L'ÉVALUATION MÉDICO-ÉCONOMIQUE

Concepts et méthodes

Claude Le Pen
Pierre Lévy

LGM *Sciences*

Ce livre est issu du cours d'évaluation économique en santé que nous avons assuré et assurons encore à l'Université Paris-Dauphine. Ce cours a été créé dès les années 80 par le Pr. Émile LEVY, un des promoteurs de l'évaluation médico-économique en France, auquel va toute notre gratitude. Il n'existerait pas sans la stimulation de nos étudiants que nous remercions.

L'ÉVALUATION MÉDICO-ÉCONOMIQUE

CONCEPTS ET MÉTHODES

Claude Le Pen, Pierre Lévy

SOMMAIRE

- *Préface* 5

Chapitre 1 : L'évaluation médico-économique : Pourquoi ? Pour qui ? 7

Chapitre 2 : Les principes de base de l'évaluation économique en santé 15

Chapitre 3 : Modèles coût-efficacité – Modèles d'impact budgétaire 29

Chapitre 4 : Identification et mesure des coûts 39

Chapitre 5 : La qualité de vie et sa mesure 53

Chapitre 6 : Les QALYs : Définition, calcul, interprétation 73

Chapitre 7 : La mise en œuvre du critère de l'efficience 93

Chapitre 8 : Évaluation médico-économique et modélisation 107

Chapitre 9 : Incertitude et analyses de sensibilité 129

Chapitre 10 : Quelles données pour l'analyse médico-économique ? 153

Chapitre 11 : Médico-économie, éthique médicale, et philosophies sociales 161

PRÉFACE

L'évaluation médico-économique s'est imposée ces dernières années, en France comme dans beaucoup d'autres pays, comme une dimension incontournable des politiques de santé. Il n'est guère de recommandations portant sur tel ou tel bien de santé ou, plus généralement, sur telle ou telle stratégie médicale à visée préventive ou thérapeutique, qui n'envisage sa dimension médico-économique, à des fins de décision ou d'information.

En 2008 en France, la Loi a élargi les missions de la Haute Autorité de Santé (HAS) en y intégrant la possibilité « *de publier des recommandations et avis médico-économiques sur les stratégies de soins, de prescription ou de prise en charge les plus efficientes* » alors que celle-ci avait été explicitement exclue trois ans plus tôt, à la création de l'Agence. Des dispositions du même ordre ont été prises dans beaucoup d'autres pays, même si cela survient dans des contextes nationaux différents.

L'ajout de la dimension économique à l'évaluation médicale n'est guère fortuit. Il correspond à la nécessité de faire face à l'inéluctable croissance des dépenses de santé qui, dans beaucoup de pays développés, absorbent environ 10 % de leur richesse nationale, voire davantage. Rechercher des « gains d'efficience », c'est-à-dire trouver les stratégies et techniques médicales les plus économiques pour atteindre un objectif sanitaire et social donné, est devenu un impératif quasi universel. Dans les pays où existent des budgets fermés – ou quasi fermés – des dépenses de santé, la recherche de l'efficience permet d'optimiser l'usage de ces ressources et d'obtenir, sous cette contrainte, les meilleurs résultats collectifs possibles en termes de santé publique.

Au-delà des aspects de politique de santé, l'émergence et le succès des approches médico-économiques correspondent à une vraie révolution conceptuelle, philosophique, éthique dans un monde de la santé qui n'a jamais fait bon ménage avec l'argent et le marché et pour qui le terme de « médico-économique » est presque un oxymore.

Or, cette nouvelle discipline académique qu'est devenue l'analyse médico-économique, avec ses méthodes, ses références, ses congrès, ses

revues scientifiques et ses enseignements, reste relativement mal connue des professionnels de santé en dehors des cercles spécialisés, au demeurant assez restreints.

L'objectif poursuivi dans cet ouvrage est de donner au non spécialiste – professionnel de la santé ou non, économiste ou non – un aperçu des concepts et méthodes qu'elle utilise.

