

XIII. COMPARABILITÉ ENTRE POPULATIONS

1. Rapport sur la santé dans le monde 2000

Lors de l'élaboration des estimations pour le Rapport sur la santé dans le monde 2000, des corrections ont été apportées aux erreurs systématiques connues les plus importantes des mesures disponibles afin d'améliorer la comparabilité entre populations – par exemple, la sous-déclaration des données de mortalité dans les données d'état civil. Le concept de cohérence interne a servi de moyen pour améliorer la validité des évaluations épidémiologiques.

2. Principaux commentaires et critiques

Les critiques portant sur les données du Rapport sur la santé dans le monde 2000 ont été plutôt sévères, mais la présente section traite uniquement de la question de la comparabilité entre pays. Il y a eu peu de débats publics et de discussions à ce sujet mis à part la reconnaissance du problème lié aux données autodéclarées.

3. Réponses et propositions de l'OMS

En examinant la mortalité auto-déclarée à partir des données d'enquête pour tous les états de l'Inde, Murray et Chen (1992) ont rapporté les résultats suivants : le Kerala a la morbidité auto-déclarée la plus élevée et le Bihar la plus faible de tous les états de l'Inde. D'un autre côté, une mesure objective de la santé – telle que la mortalité – révèle que le Kerala a une espérance de vie beaucoup plus élevée que le Bihar. Ensuite, une comparaison entre les États-Unis et le Kerala indique que la morbidité auto-déclarée des États-Unis est beaucoup plus importante que celle du Kerala, bien que l'espérance de vie soit plus élevée aux États-Unis qu'au Kerala.

Que se passe-t-il ? Sont-ce les caractéristiques de l'environnement – en matière d'éducation, de soins (par ex. la fréquence d'exposition au système de santé), de revenus, etc. – qui peuvent expliquer ces résultats apparemment incohérents ? Dans un article de *Philosophy and Public Affairs*, Amartya Sen (1992) a tenté d'expliquer ces résultats par ce qu'il a appelé « objectivité positionnelle » (positional objectivity) : la « position » de l'individu (en matière d'éducation, de revenus, etc.) influence la réponse donnée – mais tous les individus dans la même position donneront la même réponse – d'où l'« objectivité positionnelle ». Dans un éditorial plus récent du

RAPPORT DU GESI SUR L'ÉVALUATION DE LA PERFORMANCE DES SYSTÈMES DE SANTÉ

British Medical Journal, il fait valoir à nouveau le fait que les données sur la morbidité auto-déclarée ont des limites qui peuvent rendre leur utilisation à des fins politiques extrêmement fallacieuse (Sen 2002).

L'OMS cherche à rendre les réponses des individus comparables (peu importe s'ils résident dans des états différents en Inde ou aux États-Unis) en *corrigeant* les données pour tenir compte des « positions » des individus dans les différents états de l'Inde et aux États-Unis. Il s'agit évidemment d'un exercice très important afin d'obtenir des informations relatives à l'état de santé d'après les données d'enquête qui soient comparables entre les pays. En outre, les données auto-déclarées sur la santé sont encore de loin la source la plus fréquente de telles informations dans le monde entier.

En réponse au manque d'informations représentatives de la population en ce qui concerne deux variables clés de l'exercice d'évaluation de la performance des systèmes de santé dans le Rapport sur la santé dans le monde 2000, l'OMS a lancé l'Enquête multi-pays sur la santé et la réactivité. Afin d'évaluer la performance des systèmes de santé, les données de cette enquête ont été utilisées pour élaborer des mesures de : (i) l'espérance de vie ajustée de l'état de santé (EVAES), et (ii) le niveau de réactivité du système de santé dans un pays. Par exemple, la mesure de l'EVAES comprend des estimations des issues non fatales qui sont dérivées, en partie, des données d'enquête dans différents domaines de la santé (par ex. la mobilité, la cognition, l'affect, etc.). De même, le niveau de réactivité du système de santé d'un pays est également basé sur des données semblables de l'enquête. Les personnes interrogées doivent évaluer leurs expériences relatives à différents domaines de réactivité du système de santé (par ex. l'autonomie, la dignité, la promptitude de la prise en charge, etc.).

Les données d'enquête présentent deux caractéristiques qui conduisent au problème de la comparabilité entre populations. Premièrement, l'information sur les domaines est obtenue par auto-déclaration. Les personnes interrogées doivent évaluer leur propre expérience (ou perception) relativement aux différents domaines de la santé et à la réactivité des systèmes de santé. Le deuxième point est que les réponses à l'enquête d'auto-déclaration sont catégorielles et classées en ordre décroissant par leur nature même.

