

« La qualité de vie : panorama et mise en perspective »

Robert LAUNOIS⁽¹⁻²⁾

Décision thérapeutique et qualité de vie. John Libbey Eurotext, Paris, 1992 :3-24

⁽¹⁾ REES France - 28, rue d'Assas - 75 006 Paris – Email : reesfrance@wanadoo.fr - Site Internet : <http://www.rees-france.com>

⁽²⁾ 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>

INTRODUCTION

La qualité du dialogue entre les pouvoirs publics et l'industrie du médicament se trouve altérée par l'incompatibilité des arguments mis en avant. Ainsi, les premiers font très souvent état d'un divorce entre le rythme d'augmentation des dépenses médicales et les progrès obtenus en matière d'allongement d'espérance de vie, alors que les seconds soulignent à juste titre que les objectifs de la médecine contemporaine consistent désormais à atténuer les conséquences de la maladie et à améliorer la qualité de vie. L'impossibilité de mettre en évidence une action positive des systèmes de santé provient du fait que l'instrument de mesure est inadapté, il convient d'en forger de nouveaux. Pour mesurer l'état de santé subjectif et son évolution, il faut trouver autre chose que les paramètres physiologiques. La première partie de notre exposé permettra de définir les concepts, dans un deuxième temps les outils disponibles seront décrits, enfin les qualités de mesure que doit présenter l'ensemble de ces instruments seront examinées.

LES CONCEPTS

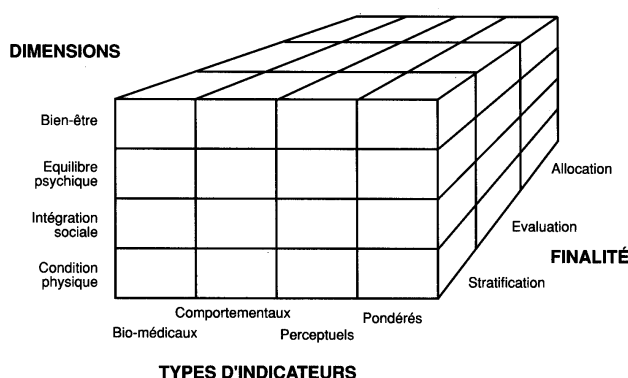
La première étape de toute étude sur la qualité de vie consiste à définir l'*univers* qui fera l'objet de l'analyse.

Une fois l'univers défini, il faut le catégoriser, c'est-à-dire définir les *dimensions* qui seront quantifiées.

Afin de pouvoir mesurer le contenu de ces dimensions, il faut avoir recours à un certain nombre de *critères* ou d'indicateurs qui permettent de les matérialiser et choisir des règles d'*étalonnage* adéquates.

Enfin, la mise en œuvre concrète d'un système d'indicateurs doit tenir compte des objectifs pour lesquels ils ont été construits, faute de quoi elle n'apportera que désillusion.

Fig. 1 - Les composants de l'analyse **qualité de vie**



□ Définition de l'univers et de ses dimensions

La qualité de vie est un concept unificateur tellement englobant que l'on pourrait à la limite tout y mettre : les facteurs d'environnement, les niveaux de revenus, les habitudes et les styles de vie. Nous en réduirons le champ en limitant notre étude à celle du retentissement de la maladie et de son traitement sur la qualité de vie des patients. Mais la vie ne peut

être évaluée en général ; le mieux que l'on puisse faire, c'est de porter un jugement sur ses différents aspects. Ce constat a un double intérêt : 1) il nous force à adopter une approche décomposée, ce qui n'est pas un mal étant donné le caractère abstrait du concept, 2) il nous incite à définir d'emblée les dimensions qui seront explorées. La catégorisation est une étape difficile, dans certains cas elle se fait *a posteriori* en fonction du positionnement respectif des dimensions de qualité de vie par rapport aux indicateurs utilisés par les malades pour les évaluer. Pour des raisons de commodité, 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 la 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.

□ **Choix des indicateurs**

Ces quatre dimensions ne sont jusqu'à présent que des concepts, c'est-à-dire des vues de l'esprit. Leur mesure doit être effectuée à travers des bases matérielles observables. Sur chacune d'entre elles, on met donc en place des critères ou des descripteurs qui serviront d'intermédiaires entre les caractéristiques abstraites qu'on cherche à cerner et les grandeurs mesurables, objectives ou subjectives, sur lesquelles ils prennent appui.

La nature des observations collectées dépend de l'approche choisie pour aborder les problèmes de santé. Pour certains, la santé, sans se réduire à la morbidité proprement dite, se limite au concept voisin d'anomalie biologique ou clinique. Pour d'autres, il convient de distinguer entre les affections étiquetées par les professionnels et les maladies exprimées en termes de comportements. Certaines définitions enfin privilégient l'abord perceptuel, c'est-à-dire fondé essentiellement sur les satisfactions ou insatisfactions des patients. A ces différents types d'approches correspondent des systèmes de recueil d'informations différents, qu'il importe d'exploiter simultanément si l'on veut étudier chaque dimension de la qualité de vie sous tous ses angles : 3 types d'indicateurs sont mis en place, bio-médicaux, comportementaux ou perceptuels.

L'intensité des symptômes, la lourdeur des incapacités ou le degré d'insatisfaction dépendent à la fois du degré de gêne et de l'importance que représente cette gêne dans la vie de tous les jours. Le poids des différents descripteurs qui permettent d'explorer une dimension devrait en principe être évalué par le patient lui-même. Mais, bien souvent, il est fixé par des observateurs extérieurs ou choisis par référence au comportement moyen de la population ; Parfois, même le problème est totalement masqué par le choix d'une équipondération implicite égale à l'unité (voir par exemple le score d'Apgar¹). Il convient de pouvoir tenir compte de l'indépendance ou de l'interdépendance de la qualité de vie. C'est pourquoi nous avons ajouté, sous la quatrième colonne, l'intitulé « pondération », que l'on pourrait qualifier d'indicateur de l'importance relative des critères. La nature exacte de ce dernier –équipondération implicite, échelle de valeur individuelle ou norme extérieure-, varie selon le type d'instrument disponible.

□ **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, ainsi que ses conditions d'utilisation, de telle sorte que la procédure puisse être répétée et ses résultats vérifiés. L'étalonnage^{2,3,4}, c'est-à-dire l'ensemble des règles qui président à l'attribution des 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 donner à l'instrument et détermine les manipulations statistiques dont celui-ci peut faire l'objet.

Lorsque les symboles numériques ne sont accompagnés d'aucune unité de mesure, ils ont un rôle purement descriptif.

Le chiffre peut alors être utilisé dans une simple procédure d'identification ; En ce cas, il correspond plus à un décompte qu'à une mesure. Les chiffres qui sont affectés aux diverses catégories pour les repérer peuvent être permutés sans dommage. Dans les résumés de sortie standardisés, par exemple, le chiffre 1 désigne les actifs, le chiffre 2, les inactifs, on ne perdrait aucune information si 1 signifiait l'inactivité et 2, l'activité.

