

Chapitre 6 : Comment interpréter un score de qualité de vie ?

L'interprétation d'une valeur d'un score (mesure unique) ou d'une différence de deux scores de QDV (mesure répétée ou mesure dans deux groupes à un instant T) doit être non seulement statistique mais également clinique. La significativité statistique est nécessaire pour l'interprétation des résultats de QDV, mais non suffisante pour attester d'un bénéfice ou d'un effet délétère pour le patient dans sa vie quotidienne.

1. Mesure unique (transversale)

Lorsque la mesure est unique correspondant à une réalité instantanée sans comparatif antérieur, la question du sens clinique à lui donner est posée. Ce sens peut être donné selon la nature de la dimension concernée :

a. La mesure porte sur la présence ou non d'un symptôme

Le plus souvent caractérisé par une dimension à item unique, il permet de dire si ce symptôme est présent ou probable chez le patient. C'est le cas par exemple de la dimension « Insomnie » du QLQ-C30 qui permet au patient de répondre par « pas du tout », « un peu », « assez », « beaucoup », à la question suivante : au cours de la semaine passée avez-vous eu des difficultés pour dormir ?

Parfois, cette dimension symptôme peut comporter plusieurs items avec un calcul de score et une définition des seuils d'interprétation de ce score. Par exemple, pour le questionnaire HADS (Hospital Anxiety and Depression Scale), sur un score brut allant de 0 à 21 pour chaque dimension anxiété ou dépression, l'interprétation clinique est la suivante : Pas d'état anxieux ou dépressif (score ≤ 7), anxiété ou dépression probable ($8 \leq \text{score} \leq 10$) et anxiété ou dépression certaine (score ≥ 11).

b. La mesure correspond à un score

Dans ce cas, même si un sens global peut être donné à un score (ex : plus le score est élevé, meilleur est la QDV), l'interprétation clinique reste problématique. L'interprétation clinique peut se faire de deux manières possibles sur la base d'utilisation de valeurs de références :

Valeurs d'une population de référence

La valeur du score obtenue peut-être comparée à une valeur de référence calculée sur la population générale (lorsqu'il s'agit d'un questionnaire générique) ou sur une population spécifique (questionnaire spécifique). Par exemple, il est possible de comparer le score de QDV mesurée par le DUKE aux valeurs de références mesurées en population générale française [1].

Seuils d'acceptabilité ou d'importance clinique

Lorsqu'il n'existe pas de valeurs de références ou qu'il n'est pas pertinent de faire une comparaison par rapport à une population de référence, l'interprétation clinique d'un score peut être basée sur l'utilisation du seuil d'acceptabilité symptomatique appelé en anglais PASS (Patient Acceptable symptomatic State) [2] ou du seuil d'importance clinique appelé en anglais TCIs (Threshold for Clinical Importance) [3]. Le PASS est défini à l'échelle d'un groupe de patients et le TCIs à l'échelle d'un patient unique. Le PASS et le TCIs sont définis comme étant le niveau de symptôme en deçà duquel les patients se considèrent comme « allant bien ».

L'estimation du PASS peut se faire par le calcul de la moyenne des scores chez les patients classés comme ayant un niveau de QDV acceptable. Le PASS peut être également déterminé en utilisant un modèle de régression linéaire avec le score comme variable à expliquer (Y) et la réponse comme variable explicative à 2 modalités (par exemple : oui/non). Le coefficient de régression correspondant à la comparaison de la modalité « oui » à la modalité « non » est défini comme étant le PASS.

L'estimation du TCIs se fait par l'analyse de la courbe ROC qui permet de choisir un seuil de score optimal ayant le meilleur compromis entre sensibilité et spécificité afin de minimiser le nombre de patients mal classés.

2. Mesures répétées (longitudinales)

Si la significativité statistique s'interprète comme une exclusion de l'absence de différence entre deux scores de QDV (aux fluctuations d'échantillonnage près), cela ne justifie pas pour autant que cette différence numérique entre deux scores sous-tende une réalité clinique - c'est-à-dire une différence d'expérience de QDV chez un patient. En effet, cette différence statistique est liée à la taille d'échantillon. Pour que l'écart des valeurs numériques entre deux estimations de score de QDV reflète une différence perceptible, il est nécessaire de déterminer la différence de score la plus faible pour laquelle le patient perçoit un changement.

