

La prise en compte des préférences des patients dans les choix de santé individuels et collectifs

Integrating individual preferences into therapeutic and collective choices

R. LAUNOIS

Université de Paris XIII, Département de Santé Publique, Centre d'Evaluation Médico-Economique, Faculté de Médecine de Bobigny, 74 rue Marcel Cachin, F 93012 Bobigny Cedex.

The objectives of contemporary medicine are basically to attenuate the consequences of chronic diseases and to improve the quality of life of the patient. All medical disciplines tend now to elicit therapeutic protocols directly from the general appreciation of the patients. The community is trying to obtain a transnosographic indicator which would allow the effects of its strategic choices between different areas of the health care system to be measured, taking into account the implications on the quality of life of the population. Individual preferences are now at the center of the decision-making process. Different experimental methods are being used to reveal them. Psychometricians use direct observation of a subject's reactions in a particular pathological situation : patients are asked to arrange the intensity of the impacts on bipolar numeric scales, but an actual metrical measure is not yet available. Traditionally, economists believe that in a market, only the consumer's choices enable us to estimate his level of satisfaction. In the health care field, where market mechanisms are not fully operational, they tend to extract the patients preferences through forced choices between hypothetical health states. A real metrical measure is thus obtained. The objective of this article is to exhibit these two methods : psychometric and economic, and to show how they have been implemented in French field research to obtain quality adjusted life years indicators : QALY's.

Quality of life. Utility approach. Psychometric approach.

Les objectifs de la médecine contemporaine consistent de plus en plus souvent à atténuer les conséquences des maladies chroniques, et à améliorer la qualité de vie des malades. Toutes les disciplines médicales cherchent aujourd'hui à départager les protocoles thérapeutiques à partir de l'appréciation des patients eux-mêmes. La collectivité cherche à disposer d'un indicateur transnosographique qui permette de mesurer les effets de ses choix stratégiques entre les secteurs du système de soins, tout en tenant compte des répercussions sur le confort de vie des populations. Les préférences individuelles sont désormais au cœur de la prise de décision. Différentes méthodes expérimentales sont utilisées pour les dégager. Les psychométriciens ont recours à l'observation directe des réactions du patient face à une situation pathologique concrète, permettant d'ordonner l'intensité de ces atteintes sur des échelles numériques bipolaires, mais ne constituant pas une véritable métrique. Les économistes,

traditionnellement, pensent que sur le marché, seuls les choix du consommateur permettent d'appréhender son niveau de satisfaction. Dans le domaine de la santé, où les règles du marché ne s'appliquent pas totalement, ils sont amenés à extraire artificiellement les préférences du patient, en le confrontant à des situations de santé hypothétiques entre lesquelles il doit choisir. On obtient alors une véritable mesure de grandeur. Cet article a pour objet de présenter ces deux démarches, psychométrique et économique, et montre comment elles ont été utilisées dans les travaux français pour obtenir un indicateur de performance pondérant la longévité par la qualité de vie : l'AVAQ (Années de vie ajustées sur la qualité).

Qualité de vie. Approche utilité. Approche psychométrique.

Il sera difficile de mettre en évidence l'exacte contribution du système de soins à l'amélioration de l'état de santé de la population aussi longtemps que ses indicateurs de performance seront définis en termes d'espérance de vie et de mortalité, alors qu'un des principaux objectifs de la médecine contemporaine est d'atténuer les conséquences quotidiennes de la maladie en améliorant la qualité de vie.

LE CONCEPT DE QUALITÉ DE VIE

La première étape de toute étude sur la qualité de vie consiste à définir *l'univers* qui fera l'objet de l'analyse. Le terme qualité de vie désigne les conséquences physiques, psychologiques et sociales d'une maladie ou d'un traitement, au delà de leurs répercussions strictement cliniques. Bien entendu, la qualité de vie dépend de multiples facteurs comme le revenu, les conditions de travail et de vie, dont certains n'ont rien à voir avec la santé. Dans l'évaluation des effets des stratégies thérapeutiques, les économistes négligent les facteurs sur lesquels les traitements ne peuvent avoir d'influence à court terme, et s'en tiennent à l'analyse de la variation de la qualité de vie due à une modification de l'état de santé (« *Health-Related Quality of Life* »).

Une fois l'univers défini, il faut le *catégoriser*, c'est-à-dire définir les dimensions qui seront quantifiées. La catégorisation est une étape difficile. Nous la définirons a priori en prenant appui sur la citation de l'OMS la plus fréquemment utilisée : *la santé n'est pas seulement l'absence de maladie ou d'infirmité mais un état total de bien-être physique, mental et social* ». La dimension bien-être, synthétisant l'ensemble des précédentes, sera

considérée comme une dimension à part entière. Une bonne qualité de vie se caractérise donc par un sentiment de bien-être, une émotivité équilibrée, une intégration sociale satisfaisante et un bon état physique.

Ces quatre dimensions ne sont jusqu'à présent que des concepts. Afin de pouvoir en mesurer le contenu, il faut avoir recours à un certain nombre de *critères* ou d'indicateurs de nature biomédicale, comportementale, et perceptuelle.

La mise en œuvre concrète d'un système d'indicateurs doit enfin tenir compte des *objectifs* pour lesquels ils ont été construits. Pour le médecin, il s'agit de dépasser une approche trop strictement biologique, certes utile en situation de détresse, mais qui peut ne jouer en définitive qu'un rôle secondaire, une fois la survie assurée. Ce qui importe au malade, c'est de pouvoir exprimer le diversité de ses attentes en fonction du poids qu'il accorde à tel ou tel aspect de sa vie. La collectivité est enfin à la recherche d'une mesure commune qui reflète les priorités de la société et lui permette d'éclairer ses choix.

FIXATION DES RÈGLES D'ÉTALONNAGE

La mesure des attributs physiologiques ou fonctionnels ne pose pas de problème lorsqu'elle s'opère directement à partir d'indicateurs physiques. Il n'en va pas de même lorsqu'on cherche à les appréhender à travers le retentissement émotionnel chez le patient. Lorsqu'on désire inférer une propriété à l'aide de mesures faites sur un indicateur perceptuel, il faut soigneusement spécifier la nature de l'instrument de mesure utilisé [1-6], ainsi que ses conditions d'utilisation, de telle sorte que

la procédure puisse être répétée et ses résultats vérifiés. L'étalonnage, c'est-à-dire l'ensemble des règles qui président à l'attribution de nombres aux différentes positions de l'indicateur, constitue un aspect majeur de la standardisation de l'instrument de mesure. Il conditionne l'interprétation empirique qu'on peut lui donner et détermine en grande partie la nature paramétrique ou non-paramétrique de l'analyse statistique qui peut être utilisée.

1 - Le nombre peut être utilisé *exclusive-ment à des fins d'identification*. Par exemple, sur un questionnaire on peut utiliser 1 = femme, 2 = homme ou l'inverse. Il s'agit d'une simple procédure d'étiquetage qui permet d'opérer des décomptes. Aucune opération mathématique sur ces nombres n'est possible.

2 - Le nombre peut également servir d'instrument de mise en rang. Il permet d'ordonner les valeurs d'un indicateur sur une échelle *ordinale*. Par exemple : vous souffrez 1 = un peu, 2 = légèrement, 3 = moyennement, 4 = beaucoup. Mais l'échelle n'étant pas étalonnée à l'aide d'une mesure constante, le classement n'apporte aucune information sur la distance qui sépare les catégories entre elles. Toute série de nombres qui préserve la relation d'ordre est aussi acceptable que la série de nombres initialement établie. La plupart des échelles cliniques de qualité de vie sont de ce type. Il est donc impossible de les utiliser dans des études coût-efficacité. Pour les mêmes raisons, la pratique qui consiste à calibrer des indices pondérés sur la base de données ordinales est illicite.

3 - Dès lors que l'échelle est calibrée à l'aide d'une unité de mesure unique et constante sur toute son étendue, les écarts de niveau peuvent être comparés entre un groupe témoin et un groupe traité par exemple, même si les rapports entre niveaux n'ont pas de sens. Ce type d'échelle, dite à *intervalles égaux ou cardinale faible*, permet donc de comparer les différences entre les entités mesurées, mais ne permet pas d'affirmer qu'une valeur est multiple d'une autre, lorsque l'origine de l'échelle est arbitraire. C'est le cas par exemple des échelles de température Celsius et Fahrenheit, qui ont deux origines différentes.