Le chiffre peut également servir d'instrument de mise en rang. Il indique la position relative de l'indicateur sur un continuum représentant l'attribut qu'on cherche à cerner, ce qui suppose que celui-ci présente une caractéristique commune à des niveaux différents : je t'aime, un peu, beaucoup. Mais l'échelle n'étant pas étalonnée, les écarts entre les graduations ne sont pas comparables. Il est donc impossible de spécifier la distance qui sépare deux paliers de l'échelle, même si les nombres qui les représentent sont également espacés. On a une mesure d'ordre, mais les écarts entre deux niveaux de l'échelle et le rapport de ces niveaux ne peuvent être établis. La plupart des échelles de qualité de vie cliniques sont de ce type. Il est donc impossible de les utiliser pour mesurer le changement. Il faut mettre sur pied des échelles comparatives à catégories spécifiques : je vais mieux, un peu mieux, rien n'a changé.

Pour être plus que descriptifs, les symboles numériques doivent être accompagnés d'une unité de mesure. Le nombre, pour devenir une véritable mesure de grandeur, doit être mis en relation avec une unité standard. 2 tout seul n'a pas de sens, 2 mètres a une signification.

Dès lors que l'échelle est calibrée à l'aide d'une unité de mesure unique et constante sur toute sa longueur, les écarts de niveau peuvent être comparés entre un groupe témoin et un groupe traité par exemple, même si les rapports de niveau sont impossibles à obtenir en l'absence d'un zéro absolu. Il est souvent difficile, voire impossible, de démontrer hors de tout doute la disparition totale de la caractéristique de qualité de vie que l'on cherche à mesurer. Même s'il existe des situations de vie pires que la mort, on a du mal à imaginer un degré zéro de qualité dans de nombreuses dimensions.

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 paliers de l'échelle, le chiffre possède la structure de l'arithmétique ; les écarts entre 2 paliers ainsi que le rapport qui les unit peuvent être établis. Cette propriété essentielle de l'échelle métrique (encore appelée échelle de rapport ou échelle de proportionnalité) permet d'affirmer qu'un état de santé est 2 fois plus sévère qu'un autre, ce qui n'était pas possible précédemment.

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 *métrique* 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.

□ **Spécificité des outils**

Si les échelles sont des instruments de mesure, encore faut-il, pour quelles soient performantes, qu'on les utilise à bon escient, en d'autres termes qu'elles mesurent ce pour quoi elles ont été faites. L'outil qui permet d'identifier un problème n'est pas forcément celui qui permet de suivre son évolution et l'instrument apte à le faire peut fort bien se révéler inutilisable dans les procédures d'allocation de ressources. Le choix d'un instrument suppose une définition préalable des besoins de l'utilisateur : identification d'un problème, évaluation du changement sous traitement ou recherche d'une plus grande cohérence dans l'allocation des ressources rares.

✓ *Indicateur d'état, indicateurs de résultat*

Les différents paliers des indicateurs mesurent des niveaux de qualité de vie ; Ce sont des *indicateurs d'état*. Les changements de paliers enregistrent les effets différentiels de la maladie et du traitement. Puisqu'ils mesurent les modifications d'états, ils correspondent à des *indicateurs de résultats*. Il n'est pas sûr, comme le soulignent Kischner et Guyatt, que des indicateurs d'état puissent être utilisés comme des indicateurs de résultat⁵.

Les indicateurs d'état sont destinés à permettre la segmentation d'une population en sous-groupes en fonction des caractéristiques spécifiques de chacun des individus qui la composent, à un moment donné du temps, lorsqu'il n'existe pas de critère de référence pour les distinguer ; Il faut choisir pour les construire des indicateurs qui soient les mêmes pour tous : patients et bien-portants ; le nombre de leurs paliers peut être limité, voire même dichotomique : la caractéristique est présente ou absente. Pour s'assurer de la fidélité de l'instrument, il suffit de vérifier la stabilité des différences inter-individuelles au cours du temps. Toute évolution systématique et parallèle des scores est donc indécélable. De toutes façons, le score n'a aucune importance car il sert à classer les sujets dans des distributions.

La vocation des indicateurs de résultat est tout autre. Ils sont destinés à mesurer les modifications différentielles quantitatives de la qualité de vie ; leurs items sont sélectionnés en fonction de leur capacité à mettre en évidence des modifications ; le choix des réponses est multiple. La fidélité de l'outil est estimée à travers la stabilité des changements intra-individuels dans le temps lorsque le traitement n'est pas modifié.

L'évaluation du niveau de qualité de vie et celle des modifications de la qualité de vie exigent donc des instruments différents. La mise en œuvre d'un indicateur d'état dans le

cadre d'un essai randomisé est d'avance condamnée à l'échec, car ce type d'outil n'a pas été conçu pour permettre d'enregistrer l'impact d'un traitement.

✓ *Structure de la qualité, qualité globale*

L'économiste a besoin de résultats globaux⁶. Il lui faut disposer d'une commune mesure pour comparer les effets des choix les plus en amont du système de santé, ceux qui portent sur les techniques et les équipements utilisables par des populations statistiques moyennes. La démarche du médecin est différente. Il recherche dans le cadre des techniques et des équipements existants à faire le maximum pour celui qui se confie à lui. Il s'agit de prendre le pas sur tous les aspects de la maladie et cela explique qu'il tienne à rester dans une problématique multidimensionnelle. Ces deux états d'esprit correspondent à deux approches différentes des méthodes d'élaboration des questionnaires. La qualité de vie peut être appréciée de deux façons : par composition, en allant du particulier au général ou par examen global, en examinant d'emblée l'utilité totale, qui intègre automatiquement les évaluations partielles sans les catégoriser.

Dans la démarche de composition, on élabore un certain nombre d'indicateurs partiels à partir d'une combinaison simple ou d'une pondération ad hoc des variables choisies. Ces indicateurs partiels peuvent ou non être synthétisés sous la forme d'un score global. Lorsqu'ils subsistent en tant qu'entités distinctes dans l'instrument d'évaluation final, on parle de profils. S'ils font l'objet d'une agrégation en un chiffre unique, on utilise le terme d'indice ou d'indicateur synthétique. Dans tous les cas, le mode de présentation choisi reflète la complexité des situations. C'est la méthode qui a été employée dès le XIX^{ème} siècle par les psychométriciens lorsqu'ils s'efforcèrent d'imposer « la discipline de la mesure et du nombre aux choses de l'esprit ». C'est également celle dont se réclame Alvan Feinstein et les psychopathologues⁷ lorsqu'ils préconisent la codification des jugements cliniques. Un médecin, dans le secret de son cabinet, n'a pas besoin de questionnaire pour s'enquérir de la qualité de vie de son patient. L'écoute de ses plaintes suffit pour les découvrir et permet de les pallier. La situation est tout autre lorsqu'il s'agit d'évaluer un traitement sur un groupe de patients. Il convient alors d'avoir des instruments de mesure standardisés et fiables. L'approche médicale est très pragmatique, il s'agit moins d'explorer toutes les dimensions de la qualité de vie liées à la santé que d'étudier le retentissement spécifique de la maladie et des options thérapeutiques. Le champ exploré est plus ou moins exhaustif en fonction des circonstances et du traitement envisagé.

