

Chapitre 5 : Comment calculer un score de qualité de vie ?

Lorsque tous les questionnaires de QDV d'une étude ont été complétés et saisis, l'analyse des données ne se fait généralement pas directement sur les réponses aux items bruts. En effet, des scores sont généralement calculés à partir des réponses données aux items permettant de résumer dans une variable unique l'information contenue dans un ou plusieurs items. Les scores des questionnaires de QDV doivent être calculés selon les recommandations en vigueur des développeurs des questionnaires. Ces scores sont issus de la validation de structure de l'échelle. La méthode de calcul pourra donc dépendre de la méthode de validation (tests classiques ou théorie de réponse à l'item). Si un questionnaire mesure plusieurs dimensions de QDV, un score peut être calculé par dimension regroupant tous les items mesurant cette dimension. Un score global peut également être calculé. Ce score global correspond généralement à une synthèse des réponses données à l'ensemble du questionnaire. Le ou les scores générés seront considérés comme une mesure quantitative tel qu'un score élevé correspondra à un haut ou bas niveau de QDV, en fonction du sens retenu pour le scoring.

1. Méthodes d'estimation de scores

Pour une large majorité de questionnaires, les scores estimés correspondent à la somme ou à la moyenne des réponses aux items considérés. Ces scores sont ensuite le plus souvent standardisés sur une échelle allant de 0 à 100, afin de faciliter leur interprétation ainsi que la comparaison entre les scores issus d'une même échelle ou entre plusieurs échelles. Certains questionnaires font exception à cette règle. Nous pouvons par exemple mentionner le questionnaire HADS mesurant l'état d'anxiété et dépression [3], le questionnaire MFI-20 mesurant la fatigue [9], ou bien les questionnaires du groupe nord-américain FACIT (<https://www.facit.org/>). Le calcul des scores pour ces questionnaires correspondra à la somme des réponses aux items. Ces calculs de scores reposant sur l'estimation de la moyenne ou de la somme sont assez simples et intuitifs et peuvent être implémentés sous tout type de logiciel statistique afin d'automatiser les calculs. En revanche, pour certains questionnaires, la méthode de scoring choisie peut être beaucoup plus sophistiquée. A titre d'exemple, pour certains questionnaires, la méthode de scoring retenue est basée sur des modèles issus de la théorie de réponse aux items. Nous pouvons par exemple mentionner les questionnaires du groupe PROMIS (Patient-Reported Outcomes Measurement Information System, <https://www.promishealth.org/57461-2/>) ainsi que le BREAST-Q (www.qportfolio.org) évaluant la QDV de patientes atteintes d'un cancer du sein [1].

L'estimation de ces scores se fait donc généralement en ligne sur un e-software dédié où les utilisateurs doivent charger leur base de données sur un site internet sécurisé.

2. Manuel de scoring

Chaque questionnaire validé doit proposer un manuel de scoring précisant les règles de calcul de ces scores et leur interprétation. Ces manuels peuvent être obtenus auprès des développeurs des questionnaires. Par exemple, lorsque l'utilisateur fait la demande d'utilisation d'un questionnaire, l'autorisation peut lui être accordée avec l'envoi du manuel de scoring afin de faciliter son utilisation et éviter toute erreur de scoring. Il est ainsi important de se procurer systématiquement la version officielle et en vigueur du manuel de scoring lorsqu'il existe et ne pas uniquement se référer à d'éventuelles informations reportées dans un article ayant utilisé ledit questionnaire. En effet, ces informations sont souvent incomplètes du fait de l'espace limité dans les articles scientifiques et peuvent ainsi être source d'erreur. Concernant les questionnaires des groupes EORTC et FACT, ce manuel est obtenu lorsque la demande d'utilisation est accordée et envoyée par email à l'utilisateur. Pour les questionnaires validés de l'EORTC, la demande est effectuée via l'URL suivant : <https://qol.eortc.org/form/#1>. Pour les questionnaires validés du groupe FACT, la demande peut être faite via l'URL suivant : <https://www.facit.org/measure-language-availability>. La plateforme SeleQt permet également d'obtenir le manuel de scoring d'un large panel de questionnaires de QDV et santé perçue (<https://seleqt.univ-lorraine.fr/>).

3. Données manquantes

Un facteur important à prendre en compte lors du calcul de ces scores, influençant potentiellement les résultats de QDV, est la prise en compte de données manquantes. En effet, il est possible que le patient ne remplisse pas intégralement le questionnaire. Ces données manquantes sont un facteur limitant pour le calcul des scores de QDV et doivent être prises en compte de façon adéquate. Généralement, un pourcentage de données manquantes est toléré pour estimer un score. Ainsi, l'EORTC recommande qu'au moins 50% des items soient renseignés pour pouvoir estimer le score de la dimension considérée. Le score sera alors calculé en considérant que l'item manquant ne diffère pas des items répondus, i.e. qu'une imputation par la moyenne des items répondus sera appliquée par patient. Autrement dit, le score sera calculé par rapport au nombre d'items répondus et non au nombre d'items mesurant la dimension. Si plus de 50% des items sont manquants pour un patient donné, alors son score sera considéré comme manquant. Néanmoins, l'EORTC

recommande d'étudier le profil des données manquantes et d'éventuelles analyses de sensibilité peuvent être appliquées afin d'évaluer l'impact des données manquantes sur les données de QDV [2]. Pour certains questionnaires, il est possible d'avoir des règles plus strictes quant au pourcentage de données manquantes toléré. A titre d'exemple, aucune donnée manquante n'est tolérée pour pouvoir calculer les scores issus du questionnaire HADS mesurant l'état d'anxiété et de dépression d'après les auteurs [3].