Issu d'un enseignement à l'Université Paris-Dauphine où les auteurs ont exercé ou exercent encore, cet ouvrage se veut pédagogique et illustratif de la démarche d'évaluation médico-économique. Il ambitionne non seulement de décrire les outils mais aussi de donner des clés de compréhension quant à leur développement, leur portée, leurs implications.

Cet ouvrage n'est ni un ouvrage de recherche – il s'efforce de rester dans les limites imposées par l'état reconnu de la discipline – ni un guide méthodologique à destination des praticiens comme peuvent l'être les guidelines de la HAS en France ou du NICE au Royaume-Uni ; ceux-ci sont, en effet, souvent plus prescriptifs et normatifs qu'explicatifs.

Le projet était de concevoir un manuel relativement simple s'efforçant toutefois d'aller au-delà de la simple initiation. Certains chapitres sont nécessairement et naturellement plus techniques que d'autres, par exemple ceux relatifs à la modélisation ou aux analyses de sensibilité, mais nous nous sommes efforcés de rester le plus pédagogique possible, tout en donnant au lecteur un aperçu de la réalité de la discipline et des débats qu'elle suscite.

Nous n'avons pas non plus esquivé les questions éthiques que peuvent parfois soulever certains choix méthodologiques, voire la démarche médico-économique en elle-même.

Enfin, nous avons fait en sorte que chaque chapitre puisse être lu indépendamment des autres, serait-ce au prix de quelques redondances. Aller chercher ici et là des éléments d'information ou de compréhension à propos d'un terme, d'un concept, d'une méthodologie, est peut-être la meilleure façon d'entrer dans l'ouvrage.

Claude Le Pen & Pierre Lévy

Chapitre 1
L'évaluation médico-économique : Pourquoi ? Pour qui ?

Qu'est-ce que l'évaluation médico-économique ?

Le principe de l'évaluation médico-économique (EME) est simple : il s'agit d'établir le bilan des coûts et des avantages de différents produits, services ou programmes de santé, sur une période de temps donnée, de manière à orienter la décision publique vers les options qui entrainent le plus grand bénéfice collectif dans un monde soumis à une contrainte de rareté des ressources.

Il s'agit donc d'établir la « rentabilité » d'une intervention publique en santé, sans que celle-ci ne soit considérée sous un angle exclusivement financier : les avantages peuvent s'exprimer sous la forme de bénéfices non monétaires (diminution de la morbi-mortalité, augmentation de l'espérance de vie, de la qualité de vie, etc.) et ces derniers incluent des dimensions comme l'accès aux soins ou l'équité.

Le fait de confronter les coûts et avantages d'un choix de consommation ou d'investissement constitue une démarche qu'effectuent spontanément – et le plus souvent de manière implicite – tous les agents économiques en économie de marché. N'ayant que des conséquences privées, cette démarche peut rester informelle. Dès lors toutefois qu'il s'agit de choix publics, pris par des décideurs au nom de la collectivité, mobilisant des ressources publiques et entrainant des conséquences collectives, l'explicitation des termes du choix apparait comme une exigence démocratique.

Il est important de souligner que le bénéfice recherché est un bénéficie collectif, mesuré à l'échelle du système de santé, et non un bénéfice individuel, mesuré à celle d'un patient donné. Comme la santé publique ou comme l'épidémiologie, l'EME s'intéresse à des populations de patients et non à

des individus. Le bénéfice et les coûts sont toujours statistiques. C'est au praticien qu'il revient la tâche délicate de passer de la règle générale au cas particulier d'un patient donné, dans un environnement médical, social et économique donné.

La nature des produits et services susceptibles d'être soumise à l'EME est très variée. Elle comprend en tout premier lieu les biens de santé (médicaments et dispositifs médicaux) pour lesquels la Loi impose, dans certaines conditions, une évaluation économique obligatoire par la Haute Autorité de Santé (HAS) pour la prise en charge par l'assurance-maladie et la tarification[1].

Mais cela peut être des politiques de santé publique (coûts et bénéfices d'un dépistage organisé du cancer de sein), des schémas organisationnels (coûts et avantages de la chirurgicale ambulatoire), des institutions (coûts et avantages des hôpitaux locaux en zone rurale).

