



Open Access Full Text Article

EDITORIAL

Editorial: The obsession with precision in economic evaluations: have we gone too far?

[Éditorial : L'obsession de la précision dans les évaluations économiques: sommes-nous allés trop loin?]

Jason R. Guertin^{1,2}

¹Département de médecine sociale et préventive, Faculté de Médecine, Université Laval

²Centre de recherche du CHU de Québec – Université Laval

Correspondence:

Jason R. Guertin, Département de médecine sociale et préventive, Faculté de Médecine, Université Laval, Pavillon Ferdinand-Vandry, 1050 avenue de la Médecine, Québec, J1V 0A6, QC, Canada

Email:

Jason.Guertin@fmed.ulaval.ca

Article received:

5 February 2020

Article accepted:

14 February 2020

Editorial: The ultimate goal of all economic evaluations is to help decision makers decide on how to best allocate limited resources in order to maximize patients' health. To do so, outcome researchers in health economics provide decision makers with a relatively simple metric: the incremental cost-effectiveness ratio (ICER). Unfortunately, like many things in life, the simplicity of this metric hides a tremendous amount of work. In order to obtain a study's ICER, we often need to combine dozens if not hundreds of parameters together and, since we know that even minimal differences in key parameters can substantially impact the ICER, we always strive to include the most accurate estimates in our evaluations. Though our end goal is noble, I often wonder if we are exaggerating the level of precision of these parameters. In my opinion, I consider it happens frequently when we look at utility score measures.

In view of their importance within cost-utility analyses, it is logical that we put so much emphasis on estimating utility scores with the utmost precision. Luckily for us, the most commonly used indirect assessment questionnaires (e.g., EQ-5D; SF-6D; HUI-3) provide us with estimates that are precise up to the third decimal point [1-4].

That being said, could anyone really affirm that the utility score they would assign to a given health state be as precise had they been asked to provide this estimate with the use of a direct assessment approach? For example, for the standard gamble approach to provide a utility score estimate that was precise to the third decimal point, a respondent would have had to provide an answer that was precise to a tenth of a percent. Although theoretically possible, I do not believe it is plausible for anyone to do so. However, this raises the question as to what should the level of precision on the utility score be.

In a broader sense, I think the level of precision on the utility score estimates we provide should be associated with the minimum clinically important differences (MCID) of the approach used to assess these estimates. Unfortunately, a "one-size fits all MCID" does not exist. Previous studies have shown that the MCID for utility scores depend on both the method used to assess these estimates and the population/clinical condition being examined. For example, Walters and Brazier previously reported that the average MCID with the SF-6D was 0.041 (range: 0.011 to 0.097) and 0.074 (range -0.011 to 0.140) with the EQ-5D [5]. Alternatively, values of 0.01 and 0.03 have both been reported as the MCID for the HUI3 [6].

Considering the range of MCID that have been reported with indirect assessment questionnaires and aligned with the level of precision one could expect to obtain with the direct assessment approaches, I think utility score estimates should be limited to two decimal points. Despite the fact that a ± 0.01 level of precision may be considered too precise in some settings, it grants a conservative cut-off between utility score estimates all while being aligned with plausible differences one could expect to obtain.

©2020 Guertin, publisher and licensee CybelePress.com. This is an Open Access article, allowing unrestricted non-commercial use, provided the original work is properly cited.

Éditorial : Le but ultime de toutes les évaluations économiques est d'aider les décideurs à choisir comment utiliser le plus adéquatement leurs ressources limitées afin de maximiser la santé des patients. Pour se faire, les experts en évaluations économiques fournissent aux décideurs une mesure relativement simple : le rapport coût-efficacité différentiel (RCED). Malheureusement, comme

plusieurs choses dans la vie, la simplicité de cette mesure cache un travail énorme. Afin d'obtenir le RCED d'une étude, nous devons souvent combiner ensemble des douzaines, voire des centaines de paramètres, et comme nous savons que même les plus petites différences chez certains paramètres clés peuvent altérer substantiellement le RCED, nous nous efforçons sans cesse d'inclure les paramètres les plus précis dans nos travaux. Malgré le fait que notre but est noble, je me questionne souvent à savoir si nous n'exagérons pas le niveau de précision de ces paramètres. De mon point de vue, j'ai souvent l'impression que ce problème est plus flagrant lorsque nous examinons les valeurs des scores d'utilité.