Un exemple tiré de l'Enquête multi-pays sur la santé et la réactivité de l'OMS en ce qui concerne le domaine de la mobilité illustre les caractéristiques des données. La question principale d'auto-déclaration demande aux personnes interrogées quel degré de difficulté elles ont éprouvé dans leurs déplacements au cours des 30 derniers jours. Les personnes interrogées doivent déterminer leur mobilité en fonction d'une échelle ordinale composée de cinq catégories allant de 1 à 5, pour laquelle 1 est « Difficulté extrême / Impossible », 2 « Sérieuse difficulté », 3 « Difficulté moyenne », 4 « Difficulté normale » et 5 « Aucune difficulté ».

La question de la comparabilité entre populations se pose à ce niveau-là. Le problème, lors de l'utilisation des données auto-déclarées pour les domaines de la santé et de la réactivité, est que les réponses ne sont pas comparables entre les pays, ou même entre les différents groupes socio-démographiques au sein des pays. Comme l'illustre la Figure 1, les réponses catégorielles peuvent être conçues comme un traçage de la carte du véritable niveau du domaine (sur cette figure, la ligne intitulée « échelle latente de mobilité ») d'après les réponses catégorielles des trois différentes populations A, B et C. Comme la figure l'illustre, une personne de la population A qui répond « Aucune difficulté » est tracée à un endroit différent de l'échelle réelle qu'une personne qui répond « difficulté légère » dans la population B et qu'une personne qui répond « difficulté modérée » dans la population C. En contrepartie, le même niveau de mobilité réelle pourrait être auto-déclaré par une personne des populations A, B et C comme ne présentant « Aucune difficulté », une « Difficulté moyenne » et une « Difficulté normale » respectivement. Les raisons pourraient être des normes, des attentes et des expériences différentes pour les personnes interrogées des différentes populations.

Ce problème a été reconnu auparavant dans la littérature sur la psychométrie relative aux tests d'aptitude (Q.I.) et, de manière plus générale, aux tests de connaissance en utilisant des tests standardisés (par ex. GRE, test d'aptitude scolaire, GMAT, etc.). Certains groupes, en fonction de leur aptitude ou de leurs connaissances, sont systématiquement meilleurs pour certains types de questions que d'autres groupes. Ce problème est connu sous le nom de « fonctionnement différentiel des items » dans la littérature sur les tests de connaissance et la psychométrie (Holland et Wainer 1993). Par exemple, dans la littérature sur la théorie de la réponse d'item, le modèle des indices partiels (qui est apparenté au modèle probit ordonné) reconnaît la probabilité de répondre en une ou deux catégories ordonnées (adjacentes) comme une fonction croissante de l'aptitude de la personne interrogée et d'une fonction décroissante de la difficulté catégorielle. Pour le même niveau d'aptitude, les difficultés peuvent être systématiquement différentes pour les différents groupes de population, ce qui entraînera des erreurs systématiques de mesure de l'aptitude. Bien que ce problème soit similaire au problème de glissements des points de découpage par catégories de réponse pour mesurer la santé ou la réactivité d'un système de santé, les méthodes de la résolution sont quelque peu différentes.