Enfin, il faut insister sur le fait que l'EME n'est pas un dispositif de maitrise des coûts de la santé. Une évaluation médico-économique peut justifier la mise en œuvre d'une intervention augmentant les dépenses de santé si cette augmentation-ci est compensée par un bénéfice collectif significativement supérieur. L'EME vise à rendre le système de santé plus « efficient » et pas nécessairement moins coûteux.

Une mise en œuvre délicate

Simple dans sa définition, l'EME l'est beaucoup moins dans sa mise en œuvre. Ni la définition ni la mesure des « avantages » (monétaires et non monétaires, immédiats et différés, médicaux ou non médicaux, attendus ou réels, etc.) ne vont de soi. Et il en va de même des coûts qui ne peuvent être appréciés que dans une « *perspective* » (un coût pour un agent peut être une recette pour une autre) ce qui nécessite parfois de distinguer entre prix et tarifs ou entre coûts proprement dits et transferts, les premiers mesurant une consommation de ressources matérielles, les seconds étant des paiements compensatoires sans contrepartie (par exemple les indemnités journalières en cas d'arrêt de travail).

À ces difficultés de base s'ajoutent celles, très discutées au sein de la profession des économistes, tenant à la prise en compte du temps (un paiement demain n'est pas équivalent au paiement de la même somme

1 On désigne parfois sous le nom de « pharmaco-économie », l'application de l'EME aux innovations pharmaceutiques. Mais cela ne vise que l'objet analysé et n'implique aucune spécificité méthodologique ou conceptuelle.

aujourd'hui) et de l'incertitude (un résultat clinique est toujours entaché d'une dose d'aléas) !

Enfin, la confrontation des coûts et des avantages passe souvent par la construction de « modèles » qui constituent des cadres de cohérence ordonnançant les différentes informations, médicales, épidémiologiques, institutionnelles, économiques, et permettant, par exemple, de tester différentes hypothèses ou niveaux d'intervention.

Au-delà de toutes ces questions techniques, l'EME n'échappe pas à des interrogations éthiques et morales, en raison notamment de ses effets redistributifs (entre « jeunes » et « vieux » par exemple) ou plus fondamentalement encore, en raison de la légitimité même du calcul économique dans un domaine, la santé, où, pour beaucoup, ne devraient prévaloir que des considérations fondées sur un impérieux devoir de solidarité, en dehors de tout calcul économique « rationnel ».

Tous ces aspects, techniques ou non, sont traités dans cet ouvrage.

Une branche sectorielle du « calcul économique »

L'EME constitue l'application à la santé de ce que les économistes appellent le « *calcul économique* », c'est-à-dire l'ensemble des concepts et techniques permettant à un agent économique rationnel d'optimiser ses décisions quand diverses options d'allocation de ressources rares lui sont offertes. Dans la santé comme dans d'autres domaines, les choix les plus fréquents sont des choix d'investissements et de tarification, mais tout choix peut faire l'objet d'un « calcul économique »[2] pour en mesurer l'impact et en évaluer la pertinence.

La plupart des questions posées ci-dessus, par exemple la définition des coûts, la prise en compte du temps et de l'incertitude ont été traitées, ont reçu des solutions générales, généralement dans des domaines d'intérêt collectif, avant d'être adaptées au cas de la santé.

Le « calcul économique » pour les choix publics ou collectifs a une longue histoire notamment en France où il a été porté par des générations d'« *ingénieurs économistes* », issus pour la plupart des grandes écoles d'ingénieurs (Polytechnique, Ponts et Chaussées, Mines, ENSAE). À la tête de grandes entreprises (SNCF, EDF) ou d'institutions publiques (Ministère des transports, Commissariat Général du Plan) ils ont eu très tôt recours, souvent

2 Pour une excellente introduction cf. Bernard Walliser : *Le calcul économique*, La Découverte, 1991.

de manière créative[3], à ces techniques pour optimiser les activités dont ils avaient la charge, essentiellement le transport et l'énergie – plus rarement la défense[4].