4 - Lorsque sur l'ensemble des modalités d'un indicateur, on peut définir simultanément un zéro naturel qui caractérise l'origine et la distance entre 2 niveaux de l'échelle, le nombre possède les propriétés habituelles de l'arithmétique. Les écarts entre 2 niveaux, ainsi que le rapport qui les unit, peuvent être établis. Cette propriété essentielle de l'échelle de *proportion* (magnitude scale), encore appelée échelle de rapport (ratio scale) ou *cardinale forte*, permet d'affirmer qu'un état de santé est 2 fois plus sévère qu'un autre, ce qui n'était pas possible précédemment. Pour prendre une autre image : un tremblement de terre de force 4 sur l'échelle de Richter, provoque des ondes sismiques deux fois plus graves que celles d'un séisme de force 2.

Ce qui caractérise le mieux chaque type d'échelle, c'est l'étendue de son invariance, c'est-à-dire l'ampleur des transformations qu'on peut lui appliquer sans provoquer de distorsions dans sa structure. Dans l'échelle *ordinale*, toutes les transformations qui préservent l'ordre ne changent pas les informations disponibles. On dit que l'échelle est unique, à une transformation monotone près. Dans une échelle *d'intervalles*, tous les nombres de l'échelle peuvent être multipliés par une constante et modifiés par un écart constant à l'origine, sans que les résultats en soient modifiés. On dit que l'échelle est unique, à une transformation *affine* près. Dans l'échelle *de proportion* enfin, les relations entre les nombres associés à l'échelle ne changent pas si on les multiplie par une même constante : elle est unique, à une transformation *linéaire* près. Plus les informations que procure l'échelle sont précises, moins grande est la liberté de modifier l'échelle sans modifier les informations qu'elle contenait initialement.

Pour l'économiste, seules les échelles *cardinales* faibles ou fortes permettent d'obtenir une mesure véritable, qui puisse être mise en rapport avec les coûts.

PROCÉDURES DE RÉVÉLATION DES PRÉFÉRENCES

Dégager les préférences d'un sujet, revient à lui présenter plusieurs états de santé en lui

demandant de les hiérarchiser, c'est-à-dire à évaluer leur désirabilité relative. Pour atteindre cet objectif, on s'efforce de créer des situations « de tension mentale » où le sujet interrogé se voit obligé, soit de se situer entre deux propositions verbales opposées, soit de procéder à un choix entre deux éventualités. Dans le premier cas, on se borne à demander aux personnes interrogées de noter le palier de l'échelle qui leur paraît rendre compte de la manière la plus adéquate de l'intensité de leurs réactions par rapport au phénomène étudié : c'est l'approche psychométrique de l'évaluation des préférences individuelles. Dans le second cas, on confronte les individus à des situations hypothétiques (prises deux à deux) dont aucune n'est pleinement satisfaisante, mais entre lesquelles un choix doit être opéré. Cette tension entre les possibles contradictions, crée une situation angoissante forte qui oblige les sujets à s'interroger en profondeur sur la nature et l'importance relative des enjeux. C'est la méthode de révélation des préférences, privilégiée par les économistes.

L'APPROCHE PSYCHOMÉTRIQUE : LES COTATIONS NUMÉRIQUES

Deux types d'échelles supposées cardinales sont utilisées dans la pratique pour mesurer les réactions des malades sous traitement : les échelles de cotation différentielle (échelles d'intervalles du type du thermomètre) et les échelles d'amplitude (échelle proportionnelle des sismographe).

Les échelles de cotation différentielle

Pour appréhender la perception d'un phénomène par un individu, on commence, selon cette méthode, par fixer deux points de référence opposés, de telle façon que le sujet interrogé soit obligé de trouver une solution entre ses contraires. La distance entre les paliers de l'échelle est censée refléter des écarts psychologiques. Ces écarts sont considérés comme équivalents dès lors que la distance qui sépare les paliers de l'échelle est constante, du haut en bas de celle-ci. Une hypothèse de linéarité des attitudes par rapport au critère étudié est donc implicitement introduite. Par exemple, l'intensité des émotions est supposée

croître dans les mêmes proportions, que l'on passe du constat d'une passion : je t'aime, je t'aime un peu, à l'exaltation des sentiments : passionnément, à la folie. En toute rigueur, il conviendrait de s'assurer de cette linéarité, avant d'utiliser, sans preuves, une véritable échelle d'intervalles égaux. Les écarts qu'on enregistre sur ce type d'échelle rappellent les cours différentiels du marché boursier, d'où l'appellation de « rating scale » retenue par les anglo-saxons pour la désigner. On distingue traditionnellement deux types d'échelles de cotation différentielle : les échelles graphiques, continues, et les échelles à catégorie de réponses prédéterminées, discontinues.

— **Les échelles graphiques** peuvent être non graduées ou graduées. Dans le premier cas, elles se présentent sous la forme d'*échelles sémantiques bipolaires* (visual analog scale) [7]. Les personnes sont invitées à indiquer, pour une sensation éprouvée, l'image qu'ils s'en font en se positionnant sur une échelle horizontale ou verticale, entre deux pôles extrêmes définis par des antonymes parfaitement opposés, absence de douleur, douleur extrême par exemple. Le répondant exprime graphiquement l'intensité des sensations qu'il perçoit, le long du continuum qui sépare les deux extrêmes de l'échelle. Il est exclusivement confronté à une paire d'adjectifs opposés. Seul l'analyste a connaissance des équivalents numériques portés au dos de l'échelle. Dans le second cas, les bornes de l'échelle, et chaque variable intermédiaire, sont décrites par un nombre. On obtient alors des *échelles à degrés équidistants* (fixed interval scale) [8]. Leur principe est le même que celui des échelles visuelles analogiques, à cette différence près que la graduation est désormais apparente. D'où le nom de « baromètre de satisfaction » qui a été utilisé par Torrance [9] pour désigner ce type d'instrument. Son emploi permet de familiariser le sujet avec les états de santé, le force indirectement à ordonner ses préférences vis-à-vis d'un état par rapport à un autre, ce qui fait apparaître les incohérences éventuelles entre les réponses. Mais aux dires même de ses utilisateurs [10], il ne permet pas d'obtenir à coup sûr une mesure cardinale.

— **Les échelles à catégories de réponses pré-déterminées** (category-rating). Le répondant, au lieu de devoir s'ajuster sur un continuum séparant les bornes extrêmes d'une échelle visuelle analogique ou d'une échelle à degrés équidistants, doit choisir parmi une série de réponses pré-définies, autrement dit se trouve en face d'une question à choix multiples [11-13]. L'échelle est dite sémantique-différentielle lorsqu'elle mesure les phénomènes à partir de mots qui discriminent les *sujets*. On parle d'échelle à points, lorsqu'on y associe des symboles numériques qui permettent de mieux cerner les différences qui séparent les *situations* soumises à son appréciation. Les sujets doivent alors classer l'ensemble des états de santé sur lesquels ils sont interrogés, en les ventilant entre un nombre donné de cases. A chaque case est affectée une note. Pour obtenir un score, il suffit de totaliser les notes obtenues dans chacune des cases. Par hypothèse, il est admis que l'écart existant entre deux nombres correspondant à ces cases, est représentatif de la différence psychologique établie par le sujet. C'est la technique qui a été utilisée par les auteurs du *Sickness Impact Profile (SIP)* [14] ou du *Quality of Well Being (QWB)* [15-18], pour déterminer la gravité relative des dysfonctionnements repérés. Pour construire le SIP, par exemple, 108 adhérents d'une « *Health Maintenance Organization (HMO)* » de Seattle et 25 professionnels de santé ont été interrogés. Il leur a été demandé d'apprécier la gravité relative des dysfonctionnements, en donnant une note de sévérité croissante allant de 0 à 15. Les poids par sous-dimension et le score global, ont été calculés en divisant la somme des scores par item par la somme maximale des scores possibles.

Les échelles d'amplitude

Dans ce système, les répondants, pour traduire leurs réactions, ont la possibilité d'utiliser n'importe quel nombre positif ou nul, la seule contrainte qui leur est imposée est qu'ils utilisent des nombres qui soient dans le même rapport que l'intensité de leurs réactions. Les valeurs numériques utilisées sont donc proportionnelles à la force de ces dernières, d'où

le nom d'échelles « d'amplitude », encore appelées échelles « de rapport » ou échelles « proportionnelles ». Ces échelles de mesure ne comportent qu'un seul point de référence [19-20], un état de santé sert de point de référence, tous les autres états sont mesurés par rapport à celui-ci.