La méthode utilisée par les économistes et les tenants de l'analyse décisionnelle (Weinstein, Sonnenberg) est radicalement différente. Elle est globale d'emblée et elle suppose qu'il existe un continuum de qualité de vie allant de la bonne santé à la mort. 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 et de l'amour de la vie 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 mutilé 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). Le produit des années ou fractions d'années passées dans un état de santé donné et du coefficient de qualité de vie correspondant transforme le temps passé en mauvaise santé en fractions équivalentes d'années de bonne santé⁸. 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 corrigé en fonction de la qualité de vie appelé QALY⁹. Il suffit alors de diviser le coût du traitement par le résultat en QALY pour disposer d'un critère au regard duquel l'impact du traitement ou d'un non-traitement ou les mérites respectifs de deux traitements alternatifs peuvent être comparés. Bien entendu, numérateur et dénominateur sont actualisés puisque deux effets de santé identiques ou deux dépenses d'un même montant n'ont pas la même valeur dès lors qu'ils surviennent à un moment différent du temps. Les raisons en sont simples, une consommation immédiate est toujours préférable pour les individus à une consommation différée. Par ailleurs, les ressources qui ne sont pas consommées aujourd'hui pourraient toujours être investies ailleurs. Les coûts et les bénéfices à venir doivent donc avoir un poids moindre que ceux qui surviennent immédiatement.

✓ *Analyse instantanée ou suivi inter-temporel*

L'approche qualité de vie traditionnelle ignore la dimension temporelle puisqu'elle n'intègre pas les arbitrages quantité-qualité. Cela ne peut se justifier que si les thérapeutiques évaluées sont strictement comparables sur 3 points : 1) le risque de mort qui leur est associé, 2) la durée de vie totale, et enfin le temps de vie passé dans les différents stades d'évolution de la maladie pendant la période d'observation. Cela suppose que les deux thérapeutiques produisent leurs effets dans le même laps de temps (t_1), que cet effet soit absolument stable durant une même période (t_2), puis que la progression de la maladie jusqu'au décès final se déroule selon un schéma absolument identique, sur une même durée (t_3). Ces hypothèses faites *a priori* paraissent excessivement fortes. Le bilan différentiel des traitements mesuré en termes d'utilité permet en revanche de prendre en compte leurs effets à long terme. Si l'on dépasse le champ de la décision clinique pour s'intéresser à l'allocation des ressources, il est *a fortiori* nécessaire d'avoir un score unique à actualiser.

Deux types de situation sont à distinguer : ou bien les traitements peuvent être mis en œuvre simultanément et en toute indépendance dans des domaines médicaux distincts ou bien les traitements sont mutuellement exclusifs dans le cas d'une même indication. Dans la première hypothèse, l'algorithme de décision consiste à hiérarchiser les projets en fonction de leur rapport coût-performances moyen jusqu'à épuisement de la contrainte budgétaire. Qui dit classement dit liste. La publication inévitable de celle-ci fait apparaître l'incohérence des choix^{10, 11}. Plus les coûts par unité d'efficacité sont élevés, moins l'investissement correspondant semble justifié. Le développement de techniques moins onéreuses permettrait, à ressources égales, d'obtenir des résultats supérieurs en termes de santé publique.

On est conduit, dans le second cas, c'est-à-dire lorsque les traitements sont incompatibles, à scinder la stratégie de recherche en deux opérations successives^{12,13}. Le premier stade consiste à isoler, parmi toutes les stratégies envisageables, un ensemble de stratégies efficientes sur la base du principe de dominance. Dans un second stade, la société choisit, parmi toutes les stratégies efficientes, celle qui lui paraît la meilleure en arrêtant la somme qu'elle est disposée à payer pour obtenir le rapport surcoût/surcroît d'efficacité qu'elle juge optimal.

LES INSTRUMENTS EXISTANTS

L'évaluation de la qualité de la vie doit répondre aux objectifs de ceux qui y procèdent. Pour le médecin, il s'agit de dépasser une approche trop strictement biologique, certes utile en situation de détresse, mais qui ne joue en définitive qu'un rôle secondaire, une fois la survie assurée. Au-delà des déviations organiques, il convient d'écouter le langage du corps, mais tout jugement de valeur demeure suspect. Le malade, quant à lui, n'aura que faire d'une description exhaustive de ses plaintes et de ses gênes. Ce qui lui importe, c'est de pouvoir exprimer la diversité de ses attentes en fonction du poids qu'il accorde à tel ou tel aspect de sa vie. La collectivité, enfin, est à la recherche d'un code de cohérence qui reflète les priorités de la société et lui permet d'éclairer ses choix. Ces trois visions, qui relèvent de schémas conceptuels différents –clinimétrie, théorie de la décision, indicateurs de santé-, ont inspiré la construction des instruments existants. La place et le rôle tenus par les éléments de pondération qui y sont introduits sont au cœur de leurs différences.

□ Le regard des soignants

✓ *Echelles d'incapacité fonctionnelle*

L'évaluation de la dépendance des personnes âgées a donné lieu, au cours de ces dernières années, à l'élaboration d'une multitude de grilles qui permettent de mesurer les performances des individus à travers un certain nombre de fonctions essentielles à la survie. Elles reposent quasiment toutes sur une mesure effectuée par les soignants.

L'échelle de Katz¹⁴ établit un score global sur 6 activités : le bain, l'habillement, la toilette, la mobilité, l'incontinence, l'alimentation. Chaque item est évalué à l'aide d'un score à 3 niveaux.

