# « Les Indicateurs de Qualité de Vie : Processus de Mesure et Validation »

*J. REBOUL-MARTY*[(1)], *Robert LAUNOIS* [(1-2)]

Cardioscopies, juin 1995, n° 33 :635-637

(1) UFR SMBH – Université de Paris XIII – 74 rue Marcel Cachin – 93017 BOBIGNY Cedex - Email : launois_ireme@smbh.univ-paris13.fr - Site web : http://smbh7.smbh.univ-paris13.fr

(2) REES France - 28, rue d'Assas - 75 006 Paris – Email : reesfrance@wanadoo.fr - Site Internet : http://www.rees-france.com

***L'analyse et les méthodes statistiques appliquées à la recherche clinique ont pris une place de plus en plus importante en médecine. Les études et les essais cliniques sont actuellement conçus avec une méthodologie très rigoureuse, seule garantie de la validité des résultats et de leur valeur par rapport à la population générale. Un des rouages important dans une étude consiste à établir des critères de jugement, par mis lesquels on retrouve les indicateurs de qualité de vie. Dans cette mise au point, les auteurs exposent les différentes modalités de ces mesures ainsi que leurs attributs : pertinence, acceptabilité, fiabilité, sensibilité, validité. Cardioscopies, 1995, 33, p. 635-637.***

Mesurer consiste à assigner des positions sur des échelles à des observations en respectant certaines règles. Mesurer, c'est affecter un score à un objet ; ce score indique avec précision la caractéristique de l'objet. Le processus de mesure requiert donc l'utilisation d'une échelle (ou d'un instrument) de mesure. Plusieurs niveaux d'échelle existent, les degrés de précision et les analyses statistiques différents en fonction du niveau.
Enfin, une fois construites pour être scientifiquement reconnues, ces échelles de mesure (ou indicateurs) doivent réunir 5 conditions : pertinence, acceptabilité, fiabilité, sensibilité et validité. Ces conditions seront détaillées dans la deuxième partie.

## LES NIVEAUX D'ECHELLES

- *La mesure de niveau nominal*

Il s'agit plus d'un processus de classification que d'une réelle mesure. Les nombres sont utilisés pour ranger l'objet dans une classe particulière et ne fournissent pas d'information sur le(s) objet(s) de cette classe. Ces mesures ne sont pas ordonnées. Dans les données médicales, une information de type sexe du patient (1 : masculin ; 2 : féminin) est de niveau nominal.
Il y a une simple relation de correspondance de un à un associant des numéros à un ensemble de classes mutuellement exclusives. Les numéros qui désignent les classes peuvent être modifiés, l'information ne changera pas. Par exemple, le sexe du patient peut être codé M, F. Les statistiques autorisées sont les fréquences, le mode et le chi2[1].

- *La mesure de niveau ordinal*

Par rapport au niveau de mesure précédent, celui-ci rajoute une information supplémentaire : il permet d'ordonner des objets qui ont en commun une même caractéristique mais à des degrés différents. Les objets sont donc classés dans des catégories qui sont hiérarchisées.
Par exemple, la question : « dans quelle mesure vos problèmes de jambes vous ont-ils gênés pour effectuer certaines tâches domestiques ?

1) aucune gêne,
2) un peu gêné,
3) modérément gêné,
4) très gêné,
5) impossible à faire.

On voit que le niveau 1 est inférieur au niveau 2, qui est lui-même inférieur au niveau 3 … etc. On peut donc ordonner les catégories, mais on ne peut spécifier la distance qui sépare deux catégories, ces distances étant inégales. En effet, quelle est la différence de pénibilité entre « un peu » et « modérément », de même, quelle distance y a-t-il entre « modérément gêné » et « très gêné » ? De la même façon, on ne peut dire de combien de fois « un peu gêné » est plus pénible que « aucune gêne », ni de combien de fois le niveau « modérément » est plus pénible que le niveau « un peu ».
Toutes les transformations qui respectent l'ordre dans lequel se trouvent les catégories sont autorisées ; l'information ne changera pas. Par exemple, les codes 1, 2, 3, 4, 5 pourront être modifiés en 4, 7, 9, 12 et 15 si l'ordre initial est respecté.
Les statistiques permises sont la médiane et les tests non paramétriques[1] (comparaison des moyennes basées sur les rangs des variables et non sur leur valeur absolue, la corrélation basée sur les rangs).