Deux échelles ont été utilisées par Rosser [21]. La première explore l'adaptation fonctionnelle de l'individu au milieu dans lequel il vit, et stratifie ses incapacités physiques. Elle s'appuie sur des « signes » objectifs : absence d'handicap et d'incapacité, perturbations mineures de la vie sociale, perturbations majeures de la vie sociale, altérations des capacités physiques, inaptitude à assurer les rôles habituels, immobilisation au fauteuil, grabatisation, inconscience. La seconde échelle appréhende le retentissement subjectif de la maladie en mesurant « la souffrance » qui l'accompagne, terme ambigu qui a été choisi à dessein pour la richesse de ses signifiés : « douleur et/ou troubles mentaux, et/ou retentissement psychologique de l'incapacité ». Les 8 niveaux d'incapacité sont combinés avec les 4 paliers de souffrance, ce qui permet d'obtenir 32 états de santé. Après élimination des combinaisons impossibles (inconscience et présence de souffrance, quelle qu'en soit l'intensité), il subsiste un système de classification des malades à 29 niveaux. Dès lors qu'une grille est disponible pour stratifier la population en fonction de la nature de ses atteintes, il est inévitable que se pose le problème de la mesure de leur sévérité relative. En 1978 [22], Rosser propose avec Kind d'étalonner le concept de gravité à l'aide d'une échelle de rapport. La procédure est simple, des entretiens en profondeur semi-structurés, d'une durée de plusieurs heures, ont été réalisés auprès de 30 professionnels de santé, 20 patients et 20 sujets sains. Six états marqueurs sont sélectionnés pour refléter la diversité des 29 réponses envisagées. Le descripteur utilisé pour explorer la dimension souffrance est la douleur physique. A aucun moment, les auteurs ne font allusion au bon ou au mauvais moral des patients. Il est demandé dans un premier temps aux évaluateurs de classer le cas-type en ordre de gravité, en commençant par le plus bénin. Rosser et Kind les

prient ensuite d'affecter un nombre positif à chacun d'entre eux, sans fixer a priori de limite supérieure à leurs possibilités de choix, sous la seule contrainte que les valeurs numériques adoptées soient dans un rapport identique à celui de leur gravité respective. La question précise est « combien de fois plus grave jugez-vous l'état 2 par rapport à l'état 1 quand ils sont l'un comme l'autre appelés à perdurer ». Pour que les juges soient bien conscients des conséquences de leur choix, Rosser et Kind en font ressortir les implications : « le rapport indique, soit le nombre minimum de cas bénins dont vous jugez la guérison équivalente à celle d'un cas grave, soit l'importance relative des moyens qu'il est, selon vous, légitime de consacrer au traitement d'un cas lourd par rapport à ceux qui le sont moins ».

La même procédure est utilisée pour évaluer les autres états marqueurs et les 23 situations intermédiaires restantes. A chaque fois, la valeur attribuée au $n^{\text{ième}}$ état est calculée par rapport au $n-1^{\text{ème}}$ sans qu'il soit nécessaire de procéder à une comparaison des états entre eux. Les relations qui donnent directement les appréciations relatives d'un état par rapport à celui qui le précède, révèlent indirectement, par simple multiplication, celles du dernier d'entre eux par rapport à l'état de santé idéal qui définit l'origine de l'échelle. L'appréciation collective du groupe est finalement obtenue en prenant la médiane des scores attribués par chacun de ses membres. La valeur de ces médianes mesure donc les pertes d'utilité provoquées par l'altération de la qualité de la vie.

La troisième version [23] de l'indicateur, publiée par Kind et Rosser en 1982, calibre ces médianes en les divisant par une valeur pivot — celle qui était attribuée au décès dans la version de 1978 — ce qui transforme l'échelle de rapports en une échelle d'intervalles dont les bornes extrêmes sont 1 pour la bonne santé et 0 pour le décès. La soustraction des scores calibrés d'altération de la qualité de la vie du score de santé idéale, permet d'obtenir les coefficients de qualité de vie correspondants.

L'APPROCHE UTILITÉ-PRÉFÉRENCES : LES CHOIX FORCÉS

Trois techniques peuvent être appliquées, soit séparément, soit simultanément : les « temps psychologiquement équivalents » (Time trade-off), le jeu de hasard idéalisé (standard gamble), les besoins socialement équivalents (Personal trade-off).

Les temps psychologiquement équivalents

On demande [24-26] au patient de choisir entre les deux situations hypothétiques : vivre en excellente santé, mais en acceptant de perdre un certain nombre d'années de vie ; ou vivre 5 ans, avec l'une des conséquences que peut entraîner un état de santé chronique, par exemple les séquelles d'une intervention pour hypertrophie bénigne de la prostate : incontinence et impuissance, détérioration en dépit de l'intervention.

On cherche par tâtonnements successifs à déterminer la durée de bonne santé pour laquelle ces deux options sont équivalentes, en faisant varier le laps de temps d'excellente santé proposé. Lorsque celui-ci est bref, le fléau de la balance penche à droite : une vie prolongée, même si elle comporte des séquelles, est préférée à un décès précoce. Lorsque la tranche de temps passée en bonne santé est allongée, le fléau penche à gauche : une vie de qualité est préférée à celle qui ne l'est pas. Le retour à l'équilibre est fonction du poids des séquelles. Plus celles-ci sont sévères, plus longue est la tranche de temps que le sujet est disposé à sacrifier pour les éviter. Plus celles-ci sont légères, moins le sujet est enclin à perdre des années de bonne santé. Pour un niveau de séquelles donné, il existe un nombre-seuil d'années de bonne santé pour lequel le patient est indifférent entre les deux options qui lui sont offertes. Le rapport du temps passé dans chacune de ces deux situations, excellente santé et état de santé dégradé, permet de transformer les années de vie passées en mauvaise santé en fractions équivalentes d'années de bonne santé. Par exemple, si un sujet estime que deux ans et demi passés en bonne santé est équivalent à 5 ans restant à vivre dans son état chronique

actuel, la valeur des années passées dans celui-ci est égale à $2,5/5$, soit $0,5$. Une année passée dans cet état ne vaut pas à ses yeux 12 mois, mais 6. Le coefficient de qualité de vie sera donc de $0,5$.

Le jeu de hasard idéalisé

La deuxième de ces méthodes est traditionnellement utilisée pour évaluer les préférences cardinales en situation d'incertitude [27-32]. Elle est, de ce fait, considérée comme particulièrement appropriée en matière médicale. Le protocole sur lequel elle s'appuie est simple dans ses principes. Trois états de santé (S1, S2, S3) sont décrits de façon détaillée et présentés à un sujet qui doit choisir entre les options suivantes : soit un traitement A qui assure de façon certaine la situation S2, soit un traitement B qui comporte deux issues possibles : l'état S1 de probabilité p , ou l'état S3 de probabilité $1-p$. Les états S1, S2 et S3 sont censés être hiérarchisés dans cet ordre, S2 occupant un rang intermédiaire entre S1 et S3. Lorsqu'on fait varier la valeur de p de 0 à 1, il existe une valeur-seuil telle que le sujet soit indifférent entre le traitement A et le traitement B. Cette valeur mesure l'utilité qui est associée à la première de ces thérapies.

Le dilemme du coronarien permet de mieux comprendre la richesse d'une telle formulation. M. X souffre d'une angine de poitrine, deux possibilités s'offrent à lui : ou il accepte une thérapie au long cours, ou il prend le risque d'un pontage. L'issue du premier choix est à court terme sans ambiguïté, il est certain de vivre. Le second choix est plus risqué, puisque les chances de succès de l'intervention ont été estimées dans son cas à 90 % par son médecin traitant. Le coronarien hésite entre deux attitudes. Il peut soit opter pour la situation à risque, qui comporte un taux d'échec non négligeable, soit jouer la sécurité en écartant a priori toute possibilité d'amélioration de son état fonctionnel. Le problème surgit parce que, en jouant la sécurité, il se place dans une situation meilleure que celle dans laquelle il pourrait être placé, si l'issue défavorable de l'alternative à risque survenait : la mort. Mais cette situation est moins

bonne que celle qui pourrait être la sienne si l'opération réussissait. Pour qu'une décision soit prise, il convient d'évaluer la désirabilité relative du maintien dans l'état de santé actuel, avec douleur angineuse, par rapport à la meilleure et à la plus défavorable des issues possibles lorsque l'éventualité à risque est choisie.

