentially emerging associations by methods used in pharmacosurveillance: the example of systemic sclerosis

Among the occupational health reports classified as work-related diseases, there were 76 cases of systemic sclerosis. Their number rose to 106 taking also into account occupational health reports with attributability nil and systemic sclerosis coded as co-morbidity. All these observations taken together generated 68 different systemic sclerosis x exposure associations, of which seven generated a signal with the PRR method. Associated systemic sclerosis x silica gives entitlement to compensation, whereas the other six associations do not. They are considered as potentially emerging.

Table 5 shows these disease-exposure associations, as well as the disease-sector of activity and disease-occupation pairs which also generate a signal. For each of these associations, the number of cases, the number of hospital university centres which reported these associations, and the years in which the emerging signal was observed are also given.

Table 5. « Systemic sclerosis x exposure », « systemic sclerosis x sector of activity », « systemic sclerosis x occupation » associations generating a signal with the PRR pharmacosurveillance method. Data from the 30 university hospital centres.

Exposure (INRS/CNAM code)		Systemic sclerosis cases		
		Nb	Nb centre	Years of signal
31160	Silica	38	18	2002-7
22224	Trichloroethylene	4	4	2003-7
31146	Hard metal dusts	3	2	2006-7
35110	Solvents	21	10	2003-7
36R00	Pesticides	3	3	2007
48000	Temperature conditions	5	5	2006-7
49000	Vibrations	3	3	2007
Sector of activity (NAF code)				
251	Rubber industry	3	3	2007
453F	Installation of heating and air conditioning equipment	3	3	2005-7
452B	Construction of miscellaneous buildings	6	6	2003-
502Z	Vehicle repair and maintenance	4	4	2006
Occupation (CITP88 code)				
321	Life sciences and health technicians	3	2	2007

821	Machine operators, mineral and metal processing	4	4	2005-7
823	Machine operators, manufacture of rubber and plastic products	6	5	2003-7
914	Building caretakers, window and related cleaners	4	4	2007
932	Labourers in manufacturing industries	3	3	2006
7122	Masons	7	7	2003-7
7136	Plumbers and pipe-fitters	4	4	2003-5

Sample: all occupational health reports, Nb: number of cases, Nb centres: number of occupational disease consultation centres which supplied the reports, Years of signal: years when the emerging signal was reported.

As an illustration, the association systemic sclerosis x installation of heating and air conditioning equipment generated a signal because three plumbers-heating engineer working for this activity sector were reported. This signal is interesting because the three cases come from different centres, and has been recently strengthened by new data (>2007, data not shown). It is noteworthy that some exposures, sectors of activity or occupations generated an emerging signal every year, whereas others have only generated a signal after the cumulative data over six years.

In the second stage of analysis of emergence, these data are compared with those of the literature and a simple classification is proposed according to the level of epidemiological evidence on the one hand and of toxicological evidence on the other. In the present case, among the associations which generate a signal, the relation with solvents, and with halogen solvents in particular, appears almost constantly in the literature.(19, 20) Some of the other associations observed have already been mentioned (vibration, welding fumes).(21, 22) The term of pesticide is too vague in itself to attempt to draw a conclusion, although this risk has recently been suggested in the literature.(23) With regard to temperature conditions, it is probable that cold has been a factor leading to diagnosis (triggering Raynaud's phenomenon).

Multiple exposures and the exposome

The 106 occupational health reports in which systemic sclerosis was the principal disease or one of the comorbidities are shown in figure 1. Each *node* is characterised by a weight equal to the total number of copies of identical occupational health problem in the database. Each link illustrates shared exposures between nodes. The nodes that do not share any exposure with the others nodes are considered as isolated and are shown separately to the right of the figure. The diagonal lines represent at least one exposure shared between nodes. The higher the density of the diagonal lines is, the more nodes share a common exposure. This revealed three clusters around exposures to silica, solvents, and more precisely chlorinated solvents. The numbers of observations in each node are shown as the sum of incident cases for 2001 to 2006, followed by the + sign and by the number of incident cases in 2007 in order to see the novelty brought by the data of the last year analysed. In 2007 there were 27 new observations of systemic sclerosis associated with 15 new exposures (black nodes). Some of these new exposures only occurred once (for example, paints, varnishes and lacquers with aqueous-phase solvents), while others were associated with pre-existing nodes (cement, temperature conditions, epoxy resins, asbestos, formaldehyde, welding fumes). The exposome can also

project occupations and sectors of activity: the three cases of systemic sclerosis in the sector of installation of heating and air conditioning equipment are displayed in the exposome as a black circle (figure 1).

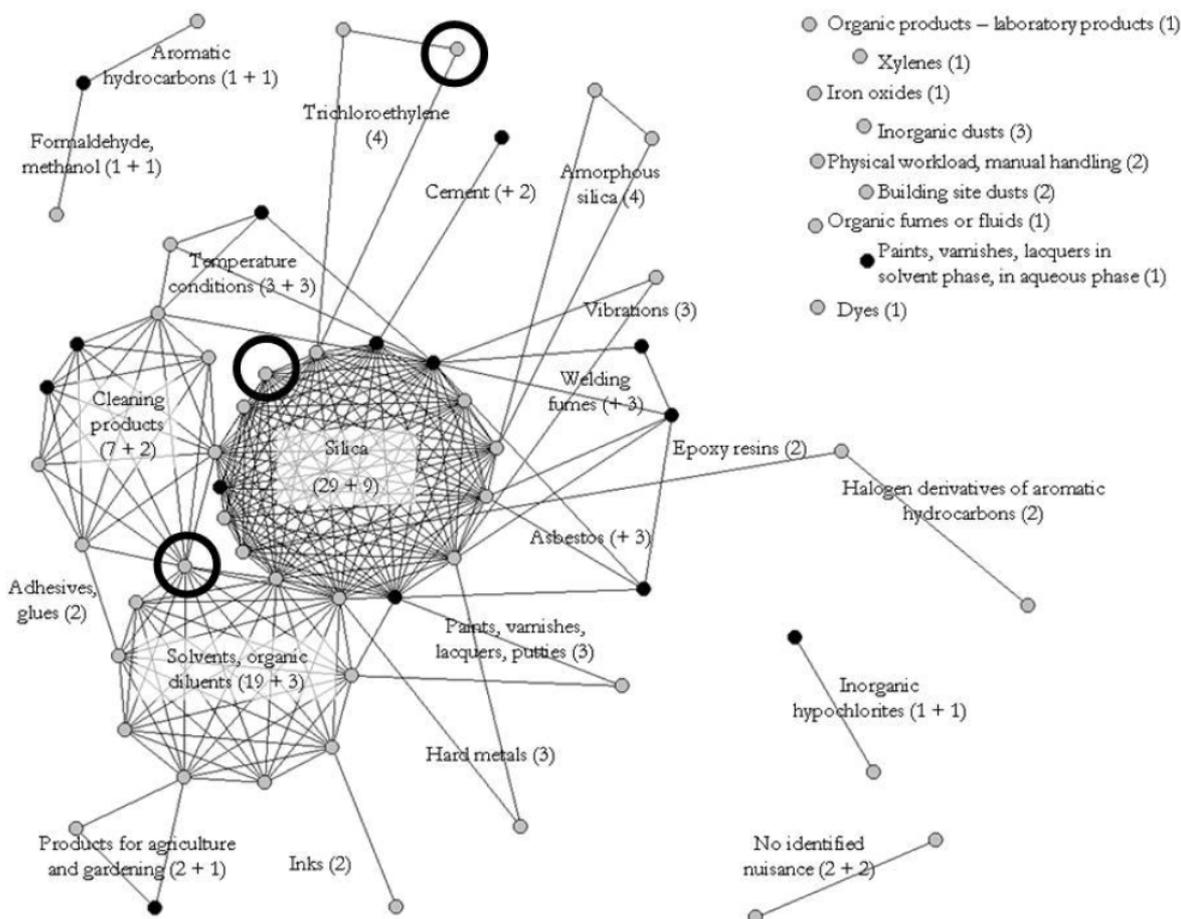


Figure 1 Exposome of the 106 cases of systemic sclerosis. Data from the 30 university hospital centres.

Legend:

- grey dots, nodes created by analysis of the data for 2001-2006;
- black dots, nodes where the number of cases increased with inclusion of the data for 2007;
- black circles, projection of the sector of activity « installation of heating and air conditioning equipment ».
- [the numbers of observations in each node are shown as the sum of incident cases for 2001 to 2006, followed by the + sign and by the number of incident cases in 2007](#)

DISCUSSION

The RNV3P intervenes in health surveillance independently of any statutory considerations related to occupational diseases. The network collects all cases where a physician, of any specialty, and throughout metropolitan France, referred a patient to a university hospital centre to establish the relation between the disease observed and one or several factors to which the patient had been exposed. This approach is particularly interesting for health surveillance and the detection of potentially emerging disease-exposure associations.

We admit that data concerning patients attending the university hospital centres is still of limited *classical* epidemiological value, as the population from which these patients come is not sufficiently known to allow the use of conventional epidemiological indicators. For example, if we look at the cases of systemic sclerosis, the mean age (at first consultation) is 51.9 (SD=10.4), with no difference between men and women), which is consistent with a French study estimating the prevalence of systemic sclerosis in a large county (n=104, mean age at onset 50.2 (SD=14.7), whereas the sex ratio we report is really unusual with two fold more men than women (69 men and 37 women), when all the prevalence studies estimate the sex ratio from 2 to 9 fold higher in women (24) (92% of female gender in the above cited French study). This inverse sex ratio confirms that physicians don't address us all their patients having a specific disease, but patients that have specific occupational exposures. Another point is that greater social concern for some diseases, like adjustment disorders, could explain part of the observed trends ("circularity of argument with referrals"), even if physicians are probably less susceptible to fads than general population.

Nevertheless, **programmed health surveillance** by these 30 centres makes it possible to follow the annual referrals to experts for pre-selected disease-exposure associations observed in sufficient number. Observation over the years of a significant increase or decrease in referrals to the expert network is a valuable *surveillance signal which deserves interpretation*: less need for expert referrals (progress in diagnostic methods, better knowledge of the disease and its causal factors) or a true impact of preventive measures. Thus, the continuous decrease in referrals for asthma and rhinitis related to exposure to organic dusts and to chemical products, other than cosmetics and cleaning products, is probably meaningful because, during these six years of surveillance there has been no substantial change in the characteristics of the 30 centres (number and type of physicians providing consultations, imaging techniques, pattern of physicians referring patients). These observations with regard to asthma are in agreement with the statistics of the national health insurance organisation (Caisse Nationale d'Assurance Maladie, CNAM) on respiratory allergies (25),(26) which also recorded a decrease of allergic asthma and rhinitis compensated the last four years (2001: n=456; 2002: n=458; 2003: n=474; 2004: n=467; 2005, n=403, 2006: n=379, 2007: n=386) and with the data of the THOR network. (27, 28). The remarkably sharp rise in mental and behavioural disorders is mainly related to psychosocial and organisational work factors, which prevalence in France, assessed by Karasek questionnaire, was recently estimated at 18% among men and 27% among women in the last national study (29).

Naturally, all such interpretations have to take into account parameters of the centres as participation in clinical studies, which modifies the recruitment pattern of these centres, but are easily characterized (e.g. the increase of lung cancers in 2006 and 2007 is mainly due to a clinical research program conducted by 5 centers). Trends will also be affected by modifications in the occupational consultation supply (e.g. if there is opening of new

consultations schedules for adjustment disorders, the observed trend could be artificially amplified due to increase in recruitment). In order to improve the quality of the health surveillance interpretations, two research projects are currently planned in the network. The first, which is ongoing, aims at better characterisation of all aspects of the centre effect by controlling for the specific characteristics of each centre that can interfere with the number and type of worker referred; the second, longer-term research programme will be planned with a sociologist and a geographer in order to be able, after taking the centre effect into account, to interpret trends of different observed diseases in relation to socioeconomic and industrial changes.

Another way to compensate this epidemiological shortcoming, was to extend the methodology to an **occupational health service (OHS)** with eight sentinel physicians in Grenoble area. This venture validated the technical feasibility of estimating annual incidence by sector of activity. Extension of data collection, which has begun with four more occupational health services in France in 2007 (Bordeaux, Clermont-Ferrand, Paris, Rouen), will reinforce these estimations of incidence by considerably increasing the number of persons followed and will cover regions with differing occupational activities. It is important to note that the very marked increase in referrals for mental and behavioural disorders is consistent with the rising incidence of these disorders observed in the occupational health services by the sentinel physicians, which indicates good concordance between the two systems of data collection used by the RNV3P. The expected goal of extension to other OHS is to provide national estimations of incidences of occupational health reports according to sectors of activity in order to interpret, with regard to employed persons, the trends in referrals to university hospital centres. In each occupational health service, these statistics are useful for the sentinel physicians so that they can evaluate their professional practices and the direction of their prevention strategies, and such knowledge makes an important contribution to their motivation. Ensuring the exhaustiveness of collection of occupational health reports is a recurrent problem with any system of declaration. The funding of one of the sentinel physicians as a coordinator for their colleagues aims to improve exhaustiveness, like the six-monthly coordination meetings with the experts of the associated university hospital centre.

Early, systematic detection of potentially emerging new occupational risks by the RNV3P is based on systematic data mining methods employed in pharmacosurveillance, in order to identify potentially emerging associations which would not have been recognised by the clinicians. To be detected, a potentially emerging disease must have been coded in a relatively uniform way for at least two of the four principle variables of interest (disease, exposures, occupation, sector of activity). The greater the role which can be attributed to an exposure in a disease, the sooner a potentially emerging disease-exposure association will generate a signal.

It is just as interesting to look at the exposures suggested by case reports in the literature but which do not generate signals with the pharmacosurveillance methods used in the RNV3P. Taking the previous example of systemic sclerosis, epoxy resins (code 36410), which have been suggested in the literature as a possible causal agent (30-32), are only reported with systemic sclerosis (M34) twice in the whole database, but they are also reported twice with localised sclerosis (L94 and L94.0) and epoxy paints (code 36217), and once again with another systemic involvement of connective tissue (M35). A more global coding of the disease under the name of connective tissue disorders generates a signal with the PRR method. There are moreover numerous observations of systemic sclerosis in which according to available data use of epoxy resins was a possibility, and the original dossier would deserve review if the exposures « paints » or « resins » were not precisely coded or in the numerous

cases relating to masons (some resin cements for specific uses contain epoxy resins). Similarly, a recent publication (33) reported a cluster of systemic sclerosis in typography workers exposed to polyvinyl-acetate glues containing up to 1% of vinyl acetate. In the RNV3P data, two cases of systemic sclerosis were reported with inks, two further cases of systemic sclerosis and two cases of lupus were reported with glues of unspecified type, and a single case of systemic sclerosis was linked with use of cyanoacrylate glues.

The exposome helps in interpreting emerging data by drawing particular attention, for selected diseases, to the occupational health reports which share the same exposure. The nodes « glues and adhesives », « epoxy resins » and « inks » are visible before a signal emerges using pharmacosurveillance methods (figure 1).

Lastly, it is interesting to consider these two methods as complementary and to display on the exposome the sectors of activity or the occupations which generate a signal. The variables exposures, occupation, and sector of activity in fact each only give a partial reflection of reality. For instance, three plumbers whose work consisted of installation of heating and air conditioning equipment are displayed in different nodes because they were reported with different exposures, the possibilities being silica, solvents, cleaning products, and trichlorethylene (figure 1). However, they may well have shared other exposures in common which the expert did not consider it useful to report or which were not recognised during the consultation (for example, welding fumes, galvanised and stainless steel, refrigerating gases, and chlorine and/or fluorine derivatives of chlorofluorocarbons and hydrochlorofluorocarbons, exposures which already appear in other nodes). However, the exposome needs further improvement in order to define thresholds of alerts and also to validate its usefulness for the experts, both for diseases which affect few patients and for those where patients are more numerous. All the emerging signals detected by the automated methods we have presented here aim to generate hypotheses. It is important to return pre-alerts to the experts with as much documentation as possible, so that after complementary investigations of the literature, they can decide either that a true alert is needed, or that close surveillance should be set up, or that no further action on this pre-alert is required. With logistic support from the AFSSET, network members are now setting up a cell for the management of alerts.

ACKNOWLEDGMENTS

The authors thank Lynda Larabi for her essential contribution to the processing and quality control of computerised data, the staff of university hospital centres who supply the RNV3P with data, the sentinel physicians of the occupational health services for recording and transmitting incident occupational health reports, the national health insurance organisation (Caisse Nationale d'Assurance Maladie, CNAM) for funding of the consultations and participation in funding of the network, the Agency for Environmental and Occupational Health Safety (AFSSET) for logistics, funding and organisation, the engineers of the regional branches of the national health insurance organisation (Caisses Régionales d'Assurance Maladie, CRAM) for their part in validating exposures and preventive actions, and Nina Crowte for assistance with translation.

COMPETING INTERESTS

None declared.

FUNDING

The RNV3P network receives financial support from the Caisse Nationale d'Assurance Maladie (CNAM), the French Agency for Environmental and Occupational Health Safety (AFSSET) and the university hospitals.