- *La mesure de niveau d'intervalles*

Ce niveau permet de ranger et d'ordonner les objets, mais aussi de spécifier avec exactitude la distance qui sépare les nombres représentant les catégories. Le niveau d'intervalles se caractérise par une unité de mesure commune et constante qui associe un nombre réel à toutes les paires d'objets ordonnés. Avec ce type d'échelle, l'origine est arbitraire et le support entre 2 intervalles quelconques est indépendant de l'unité de mesure et de l'origine de l'échelle. L'exemple le plus courant est l'échelle de la température. L'origine de l'échelle de Celsius est à 0° (point de congélation), ce qui correspond à 32° sur l'échelle Fahrenheit. Le rapport entre 2 intervalles quelconques de l'échelle Celsius est identique au rapport entre 2 intervalles équivalents de l'échelle Fahrenheit. De plus, ce rapport est indépendant de l'origine puisque celui-ci diffère en fonction de l'échelle.
Par exemple, le rapport (20°C – 10°C)/(10°C – 0°C) = 1 ; l'équivalent Fahrenheit est (68°F – 50°)/(50°F – 32°F) = 1.
Par contre, comme l'origine est arbitraire, on ne pourra jamais dire que 40°F est 2 fois plus chaud que 20°F. En effet, la conversion en degrés Celsius (C = 5/9*(F-32)) donne les températures de 4,4°C et –6,6°C, qui est loin du rapport de 2.
En médecine, certains indicateurs de qualité de vie utilisent ce type d'échelle. Comme les catégories de l'échelle sont présentées à intervalles égaux numériquement et physiquement, on fait l'hypothèse que les réponses sont de niveau d'intervalles.
Par exemple, la question : « dans quelle mesure vos problèmes de jambes vous ont-ils gênés pour effectuer certaines tâches domestiques ? »

1) aucune gêne,
2) un peu gêné,
3) modérément gêné,
4) très gêné,
5) impossible à faire.

Si l'on reprend l'exemple ci-dessus, on postule que la différence entre « aucune gêne » et « un peu gêné » est la même que celle qui existe entre « modérément » et « très gêné ». Cette hypothèse de travail est très forte et pas forcément démontrée dans les indicateurs de qualité de vie publiés. Mais l'utilisation de cette hypothèse permet d'utiliser des analyses multivariées telles l'analyse factorielle ou l'analyse en composantes principales lors de la construction de l'échelle. Les transformations permises sont celles qui respectent l'ordre, mais aussi les écarts relatifs entre objets. On peut donc multiplier tous les nombres par une constante et leur rajouter une constante (l'origine étant arbitraire, on peut la changer).

Les statistiques permises sont la moyenne arithmétique, l'écart-type et les tests paramétriques : le test t de Student ou l'analyse de variance pour la comparaison de moyennes ou le calcul du coefficient de corrélation de Pearson[1].

- *La mesure de niveau de rapports*

Ce niveau a toutes les caractéristiques du niveau d'intervalles avec en plus une origine absolue. Le rapport entre 2 points de l'échelle est indépendant de l'unité de mesure ; donc, on peut dire que 4 est 2 fois supérieur à 2. Les exemples sont nombreux : la longueur, le poids, etc … Le rapport des longueurs de 2 objets sera le même, quelle que soit l'unité de mesure (pouces, centimètres).
Ce niveau permet une relation d'ordre, la possibilité de comparer les écarts et les rapports entre 2 niveaux de l'échelle.
On peut transformer l'échelle en multipliant les nombres associés à cette échelle par une constante positive sans la modifier.
Les statistiques permises sont identiques à celles autorisées pour le niveau d'intervalles.
Les mesures utilisant les niveaux d'intervalles ou de rapports sont également appelées des mesures cardinales.

## VALIDATION D'UN INDICATEUR

Pour cette validation, 5 caractéristiques sont exigées[2,3,4,5,6].

- *La crédibilité ou « face validity »*

Elle dépend de la formulation des questions. En effet, celles-ci doivent être comprises sans aucune ambiguïté par les sujets auxquels les questions s'adressent. Elle est supposée être acquise si la formulation des items a été faite à partir des verbatims des patients.