Cette différence est appelée DMI (Différence Minimale Importante, ou MID en anglais - Minimum Important Difference). Celle-ci est définie comme « *la plus petite différence de score dans le domaine d'intérêt que le patient perçoit comme bénéfique ou délétère et qui conduirait, en l'absence d'effet secondaire notable et de coût excessif, à un changement dans la prise en charge du patient* » [4]. Le MID est défini à l'échelle d'un groupe de patients.

Cette différence peut également être exprimée à l'échelle d'un patient, dans ce cas elle est appelée la « définition du répondeur » ou « responder definition (RD) » [5]. Cette différence n'est pas une valeur fixe pour un instrument de mesure de la QDV donné. Sa variabilité dépend du score de QDV initial dans l'étude, du contexte de l'étude, et de la méthode employée pour sa détermination [6]. Bien que les développeurs de questionnaires de QDV indiquent parfois une valeur de MID ou du RD pour orienter les futurs utilisateurs, il ne s'agit pas du cas le plus fréquent. Pour certains questionnaires de QDV, il existe des recommandations générales concernant l'interprétation des scores [7, 8]. Néanmoins celles-ci ne sont pas applicables à toutes les situations, et il est souvent souhaitable de savoir comment elles ont été élaborées.

Les deux principales méthodes utilisées pour la détermination d'une MID ou RD sont les méthodes basées sur l'ancrage et celles basées sur la distribution.

a. Méthodes basées sur l'ancrage

Un ancrage est défini comme une mesure de changement facilitant l'interprétation du résultat obtenu avec le questionnaire de QDV. Différents types d'ancrages existent : il peut s'agir de mesures observées, ou de mesures perçues (ces dernières étant les plus utilisées en QDV).

Lorsque la mesure d'ancrage est perçue, la perspective adoptée est variable et doit être identifiée : est-ce celle du patient ? de l'entourage familial ? du clinicien ? de la société ? Cette perspective de l'ancrage doit être appréhendée de manière critique au regard du contexte de l'étude dans laquelle on souhaite utiliser la MID/RD ; dans le champ de la QDV, il semble logique de choisir une MID/RD dont l'ancrage dépend de la perspective du patient. Trois étapes sont nécessaires à la détermination d'une MID/RD basées sur l'ancrage [9]:

Choix de l'ancrage

Malgré qu'il n'y ait actuellement pas de recommandation pour choisir un ancrage approprié, il existe un consensus sur les critères définissant un bon ancrage. Celui-ci doit être :

- Facilement interprétable
- Fréquemment utilisé
- Au moins partiellement corrélé avec la mesure du questionnaire (ici, la QDV)

Différents ancres sont susceptibles de produire différentes estimations de MID pour le même instrument de mesure [10].

Détermination d'un seuil définissant un changement important sur l'ancrage

D'une étude à l'autre, le seuil choisi sur l'échelle d'ancrage pour la détermination d'une MID/RD peut varier. Le choix du seuil de l'ancrage est d'importance majeure et ne doit pas être réalisé arbitrairement mais correspondre à un véritable changement clinique.

Choix de l'approche statistique pour estimer la MID/RD

L'estimation du MID peut se faire par le calcul de la moyenne de changement de score chez les patients classés comme ayant amélioré ou détérioré leur QDV en fonction de l'ancre. Le MID peut être également déterminé en utilisant un modèle de régression linéaire avec le changement de score comme variable à expliquer (Y) et les niveaux de l'ancre comme variable explicative à 3 modalités (x : Amélioration, stable et détérioration). Le coefficient de régression correspondant à la comparaison de la modalité amélioration ou détérioration versus stable est défini comme étant le MID.

L'estimation du RD se fait par l'analyse de la courbe ROC qui permet de choisir un seuil de changement optimal ayant le meilleur compromis entre sensibilité et spécificité afin de minimiser le nombre de patients mal classés.

b. Méthodes basées sur la distribution

Elles sont fondées sur des indicateurs de variabilité statistique dans l'échantillon pour estimer une MID. Leur utilisation seule n'est pas recommandée, car elles ont l'inconvénient de ne pas prendre en compte le point de vue du patient (elles ne sont pas liées à des résultats cliniques observables, donne les mêmes seuils pour la détérioration et l'amélioration et dépendent de la composition de l'échantillon analysé). Elles peuvent servir en complément d'une méthode basée sur l'ancrage car elles renseignent sur le fait qu'une différence de scores excède ou non l'erreur de mesure standard, ce qui signifie qu'elle représente un signal d'amplitude supérieure au bruit (erreur) de la mesure.