LEGAL

"The Corresponding Author has the right to grant on behalf of all authors and does grant on behalf of all authors, an exclusive licence on a worldwide basis to the BMJ Publishing Group Ltd and its Licensees to permit this article to be published in Occupational and Environmental Medicine editions and any other BMJPG products to exploit all subsidiary rights, as set out in our licence (<http://oem.bmjournals.com/misc/ifora/licenceform.shtml>)."

RÉFÉRENCES

1. European Agency For Safety and Health at Work. Monitoring occupational safety and health in the European Union, Combining data systems. *FORUM*. 2002;11:1-12
2. Fan ZJ, Bonauto DK, Foley MP, Silverstein BA. Underreporting of work-related injury or illness to workers' compensation: individual and industry factors. *J Occup Env Med*. 2006;48:914-922.
3. Rosenman KD, Kalush A, Reilly MJ, C. GJ, Reeves M, Luo Z. How much work-related injury and illness is missed by the current national surveillance system? *J Occup Env Med*. 2006;48:357-365.
4. EUROGIP. Survey on under-reporting of occupational diseases in Europe. Paris: Eurogip 2002:28 pages.
5. Bilger P, Badouin O, Bonnet P, Laroze M. Cancers professionnels: peut on améliorer le repérage et la déclaration à partir des données médicalisées de l'Assurance Maladie ? *Revue médicale de l'assurance maladie*. 2004;35:251-257.
6. Bonnetterre V, Bicout D, Bernardet C, et al. The National Occupational illness surveillance and Prevention Network (RNV3P) and health monitoring. *Sante Publique*. 2008 20:S201-210.
7. Chamoux. Positionnement du Réseau national de vigilance et de prévention des pathologies professionnelles (RNV3P). *Arch Mal Prof Env*. 2006;5:715-717.
8. Celse M, Bonnetterre V, Parson M, Bernardet C, Perdrix A, de Gaudemaris R. Extension du réseau national de vigilance des pathologies professionnelles en service inter entreprise (SIE): un outil épidémiologique pour la vigilance des risques professionnels. *Arch Mal Prof Env*. 2006;5:725-730.
9. Hauben M, Patadia V, Gerrits C, Walsh L, Reich L. Data mining in pharmacovigilance: the need for a balanced perspective. *Drug Saf*. 2005;28:835-842.
10. Roux E, Thiessard F, Fourrier A, Begaud B, Tubert-Bitter P. Evaluation of statistical association measures for the automatic signal generation in pharmacovigilance. *IEEE Trans Inf Technol Biomed*. 2005;9:518-527.
11. Evans SJ, Waller PC, Davis S. Use of proportional reporting ratios (PRRs) for signal generation from spontaneous adverse drug reaction reports. *Pharmacoepidemiol Drug Saf*. 2001;10:483-486.
12. Bonnetterre V, Bicout DJ, Larabi L, et al. Detection of emerging diseases in occupational health: usefulness and limitations of application of pharmacovigilance methods to the database of the French national occupational disease surveillance and prevention network (RNV3P). *Occup Environ med*. 2008;65:32-37.
13. Girvan M, Newman ME. Community structure in social and biological networks. *Proc Natl Acad Sci U S A*. 2002;99:7821-7826.
14. Palla G, Barabasi AL, Vicsek T. Quantifying social group evolution. *Nature*. 2007;446:664-667.
15. Adamcsek B, Palla G, Farkas I, Derenyi I, Vicsek T. CFinder : Locating cliques and overlapping modules in biological networks. *Bioinformatics*. 2006;22:1021-1023.
16. Faisandier L, Bonnetterre V, De Gaudemaris R, Bicout DJ. Elaboration d'une méthode statistique pour la détection d'événements émergents: application au Réseau National de Vigilance et de Prévention des Pathologies Professionnelles (RNV3P). *Epidémiol et santé anim*. 2007;51:111-118.
17. Borgatti S, P, Everett M, G, Freeman L, C. Ucinet for Windows: Software for Social Network Analysis. *Harvard, MA: Analytic Technologies* 2002.

18. Borgatti SP. NetDraw: Graph Visualization Software. *Harvard: Analytic Technologies*. 2002.
19. Kettaneh A, Al Moufti O, Tiev K, P, et al. Occupational exposure to solvents and gender-related risk of systemic sclerosis: a metaanalysis of case-control studies. *J Rheumatol*. 2007;34:97-103.
20. Maitre A, Hours M, Bonnetterre V, et al. Systemic sclerosis and occupational Risk factors : role of solvents and cleaning products. *J Rheumatol* 2004;31:2395-2401.
21. Diot E, Lesire V, Guilmot JL, Metzger MD, Rogier S, Stadler M. Systemic sclerosis and occupational risk-factors: a case-control study. *Occup Environ Med*. 2002;59:545-549.
22. Magnant J, Diot E. Systemic sclerosis: epidemiology and environmental factors. *Presse Med*. 2006;35:1894-1901.
23. Gold LS, Ward MH, Dosemeci M, De Roos AJ. Systemic autoimmune disease mortality and occupational exposures. *Arthritis Rheum*. 2007;56:3189-3201.
24. Le Guern V, Mahr A, Mouthon L, Jeanneret D, Carzon M, Guillevin L. Prevalence of systemic sclerosis in a French multi-ethnic county. *Rheumatology (Oxford)*. 2004;43:1129-1137.
25. CNAMTS-DRP. Dénombrement des maladies professionnelles déclarées et reconnues par le régime général de 2004 à 2007. 2009:29/39 (Tableau 66A).
26. Ministère-du-Travail-Conseil-supérieur-de-la-prévention-des-risques-professionnels. Les accidents du travail et les maladies professionnelles (Chap 19). *Bilan 2007 des conditions de travail*. Paris: Ministère du travail; 2008:385-424 (page cited : 413).
27. McNamee R, Carder M, Chen Y, Agius R. Measurement of trends in incidence of work-related skin and respiratory diseases, UK 1996-2005. *Occup Environ Med*. 2008;65:808-814.
28. McDonald JC, Chen Y, Zekveld C, Cherry NM. Incidence by occupation and industry of acute work related respiratory diseases in the UK, 1992-2001. *Occup Environ Med*. 2005;62:836-842.
29. Ministère-du-travail-DARES. Les facteurs psychosociaux au travail. Une évaluation par le questionnaire de Karasek dans l'enquête SUMER 2003. Premières Informations Premières synthèses. Paris: Ministère du Travail; mai 2008:8p.
30. Inachi S, Mizutani H, Ando Y, Shimizu M. Progressive systemic sclerosis sine scleroderma which developed after exposure to epoxy resin polymerization. *J Dermatol*. 1996;23:344-346.
31. Ishikawa O, Warita S, Tamura A, Miyachi Y. Occupational scleroderma. A 17-year follow-up study. *Br J Dermatol*. 1995;133:786-789.
32. Yamakage A, Ishikawa H, Saito Y, Hattori A. Occupational scleroderma-like disorder occurring in men engaged in the polymerization of epoxy resins. *Dermatologica*. 1980;161:33-44.
33. Magnavita N. Scleroderma cluster among type-setters [Article in Italian]. *G Ital Med Lav Ergon*. 2007;29:440-441.

"WHAT THIS PAPER ADDS" BOX

'WHAT IS ALREADY KNOWN ON THIS SUBJECT?'

Association of the occupational and environmental disease consultation centres of French university hospitals with national organisations responsible for disease prevention (health insurance organisations, national agencies for health and safety) has led to the creation, in 2001, of the national occupational disease surveillance and prevention network (RNV3P), which collects systematic and standardized records of the patients referred to its consultation centres.

'WHAT DOES THIS STUDY ADD?'

This paper shows how the RNV3P is contributing to occupational health surveillance: not only does the RNV3P monitor referrals to experts in occupational diseases at the French national level (follow up of predefined disease-exposure associations), but it is also capable of drawing attention to emerging signals with the application of pharmacosurveillance methods but also with the development of an original network-based approach, taking into account multi-exposure, and called the exposome. These hypotheses (or signals) have to be further investigated by the experts and partners of the network for validation and further action.

ELABORATION D'UNE METHODE STATISTIQUE POUR LA DETECTION D'EVENEMENTS EMERGENTS : APPLICATION AU RESEAU NATIONAL DE VIGILANCE ET DE PREVENTION DES PATHOLOGIES PROFESSIONNELLES (RNV3P)

Laurie Faisandier¹, Vincent Bonnetterre^{1,2},
Régis De Gaudemar^{1,2} et Dominique J Bicout¹

RESUME

Introduction : La surveillance des risques sanitaires occupe une place importante dans le domaine Environnement-Santé. En France, 30 centres de consultations des pathologies professionnelles enregistrent depuis 2001 de façon systématique et standardisée l'ensemble de leurs observations au sein du Réseau National de Vigilance et de Prévention des Pathologies Professionnelles (RNV3P).

Objectif : L'objectif de ce travail est de mettre au point, et de tester sur le RNV3P, une méthode capable d'identifier et de décrire les différents groupes de « cas pathologiques » (ici associations pathologie x nuisances professionnelles), ainsi que leur dynamique temporelle à des fins de vigilance et pour la détection d'événements émergents.

Méthode : La méthode utilisée identifie les différentes associations pathologie x nuisances professionnelles, leur nombre d'occurrences, et connecte entre elles les associations partageant au moins une exposition. Le réseau de cas pathologiques partageant au moins une exposition ainsi construit est nommé exposome. La projection ou comparaison de ces exposomes à des temps différents permet à la fois de suivre la dynamique temporelle des cas pathologiques déjà observés, mais également de mettre en évidence la survenue de cas pathologiques non rapportés jusqu'alors, ainsi que leurs liens avec les cas pré-existants.

Résultat : L'intérêt de la méthode est illustré à partir d'un exemple concret concernant la maladie « sclérodémie systémique ». Les agrégats de cas pathologiques générés pour les données 2002-2005 sont présentés, ainsi que la projection des données 2006.

Conclusion : Il s'agit d'une nouvelle méthode de caractérisation de données, dont l'interface graphique fournit une visualisation utile des agrégats de cas et de leur dynamique temporelle. Cette méthode devrait permettre d'optimiser l'analyse de certaines bases de données dans le domaine Santé-Environnement.

Mots clefs : Maladies professionnelles, maladies émergentes, réseau de surveillance, interactome ou exposome.

* Texte de la communication orale présentée lors de la Journée AEEMA, 1^{er} juin 2007

¹ Laboratoire Environnement et Prédiction de la Santé des Populations TIMC UMR CNRS 5525 Faculté de Médecine, Domaine de la Merci, 38706 LA TRONCHE Cedex, France

² Service de Médecine et Santé au Travail Centre Hospitalier Universitaire de Grenoble B.P. 217, 38043 GRENOBLE Cedex 09, France

Correspondant : Dominique.Bicout@imag.fr, d.bicout@vet-lyon.fr

SUMMARY

Introduction: Monitoring of the medical risks occupies an important place in the field of the environment health. In France, since 2001, 30 centers of consultations of occupational diseases record in a systematic and standardised way all occupational health problems in the National Occupational Disease Surveillance and Prevention Network (RNV3P).

Objective: The aim of this work is to develop and test on RNV3P, a method allowing to identify and describe different groups of "pathological cases" (disease x occupational nuisances associations) with their time dynamics, for ends of vigilance and detection of emerging associations.

Method: This method identifies different disease x occupational nuisances associations, their numbers of events and connects associations sharing at least an exposure in common. We named exposome such a network of pathological cases. The projection or comparison of these exposomes at different times allow both to follow the time dynamics of already observed pathological cases but also to single out new pathological cases never observed before.

Results: As an illustrating example, the method is applied to the disease of "systemic sclerosis". The groups of pathological cases generated from 2002-2005 data are reported and compared with the 2006 exposome.

Conclusion: This new method of characterizing data is very promising in optimising analyses of databases in the field of Environnement-Health.

Keywords: Occupational diseases, emerging diseases, occupational network, exposition network or exposome.



I - INTRODUCTION

L'identification des risques sanitaires liés à l'environnement occupe une place prépondérante dans le domaine de la santé publique. Certains réseaux de surveillance, comme le Réseau National de Vigilance et de Prévention des Pathologies Professionnelles (RNV3P) participent à cette action de veille sanitaire par un enregistrement systématique de cas pathologiques et de données

d'exposition. L'objectif de ce travail est de présenter une méthode statistique permettant de caractériser ce type de bases de données, et de détecter d'éventuels phénomènes émergents. Cette méthode est appliquée au RNV3P afin de discuter son aptitude à optimiser la vigilance dans le domaine santé environnement.

II - MATERIEL ET METHODE

L'existence d'interactions entre les associations « pathologie x nuisances professionnelles » justifie le recours à l'analyse sous forme de réseau (au sens de connections dans un graphe) et à la représentation graphique du RNV3P. Les réseaux permettent de décrire des systèmes complexes de relations entre différents acteurs. Cette approche est appliquée à des champs

disciplinaires variés comme dans le domaine de la sociologie ou encore de la biologie moléculaire [Girvan et Newman, 2002 ; Palla *et al.*, 2007]. Par exemple, le réseau d'interactions entre protéines ou acides aminés au sein d'une cellule peut être représenté par un interactome [Adamcsek *et al.*, 2006]. Dans le cadre du réseau de surveillance RNV3P, nous avons baptisé

exposome le réseau des « cas pathologiques » partageant au moins une exposition professionnelle. Dans la suite, nous allons décrire le concept générique d'exposome puis analyser l'application aux données du RNV3P. La construction puis l'exploitation de l'exposome de RNV3P se fera en deux étapes : la première consiste à identifier les groupes de « cas pathologiques » et à relier entre eux ceux qui partagent au moins une exposition (statistique descriptive), et la deuxième étape consiste en l'analyse de la dynamique temporelle de ces cas et

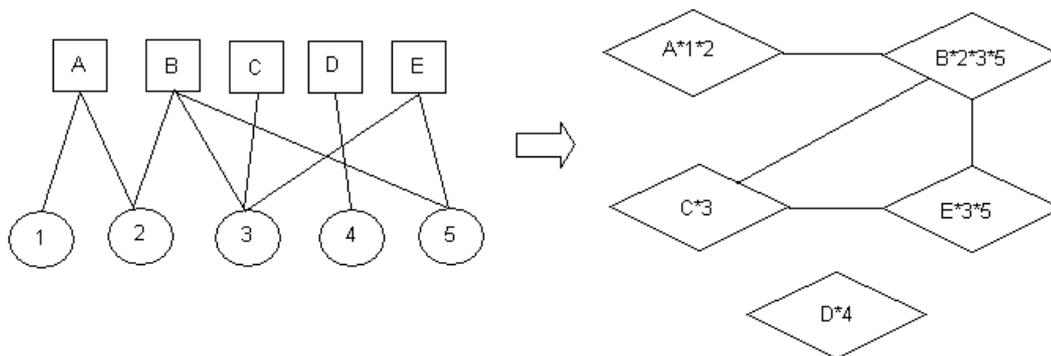
notamment la recherche de cas émergents (statistique exploratoire).

1. LE RESEAU BIPARTITE

Une observation est l'association entre une exposition et une réponse. Elle caractérise l'état d'un individu. Sur la figure 1, les cercles et les chiffres représentent les facteurs relatifs à l'exposition, et les carrés et les lettres les facteurs réponses. Une réponse peut être associée à un ou plusieurs facteurs d'exposition. C'est par exemple le cas du carré B qui est associé aux expositions 2, 3 et 5.

Figure 1

Du réseau bipartite à l'exposome



2. DEFINITION ET PRINCIPE DE L'EXPOSOME

En s'appuyant sur la figure 1, le carré A est associé aux expositions 1 et 2, leur association forme le nœud A*1*2. Chaque nœud est unique et contient des observations identiques. Par exemple, si n associations de A avec les expositions 1 et 2 sont observées, alors le nœud A*1*2 contient un effectif égal à n mais n'apparaît qu'une seule fois au sein l'exposome. La représentation graphique des relations entre ces nœuds permet de visualiser des groupes [Everett et Borgatti, 1998 ; Newman et Girvan, 2004] qui ont des expositions connues et communes et de les suivre dans le temps [Bernardet *et al.*, 2007]. Deux nœuds sont connectés s'ils partagent au minimum une exposition. Sur la figure 1, les nœuds A*1*2 et B*2*3*5 sont connectés car ils partagent l'exposition 2. Les nœuds non connectés sont ceux qui n'ont aucune exposition en commun. Des nœuds isolés (ou indépendants) sont des nœuds qui ne partagent pas d'expositions avec les autres

nœuds. Sur la figure 1, le nœud D*4 est isolé parce qu'aucun autre nœud n'est associé à l'exposition 4. Au sein de l'exposome, la taille des nœuds est proportionnelle au nombre d'observations qu'ils englobent. De plus, pour chaque représentation graphique, on définit un seuil λ en dessous duquel le nœud n'est pas considéré dans l'exposome. Tous les nœuds apparaissent dans l'exposome lorsque λ = 1 tandis que pour λ égal à 2, le réseau sera constitué des nœuds qui contiennent au minimum deux observations.