L'application à la santé a été plus tardive. En France, un élément fondateur a été le premier « Plan Périnatalité » en 1971, élaboré conjointement par le Commissariat général du plan et la « Mission R.C.B. » du Ministère de la santé[5] et qui comportait un objectif chiffré de réduction de 50 % de la prématurité et de la mortalité périnatale. Les mesures prises, essentiellement des mesures de suivi des grossesses et d'amélioration des conditions d'accouchement, avaient été sélectionnées sur la base d'une analyse coût-bénéfice très détaillée compte tenu des sources d'information alors disponibles.

Bien que ce plan fût un succès (le taux de « grands » et de « très grands » prématurés, moins de 32 semaines d'aménorrhée, est ainsi passé de 1,6 % des naissances vivantes en 1971 à 0,5 % en 1991, ce qui situe la France dans le « top 10 » du palmarès des nations[6]), le calcul économique en santé n'a guère prospéré, pas plus que dans d'autres domaines de l'action publique, jusqu'à son renouveau à partir de la toute fin du siècle dernier.

La crise de 1974, la volatilité accrue des marchés et de la conjoncture, l'internationalisation et la complexification des mécanismes économiques, l'échec des économies planifiées, les interrogations sur la notion « d'intérêt général » expliquent sans doute l'effacement d'une doctrine qui visait à l'efficacité de l'action publique à une époque où la légitimité même de cette dernière était mise en question au profit de politiques de dérégulation et de privatisation.

---

3 Certains ont marqué l'histoire de la science économique comme Augustin Cournot (1801-1877), Jules Dupuit (1804-1866), Léon Walras, (1834-1910), Clément Colson (1853-1939), René Roy (1894-1977), Pierre Massé (1898-1987), Marcel Boiteux (1922-), Edmond Malinvaud (1923-2015), Jacques Lesourne (1928-). Les deux prix Nobel d'économie français Maurice Allais (1911-2010) et Jean Tirole (1953-) s'y rattachent également.

4 Nous n'envisageons ici que le calcul économique public, sachant qu'en économie de marché, chaque agent fait individuellement et de manière décentralisée, parfois sans le savoir, un « calcul économique », dès lors qu'il est confronté à un choix économique sous contrainte budgétaire, qui peut être implicite et informel (dans la vie courante) ou explicite et formel (dans le mode des affaires).

5 La « Rationalisation des choix budgétaires » (RCB) était la démarche administrative qui a porté l'introduction du calcul économique dans les choix publics à la fin des années 60.

6 Deux autres plans « périnatalité 1994 » et « périnatalité 2005 » ont été depuis lancés pour reprendre une progression qui s'était stabilisée et pour traiter de « nouveaux » risques négligés jusqu'alors comme la morbidité psycho-parentale.

Le Royaume-Uni pionnier de l'EME

Pour différentes raisons institutionnelles et philosophiques, l'Angleterre, qui paradoxalement a été un des pays d'Europe qui a le plus dérégulé ses services publics, a développé et maintenu une large pratique de l'évaluation économique en santé, à la fois au plan théorique et pratique.

C'est d'Angleterre que sont venus, au tout début des années 80, les premiers travaux appliquant les techniques du calcul économique aux choix thérapeutiques, avec notamment l'article séminal de Anthony Culyer et de Alan Maynard montrant qu'un traitement par la cimétidine était plus « coût-efficace » que la chirurgie dans le traitement de l'ulcère gastroduodénal[7]. C'est en Angleterre qu'ont été popularisés – sinon inventés[8] – des concepts qui allaient faire référence comme celui de QALY (« *quality adjusted life years* »). C'est en Angleterre qu'a été créé, dès 1999, la toute première agence à vocation médico-économique, le *National Institute for Clinical Excellence* (NICE), dont les recommandations ont défini le « gold standard » de l'EME et dont le modèle a inspiré les agences de même nature créées par la suite en France (la HAS, en 2004) et en Allemagne (l'Institute für Qualität und Wirtschaftlichkeit in Gesundheitswesen, IQWIG, en 2004 également).