La solution du dilemme passe par l'introduction d'un jeu de hasard idéalisé obéissant à la loi des grands nombres. La structure du jeu est la même que celle du problème primitif. Le choix se limite à une éventualité certaine et à une éventualité à risque, la survie sans séquelle ou la mort. Deux différences cependant par rapport au dilemme initial : a) la décision se rapporte à une situation hypothétique, ce qui élimine la connotation émotionnelle que comportait le problème initial, b) le calcul des risques ne repose pas sur une estimation personnelle mais sur une mesure objective. En faisant varier la probabilité de l'issue la plus favorable de la situation à risque, on peut évaluer la valeur psychologique que le sujet attache à la situation certaine. Lorsque cette issue favorable a une chance sur cent de se produire, le malade doit choisir entre la certitude de vivre avec une douleur angineuse ou le risque de subir une intervention dont les taux de succès sont faibles. Le jeu n'en vaut pas la chandelle, il choisit la sécurité. Si au contraire le taux de décès opératoire est faible (1 %), la probabilité de survivre à l'intervention est élevée et le malade opte pour la solution aléatoire.

Lorsque la chance de succès est faible, c'est l'éventualité certaine qui a la faveur du malade. Dans le cas contraire, c'est vers l'éventualité à risque que vont naturellement ses préférences. La seule différence entre les deux situations tient à la valeur de la probabilité de succès. Au fur et à mesure que celle-ci augmente, le sujet est de moins en moins enclin à jouer la sécurité et de plus en plus tenté par la solution à risque. Finalement, il existe une valeur-seuil de ce coefficient, pour laquelle les deux options offertes au patient lui sont indifférentes. Cette valeur mesure la qualité de vie actuelle du patient. Si les douleurs sont sévères ou répétées, la valeur du

coefficient critique est faible. Le malade est prêt à tout pour échapper à sa condition présente, l'intervention est acceptée, même lorsque ses chances de succès sont limitées, ce qui prouve la dégradation de son état de santé. Si les douleurs sont modérées, la valeur du coefficient critique est élevée, l'utilité qui est associée à la situation présente est proche de celle que procure une bonne santé, le malade n'accepte l'intervention que s'il est quasiment certain que celle-ci réussisse.

Au lieu de rechercher la valeur-seuil de la probabilité de succès pour laquelle l'intervention aléatoire et le traitement sans risque, lui sont indifférents (*méthode de la probabilité équivalente*), on peut s'efforcer d'établir quel gain thérapeutique minimum doit offrir un traitement dépourvu de toute dangerosité pour qu'il puisse être jugé comme équivalent à un traitement aléatoire plus prometteur (*méthode de l'équivalent certain*) [33]. La première méthode fait varier la probabilité de succès ou d'échec du traitement aléatoire, en supposant donnés les gains thérapeutiques associés à chaque type de traitement. La seconde module le gain thérapeutique attendu à coup sûr de l'administration du traitement conservateur, en considérant les probabilités et les résultats de chacune des issues possibles de la solution à risque comme données.

Le processus itératif qui conduit à la valeur-seuil, est le même dans les deux cas, mais la nature des enjeux est moins clairement perceptible dans le premier que dans le second. Jongler avec les probabilités est un exercice difficile pour beaucoup de sujets, d'autant qu'ils sont contraints de garder en permanence à l'esprit les deux faces complémentaires de leur évolution : succès, échec. Une modification de l'équivalent certain, le nombre d'années à vivre par exemple, est beaucoup plus facile à comprendre et à maîtriser, d'autant que l'éventail des probabilités envisageables se trouve réduit à la forme de paris : le plus simple auquel on puisse songer est le jeu de pile ou face, avec 50 % de chances de chaque côté. Le calcul de l'utilité de l'équivalent temporel certain est alors immédiat, dès lors que la meilleure des issues possibles du traitement aléatoire est cotée 1, et la pire, 0.

Les besoins socialement équivalents

On demande à la personne interrogée [34, 35] de préciser le nombre d'individus gravement atteints, dont le maintien en vie lui semble aussi souhaitable que la préservation de la vie en bonne santé d'individus sains. On peut reformuler la même question sous une forme légèrement différente, en demandant à la personne interrogée de fixer le nombre minimum de cas graves E_D dont elle juge la guérison socialement équivalente à celle d'un nombre E_B donné de cas bénins. Le point d'indifférence entre ces deux groupes est obtenu en faisant varier le nombre d'individus dont l'état de santé est gravement détérioré. Il existe un point d'équilibre, où les deux propositions sont psychologiquement équivalentes aux yeux de la personne interrogée. L'utilité de l'état de santé dégradé U_D , peut être calculée à partir de l'égalité suivante :

$$U_D E_D = U_B E_B$$

Si l'on admet que l'état de santé parfait a une utilité égale à 1 ($U_B = 1$), alors l'utilité de l'état de santé dégradé est égal à

$$U_D = E_B/E_D$$

Intérêt de l'approche utilité-préférences

Cette méthode, présente six avantages. Tout d'abord elle permet de disposer d'une mesure exhaustive, combinant en un score unique, à la fois les symptômes de la maladie, ses diverses répercussions physiques, sensorielles, socio-émotionnelles, cognitives, et les effets indésirables du traitement. Elle permet de calculer une espérance de vie pondérée en fonction de la qualité de vie, ce qui n'est pas le cas des profils qui se contentent d'étudier le retentissement multiforme de la maladie de façon intemporelle. Le rapprochement des résultats et des coûts est possible, dans la mesure où ces deux paramètres ont une dimension cardinale faible. Le score obtenu reflète directement les préférences du patient, il n'est pas dérivé à partir de pondérations définies par les techniciens du soin ou par une population de bien-portants. Par ailleurs, l'instrument peut être considéré comme spécifique de la maladie si les descriptifs appro-

priés sont choisis pour en rédiger les scénarios. Enfin, la méthode a un fondement scientifique incontestable : la théorie de la décision en régime d'incertitude de Von Neumann et Morgenstern. Malgré les indéniables qualités de l'outil, il convient de ne pas dissimuler les difficultés inhérentes à sa mise en œuvre : d'une part, les réponses obtenues varient en fonction du contexte dans lequel sont posées les questions ; d'autre part, il est très difficile d'identifier les variables cliniques qui sont à l'origine du score global ; enfin, la sensibilité d'un tel indicateur reste à démontrer dans les différentes pathologies.

AVAQ ET GESTION DES RESSOURCES COLLECTIVES

Le défi, désormais lancé aux décideurs [36], est d'améliorer l'état de santé de la population, tout en parvenant à maîtriser les dépenses. Des tensions doivent fatalement apparaître. Si des financements conséquents sont consacrés au remboursement d'une procédure, l'utilisation de celle-ci doit permettre de dégager des gains de santé, au moins aussi importants que ceux qui auraient pu être obtenus en consacrant les deniers de l'assurance-maladie au financement d'autres soins. Inversement, si une politique drastique de maîtrise est mise en œuvre, il faut s'assurer que la réduction des dépenses mises à la charge de l'assurance-maladie ne se fasse pas au prix d'une détérioration inacceptable de l'état de santé de la population. Le pilotage à vue dans une situation aussi difficile est manifestement dépassé. L'analyse coût-utilité [37-38], permettant de mettre en évidence les relations existantes entre les coûts additionnels et l'amélioration observée dans le service médical rendu, doit être intégrée d'emblée.

La façon la plus simple d'introduire le concept, est d'imaginer que les termes de la définition de la santé par l'OMS puissent être hiérarchisés sur une échelle strictement ordonnée, allant du bien-être parfait au décès, en passant successivement par la présence de signes et symptômes, la décrépitude physique, l'altération des capacités mentales et le retrait social. Cette approche heuristique permet de comprendre la nature unidimensionnelle de

l'échelle, mais elle mutile la réalité, puisqu'elle ne hiérarchise que des perturbations isolées. Or, les dysfonctionnements se présentent le plus souvent en association. C'est donc l'ensemble du tableau des altérations de qualité de vie qu'il convient de situer sur l'échelle, en raisonnant en termes de scénarios ou en s'appuyant sur des systèmes de classification d'états de santé.