3. APPLICATION AU RNV3P

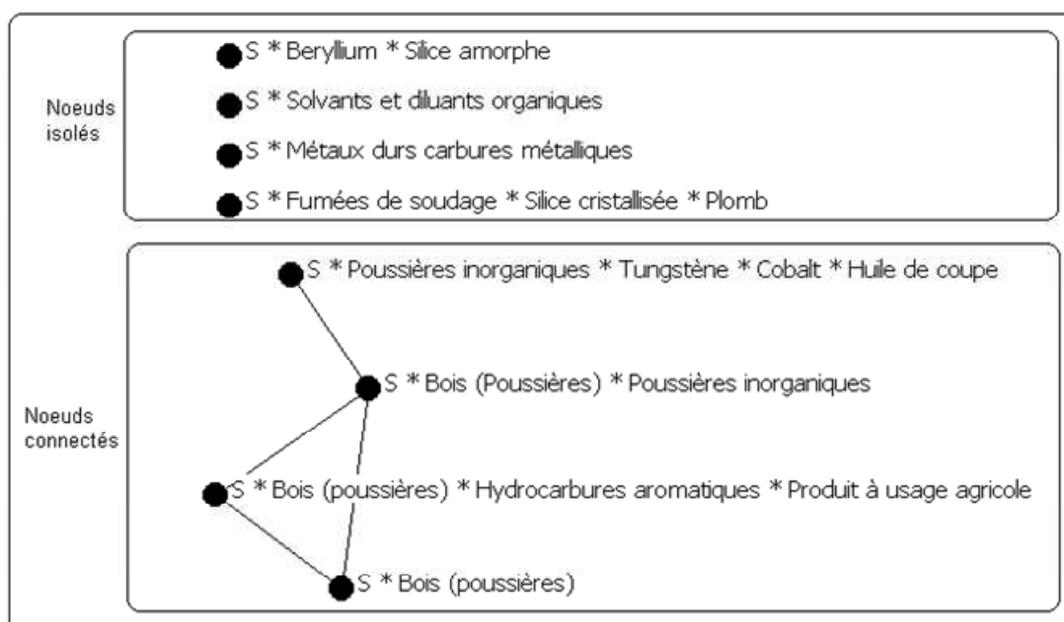
Le RNV3P est un réseau national d'experts qui recense de façon systématique et standardisée l'ensemble des problèmes de santé au travail à l'occasion des consultations de pathologies professionnelles [Bonnetterre *et al.*, 2006 ; Celse *et al.*, 2006]. Ce réseau est opérationnel depuis 2001 sous la forme d'une base de données où chaque observation correspond à un problème de santé au travail

décrit par une pathologie principale (code Classification Internationale des Maladies et des problèmes de santé connexes – CIM10), une à cinq nuisances professionnelles incriminées (code Institut National de Recherche et de Sécurité - INRS) avec leur degré d'imputabilité (faible, moyen, fort), un métier (code Classification Internationale Type des Professions - CITP) et un secteur d'activité (code Nomenclature des Activités Françaises - NAF).

Pour appliquer le principe de l'exposome au RNV3P, un cas pathologique (représenté par un nœud) correspond à un code pathologie associé à un cortège de cinq codes nuisances maximum et englobe n observations identiques. Ces nœuds sont connectés s'ils partagent au minimum une nuisance professionnelle. Des agrégats de nœuds sont identifiés ainsi que leurs structures. La figure 2 illustre un exemple d'exposome concernant le diagnostic principal de la sarcoïdose.

Figure 2

Exposome de la sarcoïdose, période 2002-2005, seuil = 1 – Logiciel Ucinet 6



La sarcoïdose (ou lymphogranulomatose bénigne) est une maladie systémique de cause inconnue qui atteint surtout les poumons mais qui peut également toucher d'autres organes. Cette pathologie est un exemple intéressant car les médecins ignorent la cause réelle de cette maladie et suspectent, au moins pour partie, une origine environnementale [Judson, 2003 ; Moller et Chen, 2002] : l'hypothèse retenue aujourd'hui est que la sarcoïdose serait une réaction immuno-pathologique associée à des antigènes variés (microbiens ou non) survenant chez des sujets génétiquement prédisposés. L'exposome de cette pathologie permet de fournir une représentation des cas observés au sein du réseau en identifiant des agrégats définis par une exposition commune. Deux types de

nœuds sont mis en évidence : les nœuds isolés et les nœuds connectés. Ces derniers forment deux agrégats d'exposition qui sont « poussières de bois » ainsi que « poussières inorganiques ». Un nœud caractérisé par ces deux nuisances permet de faire le lien entre ces agrégats. Les nœuds isolés sont « indépendants » des nœuds connectés parce qu'aucun d'entre eux ne partage de nuisance professionnelle. On remarquera toutefois que l'exposition à des poussières de métaux, ou à la silice sont des cas particuliers d'exposition à des particules inorganiques.

Les logiciels utilisés sont Splus 6.1 pour l'analyse statistique des données, CFinder 1.21 pour la conception graphique de la structure des réseaux, Visual basic pour la programmation des fichiers d'entrée des

logiciels spécifiques de représentation des réseaux, et enfin Ucinet 6. et ses fonctionnalités (Netdraw, pajek...) pour représenter l'exposome dans son ensemble.

III - RESULTATS

1. CARACTERISATION DU RESEAU BIPARTITE

Sur la période 2002-2005, 491 codes distincts (3 digits) ont été utilisés pour coder la pathologie principale et 962 codes distincts (5 digits) pour l'ensemble du cortège des nuisances professionnelles.

La figure 3a illustre la distribution du nombre de codes nuisances distincts associés au nombre de codes pathologies distincts (longueurs des codes pathologies). Ainsi, au

maximum, on a observé une pathologie associée à 346 codes nuisances distincts. La majorité des codes pathologies n'a cependant été rapportée qu'avec un faible nombre de nuisances différent. La figure 3b illustre de façon symétrique la distribution des codes pathologies par code nuisance. Au maximum, un code nuisance est associé à 152 codes pathologies distincts, toutefois la majorité des codes nuisances est associée à un faible nombre de pathologies.

Figure 3a
Distributions des longueurs des codes pathologies

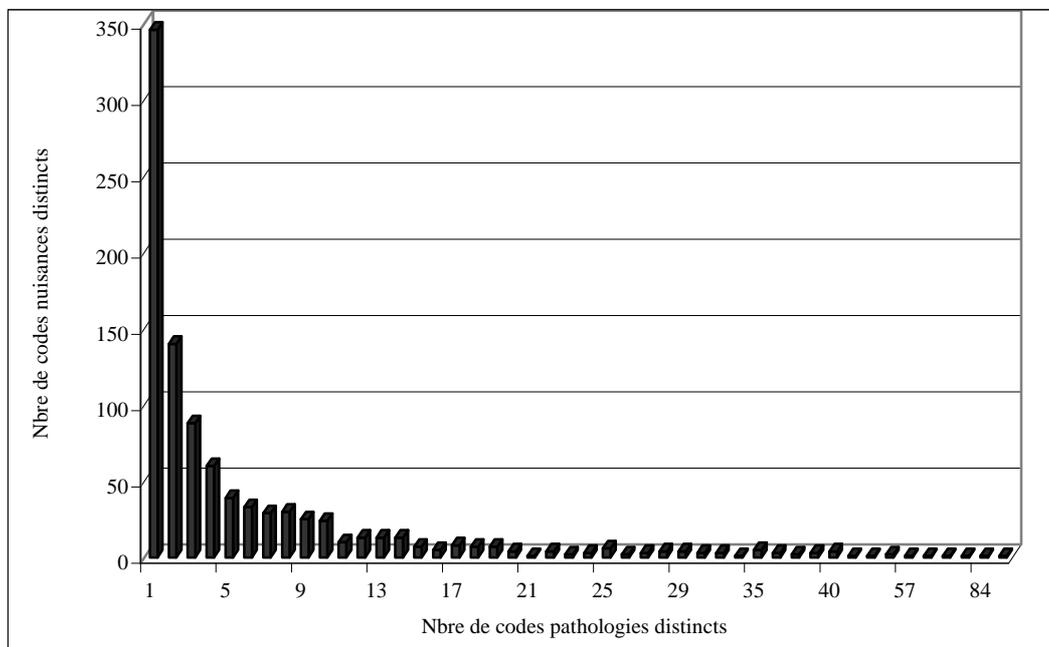
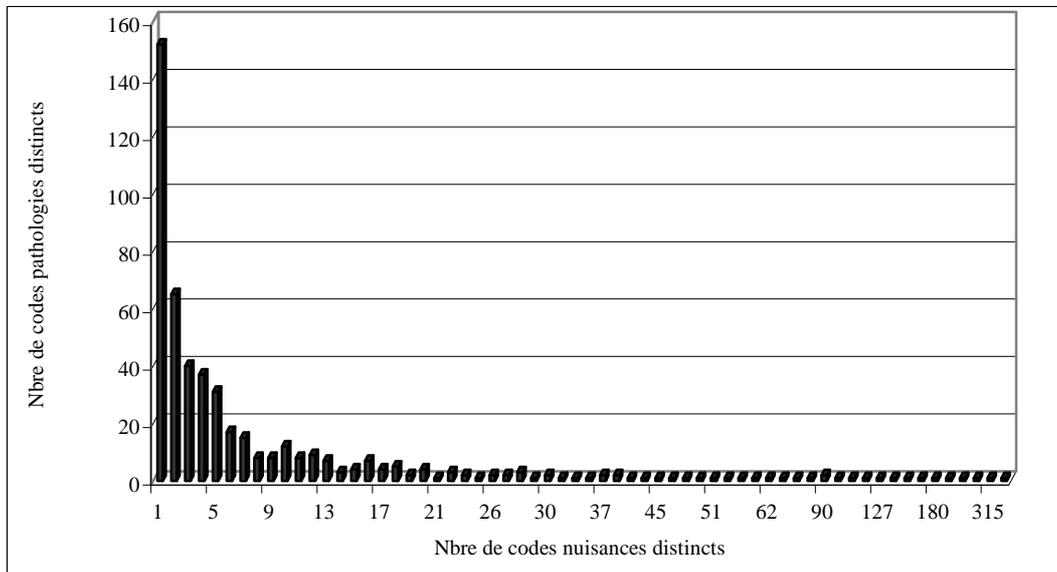


Figure 3b
Distributions des longueurs des codes nuisances



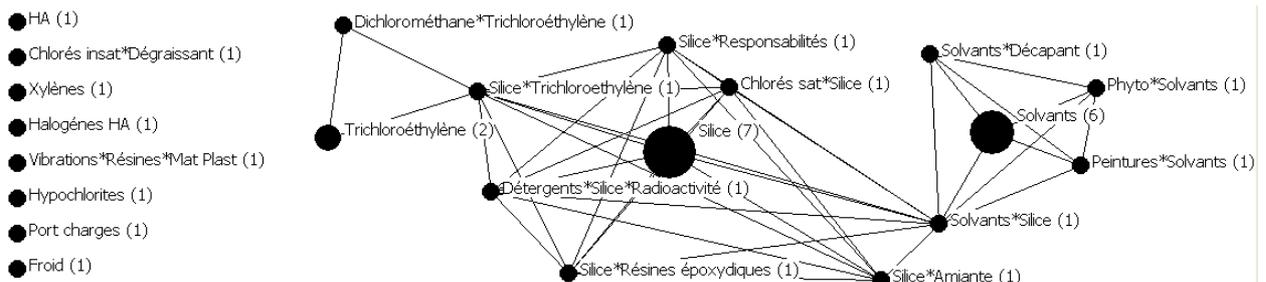
2. ELABORATION DE L'EXPOSOME

L'intérêt de l'application de l'exposome sur les données du RNV3P est illustré au moyen d'un exemple concret : la méthode a été appliquée à un échantillon de données correspondant à l'ensemble des observations du réseau correspondant à une pathologie donnée, la « sclérodémie systémique ». Il s'agit d'une maladie systémique auto-immune caractérisée par l'atteinte des vaisseaux de petit calibre, dont le retentissement est notamment cutané (doigts boudinés par exemple) mais qui peut atteindre d'autres organes comme le poumon. L'association silice et sclérodémie systémique est reconnue historiquement comme une

maladie professionnelle et est indemnisable chez les sujets exposés plus de 10 ans à la silice, mais d'autres nuisances ont été récemment pointées du doigt à la suite d'études épidémiologiques, dont les solvants chlorés [Kettaneh *et al.*, 2007 ; Maitre *et al.*, 2004].

La figure 4 illustre pour $\lambda = 1$ le schéma simplifié de l'exposome de 2002-2005 et trois agrégats de pathologies sont identifiés : les pathologies professionnelles associées aux solvants et diluants organiques, celles liées à la silice et le groupe des solvants chlorés.

Figure 4
Exposome de la sclérodémie, période 2002-2005, seuil = 1 – Logiciel Ucinet 6



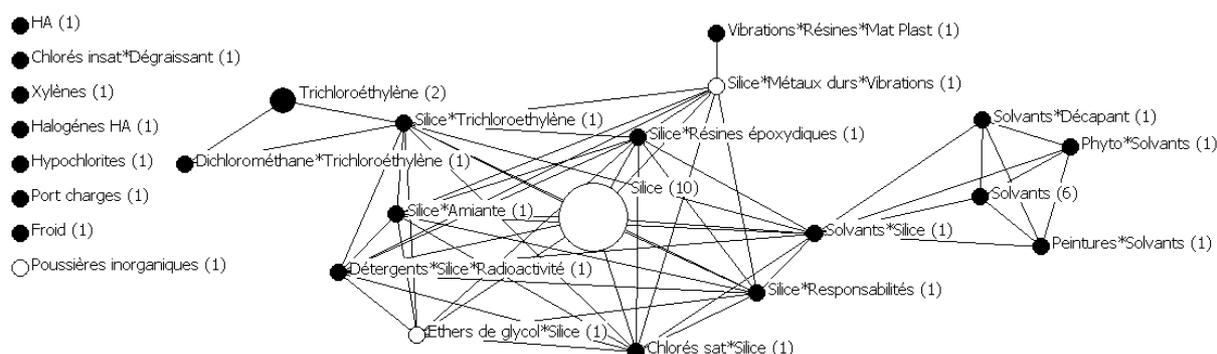
Cet exposome permet de regrouper des pathologies selon leurs expositions professionnelles et trois agrégats de pathologies sont mis en évidence autour des nuisances « silice », « solvants » et plus précisément « solvants chlorés ». La suite serait de suivre par exemple cet exposome dans le temps afin de détecter des associations « pathologie x nuisances » potentiellement émergentes [Bonnetterre *et al.*, 2007]. Après comparaison des observations de 2006 par rapport à l'exposome de la période 2002-2005 (figure 5), des évolutions sont observées au niveau de l'agrégat silice avec

notamment trois nouveaux cas de sclérodémie x silice qui se greffent sur l'agrégat silice ; le nœud sclérodémie x éthers de glycols x silice alimente également cet agrégat.

Un nouveau groupe d'exposition apparaît autour de l'exposition professionnelle « vibrations » grâce au nœud « sclérodémie x silice x métaux durs x vibrations » qui connecte un nœud présent dans l'exposome précédent (« sclérodémie x Vibrations x Résines x Mat Plast »), mais qui était isolé.

Figure 5

Exposome de la sclérodémie, projection des données 2006 sur la période 2002-2005, seuil = 1 – Logiciel Ucinet 6



IV - DISCUSSION – CONCLUSION

Cette méthode vise à faciliter la lecture des interactions entre les différentes observations enregistrées au sein d'un réseau de vigilance, ici le RNV3P. L'interaction entre plusieurs cas pathologiques a ici été définie comme le partage d'au moins une exposition professionnelle, ce qui nous a conduit à baptiser ce réseau *exposome*. La comparaison dans le temps des *exposomes*, permet une lecture visuelle, facilement accessible et communicable, de la dynamique temporelle du réseau. Toutefois, elle peut également s'effectuer de façon statistique nécessitant alors la génération d'algorithmes (travail en cours), en particulier lorsque le nombre de nœuds considérés est trop important rendant la lisibilité de l'*exposome* plus difficile.

Le but de cette méthode est de proposer une lecture nouvelle des données de bases semblables au RNV3P, prenant en compte l'exposition conjointe à des nuisances multiples, ainsi que l'évolution des cas pathologiques au cours du temps. Les résultats obtenus sont retournés aux experts pour évaluation de leur pertinence. Ceux-ci sont interrogés en outre sur les nouvelles associations et celles dont le nombre de cas augmente statistiquement : s'agit-il de fluctuations liées à la dynamique du réseau ou de l'apparition de nouveaux cas pathologiques non connus et susceptibles de présenter un risque non repéré jusqu'alors ? Les experts peuvent alors décider d'une surveillance particulière de ces agrégats, voire de la mise en place d'un système d'alerte.