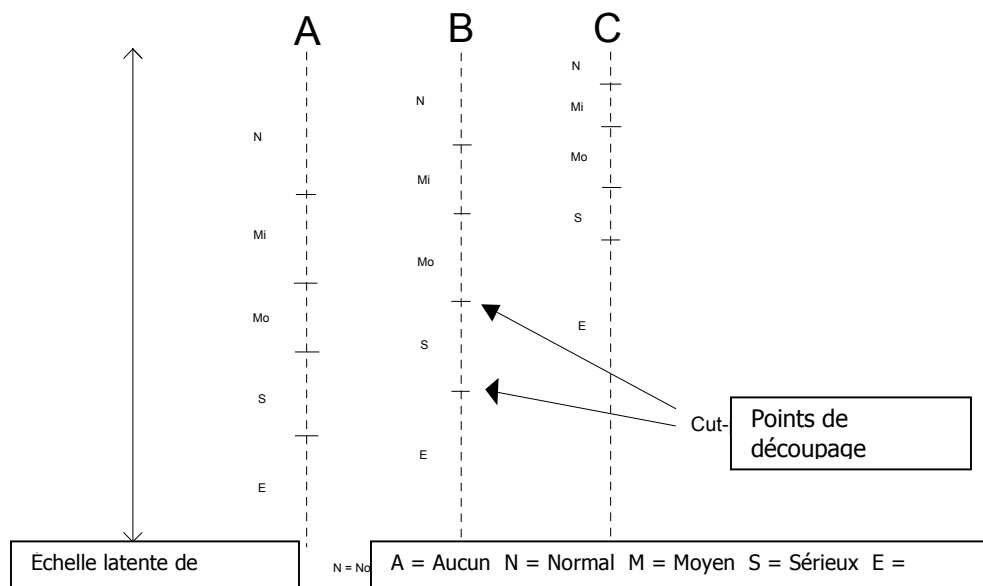


Figure 1

Il existe essentiellement deux stratégies que l'OMS a élaborées pour corriger les réponses aux enquêtes en fonction des différences systématiques d'attitudes des individus. Les deux stratégies impliquent l'utilisation d'un modèle statistique -- le modèle de probit ordonné (HOPIT). La première stratégie est d'utiliser le modèle HOPIT avec des « vignettes ». La deuxième stratégie est d'utiliser le modèle HOPIT avec des tests mesurés. Elles sont décrites l'une après l'autre.

Une vignette est une description d'un niveau d'aptitude pour un domaine donné que les personnes interrogées doivent évaluer en fonction de la même question et de la même échelle de réponses catégorielles que la question principale d'auto-déclaration. Une vignette décrit un niveau précis d'aptitude pour un domaine donné, afin que pour cette vignette, les différences de réponses dans tous les pays ou pour tous les groupes socio-démographiques puissent être attribuées au décalage des points de découpage des catégories de réponse. Les points de découpage des catégories de réponse sont estimés en utilisant le modèle HOPIT par le biais d'une procédure de vraisemblance maximale. Ces estimations des points de découpage sont utilisées pour calibrer l'auto-déclaration de la personne interrogée et les rendre comparables au sein de différentes populations. Si, par exemple, les personnes interrogées d'un certain groupe de population donnent systématiquement des réponses catégorielles plus élevées aux vignettes que les personnes interrogées d'un autre groupe, cela se manifestera par un point de découpage inférieur pour le premier groupe de l'estimation HOPIT.

Une deuxième stratégie est de calibrer les réponses de l'auto-déclaration à l'aide des tests mesurés (au lieu des vignettes) en corrélation avec le modèle HOPIT. Les tests mesurés sont les tests du niveau d'aptitude de la variable

latente sous-jacente pour un domaine de la santé. Les exemples incluent le test posturo-cinétique pour la mobilité et l'optotype de Snellen pour le domaine de la vision. De tels tests mesurés sont utilisés pour estimer les décalages des points de découpage entre les groupes de population pour le calibrage des réponses auto-déclarées comparables entre les populations. L'élaboration du modèle HOPIT avec les tests mesurés est relativement simple. Le modèle est fondé sur l'hypothèse que le test mesuré est en corrélation avec la variable latente sous-jacente pour un domaine et les points de découpage pour une réponse catégorielle donnée peuvent varier en fonction des groupes de population.

Pour diverses raisons, y compris celles reliées à l'erreur de la mesure, il semble que les vignettes représentent un mécanisme supérieur aux tests mesurés en matière de calibrage des réponses auto-déclarées. Par conséquent, les estimations actuelles de l'OMS relatives aux mesures des résultats en matière de santé et de réactivité sont essentiellement liées à l'utilisation de vignettes en tant que stratégie de calibrage.

4. Commentaires et recommandations du GESI

Le GESI a accueilli favorablement le travail de l'OMS dans ce domaine et a reconnu qu'il était important de veiller à ce que les données utilisées pour l'évaluation de la performance des systèmes de santé soient comparables d'une population à l'autre.

- (i) La méthode HOPIT dépend de manière cruciale de l'hypothèse de départ selon laquelle les réponses catégorielles correspondent à une dimension (ou un attribut) unique, que l'on peut classer sur une échelle unilinéaire. Le GESI a fait remarquer que les réponses ne doivent pas être fondées sur des appariements par les personnes interrogées qui impliqueraient des comparaisons dans plus d'une dimension (de régions planaires ou d'un ordre plus élevé correspondant aux cinq catégories). Pour pouvoir appliquer cette méthode, il faut que le domaine de référence de chaque question faisant l'objet d'une auto-déclaration soit précisé de manière très étroite et non ambiguë.
- (ii) Une voie prometteuse pour la recherche serait d'élaborer des méthodes statistiques qui regroupent les informations des vignettes et des tests mesurés dans une procédure conjointe d'estimation. Ces méthodes, qui sont apparentées aux modèles à indicateurs multiples et à cause complexe dans la littérature statistique, ont l'avantage de tenir entièrement compte de toutes les informations disponibles pour un individu donné. Dans ce cas, les sources multiples sont l'auto-déclaration par l'individu (calibrée à l'aide des vignettes), ainsi que

son test mesuré. Ces types de méthodes peuvent également entraîner différentes erreurs statistiques dans l'information auto-déclarée et dans l'information obtenue des tests mesurés.