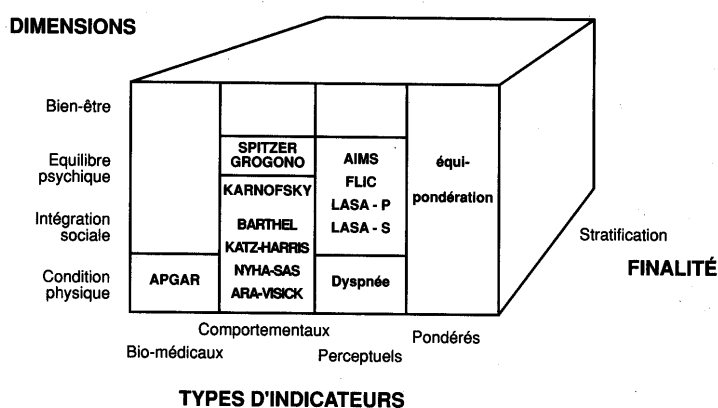
L'échelle de Harris¹⁵ explore la capacité de la personne âgée à effectuer les actes de la vie quotidienne : manger, boutonner ses vêtements, se déplacer, se coucher, prendre un bain, se laver, s'habiller, lacer ses souliers, se peigner. Deux types d'activité sont distingués, les activités principales et les activités secondaires. Cinq niveaux de sévérité sont fixés en fonction de la gêne ou de l'aide apportée. Gestes effectués 1) seul et sans difficulté ; 2) seul avec gêne ; 3) nécessairement avec une aide ; 4) impossibles à effectuer même avec une aide ; 5) avec gêne sans que l'on sache si c'est avec aide ou sans aide. Chaque item est coté en fonction de son niveau de sévérité, de 0 à 6 pour les activités principales, de 0 à 3 pour les activités secondaires. Les scores obtenus sont additionnés pour obtenir un indicateur global. L'accumulation des incapacités n'est donc pas jugée plus grave que le total de leurs inconvénients, alors qu'une défaillance simultanée dans plusieurs dimensions est toujours plus handicapante que la somme de celles qui apparaissent dans chacune d'elles.

La classification NYHA¹⁶ proposée par l'association des cardiologues de New York stratifie les pathologies du cœur et des vaisseaux en fonction de l'intensité des symptômes déclenchés par la pratique d'une activité courante ou inférieure à la normale. Quatre classes sont distinguées : absence de symptômes lors de l'activité courante (1), symptômes modérés lors de l'activité courante (2), fatigue, dyspnée, palpitations, douleur angineuse apparaissant pour des efforts plus faibles que ceux nécessités par une activité courante (3), symptômes se manifestant au repos (4). Très largement utilisée en pratique clinique et dans les essais randomisés, cette échelle ordinale laisse une part importante au jugement du médecin. Certains auteurs ont mis en évidence sa reproductibilité limitée. Sa validité a été également mise en cause, la corrélation avec la capacité fonctionnelle est

faible. Goldman a enfin montré qu'elle était très souvent entachée de subjectivité. Le classement NYHA s'améliore simplement parce que le patient renonce à des activités fatigantes. Pour pallier ces défauts, le même auteur propose un nouvel instrument : la Specific Activity Scale (SAS)¹⁷, les signes objectifs y sont privilégiés au détriment des symptômes. La capacité fonctionnelle des malades par rapport à certaines activités représentatives de l'univers quotidien de vie est calibrée en équivalents métaboliques ou « mets ».

Un des index de fonctionnement le plus souvent utilisé est celui de Karnofsky¹⁸. Il répond à trois questions : le malade a-t-il ou non conservé son travail ? Peut-il ou non vaquer à ses occupations habituelles ? Effectue-t-il ou non les gestes élémentaires de la vie quotidienne ? Les combinaisons de réponse déterminent trois stades de performances, étalonnées en 11 niveaux allant d'une activité normale (10) au décès (0). Les états fonctionnels repérés ne sont ni exclusifs ni exhaustifs, il en résulte des chevauchements et des situations impossibles à classer. Son ancienneté et une large diffusion en milieu médical expliquent qu'il soit encore mis en œuvre en dépit d'insuffisances démontrées.

Fig. 2 – Le regard des soignants



✓ *Les indicateurs spécifiques de morbidité étendue*

Tous les indicateurs de restrictions d'activité, pour les actes essentiels de la vie quotidienne, prennent en compte des niveaux de pertes d'autonomie trop sévères ou trop rares pour servir à l'ensemble de la population. Pour Stewart, 80 % de la population non institutionnalisée sont indemnes de toute atteinte fonctionnelle, alors que pour Kaplan et Bush, 50 % des personnes interrogées dans le cadre de l'enquête de San Diego présentent des troubles mineurs qui altèrent leur qualité de vie sans limiter pour autant leur autonomie ou leur mobilité. Pour apprécier les effets indésirables d'une maladie ou d'un traitement, il convient de retenir un concept de qualité de vie beaucoup plus large, intégrant à la fois les dimensions psychologiques et sociales et reposant sur les déclarations des malades eux-mêmes et non sur celles des médecins qui les traitent.

Priestman et Baum¹⁹ s'efforcent d'apprécier le retentissement d'un traitement du cancer du sein sur de telles bases. C'est la première tentative en oncologie d'utilisation d'échelles visuelles analogiques pour parvenir à une auto-évaluation de la qualité de vie par le patient (LASA-P). On demande au patient de se situer sur un segment horizontal ou

vertical dont les deux extrémités correspondent à l'absence ou à l'intensité maximale des phénomènes. « Le sujet hésitant entre les possibles se voit obligé de trouver une solution entre les tensions qu'exerce la représentation. Il doit, pour répondre, construire une sorte de « physionomie du phénomène » et par-là il le découvre et le révèle en même temps »²⁰. La technique a été appliquée sur des items de morbidité ressentie : humeur, anxiété, soulagement provoqué par le traitement. Chaque réponse est notée sur 10, le score global est égal à 100. Mise en œuvre dans un essai comparant un traitement hormonal et un traitement cytotoxique, elle a permis de montrer que les effets indésirables étaient plus élevés sous traitement cytotoxique que sous traitement hormonal, mais que la qualité de vie était meilleure dans le premier cas que dans le second du fait d'une réduction plus importante du volume tumoral.

Le *functional living index in cancer* (FLIC)²¹ répond aux mêmes objectifs : évaluer l'évolution des patients atteints d'un cancer, en dépassant l'approche fonctionnelle traditionnelle. L'élaboration du questionnaire a été extrêmement minutieuse. Des entretiens semi-structurés auprès d'un panel d'experts, comportant des malades et leurs conjoints, des médecins, des infirmières et un pasteur ont permis d'établir une liste de 250 questions. Après élimination des énoncés redondants ou mal formulés, un premier questionnaire de 92 items a été administré à 175 patients. Une analyse factorielle permit alors d'isoler les principales dimensions d'atteintes et d'éliminer 52 questions superflues. Le second questionnaire comportait 40 items, il fut testé auprès de 312 patients. A l'issue de ce passage, une autre analyse a été faite pour s'assurer de la stabilité de la structure factorielle. Elle aboutit à l'élaboration d'un troisième questionnaire comportant 20 items. 175 patients furent interrogés. Finalement, deux items supplémentaires ont été introduits et la version actuelle comporte 22 questions équipondérées. Les règles d'étalonnage retenues s'appuient sur l'utilisation conjointe d'échelles analogiques et d'échelles à catégories spécifiques. A chaque question correspond une échelle divisée en un certain nombre de cases, 4 à 7 selon les questions. Chaque sujet doit marquer d'un trait vertical l'emplacement qui lui paraît refléter le mieux sa situation actuelle. La valeur attribuée à la case qui est la plus proche du trait relevé sur l'échelle visuelle détermine le score obtenu pour l'énoncé. Le score global est égal à la somme des scores partiels relevés dans des dimensions hétérogènes, il est de nature ordinale.