BIBLIOGRAPHIE

- Adamcsek B., Palla G., Farkas I., Derényi I., Vicsek T. - CFinder : locating cliques and overlapping modules in biological networks. *Bioinformatics*, 2006, 22, 1021-1023.
- Bernardet C., Bonneterre V., Bicout D.J., Larabi L., De Gaudemaris R. - Analyse temporelle des couples nuisance/pathologie principale relevant des maladies professionnelles au sein du Réseau National de Vigilance et de Prévention des Pathologies Professionnelles (RNV3P). *Arch. Mal. Prof. Env.* : accepté pour publication.
- Bonneterre V., Bicout D.J., Larabi L., Bernardet C., Maitre A., Tubert-Bitter P., De Gaudemaris R. - Detection of emerging diseases in occupational health: usefulness and limitations of application of pharmacovigilance methods to the database of the French national occupational disease surveillance and prevention network (RNV3P). *OEM* accepté, sous presse.
- Bonneterre V., De Gaudemaris R., Celse M.H., Chamoux A. - Réseau national de vigilance des pathologies professionnelles : premier bilan et perspectives. *Arch Mal Prof*, 2006, 65(7) : 719-723.
- Celse M.H., Bonneterre V., Parson M., Bernardet C., Perdrix A., De Gaudemaris R. - Extension du réseau national de vigilance des pathologies professionnelles : un outil épidémiologique pour la vigilance des risques professionnels. *Arch. Mal. Prof.*, 2006, 65(7), 725-730.
- Everett M.G., Borgatti S.P. - Analyzing clique overlap. *Connections*, 1998, 21, 49-61.
- Girvan M., Newman M.E.J. - Community structure in social and biological networks. *PNAS*, 2002, 99, 7821-7826.
- Judson M.A. - The etiologic agent of Sarcoidosis, What if there isn't one ? *Chest*. 2003, 124, 6-8.
- Kettaneh A., Al Moufti O., Tiev K.P., Chayet C., Toledano C., Fabre B., Fardet L., Cabane J. - Occupational exposure to solvents and gender-related risk of systemic sclerosis : a metaanalysis of case-control studies. *J Rheumatol.*, 2007, 34, 97-103.
- Maitre A., Hours M., BonneterreV., Arnaud J., Tulin Arsan M., Caprentier P., Bergeret A., De Gaudemaris R. - Systemic sclerosis and occupational risk factors : role of solvents and cleaning products. *J Rheumatol*, 2004, 31 : 2395-2401.
- Moller D.R., Chen E.S. - What causes sarcoidosis ? *Curr. Opin. Pulm. Med.*, 2002, 8(5), 429-34.
- Newman M.E.J., Girvan M. - Finding and evaluating community structure in networks. *Physical. Review. E.*, 2004, 69, 026113-28.
- Palla G., Barabasi A.L., Vicsek T. - Quantifying social group evolution. *Nature*, 2007, 446, 664-667.



Occupational Health Problem Network : the Exposome

Online access : <http://fr.arxiv.org/abs/0907.3410>

Laurie Faisandier¹, Régis De Gaudemaris^{1,2}, Dominique J Bicout³

¹ Laboratoire Environnement et Prédiction de la Santé des Populations-TIMC, Université Joseph Fourier, Grenoble, Domaine de la Merci, 38706 La Tronche, France

² Service de Médecine et Santé au Travail, CHU Grenoble, BP 217, 38043 Grenoble cedex 09 France

³ Biomathématiques et Epidémiologie, Laboratoire Environnement et Prédiction de la Santé des Populations-TIMC, UMR CNRS 5525 Université Joseph Fourier, Ecole Nationale Vétérinaire de Lyon, 69280 Marcy l'Etoile, France

Corresponding authors: LFaisandier@chu-grenoble.fr, d.bicout@vet-lyon.fr

In the past two decades, industrialisation, urbanisation, the booming development of the transport industry and intensive farming have all resulted in considerable changes at both an individual and societal level. The professional environment has in particular undergone technological advances with the discovery of new substances and the modernisation of manufacturing processes. With an increase in the incidence of pathologies whose root causes are poorly understood and sometimes not even known, and a growing demand by the public for the improvement of working conditions, these developments and changes raise general questions about the relationship between the environment (professional) and health.

Initiated in 2001, the Réseau National de Vigilance et de prévention des Pathologies Professionnelles (RNV3P) is a computer-based network which collates information on health problems in the work place, issued by the centres of occupational diseases throughout France. The observations are reported by doctors – expert database findings grow each year by, on average, 5,000 occupational health problems. The objectives of the RNV3P are based on two main principals: the identification of professional situations that expose people to risk and the aetiology of emerging medical events associated with problems of health at work. With a substantial number of observations reported, the RNV3P constitutes an essential resource in the prospective epidemiological surveillance of occupational diseases.

An occupational-health problem (OHP) is seen in the situation where a patient presents with a principal pathology associated with a cascade of up to 5 noxious agents that are related to the place of work, the profession (or job) and the patient's particular activity. A degree of responsibility is attributed to each noxious substance and reflects the intensity of the causal-effect relationship of the substance on the presentation of disease. Thus two OHPs may differ either by principle pathology and/or an occupational exposure defined by noxious agent, profession or sector of activity. Likewise, several patients can present with exactly the same OHP.

In the face of so many individual factors and complex professional situations where exposures of diverse origins and variable intensity combine over time to have an effect on health, it may be helpful to investigate characteristics or traits which gather or separate OHPs. In effect, many diseases are of multifactorial origin due to a prolonged and regular exposure to a range of noxious products. Because of this, several distinct OHPs are likely to have in common similar professional situations. Consequently, it would be convenient to develop approaches capable of analysing all OHPs in a global manner for the study of pathology-exposure relationships. For this reason, we have developed the concept of “exposome”, defined as a

relational network of OHPs having in common an element at least of an occupational exposure: the noxious substances, job or sector of activity [1]. Thus, the RNV3P database is structured by groups of OHPs sharing characteristics in common. In this way, exposomes represent a tripartite occupational exposure – pathology network, founded upon presentation of disease.

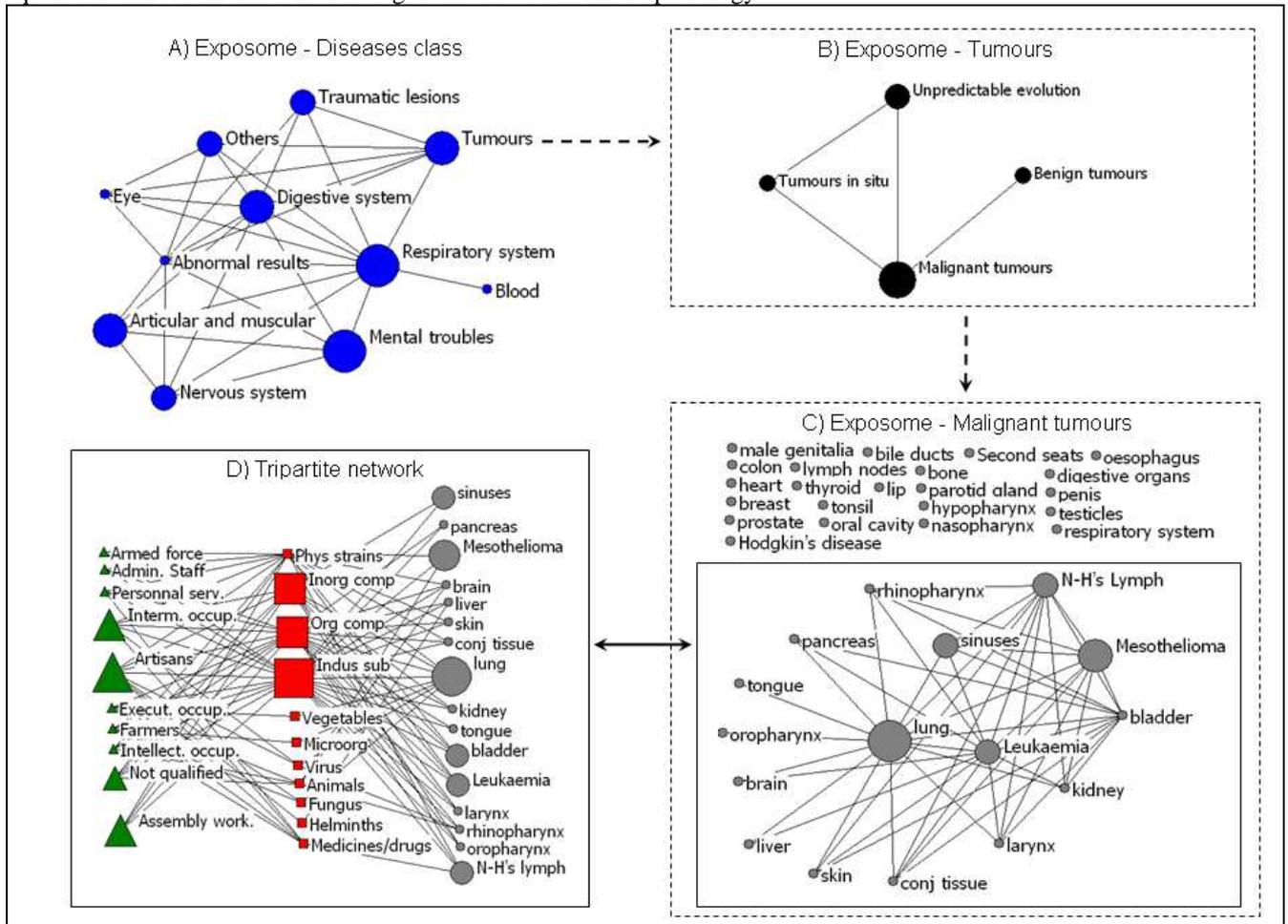
Figure 1 illustrates the structure encased in the network with the exposomes at different levels of organisation, indicating the relational complexity between OHPs. Each link in the exposome represents an OHP and the size of the link is proportional to the number of identical OHPs. Two links are connected when they have at least one element of occupational exposure in common. The number of connections between links indicates both the number of different exposures (multi-exposures) and the diversity of connections from the number of different neighbouring OHPs. In this way, the organisation of OHPs in the exposome allows identification of groups of OHPs that share similar occupational exposures and inter-connected or linked OHPs belonging to several groups. This exposome network is then used to conduct an epidemiological survey by the programmed prospective surveillance of OHP groups (or associations pathology x occupational exposures) identified and of a prospective surveillance by the detection of emerging events as new links and/or new connections appear [2]. By definition, these exposomes should also enable study of synergistic relationships of occupational exposures on the OHPs and to construct hypotheses on the aetiology of pathologies.

The term exposome has already been used in the literature. In effect, just as for health risk analysis, where the toxicological reference values are established from lifetime exposure, C P Wild defines exposome as the collection and succession of individual and environmental exposures encountered by an individual during his/her lifetime [3]. This author proposes reconstructing an individual's given network of exposures in order to better understand the role of each exposure and thus generate research hypotheses as for the aetiology of diseases. In a similar way, Barabasi has constructed a "diseasome" in order to illustrate under the form of a network the environmental and social factors which would have a potential role in the origins of obesity [4]; and Goh et al. have explored a network of human diseases that implicate similar genetic mutations [5]. What all these analyses have in common is the research of similar characteristics for the understanding of mechanisms and the identification of key factors in the development and manifestation of diseases.

The exposome approach that we are in the process of developing has also an immediate application in health surveillance and the detection of emerging events for the RNV3P. This approach could be applied to a global analysis of health problems that would include in the RNV3P database other types of findings, variables and descriptors such as those used in the diseasome and exposome of the Wild type. A step or enterprise such as this greatly exceeds the simple framework, already complex in itself, of the RNV3P and necessitates going into and/or combining other databases.

Figure 1: Structure of the disease exposome.

Each tie represents a pathology for which the size is proportional to the number of related OHPs. The inter-connected links have at least one occupational exposure in common, such as the noxious substance. A) Exposome of disease categories of tumours, reported in the RNV3P database. B) Exposome of tumours. This category of diseases represents 4 sub-groups: malignant tumours, tumours of an unpredictable evolution, in-situ tumours and benign tumours. C) 38 diseases are distinguished amongst the malignant tumours, of which 22 isolated links do not have any noxious agents in common and the remaining 16 inter-connected links form a network as shown in fig. 1D. D) The tripartite network: disease–noxious agent–occupation is the origin of the exposome formed by 16 malignant tumours. The circles, squares and triangles represent malignant tumours, noxious agents and occupation, respectively, relative to each observation. The links between squares and triangles indicate noxious agents associated with the occupation for the OHP considered, and the links between squares and circles indicate noxious agents associated with the pathology.



References :

1. Faisandier L, Bonnetterre V, De Gaudemaris R et al. Elaboration d'une méthode statistique pour la détection d'événements émergents: application au Réseau National de Vigilance et de Prévention des Pathologies Professionnelles (RNV3P). *Epidémiol. et santé anim.* 2007;51:111-8.
2. Bonnetterre V, Bicout DJ, Larabi L et al. Detection of emerging diseases in occupational health : usefulness and limitations of application of pharmacovigilance methods to the database of the French national occupational disease surveillance and prevention network (RNV3P). *Occup Environ Med.* 2008;65:32-7.
3. Wild CP Complementing the Genome with an "exposome": the outstanding challenge of environmental exposure measurement in molecular epidemiology. *Cancer Epidemiol Biomarkers Prev* 2005;14:184-8.
4. Barabasi A-L Network Medecine – From Obesity to the « Diseasesome ». *N. Engl. J. Med.* 2007;357:404-7.
5. Goh K-I, Cusick ME, Valle D et al. The human disease network. *Proc. Natl. Acad. Sci. U.S.A.* 2007;104:8685-90.

Réseau de Vigilance des Pathologies Professionnelles -2 Prise en compte de la multiplicité des expositions

L Faisandier¹, V Bonnetterre^{1,2}, R De Gaudemaris^{1,2}, DJ Bicout¹

- 1- Equipe EPSP, Environnement et Prédiction de la Santé des Populations, Laboratoire TIMC, Faculté de Médecine, Domaine de la Merci, 38706 La Tronche Cedex
2- Equipe médecine et santé au travail, hôpital Michallon, CHU Grenoble, BP 217, 38043 Grenoble cedex 9

INTRODUCTION

CONTEXTE Depuis 2001, le Réseau National de Vigilance et de Prévention des Pathologies Professionnelles (RNV3P) enregistre en moyenne 5000 problèmes de santé au travail (PST) en France. Ces observations sont rapportées de manière standardisée dans une base de données. Chacune est décrite par une pathologie principale, 1 à 5 nuisances/contraintes présentes dans l'environnement professionnel, un métier ainsi qu'un secteur d'activité.

OBJECTIFS Développer un outil capable de (i) surveiller en continu les associations pathologie-expositions et de (ii) détecter des événements potentiellement émergents (iii) en prenant en compte la multiplicité des expositions professionnelles.

MATERIEL



Un PST décrit la situation d'un patient :

- ✓ Une pathologie principale
- ✓ Un cortège de 1 à 5 nuisances/contraintes présentes dans l'environnement professionnel,
- ✓ Une profession et un secteur d'activité.

METHODE

A l'image des réseaux sociaux qui décrivent des interactions entre individus, le RNV3P peut être assimilé à un réseau de PST, connectés les uns aux autres à condition qu'ils partagent une ou plusieurs nuisances/contraintes professionnelles. n PST identiques sont inclus dans un nœud. La taille des nœuds est proportionnelle au nombre de PST.



L'ensemble des PST forme ainsi un réseau nommé « *exposome* ».

RESULTATS – ILLUSTRATIONS

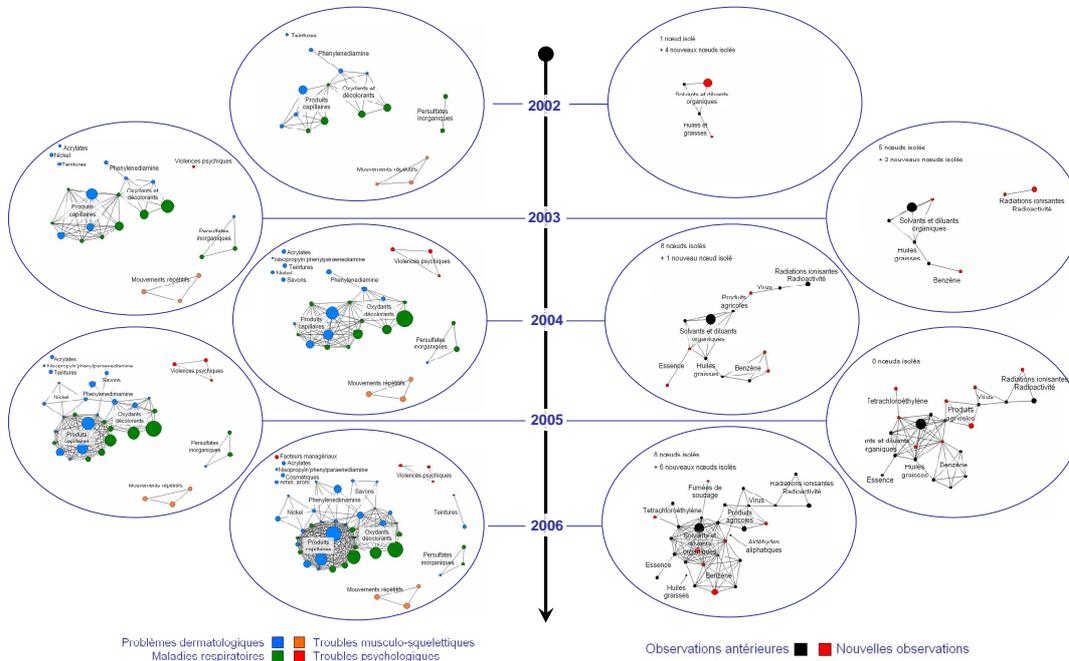
L'exposome peut être focalisé sur un métier, un secteur d'activité, une pathologie ou une nuisance/contrainte spécifique. La surveillance prospective des associations entre une pathologie et ses expositions professionnelles consiste à observer des événements nouveaux comme :

- ✓ Des nouveaux *nœuds* (une nouvelle association entre une pathologie et ses expositions) connectés ou isolés,
- ✓ De nouvelles connections (au moins 2 *nœuds* partagent au moins 1 nuisance/contrainte),
- ✓ De nouveaux agrégats d'exposition (des nœuds qui partagent au moins 1 exposition).