- *La pertinence/exhaustivité ou « content validity »*

Pour qu'un instrument soit pertinent, deux conditions doivent être réunies : l'ensemble des dimensions ou concepts explorés doivent être intégrés (exhaustivité) dans la mesure et la représentativité des énoncés doit être vérifiée. Les concepts peuvent être dégagés soit par consultation de la littérature et/ou des experts, soit en recueillant les plaintes directement auprès des malades. Ces interviews permettent de dégager un nombre important d'énoncés. Ces derniers ne pouvant être tous retenus dans l'échelle de mesure, il convient d'en éliminer et de ne garder que les énoncés les plus représentatifs et non redondants de cette banque de données. Les indicateurs spécifiques ont souvent utilisé la deuxième méthode pour faire émerger les plaintes et les dimensions d'atteintes. Ces dimensions peuvent être par exemple le retentissement de la pathologie sur les fonctions physique, psychique ou sociale. La dimension physique peut elle-même être décomposée en mobilité, soins personnels, etc …

- *La fiabilité*

L'indicateur doit être fiable, c'est-à-dire rendre compte avec précision du phénomène mesuré. Les méthodes les plus utilisées pour apprécier cette qualité sont les suivantes :

- La reproductibilité du questionnaire

Elle est vérifiée par le « test-retest » sur des patients dont l'état clinique est stable. On suppose que le phénomène mesuré chez le sujet est stable et ne varie pas entre les différentes mesures effectuées. Les mêmes mesures sont répétées à deux reprises chez les mêmes sujets et dans les mêmes conditions. En général, un intervalle de 15 jours entre les « passages » est retenu. Quinze jours semblent être un délai suffisant pour qu'il n'y ait pas d'effet d'apprentissage pouvant biaiser les réponses (les patients pourraient se souvenir de leurs réponses antérieures) et pour que l'état du sujet ne soit pas susceptible de changer. On évalue la fiabilité par le cœfficient de corrélation de Pearson ; le cœfficient varie de – 1 à + 1 ; sous certaines conditions, un cœfficient proche de 1 en valeur absolue est signe d'une forte liaison entre les 2 variables.

- La concordance entre juges

On peut également mesurer la fiabilité par un test de concordance entre les deux mesures. Ce test peut être le test du Kappa[7] si on travaille sur des données qualitatives ou le cœfficient de corrélation intra-classe avec des données quantitatives. Le test de concordance est surtout utilisé si l'on cherche à apprécier l'accord entre les jugements provenant de 2 personnes différentes. Une valeur proche de 1 pour les tests de concordance est signe d'une bonne fiabilité.

- La cohérence interne

A l'intérieur d'une même dimension, les différents items qui la caractérisent doivent être homogènes car ils représentent le même concept mais avec des formulations différentes. Cette homogénéité ou cohérence interne est testée par le cœfficient Alpha de Cronbach. Il prend toutes les valeurs échelonnées de 0 à 1. Un cœfficient proche de 1 indique une bonne cohérence interne. En général, un cœfficient supérieur à 0,70 est jugé acceptable pour les mesures psychométriques.

- *La validité du construit ou de structure « construct validity »*

Une mesure valide doit bien mesurer ce qu'elle est censée mesurer :

- la validité de critère ou « criterion validity » : elle est appréciée par la confrontation de l'échelle développée à un étalon de référence (ou critère). Elle est alors jugée sur le degré de corrélation existant entre l'échelle et la référence. Si la mesure proposée correspond à un critère mesuré simultanément (par exemple, une tension artérielle prise au manomètre correspondant à la pression intra-artérielle mesurée au même moment), on parlera alors de validité concurrente ou concomitante. Si la mesure proposée prédit un critère futur (par exemple, les tests de QI permettraient de prédire l'intelligence), on parlera alors de validité prédictive. Par définition, le critère ou la référence doit être « supérieur », fournir une mesure plus précise du phénomène que l'on cherche à quantifier.