En rhumatologie, la littérature anglo-saxonne a développé au cours des dernières années toute une série d'instruments qui permettent de mesurer le retentissement fonctionnel et psychique de la polyarthrite rhumatoïde : « Health Assessment Questionnaire » (HAQ)²² et « Arthritis Impact Measurement Scale » (AIMS)²³. L'échelle AIMS explore à l'aide de 55 questions 9 domaines : mobilité, autonomie locomotrice, dextérité manuelle, activités domestiques, soins personnels, vie relationnelle, anxiété, dépression, douleur. L'impact de la maladie se décompose en trois dimensions non superposables : incapacité physique, « état moral » et douleur. Aucun score global n'est établi.

✓ *Les batteries d'indicateurs*

Cette méthode est celle qui a été retenue notamment dans le cadre d'un essai en double aveugle ²⁴ qui se proposait de comparer trois traitements anti-hypertenseurs. Les investigateurs ont choisi *a priori* 5 dimensions de la qualité de vie, l'état physique, l'état émotionnel, l'efficacité intellectuelle, l'intégration sociale, le sentiment général de bien-être. Ces dimensions ont été explorées à l'aide d'indicateurs multidimensionnels indépendants. Une telle approche, bien qu'exhaustive, ne va pas sans poser un certain nombre de problèmes. Pour la mettre en œuvre, il faut tout d'abord découvrir des indicateurs validés et sensibles pour mesurer chacune des dimensions. Deuxièmement,

l'interprétation des résultats peut être délicate, en l'absence de critère principal de jugement, lorsque les performances sur des dimensions différentes ne varient pas uniformément dans le même sens. La mise en batterie d'indicateurs multiples est lourde de controverses potentielles.

□ **Le vécu des patients**

Les préférences des patients peuvent s'exprimer selon Goldberg soit « par l'affectation de pondérations aux résultats de mesures partielles, suivie ou non d'une agrégation par sous-dimensions ou dimensions, soit par la constitution *a priori* d'une liste ordonnée de scénarios de qualité de vie. Gérin^{25,26}, Guyatt²⁷ et Detsky²⁸ se rattachent au premier courant alors que les travaux de Torrance illustrent le second.

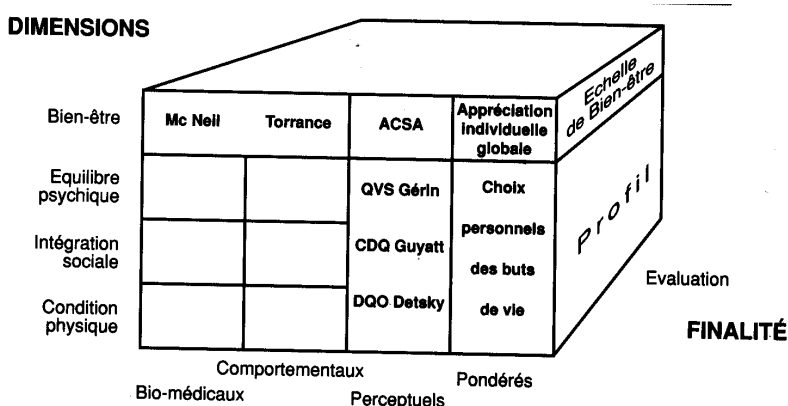
✓ *Profils de qualité de vie subjective*

Les tenants de la première école défendent une approche décomposée de la qualité de vie. Ils préconisent d'abord d'identifier les dimensions pertinentes à partir des données de la littérature et d'interviews d'experts. Les signes et symptômes repérés dans les essais et les verbatims des patients dont le corps médical se fait écho permettent de cerner l'impact d'une pathologie sur le vécu quotidien des malades. Mais ils ne sont que des descripteurs empiriques non opérationnels. Pour les rendre quantifiables, il convient de les étalonner en choisissant une échelle appropriée. Un questionnaire préliminaire doit alors, selon Guyatt²⁷, être construit. Sa rédaction répond à un double objectif, étalonner les indicateurs, sélectionner les plus pertinents d'entre eux. Compte-tenu du rôle instrumental qu'on lui fait jouer dans la préparation du questionnaire final, il doit comporter un nombre d'items bien supérieur à celui qu'il convient de retenir dans la version définitive. Pour chaque item, deux types de questions sont posées, la première se rapporte à la présence et à l'intensité de la gêne, la seconde à l'importance que lui accorde le patient. Deux méthodes d'analyse permettent de sélectionner les items pertinents. L'une consiste à retenir les items pour lesquels le produit de la fréquence et de l'importance des gênes ressenties est le plus élevé. L'autre permet d'éliminer, par une analyse en composantes principales, les items redondants, et de regrouper les énoncés en fonction de leur contribution aux différents facteurs dégagés. Mais une corrélation aussi forte soit-elle n'a par définition aucun contenu normatif. La première démarche est la plus appropriée, dès lors que le but de la recherche est de « connaître sur quoi se fondent les sujets pour apprécier la qualité de leur vie »²⁵. Guyatt opérationnalise la distinction faite par Gérin entre « valeurs centrales » en fonction desquelles le malade oriente sa vie, et « valeurs périphériques ». Seuls les items qui reflètent les premiers sont retenus dans le questionnaire final, tandis que les autres disparaissent. L'instrument d'évaluation intègre implicitement les préférences des patients, puisque ceux-ci sont à l'origine de sa structure, du choix de ses composantes et du poids respectif de celles-ci dans chacune des sous-dimensions explorées.

✓ *Echelle de bien-être personnel*

Selon Torrance²⁹, il convient au contraire d'obtenir des malades qu'ils explicitent directement leurs préférences par rapport à un éventail d'états de santé intégrant les diverses dimensions d'atteintes de la qualité de vie. Les outils de mesure utilisés pour les amener à révéler leurs préférences peuvent être très divers^{30,31,32} : jeux de hasard idéalisés (standard gamble), temps psychologiquement équivalents (time trade-off) et échelles à intervalles apparemment égaux (category rating).

Fig. 3 – Le vécu des patients



La première de ces méthodes est traditionnellement utilisée pour évaluer les préférences cardinales en situation d'incertitude. 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 risques 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 risques est choisie.

La solution du dilemme passe par l'introduction d'un jeu de hasard idéalisé obéissant à la loi de 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 risques, la survie sans séquelles 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 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 du coefficient de 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.

