



UNIVERSITÉ DE NANTES



Offre de stage M2 Biostatistique  
Début : Janvier/avril 2022 - 6 mois

## EVALUATION DE L'IMPACT DES DONNEES MANQUANTES SUR LES MODELES DE RASCH ET DE LA THEORIE DE REPONSE A L'ITEM

### IMPACT OF MISSING DATA ON RASCH AND ITEM RESPONSE THEORY MODELS

#### Contexte

Mieux appréhender le vécu et le ressenti des patients est nécessaire dans de nombreuses pathologies où l'évaluation des concepts subjectifs ou Patient Reported Outcomes (PRO, par exemple : qualité de vie, bien-être, fatigue, anxiété) est une étape indispensable pour évaluer les conséquences des traitements et des prises en charge.

La psychométrie est la discipline qui permet de mesurer des concepts subjectifs en utilisant le plus souvent des réponses à des items d'un questionnaire et en construisant des modèles de mesure pour définir une mesure la plus valide et fiable possible à partir de ces réponses. Parmi les différents modèles de la psychométrie, les modèles dits de la famille de Rasch et les modèles de la théorie de réponse à l'item, qui appartiennent le plus souvent à la famille des modèles linéaires généralisés à effets aléatoires, constituent une approche intéressante. Ils décrivent la probabilité de choisir une modalité donnée à un item donné en fonction du concept subjectif que l'on cherche à mesurer, généralement à travers un lien logistique ou probit, en tenant également compte des caractéristiques psychométriques du questionnaire utilisé.

Par ailleurs, les données longitudinales sont très fréquemment recueillies en recherche clinique et épidémiologique car elles permettent notamment l'évaluation de l'évolution du critère d'intérêt. Or, il est probable que certains questionnaires puissent comporter des données manquantes à certaines visites ou que certaines visites soient manquantes. De plus, ces données manquantes sont potentiellement informatives. Au moment de la modélisation, il est nécessaire de savoir quel est l'impact des données manquantes sur les résultats de la méthode utilisée.

Les modèles de Rasch pourraient permettre une estimation non biaisée dans certains contextes grâce à leur propriété d'objectivité spécifique. En effet, cette propriété stipule que les estimations de la variable latente, représentant le concept que l'on cherche à mesurer, ne dépendent pas des items auxquels les personnes ont répondu. Des travaux ont déjà été menés pour évaluer l'impact des données manquantes sur les performances statistiques du modèle de Rasch dans le cadre d'items dichotomiques (à deux modalités de réponse). Il reste à l'évaluer pour des items polytomiques. Enfin, les modèles de la théorie de réponse à l'item ne possèdent pas la propriété d'objectivité spécifique et pourraient avoir de moins bonnes performances que les modèles de la famille de Rasch en cas de données incomplètes.

#### Objectif

Ce stage vise à comparer les performances des modèles de la famille de Rasch et des modèles de la théorie de réponse à l'item en cas de données manquantes dans des études longitudinales.

#### Méthodes

L'évaluation et la comparaison des performances des différentes méthodes sera basée sur une étude de simulation. Ce type d'études permet de répondre à un large panel de questions : l'évaluation de la précision des mesures avec les erreurs de type I et de type II, le biais des estimations ou la robustesse des méthodes statistiques dans des contextes particuliers (déviations aux hypothèses du modèle, données manquantes). Ces évaluations sont possibles car les méthodes sont appliquées non pas sur des données réelles mais sur des données simulées dont on connaît ainsi le modèle sous-jacent et les

valeurs des paramètres. Les études de simulation sont donc un outil puissant lorsque les performances d'une méthode ne peuvent être évaluées de manière analytique.

### **Missions**

L'étudiant devra mettre en place une étude de simulation permettant d'évaluer l'impact des données manquantes sur les performances des deux méthodes. Différentes situations seront envisagées selon : la taille des échantillons et du questionnaire, la quantité et le type de données manquantes (MCAR ou MNAR). Il s'agira de simuler les données selon un plan de simulation, de les analyser de manière automatisée avec les deux méthodes et d'évaluer les performances selon des critères à définir. L'étudiant aura accès au Centre de Calcul des Pays de la Loire.

### **Références**

Blanchin M, Hardouin J-B, Le Néel T, Kubis G, Sébille V. Analysis of longitudinal Patient-Reported Outcomes with informative and non-informative dropout: Comparison of CTT and Rasch-based methods. International Journal of Applied Mathematics & Statistics 2011;24:107-24.

Burton A, Altman DG, Royston P, Holder RL. The design of simulation studies in medical statistics. Statistics in Medicine 2006;25:4279-92.

de Bock É, Hardouin J-B, Blanchin M, Le Neel T, Kubis G, Bonnaud-Antignac A, et al. Rasch-family models are more valuable than score-based approaches for analysing longitudinal patient-reported outcomes with missing data. Stat Methods Med Res 2016;25:2067-87.

### **Profil attendu**

Le candidat devra être étudiant en Master 2 Biostatistique ou équivalent. Il devra avoir une bonne connaissance des modèles linéaires généralisés à effets aléatoires et de la typologie des données manquantes. Il devra être à l'aise avec les langages de programmation de logiciels et avoir une appétence pour la programmation. Des connaissances en psychométrie ou en études de simulation seraient appréciées.

**Structure d'accueil:** U1246 SPHERE « methodS in Patient-centered outcomes and HEalth REsearch »

<https://sphere-inserm.fr/>

L'unité Inserm U1246 SPHERE est une équipe de recherche des Universités de Nantes et de Tours. Cette équipe de recherche pluridisciplinaire centre ses travaux sur le développement, la validation et l'application de méthodes pour répondre à une variété de défis rencontrés dans la recherche et la prise de décision en matière de santé, en tenant compte à la fois des environnements et des perceptions qu'ont les patients de leur état de santé. L'étudiant stagiaire collaborera avec les membres de l'équipe particulièrement concernés par l'axe « Methods for the measurement and interpretation of Self-Reported Outcomes », qui sont principalement localisés à l'IRS 2 de Nantes.

Lieu de stage : Institut de Recherche en Santé 2 (IRS2) - 22, Boulevard Bénoni-Goullin, 44200 Nantes

Encadrants : Myriam Blanchin, Ingénieure de recherche et Véronique Sébille, PU-PH

Rémunération : Selon le minimum légal en vigueur au moment de la période du stage

Durée souhaitée : 6 mois à partir de janvier/avril 2022

**Merci d'adresser votre CV et lettre de motivation à**

Myriam Blanchin

[myriam.blanchin@univ-nantes.fr](mailto:myriam.blanchin@univ-nantes.fr)