Une revue de la littérature a montré que la différence minimale interprétable des questionnaires de QDV dans le cas des maladies chroniques avoisinait le plus souvent 0,5 écart-type [11].

3. Conclusion

L'évaluation de l'adéquation entre l'objectif de la mesure de la QDV, la méthode de détermination de la significativité clinique (PASS, TCIs, MID, RD) et l'application envisagée pour cette valeur est essentielle. La connaissance de ces indicateurs est d'une importance

capitale notamment pour les cliniciens qui se posent souvent la question de la pertinence clinique des scores de QDV qui sont calculés.

Caroline MASUREL

Interne de santé publique, CIC-EC, CHRU de Nancy

Abdou OMOROU

Médecin épidémiologiste, CIC-EC, CHRU de Nancy

4. Références

1. Baumann C, Erpelding M-L, Perret-Guillaume C, Gautier A, Régat S, Collin J-F, et al. Health-related quality of life in French adolescents and adults: norms for the DUKE Health Profile. *BMC Public Health*. 2011;11:401.
2. Tubach F, Ravaud P, Baron G, Falissard B, Logeart I, Bellamy N, et al. Evaluation of clinically relevant states in patient reported outcomes in knee and hip osteoarthritis: the patient acceptable symptom state. *Ann Rheum Dis*. 2005;64:34-7.
3. Giesinger JM, Loth FLC, Aaronson NK, Arraras JI, Caocci G, Efficace F, et al. Thresholds for clinical importance were established to improve interpretation of the EORTC QLQ-C30 in clinical practice and research. *J Clin Epidemiol*. 2020;118:1-8.
4. Jaeschke R, Singer J, Guyatt GH. Measurement of health status: Ascertaining the minimal clinically important difference. *Control Clin Trials*. 1989;10:407-15.
5. Sully K, Trigg A, Bonner N, Moreno-Koehler A, Trennery C, Shah N, et al. Estimation of minimally important differences and responder definitions for EORTC QLQ-MY20 scores in multiple myeloma patients. *Eur J Haematol*. 2019;103:500-9.
6. Beaton DE, Boers M, Wells GA. Many faces of the minimal clinically important difference (MCID): a literature review and directions for future research. *Curr Opin Rheumatol*. 2002;14:109-114.
7. Cocks K, King MT, Velikova G, de Castro G, Martyn St-James M, Fayers PM, et al. Evidence-based guidelines for interpreting change scores for the European Organisation for the Research and Treatment of Cancer Quality of Life Questionnaire Core 30. *Eur J Cancer*. 2012;48:1713-21.
8. Quinten C, Kenis C, Decoster L, Debruyne PR, De Groof I, Focan C, et al. Determining clinically important differences in health-related quality of life in older patients with cancer undergoing chemotherapy or surgery. *Qual Life Res*. 2019;28:663.
9. Engel L, Beaton DE, Touma Z. Minimal Clinically Important Difference: A Review of Outcome Measure Score Interpretation. *Rheum Dis Clin N Am*. 2018;44:177-88.
10. Jayadevappa R, Cook R, Chhatre S. Minimal important difference to infer changes in health-related quality of life—a systematic review. *J Clin Epidemiol*. 2017;89:188-98.
11. Norman GR, Sloan JA, Wyrwich KW. Interpretation of Changes in Health-related Quality of Life: The Remarkable Universality of Half a Standard Deviation. *Med Care*. 2003;41:582-592.

5. Points clés

POINTS CLES

En plus de l'interprétation statistique d'une valeur ou d'un changement de score de QDV, le sens clinique reste un point clé dans l'utilisation de cet indicateur.

Le sens clinique dépend du caractère unique (transversal) ou répété (longitudinal) de la mesure.

- **En cas de mesure unique, l'interprétation peut se faire :**

- A l'échelle du patient par :

La présence d'un symptôme

Le seuil d'importance clinique (TCIs : Threshold of Clinical Importance)

- A l'échelle du groupe de patient par :

La présence d'un symptôme

Le seuil d'acceptabilité symptomatique (PASS : Patient Acceptable Symptomatic State)

- **En cas de mesure répétée, l'interprétation peut se faire :**

- A l'échelle du patient par :

La définition du répondeur (RD : Responder Definition)

- A l'échelle du groupe de patients par :

La différence minimale importante (MID : Minimal Important Difference)