Un agrégat d'exposition est un groupe de PST qui partagent au moins une nuisance/contrainte présente sur le lieu de travail.

1) Exposome basé sur le métier des **coiffeurs** (données cumulées depuis 2001). Un *nœud* contient au minimum 4 PST identiques.

2) Exposome basé sur la pathologie des **lymphomes non hodgkiniens** (données cumulées depuis 2001). Un *nœud* contient au minimum 1 PST.



DISCUSSION

A l'aide de l'exposome, le suivi des groupes d'exposition dans le temps permet à la fois la surveillance chronologique des associations mais aussi la détection des événements potentiellement émergents.



A terme, un système d'alerte sera enclenché si les événements sont validés comme émergents (non connus auparavant et en quantité significative) ou ré-émergents par les experts du réseau.

Le concept des réseaux relationnels comme une approche pour caractériser et surveiller des Problèmes de Santé au Travail

Laurie Faisandier¹, Régis De Gaudemaris^{1,2}, Dominique J Bicout¹

1- Equipe EPSP, Environnement et Prédiction de la Santé des Populations, Laboratoire TIMC, Faculté de Médecine, Domaine de la Merci, 38706 La Tronche Cedex

2- Equipe médecine et santé au travail, hôpital Michallon, CHU Grenoble, BP 217, 38043 Grenoble cedex 9

Introduction

Contexte - Le Réseau National de Vigilance et de Prévention des Pathologies Professionnelles (RNV3P) enregistre des Problèmes de Santé au Travail définis comme des associations entre une pathologie et un ensemble d'expositions professionnelles composites potentiellement étiologiques (1 à 5 nuisances, un métier et un secteur d'activité).

Problématique - La veille sanitaire en santé travail porte classiquement sur le suivi épidémiologique d'indicateurs en population générale et sur des études de cohortes ou cas-témoins spécifiques. Notre approche doit faire face à plusieurs défis méthodologiques : prendre en compte toutes les composantes de l'exposition professionnelle, couvrir des pathologies dont les effectifs sont insuffisants pour être abordés par l'épidémiologie conventionnelle et pointer des associations pathologie-expositions composites potentiellement émergentes.

Objectifs - Développer une approche statistique pour mieux comprendre et caractériser la multi-exposition, établir leur suivi temporel et générer de nouvelles hypothèses étiologiques.

Matériel



Un Problème de Santé au Travail (PST) est l'association entre une pathologie principale et une exposition professionnelle composite dont une ou plusieurs composantes sont potentiellement étiologiques (un cortège de 1 à 5 nuisances, un métier et/ou un secteur d'activité).

Méthodologie

Approche conceptuelle basée sur les réseaux relationnels - « exposome » - qui consiste à mettre en relation des Problèmes de Santé au Travail ayant en commun une ou plusieurs expositions professionnelles communes (nuisances, métier et/ou secteur d'activité).

Illustration

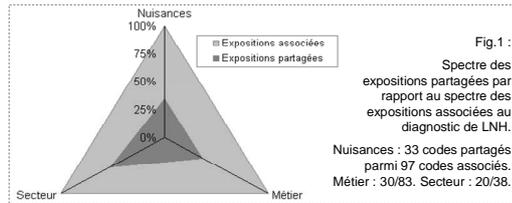
Les experts du RNV3P ont enregistré 136 cas de Lymphomes Non Hodgkiniens (LNH) au cours de la période 2001-2007.

Un nœud représente l'association entre une pathologie et un ensemble d'exposition professionnelle composite dont la taille correspond au nombre de patients décrits.

Deux nœuds sont connectés s'ils partagent au moins un élément de l'exposition professionnelle composite.

Un groupe d'exposition est un ensemble de nœuds qui partagent exactement une même nuisance.

Un nœud-passerelle est un nœud qui appartient à au moins 2 groupes d'exposition (qui partagent au moins 2 éléments de l'exposition avec les autres nœuds des groupes).



2 patients porteurs de Lymphomes Non Hodgkiniens ont été exposé à la même nuisance (produits phytosanitaires) ont exercé le même type de métier (arboriculteur) dans le même secteur d'activité (agriculture, chasse, services annexes).

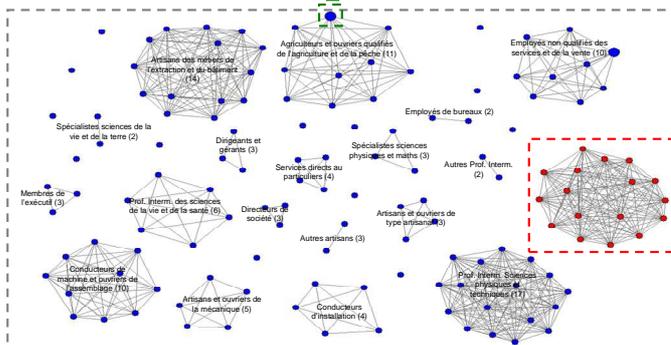


Fig. 2 : Exposome des 134 nœuds rassemblant 136 cas de LNH observés au cours de la période 2001 à 2007, reliés par les types de métiers exercés.

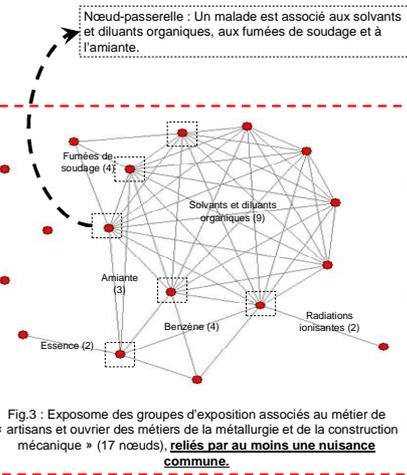


Fig. 3 : Exposome des groupes d'exposition associés au métier de « artisans et ouvriers des métiers de la métallurgie et de la construction mécanique » (17 nœuds) reliés par au moins une nuisance commune.

Conclusion

Intérêt médical - L'approche par exposome permet d'identifier et de suivre dans le temps des groupes d'exposition composés de malades exposés aux mêmes nuisances et ayant exercé la même activité professionnelle. Avec pour premières applications le champ santé travail, cette approche constitue un outil à valider pour la surveillance de la multi-exposition et la détection des risques potentiellement émergents.

Enjeux - Le principal enjeu de santé publique est de soulever un questionnement scientifique à partir des événements identifiés comme potentiellement émergents. Des études épidémiologiques spécifiques en population générale pourraient être amorcées à la suite des hypothèses étiologiques générées par l'exposome. Après validation par les experts, ces observations pourraient faire l'objet d'un renforcement de la prévention auprès des entreprises ou d'un bulletin d'alerte au niveau national.

Annexes

Annexes A. Recensement des dispositifs contribuant à la surveillance temporelle et à la détection des événements sanitaires émergents

Annexe A-1. Les réseaux de recherche fondés sur les pratiques

Réseau de cliniciens	Date création	Missions	Publications
American Psychiatric Institute for Research and Education's Practice Research Network (USA)	1993	Compléter les méthodes traditionnelles de recherche en produisant de l'information à partir d'un large éventail de patients, de traitements, et de configurations de traitement	Zarin 1997
Northern Practice Research Network (Angleterre)	-	1. Standardiser le recueil de données en routine à travers tous les sites participant 2. Utiliser la base de données résultante pour comparer et différencier les pratiques dans leur contexte, le profil des cas et la pratique standard	Audin 2001 ; Barkham 2003
Pennsylvania Practice Research Network (USA)	2001	Construire des programmes qui maximisent à la fois la validité interne par les choix de méthodologies scientifiques rigoureuses et la validité externe en conduisant la recherche en psychothérapie dans les configurations actuelles où elle est appliquée	Borkovec 2001 ; Castonguay 2006
Réseau de Recherche Fondé sur les Pratiques Psychothérapeutiques (France)	2008	1. Assurer une meilleure connaissance des pratiques psychothérapeutiques dans les différentes conditions et contextes de leur exercice 2. Développer l'évaluation des psychothérapies en conditions naturelles 3. Approfondir les connaissances sur les configurations et les mécanismes de changement au cours du processus psychothérapeutique	Audin 2001 ; Barkham 2003
Réseau de pédopsychiatrie communautaire de San Diego (USA)	2006	Etudier les psychothérapies pour des enfants âgés de 4 à 13 ans, présentant des problèmes de troubles des conduites dans 6 centres de soins publics	Garland 2006
ENIVD - European network for the diagnosis of "imported" viral diseases	1998	Encourager les échanges sur des diagnostics simples, des méthodes, et des informations pour améliorer les diagnostics	Niedrig 2007

Annexe A-2. Les réseaux de vigilance des produits de santé

Réseau de vigilance	Missions	Déclarants
Biovigilance	Surveillance des organes et tissus prélevés pour la qualité et le respect des pratiques de soins jusqu'au suivi du greffé.	Médecins, pharmaciens, chirurgiens dentistes, biologistes, sages-femmes, infirmières
Hémovigilance	Surveillance des incidents et effets indésirables survenant chez les receveurs de produits sanguins labiles	Ensemble des professionnels de santé
Pharmacovigilance	Surveillance du risque d'effet indésirable résultant de l'utilisation des médicaments et produits à usage humain	Ensemble des professionnels de santé + Entreprises ou organismes exploitant un médicament
Pharmacodépendance	Surveillance des cas d'abus de dépendance liées à la prise de substances ou plantes ayant un effet psychoactif et tout médicament ou produits en contenant	Ensemble des professionnels de santé + Entreprises ou organismes exploitant un médicament
Matérovigilance	Surveillance des incidents ou des risques d'incidents résultant de l'utilisation des dispositifs médicaux	Utilisateurs ou tiers + Fabricant
Réactovigilance	Surveillance des incidents et des risques d'incidents liés à l'utilisation d'un appareil ou d'un réactif de laboratoire, d'un tube de prélèvement, voire d'un appareil d'auto diagnostic	Ensemble des professionnels de santé + Utilisateur ou non utilisateur + Fabricant et distributeurs
Cosmétovigilance	Surveillance des effets indésirables résultant de l'utilisation des produits cosmétiques après leur mise sur le marché	Ensemble des professionnels de santé

Annexes B. Recensement des programmes pour la surveillance et la détection des alertes sanitaires

Annexe B-1. Les dispositifs contribuant à l'alerte sanitaire hors France.

Dispositifs	Date de création	Missions	Outils utilisés	Type de données	Publications
The Health and Occupation Reporting THOR-Extra (Angleterre)	2002	Détecter les problèmes de santé au travail inhabituels, avec les déclarations spontanées des médecins du travail et autres spécialistes	Avis d'experts	Professionnel	Walsh 2005 ; Chen 2005 ; Turner 2005
Netherlands Center for Occupational Diseases (Hollande)	2002	Améliorer la qualité de la prévention et le diagnostic précoce des pathologies susceptibles d'être causées par l'environnement professionnel	Avis d'experts	Professionnel	Pal 2009
GOARN - Global Outbreak Alert and Response Network (international)	2000	Détecter, investiguer et prévenir les épidémies d'agents pathogènes infectieux	GPHIN - Global Public Health Intelligence Network développé en 1997 par le WHO, qui explore continuellement et systématiquement les sites internet, les journaux locaux	Infectieux	Heymann 2001
FluNet (international)	1995	Site internet d'échange d'information dans le cadre de la surveillance mondiale de la grippe	Cartographies	Infectieux	Flahault 2006 ; Finkelman 2007
Observatoire National des Maladies Nouvelles et Emergentes ONMNE (Tunisie)	2007	1. Surveillance épidémiologique des maladies émergentes à potentiel prolifératif 2. Identification des mesures susceptibles de limiter les risques liés à ces maladies	En cours de développement	Tout environnement	Non renseigné
SENSOR - Sentinel Event Notification System for Occupational Risks (USA)	1987	Utiliser des sources sentinelles pour reconnaître et signaler certains troubles professionnels dans un centre de surveillance de pathologies	Analyse de tendances à partir de données recueillies par des médecins sentinelles	Surveillance des maladies et des accidents du travail	Baker 1988 ; Baker 1989 ; Maizlish 1995 ; Maizlish 1995 ; Calvert 2004

<p>CDC : Emerging Infections Programs (USA)</p>	<p>2001</p>	<p>1. Effectuer une surveillance active des maladies infectieuses en population générale 2. Conduire des études épidémiologiques et de la recherche en laboratoire 3. Mettre en œuvre et évaluer les moyens de prévention</p>	<p>Incidence, mortalité, morbidité</p>	<p>Infectieux</p>	<p>Pinner 2003</p>
<p>CDC : PulsNet, the molecular subtyping network for foodborne bacterial disease surveillance (USA)</p>	<p>1996</p>	<p>Faciliter l'identification de bactéries pathogènes à des fins alimentaires</p>	<p>Epidémiologie moléculaire</p>	<p>Bactérien</p>	<p>Swaminathan 2001</p>
<p>CDC : Bioterrorism surveillance (USA)</p>	<p>1996</p>	<p>1. Permettre de détecter le plus précocement possible des épidémies 2. Fournir une réponse rapide quelques jours ou quelques heures avant que des clusters soient cliniquement identifiés ou avant que les diagnostics soient faits et rapportés aux autorités de santé publique</p>	<p>Epidémiologie moléculaire</p>	<p>Bioterrorisme</p>	<p>Buehler 2003</p>

Annexe B-2. Les programmes de surveillance et de détection des alertes sanitaire en France.

Programme	Date de création	Missions	Outils utilisés	Publications
RNV3P - Réseau National de Vigilance et de Prévention des Pathologies Professionnelles	2001	1. Surveiller des Problèmes de Santé au Travail 2. Détecter des PST potentiellement émergents	Fouille de données ; méthodes de pharmacovigilance	Bonneterre 2008 ; Bonneterre 2010
ONAP - Observatoire National de l'Asthme Professionnel	1996	Estimer l'incidence et identifier les caractéristiques de l'asthme professionnel	Estimation de l'incidence ; Tendances temporelles	Ameille 2003 ; Popin 2008
Phyt'attitude	1991	Recenser les troubles de la santé signalés par tous les professionnels exposés aux produits phytosanitaires	Statistiques descriptives	Dupupet 2007
PNSM - Programme National de Surveillance du Mésothéliome	1998	Fournir aux autorités sanitaires des informations scientifiques fiables à l'heure où les conséquences des expositions passées à l'amiante	Estimation de l'incidence des mésothéliomes en France à partir de l'enregistrement des cas dans les départements	Rapport activité
SACS - Système d'Alerte Canicule et Santé	2004	1. Renseigner la surmortalité de toutes causes attribuables à la canicule 2. Identifier les facteurs de risques de surmortalité chez les personnes âgées 3. Connaître la part attribuable à l'ozone et la température	En cas de vague de chaleur : décès enregistrés par les bureaux d'état-civil, interventions du Samu, fréquentation des services d'urgences hospitaliers	Rapport activité
Programme de surveillance des intoxications au monoxyde de carbone (CO)	2004	Surveillance des effets et des expositions dans les situations d'intoxications avérées au monoxyde de carbone et amélioration de la prévention collective	Statistiques descriptives	Rapport activité par région
PSAS - Programme de Surveillance Air et Santé	1997	Relier les variations temporelles à court terme (d'un jour à l'autre) d'un indicateur de l'état de santé d'une population à celles d'un indicateur d'exposition de cette population à la pollution atmosphérique afin d'établir une relation exposition/risque	Mesure des indicateurs atmosphériques (polluants) ; Mesure de l'état de santé d'une population (mortalité, admissions hospitalières)	Rapport activité
SISE-EAUX Système d'Information en Santé Environnement sur la qualité des eaux d'alimentation	1993	Relever, surveiller et détecter des dépassements de seuils de la qualité de l'eau	Indicateurs physico-chimique des propriétés de l'eau ; Indicateurs statistiques (Nombre d'Unités de distribution (UDI) non conformes / Population concernée)	Rapport d'activité par département

<p>OPERA - Observatoire PERmanent de la Radioactivité de l'environnement</p>	<p>1959</p>	<p>1. Observer la radioactivité d'origine naturelle et artificielle présente dans l'environnement 2. Modéliser les flux de transferts globaux des radionucléides entre les différents compartiments de l'environnement</p>	<p>Mesure des radionucléides présents à l'état de traces dans des échantillons de l'environnement ; Séries chronologiques sur la variabilité des niveaux observés</p>	<p>Mesures communiquées en ligne au public</p>
<p>RNSA - Réseau National de Surveillance Aérobiologique</p>	<p>1996</p>	<p>Etudier le contenu de l'air en particules biologiques pouvant avoir une incidence sur le risque allergique pour la population</p>	<p>Prélèvements atmosphériques et suivi temporel des valeurs d'exposition</p>	<p>Thibaudon 2002 ; Thibaudon 2004</p>

Annexes C. Bordereaux de saisie.