L'approche utilité/préférences présente plusieurs avantages. Tout d'abord cette méthode permet de disposer d'une mesure exhaustive combinant en un score unique à la fois la mortalité, la morbidité, ses diverses répercussions physiques, sensorielles, socio-émotionnelles, cognitives ainsi que les symptômes de la maladie 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 spécifiques 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. En second lieu, le score obtenu reflète directement les préférences du patient, il n'est pas dérivé à partir des 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 appropriés 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 n'est pas toujours aisé 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.

□ **Les préférences de la collectivité**

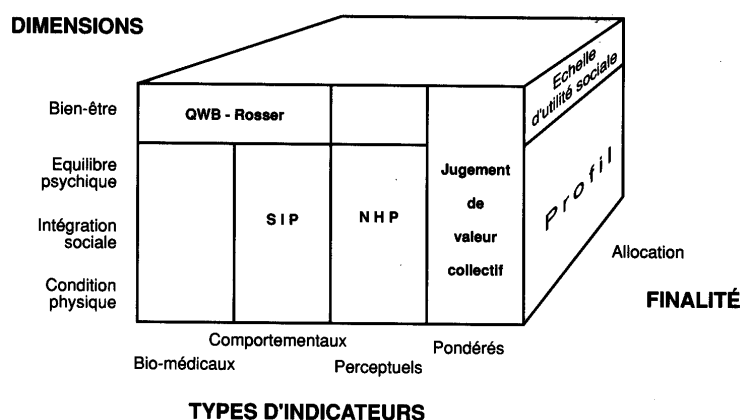
La mesure des préférences collectives fait appel à un groupe d'individus supposés représentatifs de l'intérêt général pour hiérarchiser les états de santé. L'intensité de la gêne est bien révélée par le malade, mais l'attention qui doit lui être prêtée est laissée au jugement des bien-portants. Quelle que soit l'approche retenue, profil de qualité ou mesure de l'utilité, l'appréciation qui est portée sur l'importance des items ou sur la désirabilité relative des états est le fait d'observateurs extérieurs.

✓ *Profils de qualité de vie normés*

Ils reposent sur l'utilisation d'un seul questionnaire auto-administré pour appréhender les divers aspects de la qualité de vie. Au contraire des multiples indicateurs qui peuvent être mis en batterie, le résultat de leur exploitation peut être agrégé en sous-score par dimension. Ce sont de surcroît des indicateurs généraux qui prétendent s'appliquer à toutes les pathologies. Les plus connus sont le Sickness Impact Profile (SIP)³³, et le Nottingham Health Profile (NHP)³⁴. Le SIP comporte 136 questions, regroupées en deux dimensions, fonctionnement physique et état psychique, et cinq catégories spécifiques indépendantes, le tout pouvant donner lieu à l'établissement d'un score global. Chaque affirmation du questionnaire décrit une modification du comportement et précise l'intensité de sa perturbation. La technique de l'échelle d'intervalles à degrés apparemment égaux a été utilisée pour déterminer la gravité relative de chacun des dysfonctionnements. Mise en œuvre en 1975 auprès de 108 adhérents d'un HMO de Seattle et de 25 professionnels de santé, elle a permis d'affecter à chaque item une note entre 0 et 15. Les scores par sous-dimension et le score global ont été calculés en divisant la somme des scores individuels par la somme maximale de notes possibles.

Le NHP utilise un questionnaire en deux parties, la première comporte 38 questions auxquelles on répond par OUI ou par NON, lesquelles se rapportent à 6 domaines : le sommeil, la mobilité physique, la douleur, les réactions affectives, l'isolement social, les réactions émotionnelles. La seconde partie rassemble 7 variables indépendantes : travail rémunéré, travail domestique, relations interpersonnelles, vie sociale, vie familiale, vie sexuelle, loisirs et vacances. Les réponses obtenues sont cotées 0 et 1. Les dimensions ne sont pas agrégées. Les items qui explorent chaque dimension sont par contre pondérés en fonction de leur gravité relative. La technique de référence utilisée est celle de la comparaison par paires. Chaque item dans une dimension a été successivement comparé à tous les autres items de la même dimension, en notant à chaque fois celui des deux qui était le plus difficile à supporter. Appliquée à un échantillon de 1 200 profanes sans formation médicale, cette méthode a permis d'observer la fréquence des items jugés plus graves que les autres. Les symptômes et les gênes ont alors été hiérarchisés à partir de la moyenne des écarts-type par rapport à ces fréquences.

Fig. 4 – Les préférences de la collectivité



Les profils ne sont pas dépourvus d'avantages, leur reproductibilité et leur validité sont bien établies. Par ailleurs, ils permettent d'appréhender les différentes dimensions de la qualité de vie en une seule fois sans avoir à recourir à de multiples échelles de mesure, ce qui est plus commode à la fois pour les investigateurs et pour les patients. Cependant, ils présentent les défauts de leurs qualités, notamment ils ne permettent pas d'étudier les conséquences spécifiques d'une pathologie donnée sur la qualité de vie. Par exemple, la dimension autonomie physique est explorée à l'aide de plusieurs énoncés qui cherchent à évaluer les déficits locomoteurs, et notamment la gêne à la marche. Or, dans les maladies veineuses, la marche, loin d'être un handicap, est considérée comme une thérapie, alors que la station debout immobile, dont il n'est jamais fait mention ni dans le NHP ni dans le SIP, est une véritable épreuve pour les insuffisants veineux. L'inadaptation des items par rapport au problème posé conduit inévitablement à un défaut de sensibilité, voire même de validité puisque les déficits fonctionnels explorés ne sont pas pertinents. L'arbitrage entre les dimensions pose par ailleurs problème. Dès lors que celles-ci n'évoluent pas dans le même sens, il faut pondérer les améliorations observées et les détériorations constatées. En l'absence d'un score global, l'appréciation d'ensemble est laissée à la subjectivité de l'évaluateur.

✓ *La mesure de l'utilité sociale*

La qualité de vie est appréciée pour chaque sujet au regard d'une norme arrêtée par un échantillon d'observateurs tirés de la population générale. La finalité de la mesure est moins d'apprécier « l'importance que chacun de nous attache à ce qu'il vit » que de disposer d'un indicateur global de morbidité et de mortalité au regard duquel les effets des actions de santé pourront être jugés. Les travaux de Bush³⁵ et de Rosser³⁶ répondent à cette logique. Ils fournissent aux économistes les outils dont ceux-ci ont besoin pour répartir des ressources données entre des projets concurrents.