L'évaluation de la qualité de vie globale consiste alors à déterminer la valeur des coefficients compris entre 0 et 1 qui peuvent être attribués à chacun de ces cas-types. Ces facteurs de pondération sont autant de facteurs d'ajustement de la quantité de vie en fonction de sa qualité, d'où leur appellation de coefficients de qualité de vie (QOL).

Les 5 méthodes précédemment présentées, dont 2 directes (les échelles de cotation différentielle et les échelles d'amplitude) et 3 indirectes (temps psychologiquement équivalents, jeu de hasard idéalisé et besoins socialement équivalents), permettent de déterminer la valeur des QOL. Le produit des années, ou fractions d'années passées dans un état de santé donné et du coefficient de qualité de vie correspondant, quelle que soit la manière dont il a été calculé, transforme le temps passé en mauvaise santé en fractions équivalentes d'années de bonne santé. Si la même opération est reconduite dans le temps pour les différents stades d'évolution de la maladie, on obtient un nombre d'Années à Vivre Ajusté sur la Qualité, l'AVAQ, que les Anglo-Saxons désignent par le sigle QALY (quality-adjusted life-years). La procédure est beaucoup moins anodine qu'il n'y paraît, puisqu'elle implique une spécification tout à fait particulière des préférences individuelles [39-41]. Elle suppose : a) l'indépendance respective des deux arguments de la fonction d'utilité, longévité et qualité, b) la constance dans le temps du coefficient de qualité de vie, c) l'absence d'attraction ou de répulsion vis-à-vis du risque. Sous ces conditions, le gain d'utilité, pour un niveau donné de qualité de vie, est proportionnel au nombre d'années restant à vivre.

Pour disposer d'un critère au regard duquel l'impact du traitement ou les mérites respectifs

de deux traitements alternatifs peuvent être comparés, il suffit de diviser le coût du traitement par le résultat exprimé en AVAQ.

Deux types de configuration sont sans ambiguïté [42-44] :

— si la variation d'utilité, exprimée en AVAQ, est strictement positive ($US > 0$), avec des coûts supplémentaires équivalents ou moindres ($CS \leq 0$), ou si les coûts sont inférieurs ($CS < 0$), avec ou sans amélioration du service médical rendu ($US \geq 0$), la procédure est coût-efficace : elle rend un service supérieur ou égal, sans alourdir les charges financières de la collectivité. Son admission au remboursement doit être clairement recommandée ;

— si la variation d'utilité est strictement négative ($US < 0$), avec ou sans augmentation de coût ($CS \geq 0$), ou si le coût est supérieur ($CS > 0$), avec ou sans détérioration du service médical rendu ($US \leq 0$) : l'admission au remboursement de la procédure doit être déconseillée, puisqu'elle accapare des ressources sans aucun profit, ou en pure perte.

Le problème se pose lorsqu'il faut prendre, en situation de *rareté*, une décision vis-à-vis d'une procédure nouvelle, plus utile et plus chère ($US > 0$, $CS > 0$) ou réévaluer une procédure ancienne, moins utile mais plus économique ($US < 0$, $CS < 0$). L'évaluateur doit être à même de répondre aux deux questions suivantes :

a) les dépenses qu'entraînent pour l'assurance-maladie l'admission au remboursement, sont-elles justifiées par l'amélioration du service médical rendu par rapport aux produits de référence : dans quelle mesure, le surcroît de dépenses reconnues et remboursées est-il justifié par un surcroît d'utilité, dans les indications retenues ? Ce rapport coût-utilité différentiel, est-il compatible avec ceux qui ont déjà été considérés comme acceptables par la tutelle, sur d'autres procédures, dans les mêmes indications ?

b) le coût-utilité « marginal », est-il comparable avec les sacrifices financiers que la collectivité considère comme acceptables pour atteindre, dans d'autres pathologies, des ni-

veaux similaires d'accroissement d'efficacité ?

On est en face d'un problème classique de recherche d'un optimum lié, et l'on sait que le calcul de cet optimum conduit à l'introduction d'un multiplicateur de Lagrange, à partir duquel la définition d'une borne supérieure pour le rapport surcoût/surcroît d'utilité est immédiate. La valeur du multiplicateur associé à la contrainte, mesure à la fois le supplément de coût que la société est amenée à consentir pour produire une unité d'utilité supplémentaire, et le prix maximal qu'elle serait disposée à payer pour obtenir un résultat identique à l'aide d'un autre traitement. Si, à partir d'un théorème d'existence, on peut affirmer que le meilleur emploi des ressources disponibles entraîne inévitablement la présence d'un rapport coût-utilité seuil, encore faut-il être à même de calculer la valeur de celui-ci pour donner un caractère opérationnel au concept.

Une des solutions possibles consiste à prendre pour guide des décisions déjà mises en œuvre pour d'autres innovations technologiques, en réinjectant a priori dans l'analyse des valeurs constatées a posteriori pour des procédures dont l'admission a déjà été prononcée. La valeur la plus élevée du rapport surcoût/surcroît d'utilité, acceptée lors de l'admission au remboursement par la tutelle dans un passé récent, mesure le supplément de coût que l'assurance-maladie a été amenée à prendre en charge. Elle nous indique du même coup le prix maximal auquel la société serait prête à acquérir une unité supplémentaire d'utilité par un autre traitement. Il suffit donc de comparer cette valeur-seuil avec les rapports C-U « marginaux » des nouvelles procédures pour savoir si l'admission de celle-ci au remboursement est légitime ou non. Tant que le quotient surcoût/surcroît d'utilité est inférieure à cette valeur-seuil, la demande est justifiée. Dans le cas contraire, la comparaison suggère qu'elle doit être rejetée.

LES TRAVAUX MENÉS EN FRANCE

En France, à notre connaissance, seules quatre évaluations ont été faites en termes

de coût par AVAQ additionnel : l'étude sur l'Erythropoïétine [45], la chimiothérapie adjuvante [46], la chimionucléolyse à 7 ans [47], et le traitement de l'infarctus aigu du myocarde [48]. Par ailleurs, six stratégies ont été évaluées en termes de coût par année de vie gagnée, sans prise en compte de la qualité

de vie. La plupart des évaluations économiques se sont limitées à évaluer des coûts de traitement par « cas », par « succès » ou « échec », ou encore par « séjour hospitalier », ne permettant pas de comparaison inter-pathologies. Les données publiées en France sont résumées dans le *tableau I*.

TABLEAU I. — *Relevé des ratios coût-efficacité marginaux et moyens publiés en France.*

Stratégies évaluées	Références et année de saisie des coûts	Année de publication	Ratio C/E Marginaux (F)
Coût marginal net par AVAQ gagné^a			
Erythropoïétine (versus dialyse)	Fagnani [45] (1988)	1990	450 000
Chimiothérapie adjuvante (versus abstention)	Livartowski [46] (1989)	1992	< 0
Chimionucléolyse (versus discectomie)	Launois [47] (1990)	1992	< 0
Infarctus aigu du myocarde < 3h (versus > 6h)	Castiel [48] (1989)	1990	5 810
Coût marginal net par année de vie gagnée^a			
Symbion Jarvick (versus enoximone)	CRESGE [49] (1988)	1989	1 700 000
Dépistage cancer prostate, H 60-69 ans (versus non dépistage). Prévalence = 11 %	Launois [50] (1989)	1990	196 690
Traitement médicamenteux hypolipémiant	Durand-Zaleski [51] (1992)	1992	28 320
SK + asp (versus rTPA + hep) : délais < 5 h.	Launois [52-53] (1988)	1990	< 0
Coût moyen brut par cas dépisté^b			
Dépistage de trisomie 21	Moatti [54] (1987)	1990	390-480.000
Radiodépistage systématique de la tuberculose	Lefaire [55] (1982)	1987	178-300.000
Dépistage Sida chez les femmes enceintes	Le Galès [56] (1987)	1990	70.000
Dépistage systématique drépanocytose	Le Galès [57] (1991)	1993	21-44.600
Dépistage hémoglobinosé	Le Galès [58] (1986)	1987	10-15.000

^a Différence de coût introduite par le remplacement d'une option thérapeutique par une autre, déduction faite des coûts additionnels dont il permet de faire l'économie en évitant un plus grand nombre de traitements, divisée par l'accroissement d'efficacité qui en résulte.

^b Quotient du coût et de l'efficacité totale d'une campagne de dépistage par rapport à une situation de référence où rien ne serait fait (coût des traitements évités non déduits).