Annexe C-1. Bordereau PATIENT

Coller ici l'étiquette administrative du patient	Nom du patient : Nom de jeune fille : Prénom : Sexe : I_I H I_I F Date de naissance : ____ / ____ / ____ Lieu de naissance : Identifiant dans l'hôpital : N° INSEE :
--	---

Adresse :

Code Postal : Ville :

Téléphone : -----

Entreprise actuelle :

Nom :

Adresse :

Code Postal : Ville :

Téléphone : **Télécopie :**

Médecin du travail :

Nom :

Adresse :

Code Postal : Ville :

Téléphone : **Télécopie :**

Service : Inter-entreprise I_I Autonome I_I

Médecin traitant ou spécialiste :

Nom :

Adresse :

Code Postal : Ville :

Téléphone : **Télécopie :**

Annexe C-2 : Bordereau CONSULTATION

Coller ici l'étiquette administrative du patient	N° Dossier : _____ Date : ____/____/____	
	Nom du médecin-consultant :	
Prise en charge : CRAM <input type="checkbox"/> Employeur <input type="checkbox"/> SS maladie <input type="checkbox"/> SS AT/MP <input type="checkbox"/> FASS <input type="checkbox"/> Autre <input type="checkbox"/> CMR <input type="checkbox"/>		
Origine de la demande :		
Médecin du travail <input type="checkbox"/> Médecin généraliste <input type="checkbox"/> Spécialiste de ville <input type="checkbox"/> Médecin hospitalier <input type="checkbox"/> Médecin scolaire <input type="checkbox"/> Médecin conseil <input type="checkbox"/> Patient <input type="checkbox"/> Autre <input type="checkbox"/>		
Statut :		
Actif CDI <input type="checkbox"/> Actif fonction publique <input type="checkbox"/> Actif CDD ou assimilé <input type="checkbox"/> Apprenti <input type="checkbox"/> Scolaire <input type="checkbox"/> Chômeur <input type="checkbox"/> Retraité <input type="checkbox"/> Artisan /Indépendant <input type="checkbox"/> Autre <input type="checkbox"/>		
Poste actuel : _____		Code BIT : /_/_/_/
Motif de Consultation :		
Diagnostic de maladie professionnelle <input type="checkbox"/> Conseil pour aptitude <input type="checkbox"/> Orientation professionnelle <input type="checkbox"/> Reclassement <input type="checkbox"/> Suivi pathologie professionnelle <input type="checkbox"/> Surveillance post professionnelle <input type="checkbox"/> Suivi de cohorte <input type="checkbox"/> Surveillance post-exposition <input type="checkbox"/> Expertise ou avis spécialisé SS <input type="checkbox"/> Interrogatoire professionnel <input type="checkbox"/> Autre <input type="checkbox"/> Duplicata de consultation <input type="checkbox"/>		
Type de Consultation (appareil concerné) :		
Addictologie <input type="checkbox"/> Allergologie <input type="checkbox"/> Cardiologie <input type="checkbox"/> Dermatologie <input type="checkbox"/> Endocrinologie <input type="checkbox"/> Hématologie <input type="checkbox"/> Hépatogastroentérologie <input type="checkbox"/> Méd. interne <input type="checkbox"/> Neurologie <input type="checkbox"/> Ophtalmo <input type="checkbox"/> ORL <input type="checkbox"/> Pneumologie <input type="checkbox"/> Psychiatrie <input type="checkbox"/> Rhumatologie <input type="checkbox"/> Traumatologie <input type="checkbox"/> Uronéphrologie <input type="checkbox"/> Reproduction <input type="checkbox"/> Autre <input type="checkbox"/>		
Examens complémentaires		<i>A adapter par chaque centre</i>
Liaisons :		
Médecin du travail <input type="checkbox"/> Médecin traitant <input type="checkbox"/> Inspection du travail <input type="checkbox"/> Assistante sociale <input type="checkbox"/> Médecin spécialiste <input type="checkbox"/> CRAM <input type="checkbox"/> COTOREP <input type="checkbox"/> FIVA <input type="checkbox"/> Autres <input type="checkbox"/>		
Recherche de composition de produit <input type="checkbox"/>		

Problème auquel se rattache la consultation :

Nouveau Ancien / n°

Annexe C-3 : Bordereau PROBLEME

Nature du Problème :

Pathologie professionnelle	I__I	Conseil pour aptitude	I__I
Groupe professionnel à risque	I__I	Interrogatoire professionnel	I__I
Pathologie environnementale	I__I		
Pathologie ni professionnelle, ni environnementale	I__I	Absence de pathologie	I__I
En attente	I__I		

Diagnostic	CIM10	Libellé
Principal		
Comorbidité1		
Comorbidité2		

Nuisance	Code	Libellé	Imputabilité *
1 Principale			
2			
3			
4			
5			

*0 = pas de relation, 1 = faible ou douteux, 2 = possible ou directe mais non essentielle, 3 = directe et essentielle, A = aptitude

Entreprise et Poste responsables de la nuisance principale :		
Entreprise actuelle I__I		Entreprise antérieure I__I
	Code	Libellé
EntrepriseNAF*		
Poste BIT*		

* à ne renseigner que si l'entreprise antérieure est responsable du problème de santé au travail

Conclusions

A.T. : Oui I__I Non I__I

Maladie Professionnelle :

⇒ **MP Indemnisable :** Alinéa 2 } **Tableau* : N°** I__I__I__I__I__I § I__I du régime :
 Alinéa 3 (CRRMP) I__I } RG RA
 Alinéa 4 (CRRMP) I__I }
 Maladie contractée en service I__I

CMI : Fait I__I Conseillé I__I Fait en dehors du CCPP I__I
Année du CMI : I__I__I__I

⇒ **Affection professionnelle non indemnisable :**

Affection hors liste et IPP<25% I__I
 MCP 461.6 I__I
 Affection professionnelle chez un non-salarié (artisan ...) I__I

⇒ **Données insuffisantes pour déclaration :** Cliniques I__I Scientifiques I__I

* à renseigner pour MPI alinéa 2 et alinéa 3

Aptitude : Apte I__I Apte avec réserve I__I Inapte temporaire I__I Inapte définitivement I__I Invalidité I__I
Mes. Prév. : Aménagement de poste I__I Changement de poste I__I EPI I__I Surveillance médicale I__I

Observation intéressante pour échange entre cliniciens * : I__I

Pour partage sur réseau national

* mémo clinique obligatoire

Exportation :

Régionale : Oui I__I Non I__I

Nationale : Oui I__I Non I__I

MEMO CLINIQUE : Résumé en clair du problème quand il est conclu. Le résumé doit obligatoirement mentionner le diagnostic, le poste de travail, la (ou les nuisances) ou la décision. Exemples : "colique de plomb chez un radiateuriste exposé aux poussières de plomb" ou "inaptitude à un poste de grutier chez salarié ayant un syndrome labyrinthique".

Validation senior Oui I__I date :

Annexes D. Résultats détaillés du test 2 - Réussite émergentielle.

Les résultats présentés dans les tableaux suivants sont les fréquences observées des séquences identifiées comme potentiellement émergentes au cours de la période 2001-2009, ainsi que les fréquences théoriques et leurs intervalle de confiance à 95%, obtenues à partir du test 2 de réussite émergentielle, pour les Fonctions Expertes itérative et par blocs.

Annexe D-1. Echantillon des Sclérodermies Systémiques.

Fonction itérative		2001	2002	2003	2004	2005	2006	2007	2008	2009
36R00 - Produits à usage agricole	fi	0	0	0	0	0	1	1	4	4
	<gi>						0,999	0,999	3,999	3,999
	gi-						0,999	0,999	3,998	3,998
	gi+						1,000	1,000	4,000	4,000
31110 - Chantier (Poussières)	fi	0	0	1	2	2	2	2	2	3
	<gi>			0,503	0,832	0,826	0,752	0,760	0,752	0,975
	gi-			0,493	0,918	0,813	0,739	0,747	0,739	0,960
	gi+			0,513	0,846	0,840	0,764	0,773	0,765	0,991
48200 - Froid	fi	0	0	0	1	1	2	2	3	3
	<gi>				0,999	0,999	1,999	1,999	2,999	2,999
	gi-				0,999	0,999	1,999	1,999	2,999	2,999
	gi+				1,000	1,000	2,000	2,000	3,000	3,000
34100 - Fumées de soudage	fi	0	0	0	0	0	0	1	1	2
	<gi>							0,999	0,999	0,740
	gi-							0,999	0,999	0,725
	gi+							1,000	1,000	0,754
35110-36200 - Solvants et diluants organiques - Peintures Vernis Laques Mastics	fi	0	0	0	0	1	1	1	2	2
	<gi>					0,999	0,999	0,999	1,999	1,999
	gi-					0,999	0,999	0,999	1,999	1,999
	gi+					1,000	1,000	1,000	2,000	2,000

Fonction par blocs		2001	2002	2003	2004	2005	2006	2007	2008	2009
36R00 - Produits à usage agricole	fi	0	0	0	0	0	1	1	4	4
	<gi>						0,999	0,672	3,396	3,400
	gi-						0,999	0,662	3,383	3,386
	gi+						1,000	0,681	3,410	3,413
31110 - Chantier (Poussières)	fi	0	0	1	2	2	2	2	2	3
	<gi>			0,490	0,669	0,667	0,649	0,644	0,642	0,788
	gi-			0,480	0,656	0,655	0,637	0,632	0,630	0,774
	gi+			0,500	0,681	0,680	0,661	0,656	0,654	0,802
48200 - Froid	fi	0	0	0	1	1	2	2	3	3
	<gi>				0,999	0,999	1,999	1,999	2,999	2,999
	gi-				0,999	0,999	0,999	1,999	2,999	2,999
	gi+				1,000	1,000	2,000	2,000	3,000	3,000
34100 - Fumées de soudage	fi	0	0	0	0	0	0	1	1	2
	<gi>							0,999	0,999	0,534
	gi-							0,999	0,999	0,521
	gi+							1,000	1,000	0,547
35110-36200 - Solvants et diluants organiques - Peintures Vernis Laques Mastics	fi	0	0	0	0	1	1	1	2	2
	<gi>					0,000	0,000	0,000	0,642	0,000
	gi-					0,000	0,000	0,000	0,630	0,000
	gi+					0,000	0,000	0,000	0,654	0,000

Annexe D-2. Echantillon des Tumeurs Malignes de la Vessie.

Fonction itérative		2001	2002	2003	2004	2005	2006	2007	2008	2009
36200 - Peintures, vernis, laques, mastics	fi	2	3	3	4	4	5	7	10	14
	<gi>	1,999	2,002	1,753	1,993	1,934	2,460	2,103	2,525	3,209
	gi-	1,999	1,988	1,740	1,978	1,918	2,441	2,086	2,504	3,184
	gi+	2,000	2,016	1,766	2,009	1,942	2,479	2,121	2,545	3,233
33210 - Diesel (gaz)	fi	0	0	0	0	0	10	14	17	19
	<gi>						7,522	7,755	8,100	9,174
	gi-						7,492	7,729	8,071	9,144
	gi+						7,551	7,781	8,128	9,204
36R00 - Produits à usage agricole	fi	0	0	0	0	0	0	2	5	8
	<gi>							1,194	1,675	1,989
	gi-							1,187	1,660	1,971
	gi+							1,202	1,689	2,007
36500-36700 - Matières plastiques caoutchoucs polymères - Matières colorantes	fi	0	1	1	1	1	1	2	2	2
	<gi>		0,999	0,506	0,489	0,494	0,330	0,662	0,568	0,575
	gi-		0,999	0,496	0,480	0,485	0,320	0,649	0,556	0,562
	gi+		1,000	0,515	0,499	0,504	0,339	0,675	0,581	0,587
21320-31146 - HAP - Métaux durs	fi	0	0	0	0	0	0	2	2	2
	<gi>							1,166	1,162	1,130
	gi-							1,159	1,155	1,123
	gi+							1,174	1,169	1,136
22224 - Trichloroethylene	fi	0	0	0	0	0	4	6	6	6
	<gi>						2,955	3,784	2,992	3,336
	gi-						2,939	3,769	2,975	3,316
	gi+						2,972	3,799	3,009	3,356
32300 - Fibres végétales textile T66A	fi	0	0	0	0	0	0	0	2	2
	<gi>								0,651	0,498
	gi-								0,638	0,486
	gi+								0,664	0,510
21320-36C20-334C0 - HAP - Huiles et graisses lubrifiantes - Produits noirs (goudrons bitumes asphaltes)	fi	0	0	0	0	0	0	0	0	2
	<gi>									0,500
	gi-									0,488
	gi+									0,512
36410 - Résines époxydiques	fi	0	0	0	0	0	0	1	1	2
	<gi>							0,098	0,106	0,427
	gi-							0,093	0,100	0,416
	gi+							0,104	0,112	0,438
2Q31Z - Autre amine aromatique	fi	0	0	0	0	0	0	1	1	2
	<gi>							0,502	0,504	1,499
	gi-							0,493	0,495	1,489
	gi+							0,512	0,514	1,509
36210-35110 - Peintures vernis laques mastics en phase solvants - Solvants et diluants organiques	fi	0	0	0	0	1	2	2	3	3
	<gi>					0,420	0,488	0,149	0,262	0,241
	gi-					0,408	0,475	0,142	0,253	0,232
	gi+					0,431	0,501	0,156	0,272	0,251
21320-33210 - HAP - Diesel (gaz)	fi	0	0	0	0	0	2	2	4	4
	<gi>						0,336	0,203	0,702	0,659
	gi-						0,326	0,195	0,688	0,646
	gi+						0,347	0,211	0,717	0,673

312B0-22224 - Suie - Trichloroethylene	fi	0	0	0	0	0	2	2	2	2
	<gi>						0,755	0,360	0,080	0,216
	gi-						0,740	0,349	0,075	0,207
	gi+						0,771	0,371	0,086	0,225
36F00 - Produits noirs (asphalte, goudrons, créosote)	fi	0	0	0	0	0	0	0	1	2
	<gi>								0,999	1,999
	gi-								0,999	1,999
	gi+								1,000	2,000
47100 - Radiations ionisantes	fi	0	0	1	1	1	1	1	2	3
	<gi>			0,999	0,999	0,999	0,999	0,999	1,999	2,999
	gi-			0,999	0,999	0,999	0,999	0,999	1,999	2,999
	gi+			1,000	1,000	1,000	1,000	1,000	2,000	3,000
36B11 - Carburants automobile diesel gas oil fioul gazole	fi	0	0	0	0	0	0	1	2	2
	<gi>							0,999	1,016	1,502
	gi-							0,999	1,013	1,492
	gi+							1,000	1,018	1,512
35110-36200 - Solvants et diluants organiques - Peintures vernis laques mastics	fi	0	1	1	1	1	1	1	2	3
	<gi>		0,506	0,243	0,26	0,401	0,609	0,052	1,015	2,011
	gi-		0,496	0,235	0,251	0,39	0,595	0,048	1,013	2,009
	gi+		0,516	0,251	0,268	0,412	0,623	0,056	1,018	2,013
36J10 - Produits capillaires	fi	0	1	1	1	1	1	1	1	3
	<gi>		0,494	0,499	0,495	0,508	0,493	0,506	0,247	0,609
	gi-		0,485	0,490	0,485	0,498	0,483	0,496	0,238	0,595
	gi+		0,504	0,509	0,505	0,518	0,503	0,516	0,255	0,623
21320-2Q150-2Q300 - HAP - Dérivés nitrosés des amines aromatiques aliphatiques - Amines aromatiques et dérivés	fi	0	0	0	0	0	0	2	2	2
	<gi>							1,208	1,202	1,200
	gi-							1,201	1,195	1,192
	gi+							1,216	1,210	1,207
36H00 - Produits d'entretien	fi	0	0	0	0	0	0	0	1	2
	<gi>								0,999	1,999
	gi-								0,999	1,999
	gi+								1,000	2,000
31220 - Bois (Poussières)	fi	0	0	0	1	1	1	1	1	2
	<gi>				0,999	0,999	0,999	0,999	0,999	1,206
	gi-				0,999	0,999	0,999	0,999	0,999	1,198
	gi+				1,000	1,000	1,000	1,000	1,000	1,214
21320-32110 - HAP - Amiante (Fibres)	fi	0	0	0	0	0	0	0	2	3
	<gi>								0,407	1,359
	gi-								0,397	1,350
	gi+								0,418	1,369
21320-34100 - HAP - Fumées de soudage	fi	0	0	0	0	1	1	1	2	3
	<gi>					0,506	0,497	0,495	1,505	0,870
	gi-					0,496	0,487	0,485	1,495	0,856
	gi+					0,516	0,507	0,505	1,515	0,884
21320-32110 - 35110 - HAP - Amiante (Fibres) - Solvants et diluants organiques	fi	0	0	0	0	2	2	2	2	2
	<gi>					1,500	0,677	0,546	0,366	0,248
	gi-					1,190	0,665	0,536	0,356	0,239
	gi+					1,510	0,689	0,557	0,766	0,256
35110-32110 - Solvants et diluants organiques - Amiante	fi	0	0	0	0	0	1	2	2	2
	<gi>						0,999	1,501	0,366	0,580
	gi-						0,999	1,491	0,356	0,568