Bush apprécie le retentissement de la maladie au regard de deux critères : l'incapacité fonctionnelle et les troubles subjectifs. L'incapacité fonctionnelle est explorée dans 3 dimensions : l'autonomie physique (PAC), la mobilité dans l'espace de vie (MOB), l'activité sociale (SAC). Les échelles correspondantes sont ordinales, elles comportent respectivement 4, 5 et 5 paliers. La première analyse l'autonomie du patient dans ses déplacements : mobilité active avec ou sans gêne, déplacement en fauteuil roulant, grabataire. Dans la seconde, l'accent est mis sur l'aire de circulation et la distance parcourue. La troisième évalue les fonctions sociales que l'individu est capable d'assumer. Bien entendu, ces fonctions varient pour chaque catégorie de la population. C'est par exemple le travail chez les actifs, la scolarité pour les moins de 15 ans, les passe-temps pour les retraités. Finalement, les 3 échelles précédentes sont combinées sans être sommées. Après élimination des cas impossibles, il subsiste 29 niveaux fonctionnels sur les 100 (4 x 5 x 5) initialement concevables. Une liste de 21 signes et symptômes vient compléter le dispositif. Elle permet d'intégrer des plaintes impossibles à rattacher à un diagnostic précis (coups de froid, fièvre), des syndromes plus ou moins vagues (maux de tête, vertiges), des affections incapacitantes, même si les individus qui en sont frappés ne les ont pas déclarées (lombalgies), des déficiences pures et simples (amputations). La combinaison des 29 niveaux fonctionnels et des 21 signes pour 5 groupes d'âge permet d'obtenir, après élimination des impossibilités, 343 cas-types. Chaque patient est affecté à un de ces cas, reste à les hiérarchiser pour obtenir les coefficients de qualité de vie afférant à chacun d'entre eux. La technique de révélation des préférences utilisée est celle de l'échelle d'intervalles apparemment égaux. 867 personnes issues du grand public ont été interrogées. On leur demandait de donner à chaque cas-type une note comprise entre 0 et 15. Après normalisation du score moyen attribué par l'ensemble des juges, on obtient

des coefficients de pondération compris entre 0 et 1 qui reflètent la désirabilité relative de chaque cas par rapport à la mort et à la bonne santé.

L'indicateur de Rosser procède des mêmes principes. Deux échelles sont utilisées, 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 de 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, immobilisations 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, 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 3 heures et demie 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 cas de figure envisagés. 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 les 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^{ième} état est calculée par rapport au n-1^{ème} sans qu'il soit nécessaire de procéder à une comparaison des n 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 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.

LES QUALITES REQUISES DES INSTRUMENTS

Pour qu'une mesure de la qualité de vie soit crédible, elle doit être pertinente, recevable, fiable, sensible et valide^{37,38,39,40,41,42}.

□ **Pertinence**

La pertinence de la teneur (« content value » des Anglo-Saxons) exige que soient remplies deux conditions : l'exhaustivité (l'ensemble de l'univers des atteintes doit être intégré) et la représentativité des items. 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 représentatifs de l'ensemble des énoncés concevables. Une mauvaise spécification de l'univers représente le premier type d'erreurs possibles, l'inadéquation entre l'instrument et l'univers qu'il est censé cerner. 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 scientifiques, la façon la plus simple d'y procéder consiste à multiplier la fréquence des gênes par leur importance. Mais on peut aussi recourir à des analyses plus sophistiquées comme l'analyse en composantes principales.

□ **Recevabilité**

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 au niveau des dimensions et sous-dimensions explorées ? 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 suffisante ?

□ **Fiabilité**

Une échelle est fiable si, en mesurant plusieurs fois le même phénomène, on obtient des résultats proches les uns des autres. Pour déterminer la fiabilité, on recherche l'importance de l'erreur aléatoire dans la mesure. Si celle-ci est faible, l'instrument mesure de façon constante l'univers exploré. Certains auteurs désignent ce critère sous l'appellation de reproductibilité, d'autres font référence à la précision de l'instrument. Il existe 3 façons d'évaluer la fiabilité : la cohérence interne, la fiabilité test-retest et la fiabilité inter-juges.

1. La cohérence interne : l'indicateur est cohérent lorsque ses différents éléments ne sont pas contradictoires. Cette cohérence est établie lorsque chaque item à l'intérieur d'un domaine et 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.

2. La fiabilité test-retest : elle est définie par la similitude des mesures successives, à des moments différents du temps, d'un même énoncé par la même méthode.
3. La fiabilité inter-juges : elle mesure le degré d'accord entre des observateurs différents d'un même événement. Les tests statistiques qui permettent d'évaluer cette fiabilité sont les coefficients Kappa pour les données ordinales et le coefficient de corrélation intra-classes pour les données continues.

□ **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. Guyatt⁴² 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 qui est utilisé doit présenter des scores pratiquement inchangés chez des sujets stables, c'est-à-dire qu'il doit être reproductible ;
- il doit être capable de mettre en évidence des 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 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 comment elle s'en démarque (validité divergente) ou s'en rapproche (validité convergente).

Trop souvent, la validation des instruments de mesure relève de l'intuition.

□ **Conclusion**

Le choix d'un indicateur dépend des réponses données aux 4 questions suivantes : A-t-on besoin d'un indicateur d'état ou d'un indicateur de résultat ? Veut-on mesurer la qualité de vie globale ou préfère-t-on s'intéresser à la structure de celle-ci ? L'instrument doit-il être utilisé pour le suivi des patients ou dans l'instant ? Quel est le point de vue retenu, celui du médecin, celui de la collectivité ou celui du patient ? Trop souvent, les outils disponibles sont utilisés à l'aveugle, sans que cette problématique ait été clairement posée.

REFERENCES

De nombreux travaux font le point de l'état de l'art. On recommandera aux lecteurs de langue française de lire en particulier : G. Petroni, P. Zweifel. « Comment mesurer la santé ». Collection Economie et Santé n° 14. Pharma-informations. Bâle, 1990.

M. Goldberg, W. Dab, J. Chapron, R. Fuhrer, F. Grémy : « Indicateurs de santé et sanométrie, les recherches récentes sur la mesure de l'état de santé d'une population ». *Revue d'épidémiologie et de Santé Publique*, 1979, 27 : 61-66 ; 133-152.

W. Dab, M. Goldberg, A. Thébaud, F. Grémy. « Un outil multicritères d'analyse des indicateurs de santé d'une population ». *Revue d'Epidémiologie et de Santé Publique*, 1982, 30 : 325-342. Ministère de l'économie et des finances. Direction de la comptabilité publique, direction de la prévision. Appréciation de la rentabilité économique des investissements. Note de méthode, janvier 1968.

¹ Apgar V. A proposal for a new method of evaluation of the new-born infants. *Curr Research Anesth Analg* ; July-Aug 1953 : 260-267.

² Stevens S. On the theory of scales and measurement. *Sciences*, 1946, 103 : 667-680.

³ Haski M., Moskowitz H. L'échelle sensorielle de Stevens-Moskowitz. *Rev Franç marketing*, 1980, 2, (8), 1:5-18.

⁴ Boss J. Quelques aspects de la mesure des attitudes : les échelles multi-dimensionnelles. *Rev Franç Marketing*, 1970, 34 : 23-44.

