

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 Bonneterre^{1, 2},
Régis de Gaudemar^{1, 2} et Dominique Bicot¹

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.Bicot@imag.fr, d.bicot@vet-lyon.fr

SUMMARY

Introduction: Monitoring of health risks takes an important place in the field of the environment / health. In France, since 2001, 30 consultation centers for 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 using RNV3P, a method designed to identify and describe various groups of "pathological cases" (in this instance disease / occupational nuisance combinations) with their time dynamics, for purposes of vigilance and detection of emerging combinations .

Method: This method identifies various disease/occupational nuisance combinations, their incidence and connects combinations sharing at least one exposure in common. We named exposome such a network of disease cases. The projection or comparison of these exposomes at different times allow both to follow the time dynamics of disease cases observed earlier but also to single out new disease cases never observed before.

Results: As an illustration, the method was applied to "systemic sclerosis". The groups of disease cases identified in 2002-2005 data were reported and compared with the 2006 exposome.

Conclusion: This new method of characterizing data appears very promising to optimise the analysis 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

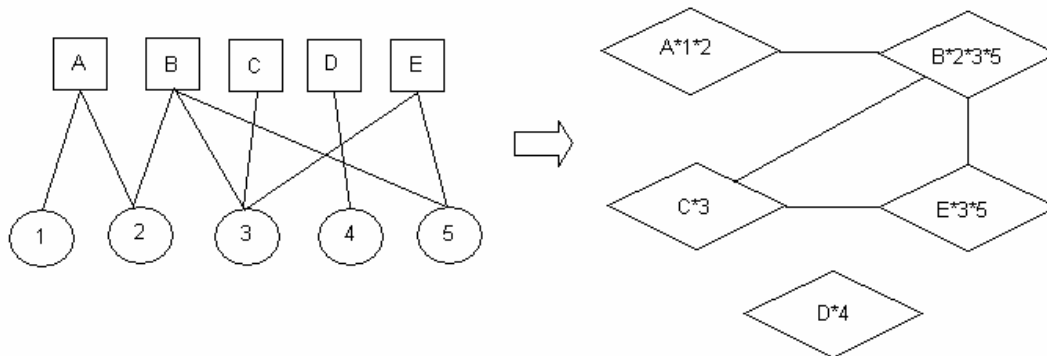
notamment la recherche de cas émergents (statistique exploratoire).

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 $\lambda = 1$ 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*

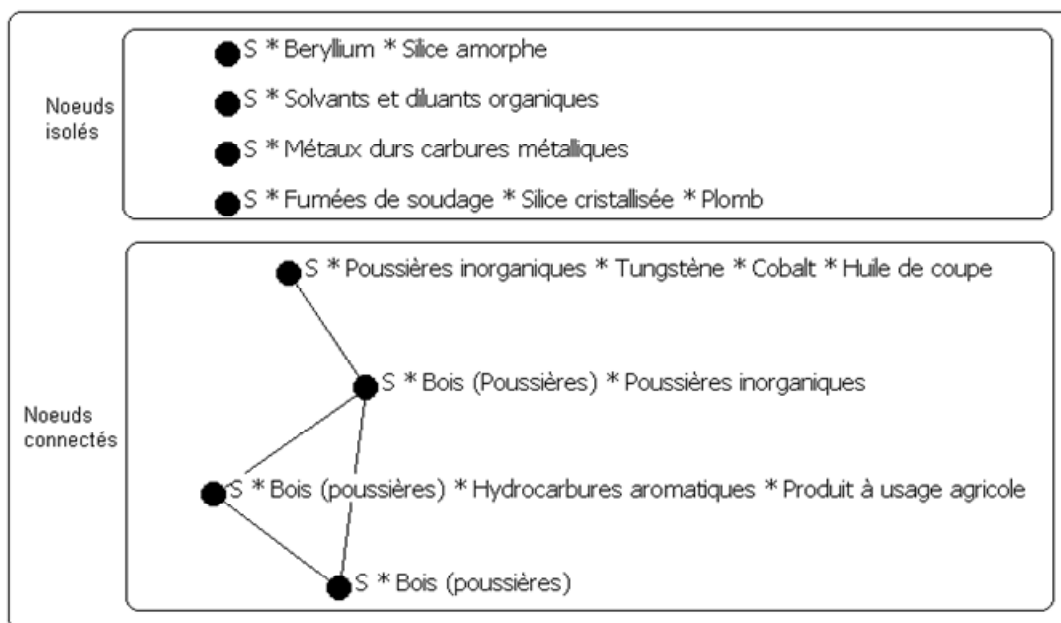
al., 2006 ; Celse *et al.*, 2006]. Ce réseau est opérationnel depuis 2001 sous la forme d'une base de données où chaque observation correspond à un problème de santé au travail décrit par une pathologie principale (code Classification Internationale des Maladies et des problèmes de santé connexes – CIM10), 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