SK = Streptokinase

asp = aspirine

rTPA = activateur tissulaire du plasminogène

hep = héparine

Bien entendu, ces ratios doivent être maniés avec prudence [59], dans la mesure où des méthodologies fort disparates sont mises en œuvre pour les calculer. Les AVAQ calculés par Castiel [48] ont été obtenus par utilisation d'une échelle visuelle analogique, ceux qui ont été publiés dans les 3 autres études ont été obtenus de façon plus homogène, par exploitation du Health Measurement Questionnaire et de l'indicateur de Rosser.

C'est surtout au niveau du calcul des coûts que les méthodes sont les plus hétérogènes. Dans le secteur ambulatoire, tous les auteurs appliquent les cotations de la nomenclature générale des actes professionnels. En revanche, au niveau hospitalier, les approches sont très différentes, et 6 méthodes ont été utilisées :

1) - la valorisation en fonction des tarifs de prestations journalières de la sécurité sociale [60] ;

2) - la valorisation sur la base d'un prix de revient journalier moyen, pondéré en fonction du type d'établissement : AP-HP (Assistance Publique - Hôpitaux de Paris), CHR (Centre Hospitalier Régional), CHG (Centre Hospitalier Général) (méthode LEGOS) [61] ;

3) - la méthode PMSI (Programme de Médicalisation du Système d'Information), soit PMSI AP-HP [62], soit PMSI « enquête sur le coût des activités médicales » [63] ;

4) - la valorisation sur la base d'un prix de journée « aménagé » : prix de revient du séjour hospitalier, hors prescription d'examen et de médicaments, plus prix de revient des examens et des traitements spécifiques, individualisés par patient (méthode CEDIT-CRESGE) [64-65, 49] ;

5) - la valorisation en coût direct : le coût direct comprend le coût médical des journées d'hospitalisation, le coût des actes effectués par d'autres services de l'hôpital (laboratoires, radios), le coût standard des actes lourds (méthode HCL, Launois) [66-67] ;

6) - l'Approche Processus de Production (méthode U 357, PRN) [68-69].

Il semble que les approches 4, 5 et 6, très sophistiquées, ne se justifient que dans le

cadre d'une analyse comparant deux attitudes thérapeutiques au sein d'un même établissement, ce qui n'est pas l'objectif recherché. Quant à l'approche 3, son utilisation est conditionnée par la publication des résultats obtenus dans les 44 établissements volontaires, prévue fin 1993. La méthode 2, qui s'appuie sur des travaux anciens du CERC (Centre d'Etudes et de Recherches sur les Coûts), est manifestement obsolète. Dans l'état actuel des informations disponibles, l'utilisation des tarifs de prestations journalières payées par la sécurité sociale permet de disposer d'une unité de mesure officielle commune à l'ensemble des soins, que ceux-ci soient dispensés en établissements hospitaliers privés ou publics, ou en milieu ambulatoire.

Aux Etats-Unis, des « league tables » similaires ont été publiées. Bien que les prix de revient des examens et des hospitalisations aux USA soient différents de ceux observés en France, nous en présentons le contenu dans le *tableau II*, afin de disposer de données de cadrage, aussi grossières soient-elles.

QUALITÉS REQUISES DES INSTRUMENTS

Pour qu'une mesure de la qualité de vie soit considérée comme établie, elle doit être intuitivement crédible, fondée quant à son contenu, fiable dans sa mesure, sensible et valide. Ces caractéristiques métrologiques ont déjà fait l'objet de nombreux développements dans la littérature [76-80], il suffira donc d'en rappeler les fondements.

ACCEPTABILITÉ

La recevabilité d'un questionnaire, appelée encore « validité de surface », ou « face value », dépend de la qualité de sa formulation : les questions sont-elles suffisamment précises pour chaque dimension et sous-dimension explorée ? Se rapportent-elles à une période de temps clairement définie ? Portent-elles sur un état de santé ou sur un changement d'état ? Sont-elles formulées en termes de capacité ou en termes de performances ? La procédure d'agrégation des différents items est-elle satisfaisante ?

TABLEAU II. — *Relevé des ratios coût-efficacité marginaux publiés aux Etats-Unis.*

Stratégies évaluées	Références et année de saisie des coûts	Ratio-Marg. C/AVAQ US \$ à l'année de la saisie	Ratio-Marg. C/AVAQ US \$ 1992 ^a	Ratio-Marg. C/AVAQ Frs 1992 ^b
Produit de contraste basse osmolarité, faible risque d'effets indésirables	Goel [70] (1986)	220 000	368 476	2 413 518
Dialyse à l'hôpital	Churchill [71] (1980)	40 200	102 030	668 296
Dialyse péritonéale	Churchill [71] (1980)	35 100	89 086	583 513
Traitement hypertension légère (Homme 40 ans, TA diastolique 90-94 mmHg)	Stason, Weinstein [72] (1976) actualisé 84 [75]	45 000 [75]	80 098	524 644
Soins intensifs prématurés 500-999 g	Boyle [73] (1978)	19 600	60 290	394 899
Traitement hypertension modérée (Homme 40 ans, TA diastolique 95-104 mmHg)	Stason, Weinstein [72] (1976)	9 880	32 951	215 829
Produit de contraste basse osmolarité, fort risque d'effets indésirables	Goel [70] (1986)	23 000	38 522	252 319
Traitement hypertension sévère (Homme 40 ans, TA diastolique > 105 mmHg)	Stason, Weinstein [72] (1976)	4 850	16 175	105 946
Soins intensifs enfants pesant 1000-1499 g	Boyle [73] (1978)	2 800	8 613	56 415
Pontage aorto-coronarien tronc commun artère coronaire gauche. Angor modéré	Weinstein [74] (1981)	3 600	8 299	54 358

^a Actualisé à la valeur 1992 sur la base de l'indice des prix à la consommation en milieu urbain des soins médicaux américains. Source : consumer price indexes for all urban consumers and for urban wage earners and clerical workers : US city, average. US Bureau of Labour Statistics : monthly labour review.

^b Après neutralisation des différences liées au système des prix américains et français par l'indice de parité des pouvoirs d'achat du Produit Intérieur Brut : 1 \$ US = 6,55 F. Source : OCDE (Communication J.P. Poullier).

PERTINENCE

La pertinence de la teneur (« content value » des anglo-saxons) exige que soient remplies deux conditions : l'ensemble de l'univers des atteintes doit être intégré et la représentativité des items retenus doit être assurée. Le contenu de l'instrument que l'on se propose de forger doit couvrir tout le champ du phénomène que l'on envisage d'étudier, et il doit renfermer un échantillon d'énoncés représentatif de l'ensemble des énoncés concevables. Une mauvaise spécification de l'univers représente le premier type d'erreurs possibles. Un deuxième biais résulterait du choix d'énoncés non adaptés ; il convient donc toujours de préciser la manière dont les énoncés

ont été choisis pour construire l'échelle. L'identification des composantes de l'instrument peut reposer sur la consultation d'experts ou sur les déclarations des malades eux-mêmes. Quant à la sélection des énoncés finaux, elle peut s'appuyer sur des méthodes plus ou moins élaborées, la façon la plus simple d'y procéder consiste à multiplier la fréquence des atteintes par leur importance. Mais on peut aussi recourir à des analyses statistiques plus sophistiquées, comme l'analyse en composantes principales.

FIABILITÉ

Pour s'assurer la fiabilité d'un instrument,

il convient de vérifier sa cohérence interne, sa reproductibilité et sa concordance.

Une échelle est dite cohérente lorsque ses diverses composantes ne sont pas contradictoires. Cette cohérence est établie lorsque chaque item à l'intérieur d'un domaine, et lorsque chaque domaine à l'intérieur de l'instrument, explorent des dimensions qui sont complémentaires sans être redondantes. Le coefficient alpha de Cronbach est la statistique la plus souvent utilisée pour en rendre compte.

Une échelle est dite reproductible si, en mesurant plusieurs fois le même phénomène, on obtient des résultats proches les uns des autres. Pour déterminer la reproductibilité, on recherche l'importance de l'erreur aléatoire dans la mesure. Si celle-ci est faible, il existe une grande similitude entre les mesures successives de l'univers exploré. Certains auteurs parlent alors de « précision » de la mesure, tandis que d'autres font référence à sa « fidélité ». Le coefficient de corrélation intra-classe est la statistique de référence.