⁵ Kischner B., Guyatt G. A methodological framework for assessing health indices. *J Chronic Dis*, 1985, 38, 1 : 27-36.

⁶ Dupuy J.P. La science de décision en matière de santé : quelques éléments de réflexion. *Economie et Santé*, 1971, 3 : 2-29.

⁷ Guelfi JD, Bobon D. Echelle d'évaluation en psychiatrie. *Encycl. Med Chir Psychiatrie* 37200 A¹⁰ 10 ; 1989 : 1-10.

⁸ Patrick DL, Bush JW, Chen M. Method for measuring levels of well-being for a health-status index. *Health Services Research*, 1973 : 229-245.

⁹ Weinstein MC, Stason WB. Foundations of cost-effectiveness analysis for health and medical practices. *N Engl J Med*, 1977, 96 : 716-721.

¹⁰ Torrance G, Zipurky A. Cost effectiveness of antemartum prevention of RH immunization. *Clinics Perinatal*, 1984, 11 : 267-81.

¹¹ Williams AH. Is it a wild use of resources ? In Oliver M, Ashley-Miller M, Woods D (eds). Screening for risk of coronary heart disease. Wiley London, 1986.

¹² Doubilet P, Wainstein M, McNeil B. Use and misuse of the term « cost effective » in medicine. *N Engl J Med*, 1986, 314, 4 : 253-255.

¹³ Eisenberg J. Clinical Economics : a guide to economic analysis of clinical practices. *JAMA*, 1989, 262, 20 : 2879-2886.

¹⁴ Katz S, Ford AB *et al.* Studies of illness in the aged : the index of ADL : a standardized measure of biological and psychological functions. *JAMA*, 1963, 185 : 314-319.

¹⁵ Harris A. Handicapped and impaired in Great Britain. London HMSO. 1971. Cité dans Culyer A. Measuring health : lessons of Ontario. Ontario Economic Council, 1978.

¹⁶ The Criteria Committee of the New York Heart Association. In Disease of the Heart and Blood Vessel. Boston Little Broun, 1964.

¹⁷ Goldman L. Comparative reproductibility and validity of system for assessing cardio-vascular functional disease : advantage of a new specific activity scale. *Circulation*, 1981, 39 : 207-210.

¹⁸ Karnofsky DA, Abelman WH *et al.* The use of nitrogen mustard in the palliative treatment of carcinoma. *Cancer*, 1948 : 634-656.

¹⁹ Priestman T, Baum M. Evaluation of quality of life in patients receiving treatment for advanced breast cancer. *Lancet*, 1976 : 899-901.

²⁰ Moles A. Les sciences de l'imprécis. Editions du Seuil, Paris, 1990.

²¹ Shipper H *et al.* Measuring the quality of life of cancer patients : the functional living index. *Journal of Clinical Oncology*, 1984, 2, 5 : 472-485.

²² Guillemin F, Briançon S, Pourel J. Mesure de la capacité fonctionnelle dans la polyarthrite rhumatoïde : adaptation française du health assessment questionnaire (HAQ). Accepté pour publication *Revue du Rhumatisme*.

-
- ²³ Meeman RF. The AIMS approach to health status measurement : conceptual background and measurement properties. *Journal of rheumatology*, 1982, 9, 5 : 785-788.
- ²⁴ Croog SW, Levine S *et al.* The effects of anti-hypertensive therapy on the quality of life. *New England Journal of Medicine*, 1986, 314, 26 : 1657-1664.
- ²⁵ Gériin P, Dazord A *et al.* L'évaluation de la qualité de vie dans les essais thérapeutiques. Aspects conceptuels et présentation d'un questionnaire. *Thérapie*, 1989, 44 : 355-364.
- ²⁶ Gérin P, Dazord A *et al.* L'évaluation de la qualité de vie dans les essais thérapeutiques. In Pharmacologie clinique : actualité et perspectives. Strauch G, Husson JM (eds). Colloque Inserm, volume 185, 1989 : 159-181.
- ²⁷ Guyatt G. The questionnaire in the assessment of cardio-respiratory disease : the McMaster approach. Workshop on the assessment of the effect of drug therapy on the quality of life in cardio-respiratory disease. Broadway, Worcestershire 25-26 april 1985.
- ²⁸ Detsky A, McLaughlin J *et al.* Quality of life of patients on long-term total parental nutrition at home. *J Intern Med*, 1986, 1 : 26-23
- ²⁹ Torrance G, Feeny. Utilities and quality-adjusted lifeyears. *International J Techno Assessment Health Care*, 1989, 5 : 559-575.
- ³⁰ Gadreau M. Une mesure de la santé. Collection de l'Institut de Mathématiques Economiques n° 17. Sirey. Paris, 1978.
- ³¹ Launois r. L'évaluation économique des stratégies thérapeutiques. *Réalités Industrielles, Annales des Mines*. Juillet-Août 1991 : 81-86.
- ³² Launois R, Orvain J, Ounis I. Apport d'une mesure des utilités : Infections respiratoires récidivantes. *Rev Epidémio Santé Publ.*, 1992, 40 : 1-10.
- ³³ Bergner M, Bobbitt RA *et al.* The Sickness Impact Profile : conceptual formalisation and methodology for the development of a health status measure. *Intern J Health Services*, 1976, 2 : 393-415.
- ³⁴ Buquet D. Indicateur de santé perceptuel de Nottingham. Manuel d'utilisation. Inserm, Unité 164. Mai 1988.
- ³⁵ Cf. réf. 8
- ³⁶ Rosser RM, Kind P. A scale of valuation of states of illness, is there a social consensus ? *Intern J Epidemiol*, 1978, 7, 4 : 347-357.
- Churchill G. A paradigm for developing better measures of marketing constructs *J Marketing Research*, 1979, 16 : 64-
- ³⁷ Kaplan R, Bush JW, Berry C. Health status : type of validity and the index of well-being. *Health Services Research*, 1976 : 478-506.
- ³⁸ Churchill G. A paradigm for developing better measures of marketing constructs *J Marketing Research*, 1979, 16 : 64-73.
- ³⁹ Ware J, Brook R *et al.* Choosing measures of health status for individuals in general populations. *Am J Public Health*, 1981, 71, 6 : 620-625.
- ⁴⁰ Ware J. Standards for validating health measures : definition and content *J Chronic Dis*. 1978, 40, 6 : 743-480.
- ⁴¹ Israel L, Waintraub. Méthodes d'évaluation psychométriques en gériatrie. Le choix d'un instrument et ses critères de fiabilité. *Press Méd.* 1983, 12, 48 : 3124-3128.
- ⁴² Guyatt G, Walter S, et al. Measuring change over time. Assessinig the usefulness of evaluative instruments. *J Chronic Dis*, 1987, 40, 2 : 171-178.