(Fibres)	gi+						1	1,511	0,376	0,593
36200-2Q300 - Peintures vernis laques mastics - Amines aromatiques et dérivés	fi	0	0	0	0	0	1	3	3	3
	<gi>						0,999	1,049	1,031	1,021
	gi-						0,999	1,045	1,028	1,018
	gi+						1	1,054	1,035	1,024
31330 - Tabac (poussières) T66A	fi	0	0	0	0	0	4	6	7	8
	<gi>						3,741	4,224	4,832	5,161
	gi-						3,72	4,208	4,819	5,144
	gi+						3,763	4,240	4,849	5,179
34100 - Fumées de soudage	fi	0	0	0	0	0	1	4	4	4
	<gi>						0,119	3,101	1,878	1,588
	gi-						0,113	3,096	1,862	1,574
	gi+						0,126	3,107	1,894	1,601
32110 - Amiante (Fibres)	fi	0	0	0	0	0	0	1	3	7
	<gi>							0,076	0,650	3,224
	gi-							0,071	0,638	3,207
	gi+							0,081	0,663	3,240

Fonction par blocs		2001	2002	2003	2004	2005	2006	2007	2008	2009
36200 - Peintures, vernis, laques, mastics	fi	2	3	3	4	4	5	7	10	14
	<gi>	1,999	2,338	1,770	1,802	1,891	2,994	2,080	2,510	3,421
	gi-	1,999	2,321	1,758	1,787	1,875	2,972	2,063	2,497	3,396
	gi+	2,000	2,355	1,788	1,818	1,908	3,017	2,097	2,538	3,447
33210 - Diesel (gaz)	fi	0	0	0	0	0	10	14	17	19
	<gi>						8,764	8,299	8,268	9,693
	gi-						8,730	8,271	8,239	9,663
	gi+						8,798	8,327	8,298	9,724
36R00 - Produits à usage agricole	fi	0	0	0	0	0	0	2	5	8
	<gi>							1,169	1,674	3,299
	gi-							1,162	1,659	3,272
	gi+							1,177	1,688	3,326
36500-36700 - Matières plastiques caoutchoucs polymères - Matières colorantes	fi	0	1	1	1	1	1	2	2	2
	<gi>		0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
	gi-		0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
	gi+		0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
21320-31146 - HAP - Métaux durs	fi	0	0	0	0	0	0	2	2	2
	<gi>							0,000	0,000	0,000
	gi-							0,000	0,000	0,000
	gi+							0,000	0,000	0,000
22224 - Trichloroethylene	fi	0	0	0	0	0	4	6	6	6
	<gi>						3,700	3,655	2,985	2,635
	gi-						3,679	3,640	2,968	2,620
	gi+						3,723	3,670	3,002	2,649
32300 - Fibres végétales textile T66A	fi	0	0	0	0	0	0	0	2	2
	<gi>								0,676	0,293
	gi-								0,663	0,283
	gi+								0,690	0,303
21320-36C20-334C0 - HAP - Huiles et graisses lubrifiantes - Produits noirs (goudrons bitumes asphaltes)	fi	0	0	0	0	0	0	0	0	2
	<gi>									0,000
	gi-									0,000
	gi+									0,000
36410 - Résines époxydiques	fi	0	0	0	0	0	0	1	1	2
	<gi>							0,042	0,100	0,156
	gi-							0,038	0,094	0,149
	gi+							0,046	0,106	0,164
2Q31Z - Autre amine aromatique	fi	0	0	0	0	0	0	1	1	2
	<gi>							0,496	0,494	1,496
	gi-							0,486	0,484	1,486
	gi+							0,505	0,504	1,505
36210-35110 - Peintures vernis laques mastics en phase solvants - Solvants et diluants organiques	fi	0	0	0	0	1	2	2	3	3
	<gi>					0,000	0,000	0,000	0,000	0,000
	gi-					0,000	0,000	0,000	0,000	0,000
	gi+					0,000	0,000	0,000	0,000	0,000
21320-33210 - HAP - Diesel (gaz)	fi	0	0	0	0	0	2	2	4	4
	<gi>						0,000	0,000	0,000	0,000
	gi-						0,000	0,000	0,000	0,000
	gi+						0,000	0,000	0,000	0,000

312B0-22224 - Suie - Trichloroethylene	fi	0	0	0	0	0	2	2	2	2
	<gi>						0,000	0,000	0,000	0,000
	gi-						0,000	0,000	0,000	0,000
	gi+						0,000	0,000	0,000	0,000
36F00 - Produits noirs (asphalte, goudrons, créosote)	fi	0	0	0	0	0	0	0	1	2
	<gi>								0,999	2,013
	gi-								0,999	2,011
	gi+								1,000	2,015
47100 - Radiations ionisantes	fi	0	0	1	1	1	1	1	2	3
	<gi>			0,999	0,999	0,999	0,999	0,999	1,999	2,999
	gi-			0,999	0,999	0,999	0,999	0,999	1,999	2,999
	gi+			1,000	1,000	1,000	1,000	1,000	2,000	3,000
36B11 - Carburants automobile diesel gas oil fioul gazole	fi	0	0	0	0	0	0	1	2	2
	<gi>							0,999	1,500	1,497
	gi-							0,999	1,490	1,487
	gi+							1,000	1,509	1,507
35110-36200 - Solvants et diluants organiques - Peintures vernis laques mastics	fi	0	1	1	1	1	1	1	2	3
	<gi>		0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
	gi-		0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
	gi+		0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
36J10 - Produits capillaires	fi	0	1	1	1	1	1	1	1	3
	<gi>		0,333	0,332	0,164	0,159	0,163	0,124	0,243	0,755
	gi-		0,324	0,323	0,157	0,152	0,156	0,117	0,234	0,740
	gi+		0,342	0,341	0,172	0,166	0,170	0,130	0,251	0,769
21320-2Q150-2Q300 - HAP - Dérivés nitrosés des amines aromatiques aliphatiques - Amines aromatiques et dérivés	fi	0	0	0	0	0	0	2	2	2
	<gi>							0,000	0,000	0,000
	gi-							0,000	0,000	0,000
	gi+							0,000	0,000	0,000
36H00 - Produits d'entretien	fi	0	0	0	0	0	0	0	1	2
	<gi>								0,999	1,999
	gi-								0,999	1,999
	gi+								1,000	2,000
31220 - Bois (Poussières)	fi	0	0	0	1	1	1	1	1	2
	<gi>				0,999	0,999	1,093	0,999	0,999	1,153
	gi-				0,999	0,999	1,087	0,999	0,999	1,146
	gi+				1,000	1,000	1,098	1,000	1,000	1,160
21320-32110 - HAP - Amiante (Fibres)	fi	0	0	0	0	0	0	0	2	3
	<gi>								0,000	0,000
	gi-								0,000	0,000
	gi+								0,000	0,000
21320-34100 - HAP - Fumées de soudage	fi	0	0	0	0	1	1	1	2	3
	<gi>					0,000	0,000	0,000	0,000	0,000
	gi-					0,000	0,000	0,000	0,000	0,000
	gi+					0,000	0,000	0,000	0,000	0,000
21320-32110 - 35110 - HAP - Amiante (Fibres) - Solvants et diluants organiques	fi	0	0	0	0	2	2	2	2	2
	<gi>					0,000	0,000	0,000	0,000	0,000
	gi-					0,000	0,000	0,000	0,000	0,000
	gi+					0,000	0,000	0,000	0,000	0,000
35110-32110 - Solvants et diluants organiques - Amiante	fi	0	0	0	0	0	1	2	2	2
	<gi>						0,000	0,000	0,000	0,000
	gi-						0,000	0,000	0,000	0,000

(Fibres)	gi+						0,000	0,000	0,000	0,000
36200-2Q300 - Peintures vernis laques mastics - Amines aromatiques et dérivés	fi	0	0	0	0	0	1	3	3	3
	<gi>						0,000	0,000	0,000	0,000
	gi-						0,000	0,000	0,000	0,000
	gi+						0,000	0,000	0,000	0,000
31330 - Tabac (poussières) T66A	fi	0	0	0	0	0	4	6	7	8
	<gi>						2,892	3,890	4,824	5,323
	gi-						2,875	3,875	4,809	5,306
	gi+						2,909	3,905	4,839	5,341
34100 - Fumées de soudage	fi	0	0	0	0	0	1	4	4	4
	<gi>						0,386	3,043	1,861	1,634
	gi-						0,376	3,039	1,845	1,621
	gi+						0,396	3,047	1,876	1,647
32110 - Amiante (Fibres)	fi	0	0	0	0	0	0	1	3	7
	<gi>							0,043	0,640	3,300
	gi-							0,039	0,628	3,283
	gi+							0,047	0,652	3,317

Annexe D-3. Echantillon des Lymphomes Non Hodgkiniens.

Fonction itérative		2001	2002	2003	2004	2005	2006	2007	2008	2009
47100 - Radiations ionisantes	fi	0	0	0	0	1	1	1	2	2
	<gi>					0,999	0,999	0,999	1,500	1,504
	gi-					0,999	0,999	0,999	1,490	1,494
	gi+					1,000	1,000	1,000	1,510	1,514
36200 - Peintures Vernis Laques Mastics	fi	0	0	1	1	1	1	1	1	2
	<gi>			0,999	0,498	0,502	0,502	0,502	0,498	1,005
	gi-			0,999	0,488	0,492	0,492	0,495	0,488	0,991
	gi+			1,000	0,508	0,512	0,512	0,511	0,508	1,019
21311 - Benzène	fi	0	1	2	2	2	2	2	3	4
	<gi>		0,999	1,999	1,999	1,999	1,999	1,999	2,506	3,504
	gi-		0,999	1,999	1,999	1,999	1,999	1,999	2,496	3,495
	gi+		1,000	2,000	2,000	2,000	2,000	2,000	2,515	3,514

Fonction par bloc		2001	2002	2003	2004	2005	2006	2007	2008	2009
47100 - Radiations ionisantes	fi	0	0	0	0	1	1	1	2	2
	<gi>					0,999	0,999	0,999	1,176	1,162
	gi-					0,999	0,999	0,999	1,169	1,155
	gi+					1,000	1,000	1,000	1,184	1,169
36200 - Peintures Vernis Laques Mastics	fi	0	0	1	1	1	1	1	1	2
	<gi>			0,999	0,999	0,204	0,204	0,200	0,198	0,532
	gi-			0,999	0,999	0,196	0,196	0,192	0,190	0,520
	gi+			1,000	1,000	0,212	0,212	0,208	0,205	0,544
21311 - Benzène	fi	0	1	2	2	2	2	2	3	4
	<gi>		0,999	1,999	1,999	1,999	1,999	1,999	2,668	3,880
	gi-		0,999	1,999	1,999	1,999	1,999	1,999	2,655	3,865
	gi+		1,000	2,000	2,000	2,000	2,000	2,000	2,680	3,894

Annexe E. " Occupational Exposome: a network-based approach for characterizing occupational health problems" - Supplementary Files.

1. Description of the RNV3P database

1.1. Code space of OHP

The RNV3P database is characterized by a four-dimensional space defined by the variables: disease, hazards, occupation, and activity sector (Fig. S1). The total number of possible distinct codes is 1716 for diseases, 6722 for hazards, 390 for occupations and 61 for activity sectors. The Fig. S1 shows that occupations and activity sectors are well represented while the used disease and hazard codes only a small part of the available space.

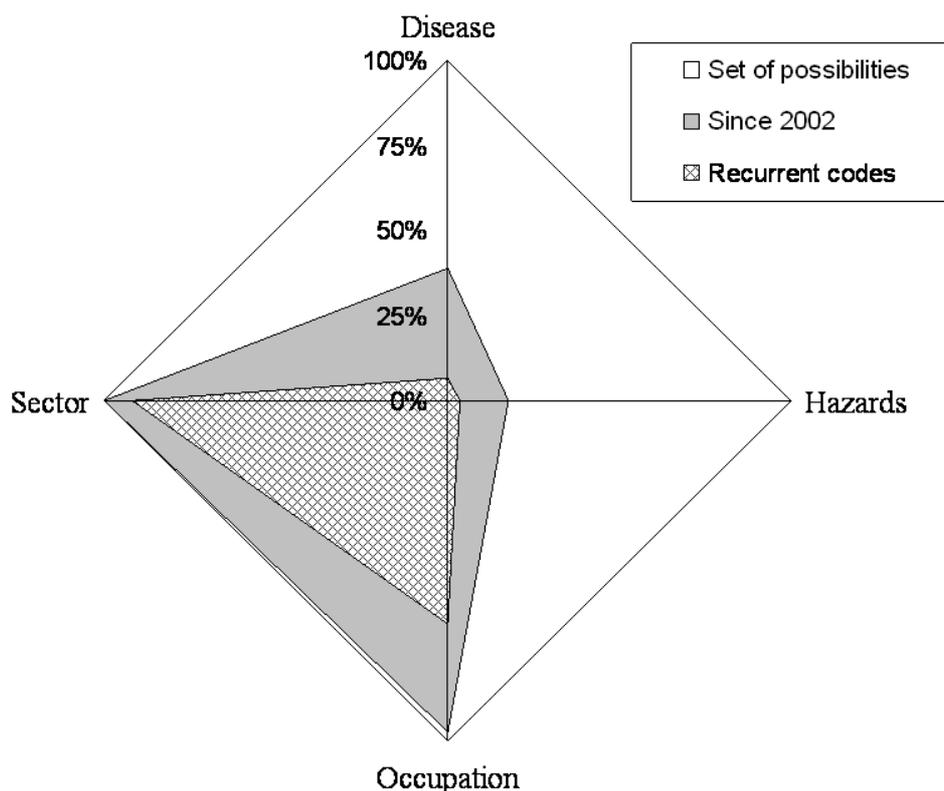


Fig. S1: Polar graph showing the proportion of space occupied by used codes, RNV3P database 2002-2007. Set of possibilities: available space of all codes. Since 2002 (gray): delimits the space of all codes already reported since 2002 until to 2007. Recurrent codes (textured gray): represents the set of codes reported every year since 2002.

2. Most reported diseases in the RNV3P

Table S1 shows the 5 most reported diseases in the RNV3P database. This illustrates the underlying diversity of the associations between diseases and occupational exposures.

Disease (ICD-10 codes)	no OHPs	no hazards	no occupations	no activity sectors
Pleural plaque (code J92)	6314 (14.0%)	102 (15.6%)	234 (61.1%)	55 (91.6)
Asthma (code J45)	3011 (6.7%)	488 (74.8%)	240 (62.9%)	54 (90.0%)

Allergic dermatitis (code L23)	2981 (6.6%)	440 (67.4%)	221 (58.0%)	52 (86.6%)
Reaction to a severe stress factor (code F43)	2876 (6.4%)	31 (4.7%)	246 (64.6%)	55 (91.6%)
Malignant neoplasms of bronchus and lung (code C34)	2613 (5.8%)	221 (33.8%)	211 (55.4%)	53 (88.3%)

Table S1: The 5 most reported diseases in the RNV3P database 2001 – 2007 (n = 57,368 OHPs, among which 813 distinct diseases associated with 1304 distinct hazards, 385 distinct occupations and 62 distinct activity sectors).

3. Exposomes

3.1. Exposome of malignant tumors

The subset of malignant tumors comprises $W = 3990$ OHPs in the RNV3P database, 2002-2007, forming $V = 195$ nodes. Most of reported malignant tumors concern the respiratory system (blue nodes in Fig. S3) like malignant tumors of the bronchus and lung (65% of malignant tumors, or 2606/3990) and mesotheliomas (13%, or 532/3990), and 210 malignant tumors of the bladder (classified as malignant urogenital tumors) among the diseases accounting for more than 5% of reports. The Fig. S2 illustrates the embedded structure of structure of malignant tumors, indicating hence the complexity of relationships between OHPs.