Une échelle est dite concordante lorsqu'elle reflète une similitude de jugements entre des observateurs différents sur le même événement. Les statistiques qui permettent d'évaluer cette concordance sont les coefficients kappa pour les données ordinales, et le coefficient de corrélation intra-classe pour les données cardinales.

SENSIBILITÉ

La sensibilité est la capacité de l'instrument à déceler des changements cliniquement importants, même s'ils sont de faible amplitude. Un indicateur a une sensibilité maximum lorsque, indépendamment de l'imprécision due aux erreurs de mesure, il reproduit toutes les variations de la variable considérée. Guyatt [9] donne une définition élargie de la sensibilité, sous l'appellation « aptitude à détecter un changement » (« responsiveness »), qui intègre à la fois les notions de reproductibilité et de sensibilité proprement dite. Deux caractéristiques doivent alors être remplies : le questionnaire utilisé doit présenter des scores pratiquement inchangés, sur des sujets stables,

c'est-à-dire qu'il doit être reproductible ; il doit également être capable de mettre en évidence les modifications qui se produisent quand l'état de santé des sujets s'améliore ou s'altère.

VALIDITÉ

On dit qu'un instrument est valide s'il mesure effectivement ce qu'il prétend appréhender (construct validity), ce qui suppose à la fois l'absence d'erreur aléatoire et de biais systématiques. La fidélité est donc une condition nécessaire, mais non suffisante, de la validité. Pour que celle-ci puisse être obtenue de façon parfaite, aucune erreur constante ne doit être présente. En l'absence d'un étalon de référence incontesté, la validité d'une échelle de mesure est obtenue en la comparant, soit à d'autres indicateurs de qualité de vie qui explorent le même domaine, soit à des indicateurs cliniques afin d'évaluer comme elle s'en démarque (validité divergente) ou s'en rapproche (validité convergente).

CONCLUSION

Le choix d'un indicateur est fonction de l'objectif poursuivi. L'outil qui permet de stratifier une population en fonction de son état de santé ne sera pas le même que celui qu'il convient de mettre en œuvre pour évaluer l'évolution de la maladie sous traitement. Enfin, les instruments clinimétriques qui permettent d'appréhender la diversité des atteintes morbides sont totalement inadaptés pour orienter les choix de la collectivité en matière d'allocations de ressources : celle-ci doit pouvoir raisonner à l'aide d'un instrument de mesure synthétique, transnosographique. Le QALY, malgré les hypothèses fortes qui en sous-tendent la construction, semble répondre à cet objectif.

RÉFÉRENCES

1. Froberg D, Kane R. Methodology for measuring health state preferences. II Scaling methods. *J Clin Epidemiol* 1989 ; 42 : 459-71.
2. Launois R. La qualité de vie, panorama et mise en perspective. In Launois R, Régnier F, eds. *Décision Thérapeutique et Qualité de Vie*, Paris, Collection

- de l'Association Française pour la Recherche Thérapeutique. John Libbey Eurotext, 1992.
3. Torrance GW. Preferences for health states. A review of measurement methods. In *Clinical and Economic Evaluation of Perinatal and Developmental Medicine* n° 20 Evanville. Mead Johnson 1982.
 4. Stevens SS. On the theory of scales of measurement. *Science* 1946; 103: 677-80.
 5. Stevens SS, Galanter EH. Ratio scales and category scales for a dozen perceptual continua. *J Experim Psychol* 1957; 54: 377-41.
 6. Stevens SS. A metric for social consensus. *Science* 1966; 151: 530-41.
 7. Scott J, Huskisson EC. Graphic representation of pain. *Pain* 1976; 2: 175-84.
 8. Joyce CRB, Zutshi DW, Mason RM. Comparison of fixed internal and visual analogue scales for rating chronic pain. *Eur J Clin Pharmacol* 1975; 8: 415-20.
 9. Feeny D, Labelle R, Torrance G. Integrating economic evaluations and Quality of Life assessments. In B. Spilker, ed. *Quality of Life Assessments in Clinical Trials*, Raven Press 1990.
 10. Nord E. The validity of a visual analogue scale in determining social utility weights for health states. *Internat J Health Planning Managem* 1991; 6: 234-42.
 11. Gadreau M. Une Mesure de la Santé. Collection de l'Institut de Mathématiques Economiques. 17. Paris, Sirey, 1978.
 12. Kaplan RM, Bush JW, Berry CC. Category rating versus magnitude estimation for measuring levels of well being. *Med Care* 1979; 17: 501-21.
 13. Kaplan RM, Ernst J. Do category rating scales produce biased preference weights for a health index. *Med Care* 1983; 21: 193-207.
 14. Carter W, Bobbitt R, Bergner M *et al.* Validation of an internal scaling: the sickness impact profil. *Health Serv Resear* 1976 (winter): 516-28.
 15. Chen M, Bush JW, Patrick D. Social indicators for health planning and policy analysis. *Policy Sci* 1975; 6: 71-88.
 16. Kaplan RM, Bush J. Health related quality of live. Measurement for evaluation research and policy analysis. *Health Psychol* 1982; 1: 61-80.
 17. Kaplan RM, Anderson J. A General Health Policy Model. Update and applications. *Health Serv Resear* 1988; 23: 203-35.
 18. Patrick D, Bush JW, Chen MM. Methods for measuring levels of well being for a healthy status index. *Health Serv Resear* 1973 (Fall): 299-45.
 19. Haski M, Moskowitz H. L'échelle sensorielle de Stevens-Moskowitz. Une nouvelle venue dans la mesure des réactions des interviewés. *Rev Fr Market* 1980; 81 (2): 5-18.
 20. Johson E, Huber G. The technology of utility assessment. *IEEE Transactions on Systems, Man and Cybernetics* 1977; SMC 7 (5): 311-23.
 21. Rosser RM, Watts V. The measurement of hospital output. *Inter J Epidemiol* 1972; 1: 361-7.
 22. Rosser RM, Kind P. A scale of valuation of states of illness. Is there a social consensus? *Inter J Epidemiol* 1978; 7: 347-57.
 23. Kind P, Rosser RM, Williams A. Valuation of quality of life: Some psychometric evidence. In: *The Value of Life and Safety*. Jones Lee M W Ed. North Holland Amsterdam 1982: 159-70.
 24. Churchill DH, Morgan S, Torrance GW. Quality of life in end stage renal disease. *Periton Dialysis Bull* 1984; 4: 20-3.
 25. Churchill DH, Torrance GW, Taylor DW *et al.* Measurement of quality of life in end stage renal disease. The time trade-off approach. *Clin Investig Med* 1987; 10: 14-26.
 26. Mohide A, Torrance GW, Scheiner DL. Measuring the well-being of family coaregivers using the time trade-off technique. *J Clin Epidem* 1988; 41: 475-82.
 27. Torrance GW. A generalized cost-effectiveness model for the evaluation of health programs. Faculty of Business Mc Master University Research Series n° 101, 1970.
 28. Torrance GW. Toward a utility theory foundation for health status index model. *Health Serv Resear* 1976; 349-69.
 29. Farquhar P. Utility assessment method. *Managem Sci* 1984; 30: 1283-300.
 30. Fischer GW. Utility models for multiple objective decisions. Do they accurately represent human preferences? *Decision Sci* 1977; 10: 451-79.
 31. Launois R, Orvain J, Ounis I. Apport d'une mesure des utilités. Le cas des infections respiratoires récidivantes. *Rev Epidemiol Santé Publ* 1992; 40: 46-55.
 32. Torrance G. Measurement of health states utilities for economic approach. A review. *J Health Econom* 1986; 5: 1-30.
 33. McNeil B. « The patient's role in assessing the value of diagnostic tests ». *Radiology* 1979; 132: 605-10.
 34. Berg RL. Establishing the values of various conditions of life for a health status index. In: Berg R L ed. *Health Status Indexes*. Chicago: Hospital Research and Educational Trust, 1973.
 35. Nord E. Methods for quality adjustment of life years. *Soc Sc Med* 1992; 34: 559-69.
 36. Launois R. L'Évaluation économique des stratégies thérapeutiques, finalités, méthodes. *Arch Mal Cœur* 1989; 82 (suppl. III): 43-7.
 37. Anderson S, Bush JW, Chen M, Dolen D. Policy space areas and properties of benefit-cost/utility analysis. *Jama* 1986; 255: 794-5.

