

## Sujet de Doctorat de Biostatistique

« Expression de l'effet cluster pour données binaires »

### Structure d'accueil

- INSERM 1246 - SPHERE – methodS in Patient-centered outcomes and HEalth ResEarch (<http://www.sphere-nantes.fr/>)  
L'unité INSERM 1246 a trois tutelles : les Universités de Nantes et Tours, ainsi que l'INSERM
- Directrice : Pr Véronique Sébille

### Directeur de thèse

- Pr Bruno Giraudeau  
Tél : 02 47 47 46 18  
e.mail : [bruno.giraudeau@univ-tours.fr](mailto:bruno.giraudeau@univ-tours.fr)

### Co-encadrant

- Pr Sandra Eldridge (Queen Mary University of London)

**Localisation** : Tours

**Prise de fonction** : rentrée universitaire 2017-2018

**Financement** : allocation de recherche du Ministère

---

**Sujet** : Expression de l'effet cluster pour données binaires

### Contexte :

L'effet cluster se définit comme la ressemblance entre les mesures faites sur deux individus qui appartiennent à la même unité « sociale », qu'on appelle classiquement un cluster. Ainsi, si le cluster est une famille, l'effet cluster correspond à la ressemblance familiale ; si le cluster est un hôpital, l'effet cluster correspond à ce qu'on appelle l'effet centre dans les études cliniques, etc... Un autre domaine dans lequel l'effet cluster est important est celui de l'essai randomisé en cluster. Pour rappel, un essai randomisé thérapeutique est une étude expérimentale dans laquelle on alloue un traitement expérimental ou contrôle à un patient par tirage au sort. Un essai randomisé en cluster, quant à lui, est un essai dans lequel on randomise non pas des individus, mais des groupes d'individus, qu'on appelle des clusters (1). Dans ce type d'essai, les réponses des individus appartenant à un même cluster sont plus semblables que celles d'individus appartenant à des clusters distincts, c'est ce qu'on appelle l'effet cluster.

Cet effet est classiquement quantifié par le coefficient de corrélation intraclasse (CCI) (2). Initialement proposé pour l'étude des ressemblances familiales, ce coefficient est aujourd'hui un des éléments clés de la planification et de l'analyse statistique des essais randomisés en cluster. Ce coefficient s'interprète de deux façons : (i) il correspond à la corrélation entre deux mesures associées à des individus appartenant à un même cluster, quels que soient les individus et quel que soit le cluster ou (ii) il correspond à la part de la variabilité totale qui est due à la variabilité inter-clusters. Estimer l'effet cluster fait en soit partie de l'analyse statistique d'un essai randomisé en cluster et cela contribue à l'interprétation des résultats de l'essai. Ainsi, dans l'étude de Chosidow et al (3) on a comparé l'ivermectine *per os* au malathion en lotion dans le traitement de la pédiculose résistante. On a randomisé 376 familles pour un total de 812 patients. Cette étude a montré que 95,2% des patients sous ivermectine étaient indemnes de poux vivants 15 jours après la randomisation contre seulement 85,0% sous malathion ( $p < 0,001$ ). Mais quand on regarde de plus près, les estimations ponctuelles de CCI étaient de 0,62 dans le

groupe ivermectine contre 0,44 dans le groupe malathion. Cette différence traduisait en fait que dans le groupe ivermectine, la plupart du temps, tous les membres de la famille étaient indemnes de poux, ou bien tous étaient toujours infestés, avec peu de familles mixant des sujets infestés et des sujets indemnes. A l'inverse il y avait bien plus de mixité intra-familiale dans le groupe malathion. Une des explications est que le traitement par ivermectine est plus simple d'administration puisqu'il s'agit d'un traitement *per os*, alors que pour le malathion, s'agissant d'un traitement topique, il faut appliquer la bonne quantité de produit, laquelle dépend de la nature des cheveux, de leur longueur, etc... et laisser agir suffisamment. Dans cet exemple, l'étude des CCI nous a donc clairement permis d'aller plus loin dans l'interprétation des résultats.

Lorsque les données sont binaires, le coefficient de corrélation intraclasse est lié à la prévalence de la réponse (i.e., il augmente avec la prévalence jusqu'à ce que celle-ci atteigne 50%, puis décroît lorsqu'elle tend vers 100%) (4), ce qui n'est pas sans gêner son interprétation. En effet, si dans un essai en cluster les taux de succès sont différents entre les groupes, les CCI associés seront différents, de façon purement artéfactuelle. D'autres coefficients existent pour mesurer l'effet cluster : le coefficient de variation (5), le coefficient Kappa, qui dépend lui aussi de la prévalence (2), ou encore le coefficient R (6), mis ce dernier coefficient est jusqu'à ce jour demeuré assez confidentiel.