- Malheureusement, dans les échelles de qualité de vie, la référence absolue n'existe pas et la validation du construit consiste plutôt à essayer de rassembler tous les éléments pouvant apporter des preuves empiriques de la « valeur » de cette mesure. Aussi, actuellement, au lieu de chercher une corrélation parfaite avec un critère putatif, on cherchera seulement à démontrer que la mesure proposée va dans le même sens qu'une autre mesure déjà validée. Ainsi, il est admis qu'un indicateur soit validé s'il est bien corrélé à des échelles de mesures cliniques ou à des indicateurs qui explorent une dimension semblable à celle prise en compte par l'indicateur en cours d'élaboration. On parlera alors de validité convergente ou « convergent validity ». Par exemple, la dimension « soins personnels » de l'indicateur proposé devrait être corrélée à certaine(s) dimension(s) d'un indicateur de dépendance.
- Selon la théorie psychométrique, il y a également validation du « construit » quand l'analyse factorielle, qui permet de dégager les principaux facteurs ou dimensions observés, retrouve les dimensions conceptuelles qui sous-tendent le domaine étudié. Cette structure factorielle doit être stable lors des différentes analyses. Cette condition ne peut être exigée que pour les indicateurs psychométriques construits sur la base de l'analyse factorielle. Les autres, notamment les indicateurs d'utilité, ne peuvent vérifier cette condition puisque leur élaboration repose sur d'autres méthodes[8].

- *La sensibilité*

Un indicateur est sensible s'il est apte à détecter des changements peu importants de la qualité de vie d'un sujet ; bien sûr, ces changements minimes n'ont de sens que s'ils sont cliniquement importants. Guyatt[9] va même plus loin : chez des sujets dont l'état clinique est stable, le score doit rester inchangé ; par contre chez des sujets dont l'état se détériore (ou s'améliore), le score doit varier. Cette qualité est importante pour les indicateurs destinés à être utilisés dans des essais thérapeutiques. En effet, un indicateur peu sensible risquerait de ne pas montrer de différence entre deux traitements parce qu'il n'a pas réussi à capter de manière fine les modifications de l'état clinique en fonction de ces deux traitements.

En médecine, différents indicateurs existent. Ils peuvent être des indicateurs cliniques (par exemple, l'échelle Apache permettant de pronostiquer l'état du patient), des indicateurs de qualité de vie (par exemple, le Nottingham Health Profile[10], ou des indicateurs d'utilité permettant de réaliser des études coût-utilité[11].

## REFERENCES

[1] Schwartz D. *Méthodes statistiques à l'usage des médecins et des biologistes*. Flammarion. Médecine Sciences. 3ème édition. France.
[2] Nunnaly J.C. *Psychométric theory. 2nd edition.* Mc Graw Hill Book Company. New York, 1978.
[3] Kaplan M., Bush J.W., Berry C. *Health status : type of validity and the index of well-being*. Health Services Research, 1979, 16, 64-73.
[4] Launois R. *La qualité de vie : panorama et mise en perspective.* In : Décision thérapeutique et qualité de vie. Launois R., *Régnier F. Eds Collection de l'Association Française pour la Recherche Thérapeutique. Editions John Libbey Eurotext. 1992, Paris.
[5] Moret L., Chwalow J., Baudouin-Balleur C. *Evaluer la qualité de vie : construction d'une échelle.* Rev. Epidém. Et Santé Publ., 1993, 41, 65-71.

[6] Launois R., Reboul-Marty J., Henry-Launois B. *Construction et validation d'un indicateur spécifique de qualité de vie : le cas de l'insuffisance veineuse chronique des membres inférieurs.* Journal d'Economie Médicale, 1994.
[7] Fleiss J.L. *Statistical methods for rates and proportions*. New York, Wiley, 1981.
[8] Launois R. *La prise en compte des préférences des patients dans les choix de santé individuels et collectifs.* Rev. Epidém. Et Santé Publ. ; 1994.
[9] Guyatt G., Walter S., Norman G. *Measuring change overtime ; assessing the usefulness of evaluative instruments*. J. Chron. Dis., 1987, 40, 171-178.
[10] Hunt S.M., McEwen J. *The development of a subjective health indicator*. Sociology of Health and Illness, 1980, 2, 231-246.
[11] Rosser R., Kind P. *A scale of valuations of states of illness : is there a social consensus ?* International Journal of Epidemiology, 1978, 7, 4, 347-358.