

## ÉTABLISSEMENT D'UN LIEN ENTRE LES RÉPONSES SUR L'APPARTENANCE À PLUSIEURS RACES LORS DU RECENSEMENT DES ÉTATS-UNIS ET LES CATÉGORIES UNIRACIALES POUR LE CALCUL DES TAUX D'ÉVÉNEMENTS DÉMOGRAPHIQUES

Nathaniel Schenker, Jennifer D. Parker et Deborah D. Ingram<sup>1</sup>

### RÉSUMÉ

En 1997, le U.S. Office of Management and Budget a publié des normes révisées concernant la collecte de renseignements sur la race dans le système statistique fédéral. L'une des révisions consiste à permettre aux individus de choisir plus d'un groupe racial lors de la réponse à des enquêtes fédérales et à d'autres activités fédérales de collecte de données. Ce changement pose des difficultés en cas d'analyse portant sur des données recueillies en vertu de l'ancien et du nouveau système de déclaration de la race, puisque ces données ne sont pas comparables.

Comme, aux termes des deux systèmes, la plupart des gens ne déclarent qu'une seule race, une solution proposée couramment est d'essayer d'établir un lien de transition par attribution d'une catégorie raciale unique à chaque personne déclarant plusieurs races en vertu du nouveau système, et de procéder aux analyses en utilisant uniquement les catégories uniraciales observées et assignées. Donc, le problème peut être considéré comme un problème de données manquantes dans lequel les valeurs pour les catégories uniraciales manquent et doivent être imputées pour les personnes qui déclarent plusieurs races.

L'Office of Management and Budget a proposé plusieurs méthodes simples d'établissement d'un lien pour résoudre ce problème des données manquantes. Schenker et Parker (*Statistics in Medicine*, à paraître) ont analysé des données provenant de la National Health Interview Survey réalisée par le U.S. National Center for Health Statistics, qui permet la déclaration de plusieurs races, mais demande aussi aux personnes qui le font de préciser une race principale, et ont constaté que l'on pourrait améliorer les méthodes d'établissement d'un lien en intégrant des covariables de niveau individuel et des covariables contextuelles dans les modèles d'établissement des liens.

Dans la présente communication, nous considérons un problème particulier d'établissement de lien qui se pose lorsqu'on calcule les taux d'événements démographiques selon la race. Ce genre de taux est utilisé fréquemment en épidémiologie et dans d'autres études. À partir de 2000, les données provenant du Recensement décennal des États-Unis, qui sont utilisées pour calculer les dénominateurs des taux, ont été recueillies en vertu du nouveau système de déclaration de la race. Par contre, pour les enregistrements de naissance et de décès, qui sont la source des données pour les numérateurs, le passage à un nouveau système se fera progressivement sur plusieurs années. Par conséquent, les données des numérateurs seront souvent disponibles conformément aux anciennes catégories uniraciales, tandis que celles des dénominateurs le seront conformément aux catégories du nouveau système.

Pour établir le lien entre les données du recensement et les données de l'état civil, nous étoffons la méthode de Schenker et Parker (*Statistics in Medicine*, à paraître) et ajustons des modèles d'établissement de lien aux données de la National Health Interview Survey. Puis, nous appliquons les résultats des méthodes d'établissement de lien aux données du Recensement de 2000 pour créer des dénominateurs dont les chiffres sont conformes à l'ancien système de déclaration de la race. Nous utiliserons les taux provisoires de natalité chez les adolescentes et de mortalité en 2000 aux États-Unis pour illustrer les résultats de la méthode et pour les comparer à ceux obtenus par des méthodes plus simples.

Alors que Schenker et Parker (*Statistics in Medicine*, à paraître) n'examinent que trois grands groupes multiraciaux, l'application décrite ici oblige à prédire les catégories uniraciales pour plusieurs petits groupes multiraciaux également. Donc, des problèmes de paucité des données se posent lors de l'ajustement des modèles d'établissement de liens. Nous les résolvons en construisant des modèles combinés pour plusieurs groupes multiraciaux, donc en puisant de la force dans chacun d'eux. Nous discuterons de ces problèmes et d'autres questions de méthodologie.

MOTS CLÉS: Régression logistique, données manquantes, imputation, enquête, données de l'état civil.

---

<sup>1</sup> U.S. National Center for Health Statistics, É.-U., nschenker@cdc.gov