Considérant leur importance au sein des analyses coût-utilité, il est normal que nous mettions autant d'emphase pour estimer les scores d'utilité avec le plus haut niveau de précision. Heureusement pour nous, les instruments d'utilité multi-attributs les plus communément utilisés (c.-à-d., EQ-5D, SF-6D, HUI3) permettent tous d'obtenir des estimés précis à la troisième décimale [1-4].

Cela étant dit, un individu est-il réellement capable d'affirmer qu'il pourrait assigner un score d'utilité à un état de santé avec autant de précision si on lui avait demandé de l'estimer via une méthode élicitation directe? Par exemple, pour qu'un score d'utilité obtenu grâce au pari standard soit précis à la troisième décimale, il faudrait qu'un répondant ait fourni une réponse précise à un dixième d'un point de pourcentage. Quoique théoriquement possible, je ne crois pas qu'il est plausible que quelqu'un le fasse dans le monde réel. Cependant, ce problème nous amène à nous questionner sur quel devrait être le niveau de précision des scores d'utilité.

Globalement, je pense que le niveau de précision des mesures de score d'utilité que nous fournissons devrait être associé aux plus petites différences cliniquement importantes (minimum clinically important difference – MCID) pouvant être déterminées par l'approche utilisée pour estimer celles-ci. Malheureusement, il n'existe pas de « règle

d'or » par rapport aux MCID. Des résultats antérieurs ont montré que le MCID des scores d'utilité dépend de la méthode utilisée pour les estimer ainsi que de la population et/ou de la condition médicale étudiée. Par exemple, Walters et Brazier ont rapporté que le MCID moyen avec le SF-6D est de 0,041 (intervalle de 0,011 à 0,097) et de 0,074 (intervalle de -0,011 à 0,140) [5]. Tandis que des valeurs de MCID de 0,01 et 0,03 ont été rapportées pour le HUI3 [6].

Étant donnée l'étendue des MCID ayant été rapportés avec les méthodes de mesures indirectes et vu le niveau de précision qu'on pourrait s'attendre à obtenir avec les méthodes de mesures directes, je pense que les mesures de scores d'utilité devraient être limitées à deux décimales. Malgré le fait qu'un niveau de précision de $\pm 0,01$ pourrait être jugé trop précis dans certains contextes, il permet un équilibre entre l'imposition d'un seuil conservateur et le niveau de précision qu'on pourrait s'attendre à obtenir.

References

- [1] Feeny D, Furlong W, Torrance GW, Goldsmith CH, Zhu Z, DePauw S, et al. Multiattribute and single-attribute utility functions for the health utilities index mark 3 system. *Med Care* 2002;40(2):113-28.
- [2] Xie F, Pullenayegum E, Gaebel K, Bansback N, Bryan S, Ohinmaa A, et al. A Time Trade-off derived Value Set of the EQ-5D-5L for Canada. *Med Care* 2016;54(1):98-105.
- [3] Brazier J, Roberts J, Deverill M. The estimation of a preference-based measure of health from the SF-36. *J Health Econ* 2002;21(2):271-92.
- [4] Poder TG, Fauteux V, He J, Brazier JE. Consistency Between Three Different Ways of Administering the Short Form 6 Dimension Version 2. *Value in Health* 2019;22(7):837-42.
- [5] Walters SJ, Brazier JE. Comparison of the minimally important difference for two health state utility measures: EQ-5D and SF-6D. *Qual Life Res* 2005;14(6):1523-32.
- [6] Drummond M. Introducing economic and quality of life measurements into clinical studies. *Annals of Medicine* 2001;33(5):344-9.