38. Launois R. Qu'est-ce qu'une régulation médicalisée? In *Evaluation et Régulation : l'évaluation économique du médicament au service d'une régulation médicalisée des dépenses de santé*. Paris, Collection Congrès et Colloques, SNIP, John Libbey Eurotext 1993 : 7-24.
39. Kaplan RM, Anderson J. The quality of well being scale : rationate for a single quality of life index. In : Walker S R, Rosser R, eds. *Quality of Life Assessment and Application*. Lancaster : MTP Press, 1988 : 51-77.
40. Torrance GW, Feeny D. Utilities and quality adjusted life years. *Int J Technol Assessm Health Care* 1989 ; 5 : 559-75.
41. Weinstein MC, Fineberg H *et al.* *Clinical Decision Analysis*. New York, NY : W B Saunders company, 1980.
42. Doubilet P, Weinstein M, McNeil B. Use and misuse of the term *cost effective* in medicine. *N Engl J Med* 1986 ; 314 : 253-5.
43. Eekhoudt L, Bauwens L, Lebrun T. Théorie de l'information et diagnostic médical : Une analyse coût-efficacité. *L'actualité Economique, Rev Anal Econom* 1987 ; 63 : 243-53.
44. Launois R. « Un exemple d'études médico-économiques : Le dépistage du cancer de la prostate ». In : *XX^e anniversaire de l'association Economie et Santé*. Paris, 1991.
45. Fagnani F, Landman R, Lafuma A. Etude coût-utilité de l'érythropoïétine dans le traitement de l'anémie du dialysé. *Quatrième colloque D.P.H.M.-INSERM*, Paris : édition INSERM, vol. 213, 1990.
46. Livartowski A, Pouillart P. Analyse coût-utilité en cancérologie, application à la chimiothérapie adjuvante dans le cancer du sein. *Med Sci* 1992 ; 10 : 1073-8.
47. Launois R, Henry B, Reboul-Marty J *et al.* Analyse coût-utilité à 7 ans du traitement de la hernie discale lombaire. *J Econom Méd* 1992 ; 10 : 267-85.
48. Castiel D. L'analyse coût-avantage : critère de choix des investissements médicaux et stratégies thérapeutiques? *Application à la thrombolyse précoce dans la prise en charge des maladies coronariennes*. Thèse Université Paris XII. Val de Marne 1991.
49. Leclercq A, Loïsançe D, Dumas Y. Analyse coût-efficacité de stratégies d'assistance pharmaceutique et mécanique dans le remplacement cardiaque en urgence. In Lebrun T, Saily J C, Amouretti M, eds. *l'Evaluation en Matière de Santé, des concepts à la pratique*. Lille : CRESGE, 1991-92.
50. Launois R. Cost-effectiveness analysis of strategies for screening prostatic cancer. In : Sweifel P, Frech III R. eds. *Second World congress on Health Economics Worldwide*. Klumer Academic publishers, 1992.
51. Durand-Zaleski I. L'actualisation dans les études coût-efficacité : difficultés méthodologiques ; application à la prévention secondaire des cardiopathies ischémiques. *J Econom med* 1992 ; 10 : 495-503.
52. Launois R, Launois B. Analyse coût-efficacité des stratégies thrombolytiques. *Arch Mal Cœur* 1989 ; 82 (III) : 55-62.
53. Launois R. Measuring benefits in hospitals : The choice of a strategy for thrombolytic therapy after acute myocardial infarction. In *Measuring the Benefit of Medicines ; the future agenda*, Londres : OHE, 1989.
54. Moatti JP, Le Galès C, Julian C *et al.* Analyse coût-bénéfice du diagnostic prénatal des anomalies chromosomiques par amniosynthèses. *Rev Epidémiol Santé Publ* 1990 ; 38 : 309-21.
55. Le Faure C, Le Galès C, Hirsch A. Radiodépistage de la tuberculose. *Concours Méd* 1987 ; 17 : 1584-8.
56. Le Galès C, Moatti JP. Cost-effectiveness of HIV screening of pregnant women in Hospitals of the Paris area. *Euro J Obs-Gyn Rep Biol* 1990 ; 37 : 25-33.
57. Le Galès G, Bougerol C. Apport d'un modèle économique de rationalité individuelle de l'évaluation des politiques publiques. L'exemple du dépistage néonatal de la drépanocytose en France métropolitaine. In *Colloque CREDES-CES : De l'Analyse Economique aux Politiques de Santé*. (Paris 16-18 déc. 1992).
58. Le Galès C, Moatti JP. Evaluation des stratégies de dépistage des hémoglobinoses en Provence-Alpes-Côte d'Azur. *J Econom Méd* 1989 ; 2 : 85-101.
59. Mason J, Drummond M, Torrance G. Some guidelines on the use of cost effectiveness league tables. *Br Med J* 1993 ; 306 : 570-1.
60. Analyse de l'Activité des Etablissements Hospitaliers en Budget global de 1987 à 1990. *Dossier et études statistiques n° 22*. CNAMTS.
61. Lévy E, Le Pen C. Le coût des maladies coronariennes. *Project Santé Futur* 1990 ; 2 : 135-48.
62. Fessler JM, Frutiger P. La Gestion Hospitalière Médicalisée, Paris : ESF, 1992.
63. Etude Nationale de Coûts par activité Médicale. Fiche 1 à 24. Architecture générale du modèle de calcul des coûts de référence. Ministère de la Santé et de l'Action Humanitaire. Direction des Hôpitaux. Mission PMSI. 1^{er} déc. 1992.
64. Viens-Bitker C, Lepoutre C, Blum-Boisgard C. Le coût de l'infection à VIH et pratiques médicales : l'exemple des patients séropositifs asymptomatiques. *Hop Paris* 1989 ; 110 : 35-7.
65. Viens-Bitker C, Blum-Boisgard C, Goldfarb B *et al.* Le coût de l'infection VIH : Méthode et résultats. *Rev Epidém Santé Publ* 1991 ; 39 : 25-36.
66. Launois R, Launois B, Reboul-Marty J *et al.* Le coût de la sévérité de la maladie. *J Econom Méd* 1990 ; 8 : 395-412.

67. Launois R. Chymonucléolyse : une alternative moins coûteuse que la discectomie. *Hospitalis Nouv* 1993; 206: 30-3.
 68. Seror V, Müller F, Moatti JP *et al.* Cost-benefit analysis of maternal serum screening for Down's syndrome using human chorionic gonadotrophin. *Prenat Diagn* (in Press).
 69. Reboul-Marty J, Till B, Launois R *et al.* Cost and nursing care of patients after long Propofol infusion versus isoflurane anaesthesia for gynaecological procedure. Laboratoire de santé publique, UFR Médecine Biologie, Bobigny, 1993.
 70. Goel V. Nonionic contrast media : economic analysis and health policy development. *Can Med Assoc J* 1989; 140: 389-95.
 71. Churchill D, Torrance G, Taylor W. Measurement of quality of life in end-stage renal disease : the time trade-off approach. *Clin Investig Med* 1987; 10: 14-20.
 72. Stason WB, Weinstein MC. Allocation of resources to manage hypertension. *N Engl J Med* 1977; 296: 732-9.
 73. Boyle M, Torrance G, Sinclair J *et al.* Economic evaluation of neo-natal intensive care of very-low-birth-weight infants. *N Engl J Med* 1981; 308: 1330-7.
 74. Weinstein MC, Stason WB. Cost-effectiveness of interventions to prevent or treat coronary heart diseases. *Annu Rev Public Health* 1985; 6: 41-63.
 75. Stason WB. Opportunities to improve the cost-effectiveness of treatment for hypertension. *Hypertension* 1991; 18: suppl. I : I-161-I-166.
 76. Moret L, Chwalow J, Baudoin-Balleur C. Evaluer la qualité de vie : construction d'une échelle. *Rev Epidem Santé Publ* 1993: 65-71.
 77. Churchill G. A paradigm for developing better measures of marketing constructs. *J Market Resear* 1979; 16: 64-73.
 78. Bombardier C, Tugwell P. Measuring disability : Guidelines for rheumatology studies. *J Rheumatol* 1983; suppl. 10: 68-73.
 79. Fermanian J. Problèmes méthodologiques et statistiques posés par l'utilisation d'échelles dans l'évaluation des troubles de la mémoire. In *Mémoire et Vieillessement : Approche Méthodologique*. Collection de l'Institut de Recherches Internationales. Servier. Doin, Paris 1989.
 80. Bombardier C, Tugwell P. A methodologic framework for developing and selecting end points in clinical trials. *J Rheumatol* 1982; 9: 758-62.
-