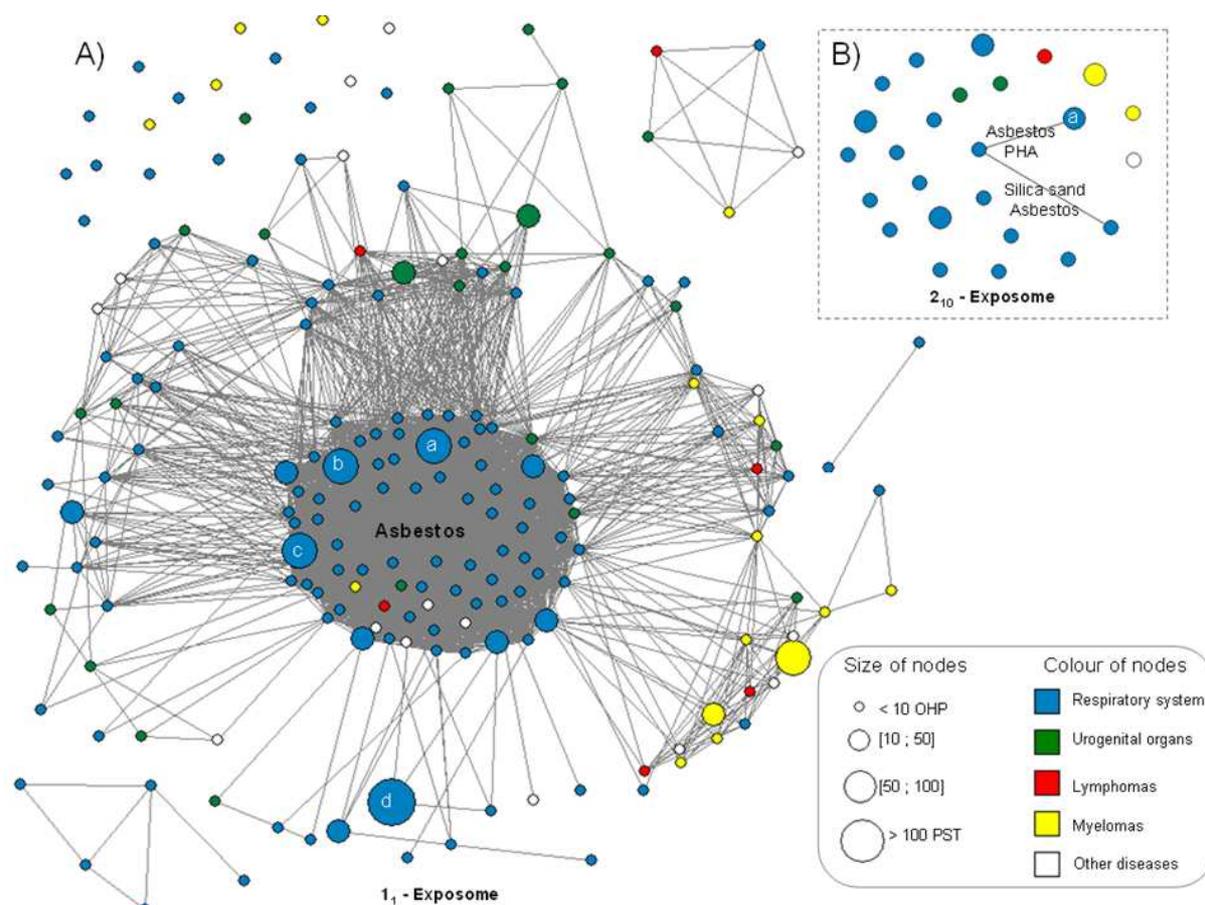


Fig. S2: Exposomes of malignant tumors, RNV3P database 2002-2007. A) – The 1_1 – exposome with $G_{MT} = \{W=3990, V=195, L=3715, D=1, \eta=1\}$. Each node represents a kind of malignant tumour with a color code, and the node size is proportional to the number of included OHPs. The central large cluster of nodes represents, in part, nodes associated with the hazard “asbestos” and other possible co-hazards. Four nodes (a, b, c and d)

comprise more than 50 OHPs: a) 58 malignant tumors of the bronchus and lung are associated with hazards “asbestos x polycyclic aromatic hydrocarbons”, b) 45 malignant tumors of the bronchus and lung and 5 mesotheliomas are associated with hazards “asbestos x welding fumes”, c) 49 bronchopulmonary cancers and 2 mesotheliomas are associated with hazards “asbestos x crystallised silica”, and d) 65 malignant tumors of the facial sinuses and 3 malignant tumors of the bronchus and lung are associated with hazards “wood dust”. B) – The 2_{10} – exposome with $G_{MT} = \{W=2648, V=25, L=2, D=2, \eta=10\}$ reveals three nodes connected with hazards “asbestos x PHA (polycyclic aromatic hydrocarbons)” and “silica sand x asbestos”.

The 1_1 - exposome of malignant tumors in Fig. S2A (obtained in including nodes with at least $\eta=1$ OHP and linked each other by at least $D=1$ common hazard) is not a very dense network ($d=0.20$) with 211 distinct hazards shared between nodes. Some nodes are unconnected, others form sparse and small groups, and most of nodes belong to a large and dense cluster, mainly concerning respiratory diseases, composed of nodes linked each other mainly by the hazard “asbestos” and other hazards. On the other hand, there is almost no connection between nodes in the 2_{10} – exposome in Fig. S2B (obtained in including only nodes with at least $\eta=10$ OHPs and linked each other by at least $D=2$ common hazards) that contains nearly 3/4 of OHPs (2648/3990) represented by 25 nodes. This exposome shows another dimension included in the structure of the 1_1 – exposome. Among the nodes of the 1_1 and 2_{10} - exposomes, we find for example lymphomas (red nodes).

3.2. NHL exposome

As shown in Fig. S3A, the distribution of node lengths (number of hazards) indicates that NHLs are more likely to be associated with one, two and three hazards, and less likely with 4 and 5 hazards. Fig. S3B and S3C illustrate how the variation of η and D can impact the shape of the NHL exposome shown in the main manuscript.

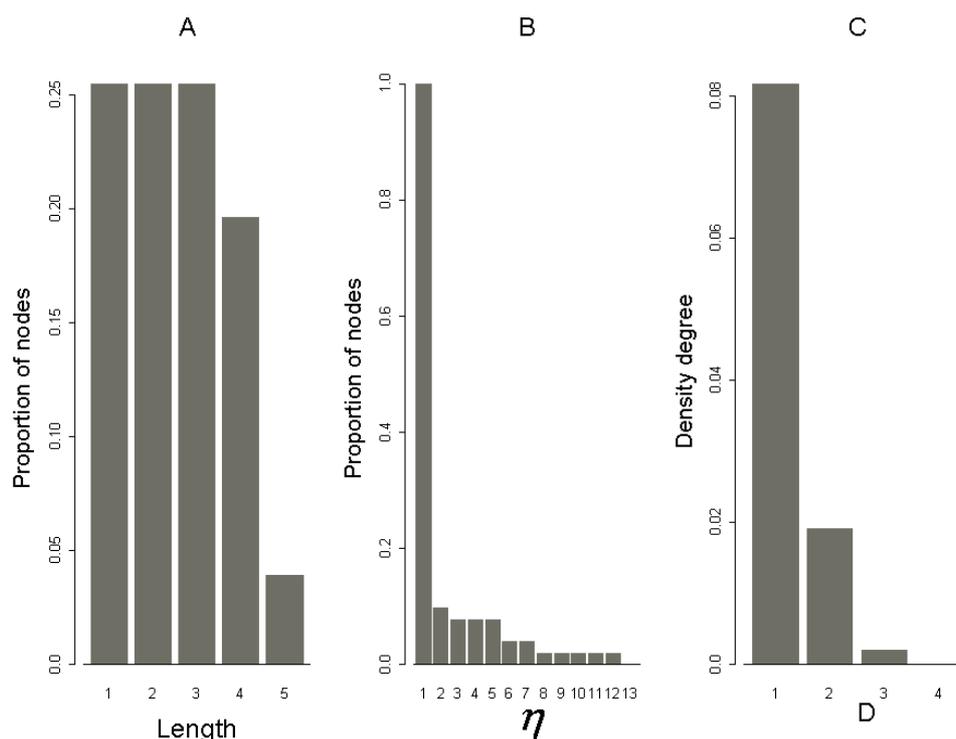


Fig. S3: A – Proportion of NHL nodes as a function hazard length (number of NHL associated hazards). B – Proportion of NHL nodes vs η values. C – Density of NHL D_1 – exposome as a function of D .

The η and D allow one to display hidden sub-layers of the exposome in focusing on heavier (several OHPs) and multi-connected (sharing several hazards) of nodes. As shown in Fig. S4B, the number of nodes dramatically decreases for $\eta \geq 2$ indicating that most of reported nodes represent only few OHPs. The largest node is associated to hazards « organic solvents and thinners » with 12 reported OHPs, i.e. 15% (12/77) of NHL OHPs. Likewise, the density of the exposome drops off with D as shown in Fig. S4C, in consistency with the overlap tree in the main manuscript (Fig. 7).

To complement the Fig. 3 in the main text, the ten highest-degree nodes and their main characteristics are detailed in Table S2. All highly connected nodes are multi-exposed, i.e., with number of hazards greater than one ($\text{length} \geq 1$), and belong to the most frequent exposure groups (cf. Fig. 8 in the main manuscript).

Table S2: Top ten highest-degree nodes and their principal characteristics (degree k , clustering coefficient c and cortege length). Gray cells indicate hazards never shared with none of exosome nodes.

k	c	Length	Hazard 1	Hazard 2	Hazard 3	Hazard 4	Hazard 5
23	0.486	3	Organic solvents and thinners	Benzene	Ionizing radiations	-	-
22	0.576	3	Organic solvents and thinners	Benzene	Trichloroethylene	-	-
20	0.616	4	Organic solvents and thinners	Benzene	Toluene	Inks	-
19	0.619	2	Organic solvents and thinners	Agricultural products	-	-	-
19	0.684	3	Organic solvents and thinners	Benzene	Diesel Gaz	-	-
19	0.655	4	Organic solvents and thinners	Trichloroethylene	Welding fumes	Lubricating oils and greases	-
19	0.643	3	Organic solvents and thinners	Trichloroethylene	Tetrachloroethylene	-	-
18	0.705	4	Organic solvents and thinners	Trichloroethylene	Asbestos	Gasoline	-
16	0.833	3	Organic solvents and thinners	Welding fumes	Asbestos	-	-
16	0.833	5	Organic solvents and thinners	Welding fumes	Asbestos	Hard metals, metallic carbons	Fluorhydric acid, hydrogen fluoride
16	0.833	4	Organic solvents and thinners	Welding fumes	Paints, varnishes, lacquers, putties	Epoxydic resins	-
15	0.904	4	Organic solvents and thinners	Welding fumes	Hard metals, metallic carbons	Glues, adhesives	-
15	0.904	2	Organic solvents and thinners	Black product wastes (asphalt, tars, creosote)	-	-	-
15	0.866	2	Organic solvents and thinners	Plastics, polymer rubbers	-	-	-
14	1.000	1	Organic solvents and thinners	-	-	-	-
12	0.515	3	Trichloroethylene	Agricultural products	Paints, varnishes, lacquers, putties	-	-

Table S3: List of hazards (aggregated codes) associated with NHL, classified by main groups of hazards (1st digit) and then by numbers in each group.

Associated exposures	Observation year						TOTAL
	2002	2003	2004	2005	2006	2007	
1 - Inorganic compounds							
Bromhydric acid, hydrogen bromide	1						1
Fluorhydric acid, hydrogen fluoride						1	1
Mercury					1		1
2 - Organic compounds							
Benzene	2	1	2	1	4	3	13
Trichloroethylene			1	3		3	7
Halogen derivatives of saturated aliphatic hydrocarbons	1		3			1	5
Tetrachloroethylene		1		1	1	1	4
Aliphatic aldehydes				2		1	3
Polycyclic aromatic hydrocarbons		1		1		1	3
Phenols and phenol derivatives	1			1			2
Toluene				1	1		2
Styrene	1				1		2
Carboxylic acids and peracids	1		1				2
Monocyclic aromatic hydrocarbons				1			1
Methanol	1						1
Diethylene glycol monobutyl ether					1		1
Other aromatic aldehydes	1						1
Aliphatic ketones	1						1
Other thioacids					1		1
Glycol esters	1						1
Acrylamide	1						1
Cyanates and isocyanates				1			1
Aromatic amines and derivatives						1	1
Aromatic nitro derivatives				1			1
3 - Industrial substances							
Organic solvents and thinners	6	1	2	5	5	7	26
Agricultural products		5	3	3	4	2	17
Welding fumes				2	1	3	6
Asbestos			1	1		2	4
Paints, varnishes, lacquers, putties						2	2
Plastics, polymer rubbers	1				1		2
Hard metals, metallic carbons						2	2
Black products (tars, bitumen, asphalts)			1			1	2
Petrol			2				2
Lubricating oils and greases			1	1			2
Cleaning products					2		2
Wood dust	1						1
Textiles, vegetal fibres						1	1
Synthetic fibres						1	1
Diesel gas					1		1
Industrial oils and greases					1		1
Heat decomposition products						1	1
Glues, adhesives						1	1
Epoxydic resins						1	1
Cloth dyes					1		1

	Inks				1		1
	Oils	1					1
	Leather treatment products		1				1
	Black product wastes (asphalt, tars, creosote)					1	1
	Capillary products	1					1
4 – Physical exposures							
	Ionizing radiations	2	2	1	2	2	9
	Work on screen		1				1
7 - Viruses							
	Viruses		1	1	1		3
9 – Animal species and substances of animal origin							
	Mammals		1				1
C – Medications and drugs							
	Antineoplastic drugs		1	1			2

Table S4: List of occupational codes (maximum 4 digits) associated with NHL classified by major unit group (first digit), and then by number of cases.

Job titles	Observation year						TOTAL
	2002	2003	2004	2005	2006	2007	
1 – Legislators, senior officials and managers							
Directors and managers in manufacturing industries		1					1
Directors and managers in construction and public works			1				1
Directors and managers in wholesale and retail trade						1	1
2 – Professionals, intellectual and scientific occupations							
Biologists, botanists, zoologists and related professionals	1			1			2
Chemists	1						1
Civil engineers		1					1
3 – Technicians and associate professionals							
Mechanical engineering technicians	1		1	1	1	1	5
Technicians in physical sciences and chemistry				1	1	1	3
Technicians in industrial chemistry	1				1		2
Electrical technicians						1	1
Physical and engineering science technicians not elsewhere classified				1			1
Agronomy and forestry technicians						1	1
Medical assistants					1		1
Dental assistants					1		1
Pharmaceutical and dispensing assistants			1				1
Nursing staff (intermediate level)		1					1
Special education teaching associate professionals					1		1
4 – Clerks							
Stock clerks			1				1
Other office workers		1					1
5 - Service workers and shop and market sales workers							
Hairdressers, barbers, beauticians and related workers	1						1
Shop salespersons and demonstrators					1		1
6 - Skilled agricultural and fishery workers							
Mixed crop growers		1	1			1	3
Tree and shrub crop growers			1		1		2

Field crop and vegetable growers				1		1
Gardeners, horticultural and nursery growers				1		1
Market-oriented crop and animal producers				1		1
Subsistence agricultural and fishery workers				1		1
7 - Craft and related trade workers						
Motor vehicle mechanics and fitters				1	3	4
Varnishers and related painters				1	1	2
Welders and flame-cutters	1			1		2
Sheet metal workers-boilermakers		1		1		2
Machine-tool setters and operators			1		1	2
Compositors, typesetters and related workers				2		2
Carpenters and joiners, construction					1	1
Floor layers and tile setters				1		1
Plumbers and pipefitters	1					1
Building and related electricians	1					1
House painters and paper hangers	1					1
Structural-metal preparers and erectors					1	1
Toolmakers and related occupations					1	1
Printing engravers, photo-engravers and etchers				1		1
Furriers and related workers					1	1
Shoe-makers and related workers				1		1
8 - Plant and machine operators and assemblers						
Chemical still and reactor operators				1	1	2
Chemical-processing-plant operators not elsewhere classified	1					1
Power-production plant operators				1		1
Mining-plant operators			1			1
Pharmaceutical- and toiletry-products machine operators	1					1
Metal finishing-, plating- and coating-machine operators		1				1
Textile-, fur- and leather-products machine operators		1				1
Electrical-equipment assemblers					1	1
Electronic-equipment assemblers	1					1
Metal-, rubber- and plastic-products assemblers					1	1
9 – Elementary occupations						
Helpers and cleaners in offices, hotels and other establishments	1	1				2
Hand-laundrers and pressers			1		1	2

Table S5: Activity sector codes (2 digits maximum) associated with NHL.

Activity sectors	Observation year						TOTAL
	2002	2003	2004	2005	2006	2007	
C – Mining and quarrying							
Mining (not specified)		1	3	2	1	1	8

Mining of uranium					1	1
D – Manufacturing						
Textiles	1					1
Fur and apparel			1			1
Leather and footwear		1				1
Wood and products of wood					1	1
Paper and paperboard					1	1
Publishing, printing, reproduction			1	2		3
Chemicals	3		1	1	1	6
Rubber and plastic products		1		1		2
Manufacture of basic metals		2	1	1	2	7
Manufacture of machines and equipment		1				1
Manufacture of office machines and computer products		1				1
Manufacture of electrical machines and appliances	1				1	2
Manufacture of motor vehicles				1	1	2
Manufacture of other transport equipment			1		1	2
E - Production and distribution of electricity, gas and water						
Production and distribution of electricity, gas and heat			1			1
F - Construction						
Construction	1		1	1	1	4
G – Trade and repair of motor vehicles and domestic appliances						
Sale and repair of motor vehicles		1	1	1	1	4
Wholesale trade and sales agents					1	1
Retail and repair of domestic appliances				1		1
I - Transports and communication						
Land transport	1		1		1	3
K – Real estate, letting and business services						
Research and development	1	1		2		4
L – Public administration						
Public administration		1	1		1	3
M - Education						
Education					1	1
N – Health and social work activities						
Health and social work activities		1	1		2	4
O – Other community, social and personal service activities						
Sewerage, road maintenance, waste management		1				1
Personal service activities	1		1	1	1	4