4. Interprétation du score

Selon l'algorithme appliqué pour calculer le score, un score élevé pourra correspondre à un haut niveau de QDV ou inversement à un bas niveau de QDV. Le choix de l'EORTC a été de considérer qu'un score élevé correspond à un haut niveau pour la dimension évaluée. Ainsi, pour les dimensions évaluant un état fonctionnel tel que la fonction physique, un haut score correspondra à un haut niveau fonctionnel, et donc un bon niveau de QDV. Inversement, pour une dimension évaluant un état symptomatique tel que la fatigue, un score élevé correspondra à un haut niveau symptomatique et donc à un faible niveau de QDV [2]. Pour les questionnaires du groupe FACT, un score élevé quel que soit l'échelle considérée et la dimension évaluée correspondra à un haut niveau de QDV [4]. Les utilisateurs doivent donc bien être au fait du sens du score évalué afin d'éviter toute erreur d'interprétation.

5. Outils statistiques disponibles pour l'estimation de scores

Différents outils ont été mis à la disposition des chercheurs pour l'estimation des scores de questionnaires sous les principaux logiciels statistiques. Par exemple, la plateforme SeleQt met à disposition le manuel de scoring ainsi que des macro-programmes sous le logiciel SAS pour l'estimation de scores d'un certain nombre de questionnaires de santé perçue (<https://seleqt.univ-lorraine.fr/>). Dans son manuel de scoring concernant le QLQ-C30 et ces premiers modules complémentaires, l'EORTC met également à disposition des macro-programmes SAS pour l'estimation des scores [2]. Un package existe également sous le logiciel R pour l'estimation des scores de la plupart des questionnaires de l'EORTC [5]. Un programme STATA a également été développé pour l'estimation des scores du questionnaire QLQ-C30 pour les utilisateurs de ce logiciel [6].

6. Conclusion

En conclusion, le calcul des scores est une étape cruciale avant de mener l'analyse des données de QDV. Cette étape doit être réalisée avec rigueur en suivant les recommandations des développeurs des questionnaires. Il est malgré cela possible que des auteurs proposent de nouveaux scores, par exemple des scores totaux résumant l'ensemble du questionnaire. Ce fut le cas du questionnaire QLQ-C30 où un score total a récemment été proposé. Dans ce cas des analyses complémentaires sont proposées afin d'étudier les propriétés de ce score et son intérêt potentiel [7,8].

Amélie ANOTA

Biostatisticienne, DRCI, Centre Léon Bérard, Lyon

7. Références

1. Pusic, A. L., Klassen, A. F., Scott, A. M., Klok, J. A., Cordeiro, P. G., & Cano, S. J. (2009). Development of a new patient-reported outcome measure for breast surgery: the BREAST-Q. *Plastic and reconstructive surgery*, 124(2), 345-353.
2. Fayers, P., Aaronson, N. K., Bjordal, K., & Sullivan, M. (1995). EORTC QLQ-C30 scoring manual. European Organisation for Research and Treatment of Cancer.
3. Zigmond, A. S., & Snaith, R. P. (1983). The hospital anxiety and depression scale. *Acta psychiatrica scandinavica*, 67(6), 361-370.
4. Cella, D. F., Tulsky, D. S., Gray, G., Sarafian, B., Linn, E., Bonomi, A., ... & Kaplan, E. (1993). The Functional Assessment of Cancer Therapy scale: development and validation of the general measure. *J Clin Oncol*, 11(3), 570-579.
5. Aota, A., Savina, M., Bascoul-Mollevi, C., & Bonnetain, F. (2017). Qolr: An r package for the longitudinal analysis of health-related quality of life in oncology. *Journal of Statistical Software*, 77(1), 1-30.
6. Bascoul-Mollevi, C., Castan, F., Azria, D., & Gourgou-Bourgade, S. (2015). EORTC QLQ-C30 descriptive analysis with the qlqc30 command. *The Stata Journal*, 15(4), 1060-1074.
7. Giesinger, J. M., Kieffer, J. M., Fayers, P. M., Groenvold, M., Petersen, M. A., Scott, N. W., ... & EORTC Quality of Life Group. (2016). Replication and validation of higher order models demonstrated that a summary score for the EORTC QLQ-C30 is robust. *Journal of clinical epidemiology*, 69, 79-88
8. Husson, O., de Rooij, B. H., Kieffer, J., Oerlemans, S., Mols, F., Aaronson, N. K., ... & van de Poll-Franse, L. V. (2020). The EORTC QLQ-C30 Summary score as prognostic factor for survival of patients with cancer in the “real-world”: Results from the population-based PROFILES registry. *The oncologist*, 25(4), e722.
9. Smets E.M.A., Garssen B., Bonke B. & De Haes C.J.M. (1995) The multidimensional Fatigue Inventory (MFI) Psychometric qualities of an instrument to assess fatigue. *Journal of Psychosom Research* 39, 315-325.

8. Points clés

POINTS CLES

- L'analyse des données de QDV se fait généralement en se basant sur des scores calculés à partir des réponses aux items.
- Chaque questionnaire validé doit proposer un manuel de scoring mis à disposition des utilisateurs indiquant la méthode de calcul des scores.
- Selon les questionnaires un pourcentage de données manquantes peut être toléré pour déterminer ces scores, correspondant à au plus 50% d'items manquants.
- Le calcul des scores doit être fait selon les recommandations des développeurs des questionnaires, précisées dans le manuel de scoring si celui-ci est disponible.