Par ailleurs, si on fait l'hypothèse que le critère binaire est issu de la binarisation d'un critère continu, il a été montré qu'il existe un lien entre les coefficients de corrélation associés aux critères continu et binaire (7). Dans le contexte des essais en cluster, ce lien persiste (8) et peut-être y a-t-il là une solution pour exprimer l'effet cluster en s'affranchissant des artefacts liés au caractère binaire des données.

### **Objectif :**

L'objectif de ce projet est de comparer différentes indices permettant d'exprimer l'effet cluster lorsqu'on s'intéresse à appréhender l'effet cluster et que le critère de jugement est binaire.

### **Méthodes :**

#### *Etape 1 :*

On cherchera en premier lieu à identifier dans la littérature les différents paramètres proposés pour étudier l'effet cluster. Outre une recherche sur mot-clés (considérant certes l'effet cluster tel qu'envisagé dans les essais en cluster, mais aussi l'effet centre, ou la ressemblance familiale...), on utilisera une approche de type boule de neige à savoir que les articles seront identifiés à partir des références des premiers articles trouvés, jusqu'à saturation du processus. Certains paramètres pourront également être proposés, tel l'application de la formule d'atténuation permettant de basculer d'une échelle binaire à une échelle continue, ainsi que précédemment évoqué (7).

#### *Etape 2*

Les différents paramètres identifiés seront comparés au moyen de simulations de Monte-Carlo. Ces études de simulation permettront d'étudier les propriétés statistiques associées aux différents paramètres, ainsi que leurs relations. Les résultats de simulation seront illustrés à partir de jeux de données disponibles dans le laboratoire.

### **Retombées :**

En termes de publication, l'étape 1 permettra de rédiger un article de type "Review" illustrant les différentes approches possiblement utilisables pour quantifier l'effet cluster. Les résultats de l'étude de simulation feront quant à eux l'objet d'un article original.

Les résultats obtenus permettront d'identifier une/des méthodes pouvant quantifier l'effet cluster en cas de critère de jugement binaire, tout en s'affranchissant de l'artéfact lié à la prévalence. Ces résultats pourront se traduire en recommandations aux auteurs d'essais en cluster.

1. Donner A. Design and analysis of cluster randomization trials in health research. London: Arnold; 2000.
2. Eldridge SM, Ukoumunne OC, Carlin JB. The Intra-Cluster Correlation Coefficient in Cluster Randomized Trials: A Review of Definitions. *Int Stat Rev.* 2009;77(3):378-94.
3. Chosidow O, Giraudeau B, Cottrell J, Izri A, Hofmann R, Mann SG, Burgess I. Oral ivermectin versus malathion lotion for difficult-to-treat head lice *N Engl J Med.* 2010 Mar 11;362(10):896-905.
4. Gulliford MC, Adams G, Ukoumunne OC, Latinovic R, Chinn S, Campbell MJ. Intraclass correlation coefficient and outcome prevalence are associated in clustered binary data. *J Clin Epidemiol.* mars 2005;58(3):246-51.
5. Hayes R. Cluster randomised trials. Boca Raton: CRC Press; 2009.
6. Crespi CM, Wong WK, Wu S. A new dependence parameter approach to improve the design of cluster randomized trials with binary outcomes. *Clin Trials Lond Engl.* déc 2011;8(6):687-98.
7. Kirk D. On the numerical approximation of the bivariate normal (tetrachoric) correlation coefficient. *Psychometrika.* 1973;38:259-68.
8. Caille A, Leyrat C, Giraudeau B. Dichotomizing a continuous outcome in cluster randomized trials: impact on power. *Stat Med.* 30 oct 2012;31(24):2822-32.

### **Compétences requises**

#### *Profil recherché :*

Le candidat devra avoir être titulaire d'un M2 de Biostatistique et montrer une appétence pour la méthodologie en recherche clinique.

#### *Profil technique :*

- Bonnes connaissances biostatistiques
- Aptitudes à conduire des études de simulation Monte-Carlo
- Maîtrise de logiciels biostatistiques (SAS, R)