- (iii) Le modèle HOPIT dépend essentiellement de la fiabilité et de la cohérence transculturelle des vignettes (par exemple les problèmes ou les erreurs de traduction ne changent pas le sens d'une question au point que ce soit une variable latente différente qui soit mesurée). Le GRPS recommande que les vignettes soient testées à nouveau dans différents contextes, y compris par la rétro-traduction.
- (iv) Il serait peut-être possible d'étudier certains problèmes relatifs au point (iii) ci-dessus en utilisant une version à coefficients aléatoires du modèle HOPIT. Contrairement à la version actuelle, un modèle à coefficients aléatoires permet à la variable latente associée à chaque vignette d'avoir sa propre variance (plutôt que d'avoir la même variance pour toutes les vignettes). Cette méthode permet de tenir compte de la possibilité que certaines vignettes soient intrinsèquement « plus bruyantes » que les autres. Cela peut être particulièrement pertinent pour les vignettes se rapportant au centre d'un domaine, c'est-à-dire pour les vignettes qui ne portent pas sur une extrémité ou l'autre du domaine.
- (v) Le GESI a reconnu que le modèle HOPIT non seulement se penche sur le problème de la comparabilité entre populations, mais aussi convertit les informations discrètes (catégorielles) de chaque domaines de la santé et de la réactivité en une *variable continue*. Pour chaque personne, l'agrégation de ces variables pour les domaines appropriés créent la distribution continue à partir de laquelle le niveau moyen et l'inégalité de la santé et de la réactivité sont estimés. Par conséquent, le modèle HOPIT permet de faire bien plus que de comparer des populations : il sert de base pour estimer quatre des cinq indicateurs intrinsèques utilisés dans l'évaluation de la performance des systèmes de santé.
- (vi) Les membres du GESI ont fait part de plusieurs commentaires techniques sur la méthodologie HOPIT. Certains sont mentionnés ci-dessous.
 - (a) Les estimations des points de découpage pour un groupe de population (par ex. un pays) dépendront de l'univers statistique des groupes inclus dans l'estimation des points de découpage. Par exemple, supposons que les points de découpage pour le groupe A sont estimés en fonction des données pour les groupes A et B. Les données du groupe C sont maintenant disponibles et les points de découpage du groupe A sont réévalués en fonction des données des trois groupes A, B et C. En règle générale, les points de découpage (et les autres estimations de paramètres) pour le groupe A changeront.

Cela pourrait rendre le classement entre, disons, les groupes A et B tributaire des autres groupes inclus dans l'estimation (surtout si l'on considère les agrégats inter-domaines). Par conséquent, il convient de faire preuve de prudence en matière d'évaluation des classements relatifs entre pays qui pourraient être tributaires de l'univers statistique.

(b) L'hypothèse utilisée dans le modèle HOPIT est que la variable latente (par ex. la mobilité) n'est pas bornée (car la distribution normale est utilisée pour le terme d'erreur). Le GESI recommande que le Secrétariat de l'OMS vérifie la robustesse de leurs résultats pour restreindre la variable latente à un intervalle fini (par ex., par le biais de l'hypothèse d'une distribution normale tronquée pour le terme d'erreur), ce qui semblerait une hypothèse plus réaliste pour les domaines en question.

(c) Les membres du GESI ont dit qu'il serait très utile d'estimer les formes fonctionnelles non linéaires pour l'équation de la variable latente (par ex. la fonction de production de la santé), ce qui peut également répondre en partie au problème mentionné au point (b). Une forme log-linéaire pour la fonction de production de la santé semble plus réaliste car elle permet de limiter les retours aux facteurs qui déterminent la santé (par exemple l'âge, l'éducation, etc.), ce qui peut être plus raisonnable que d'assumer des retours constants pour chaque facteur. Dans tous les cas, il serait utile de vérifier la réactivité des résultats du modèle HOPIT actuel en fonction des hypothèses émises sur la forme fonctionnelle.

5. Références

Holland, P.W. et H. Weiner (1993): *Differential Item Functioning*. Hillsdale, New Jersey: Lawrence Erlbaum Associates, Publishers.

Murray, C.J.L. et L.C. Chen (1992): "Understanding Morbidity Change", *Population and Development Review*, vol.18(3), sept.: 481-503.

Sen, A.K. (1993): "Positional Objectivity", *Philosophy and Public Affairs*, vol. 22: 126-45.

Sen, A.K. (2002): "Health: Perception versus Observation", *British Medical Journal*, vol.324:860-861.