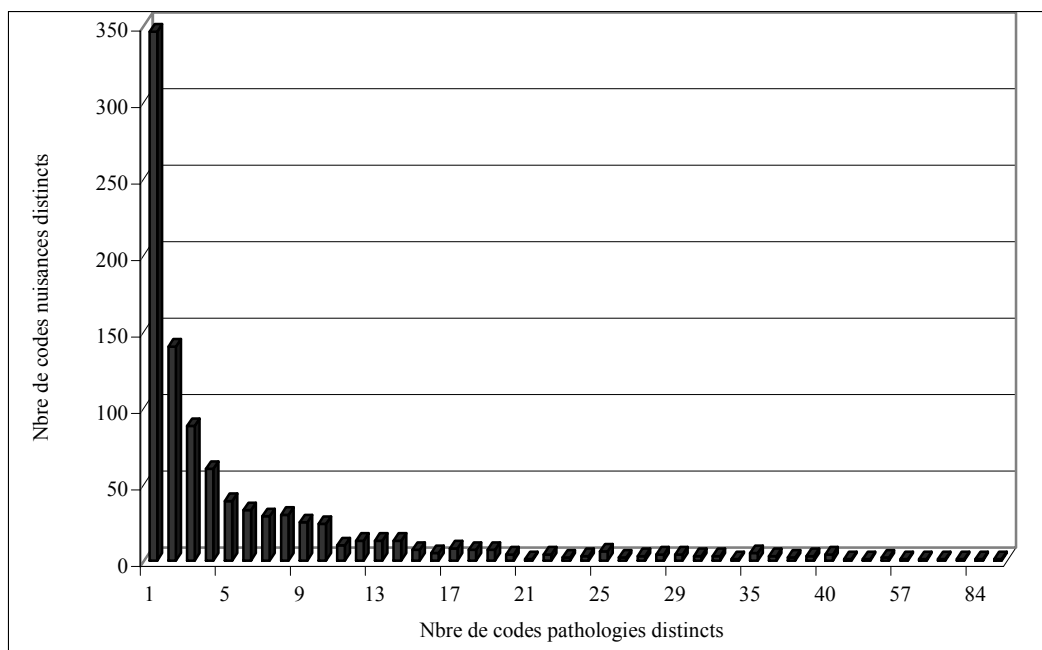
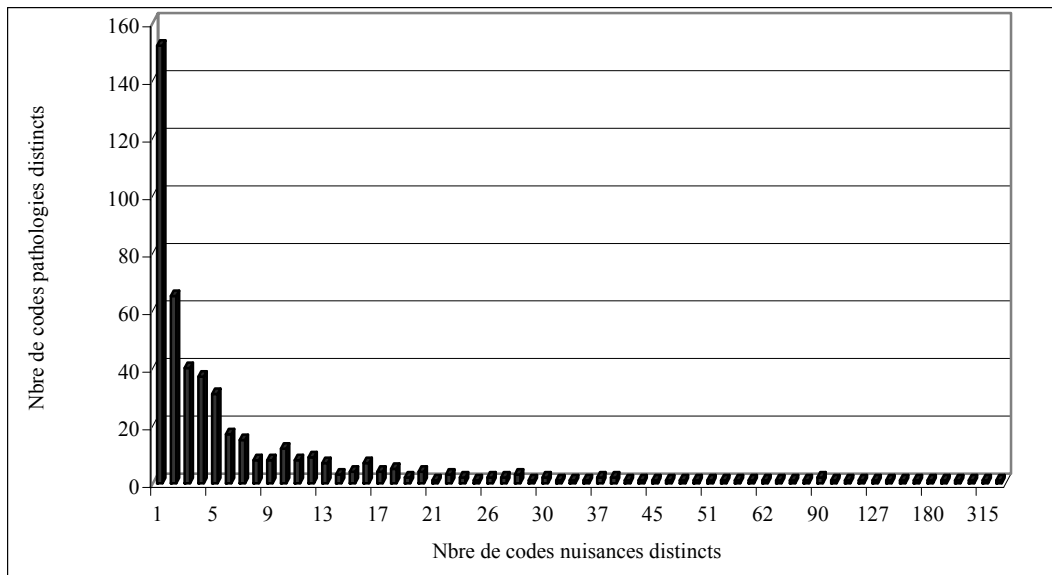


