



LES RENCONTRES DE STATISTIQUE APPLIQUÉE

Institut National d'Études
Démographiques
(Salle Alfred Sauvy)



Cohortes : De la constitution à l'analyse

Vendredi 5 avril 2013, 14h – 17h30

Les études de cohorte permettent d'analyser sur une population suivie dans le temps des liens entre survenue potentielle d'évènements et facteurs d'exposition (que ce soit à portée épidémiologique, environnementale, démographique ou économique). Elles nécessitent des protocoles spécifiques de collecte, sont souvent sujettes à des biais de sélections et requièrent des méthodologies particulières d'analyse. Ces différents points seront abordés lors de cette séance.

Résumés

14h00 : **Xavier THIERRY** (Ined–Unité mixte Elfe) • **Introduction de la séance**

14h15 : **Marie ZINS** (Inserm) et **Alice GUEGUEN** (Inserm)
La cohorte Constances : généralité et effets de sélection
(résumé à venir)

14h50 : **Nathalie De PARSEVAL** (IReSP– AVIESAN, ITMO Santé Publique) et **Marie LHOSMOT** (AVIESAN, ITMO Santé Publique) • **Les cohortes françaises recensées dans le portail Epidémiologie-France**

Le portail « Epidémiologie – France » est l'une des actions prioritaires préconisées par le Conseil Stratégique des Industries de Santé (CSIS), pour accélérer le développement de la recherche en épidémiologie et en santé publique. Sa mise en œuvre a été confiée à l'Inserm - Institut thématique Santé Publique en janvier 2010. Ouvert en avril 2011, ce site web à vocation pérenne vise à décrire l'ensemble des bases de données de santé existant en France (bases médico-administratives, registres, cohortes, études cas-témoin, panels, ...).

Les objectifs de ce recensement sont les suivants :

- Favoriser la mise à disposition d'informations et la transparence autour des données de santé pour les professionnels et le grand public.
- Optimiser le repérage thématique des données et des compétences existantes, rendre plus visible l'expertise française dans ce domaine, au niveau national et international.
- Contribuer à la coopération entre équipes, favoriser les partenariats et stimuler la génération de nouveaux projets de recherche.
- Renforcer la qualité des recherches.

Actuellement, 448 bases de données sont décrites dans le portail, et l'on estime qu'il en existe au moins autant à recenser. Les cohortes représentent une grande partie des bases de données décrites : 236 cohortes sont recensées dans le portail. On distingue les cohortes "thématisées" qui portent sur une pathologie ou une famille de

pathologies (soit dans les critères d'inclusion soit dans les événements de santé observés) des cohortes "généralistes" dont le recrutement s'opère en population générale et qui couvrent un champ ouvert de problèmes de santé.

Le portail va par ailleurs prochainement héberger un espace dédié à la mutualisation de ressources épidémiologiques utiles aux cohortes et à d'autres types d'études.

15h40 : **Pascale ROUAUD** (Cereq) • **Non réponse et attrition : comment limiter les biais ? Un exemple avec « Génération », dispositif d'enquêtes longitudinales du Cereq**

Les enquêtes « Génération » s'intéressent à l'insertion des jeunes sortant du système éducatif une année donnée. Les Générations 1998 et 2004 ont fait l'objet de trois interrogations, permettant d'analyser leurs sept premières années dans la vie active. Le taux de réponse est d'environ 30% sur la première interrogation et de 70% sur les suivantes.

Nous abordons les actions destinées à limiter la non réponse et nous évoquons les méthodes de redressements mises en œuvre en précisant quels paramètres sont les plus liés à la probabilité de répondre. Nous étudions enfin l'impact des redressements à travers l'exemple de quelques indicateurs « clés ».

16h15 : **Lidia PANICO** (Ined) • **Exemples d'analyses en classes latentes et équations structurelles appliquées à des données de cohortes (cohorte Millenium)**

Les cohortes de naissance sont utilisées pour répondre à de nombreuses questions de recherche qui s'appliquent à toute la durée de l'étude, ou uniquement à des périodes longues (par exemple de l'enfance à l'adolescence). Des questions d'analyse peuvent également se poser sur des périodes de temps plus courtes.

La nature longitudinale des données permet de mettre en œuvre de nouvelles méthodes d'analyse.

Dans cette présentation, je montrerai quelques avantages, mises en garde et obstacles à la mise en œuvre d'analyses longitudinales utilisant des données sur l'enfance à partir d'une cohorte récente, The UK Millennium Cohort Study (MCS). La MCS porte sur environ 19 000 enfants nés entre 2000 et 2002 et vivant au Royaume-Uni juste après leur naissance.

Je présenterai trois exemples recourant à différentes méthodes d'analyses longitudinales : Analyse longitudinale en classes latentes (LCA), « Cross-lagged structural equation models » et « Graphical chain models ».

Cette communication aura pour objectif de donner des exemples d'analyses longitudinales possibles avec seulement quelques vagues et soulignera les mises en garde à faire quand on utilise des données issues d'une grande cohorte généraliste portant sur de jeunes enfants.

16h50 : Discussion et conclusion

Chaque trimestre, le Service Méthodes Statistiques de l'Ined propose un séminaire de statistique appliquée. Le séminaire est ouvert à tous, sans frais de participation.

Pour une bonne organisation, nous demandons aux personnes désirant y assister de s'inscrire à l'adresse suivante : http://www.ined.fr/fr/rendez_vous/rencontres_statistique_appliquee/

Pour tout autre renseignement contacter Bénédicte Garnier (benedicte.garnier@ined.fr)

Ined : 133, bd Davout, Paris 20e • Standard: 01 56 06 20 00

Métro : L9 (Porte de Montreuil) ou L3 (Porte de Bagnolet)

Bus 57 ou Tram T3b (Marie de Miribel)