Figure 3b
Distributions des longueurs des codes nuisances



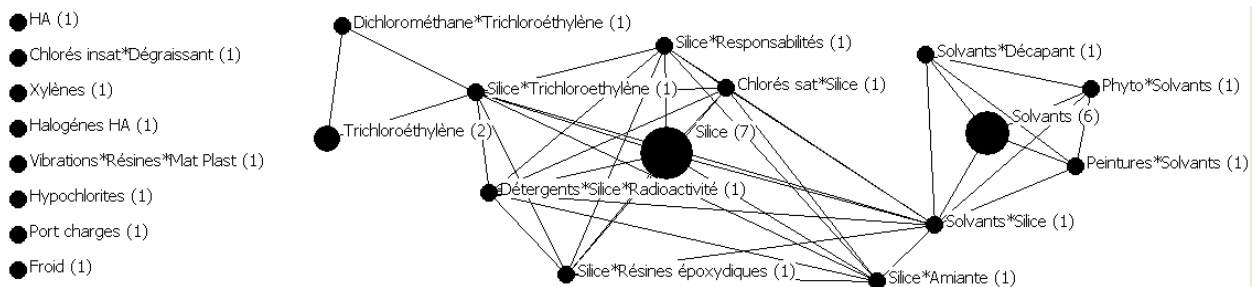
2. ELABORATION DE L'EXPOSOME

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



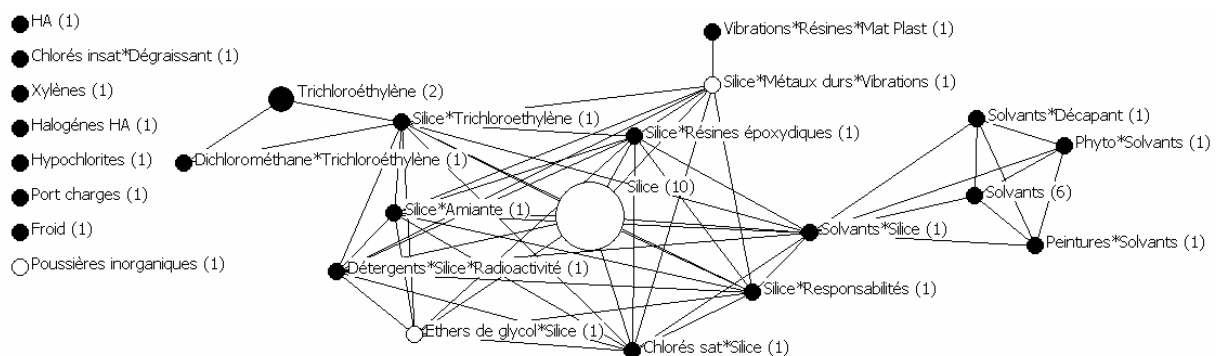
Cet exposome permet de regrouper des pathologies selon leurs expositions professionnelles et trois agrégats de pathologies sont mis en évidence autour des nuisances « silice », « solvants » et plus précisément « solvants chlorés ». La suite serait de suivre par exemple cet exposome dans le temps afin 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 sacroidosis ? *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.