Plusieurs raisons expliquent cette antériorité britannique. La nature même du NHS, qui fonctionne dans une logique de service public (et non dans une logique assurantielle comme en Allemagne ou en France), une tradition planificatrice remontant aux origines mêmes du NHS, la philosophie utilitariste et « welfariste » qui domine la sphère sociale en Grande-Bretagne, la longue et forte liaison entre le NHS et les milieux académiques, avec notamment l'Université de York, d'où sont issus de nombreux auteurs ayant contribué à l'élaboration doctrinale et pratique de l'EME (Alan Maynard, Alan Williams, Michael Drummond, Anthony Culyer).

Au-delà du cas britannique, l'EME s'est développée à partir des années 90 dans beaucoup de pays développés, y compris ceux où étaient mis en place dans d'autres secteurs, des politiques de dérégulation. La raison en est la croissance forte et continue des dépenses publiques de santé qu'ont connue ces pays et le défi que cette croissance posait au financement public. Il est apparu qu'il était possible d'optimiser dans une large mesure les

7 AJ Culyer, AK Maynard. Treating ulcers with cimetidine can be more cost-effective than surgery. *Medeconomics* 1980, 1, 12-14 et AJ Culyer, AK Maynard : Cost-effectiveness of duodenal ulcer treatment, *Social Science and Medicine*, 5C, 3-11, 1981.

8 Milton C. Weinstein est souvent crédité de l'invention des QALYs (Cf. Weinstein MC, Stason WB. Foundations of cost-effectiveness analysis for health and medical practices. *N Engl J Med* 1977 ; 296 : 716-21).

dépenses en sélectionnant ou en recommandant les interventions médicales les plus « efficientes ».

Un concept-clé : l'efficience

Le concept « d'efficience » est le concept clé de l'EME. Il désigne le rapport entre les résultats et les coûts d'une intervention médicale ; on montre en théorie – mais cela est assez intuitif – que la sélection des interventions les plus efficientes permet d'obtenir un bénéfice collectif maximal à partir d'un engagement donné de ressources. Plus d'avantages avec un montant donné de ressources ou moins de ressources pour un avantage donné.

L'efficience n'est pas l'efficacité : une intervention efficace médicalement peut ne pas être « efficiente » si son coût apparaît prohibitif. Inversement une intervention efficiente peut ne pas être très efficace : donner à un patient le conseil d'arrêter de fumer est une intervention dont le coût est pratiquement nul et il suffit qu'un tout petit nombre de patients le suive pour que le rapport résultats/coût soit pratiquement infini.

L'efficience ne garantit pas non plus la solvabilité du financeur face à une intervention médicale. Ces dernières années ont vu apparaitre des innovations pharmaceutiques offrant la triple caractéristique d'être beaucoup plus efficaces que le « standard of care », de coûter beaucoup plus cher et de concerner un très grand nombre de patients[9]. Quoique le rapport coût/bénéfices (mesuré en termes de coût par année de vie sauvée) de ces interventions soit généralement considéré comme tout à fait acceptable, les sommes à engager pour généraliser le traitement à la population cible peuvent apparaitre excessives au regard des ressources disponibles.

En tant que rapport entre deux grandeurs, l'efficience est en effet un concept sans dimension. Comme un pourcentage est indépendant du nombre auquel il s'applique, une intervention médicale peut être efficiente sans considération du nombre de patients concernés.

C'est la raison pour laquelle l'EME tend de plus à traiter conjointement les questions d'efficience avec des modèles de type coût-efficacité et de solvabilité avec des études d'impacts budgétaires de manière à capturer les deux aspects de l'efficience et de la solvabilité. Nous y reviendrions dans le cours de cet ouvrage.

9 On pense bien entendu aux antiviraux d'action directe dans l'indication de l'hépatite C.

D'une discipline académique à une pratique administrative

L'EME offre cette caractéristique d'être simultanément une discipline académique et une pratique administrative. De ce point de vue, elle ne se distingue guère de la recherche clinique où ces deux dimensions sont également présentes.

Scientifiquement, l'EME s'est développée à partir des années 80, en une véritable discipline académique qui s'est progressivement autonomisée au sein même de l'économie de la santé, avec ses spécialistes, ses enseignements, ses congrès nationaux et internationaux, ses revues scientifiques à comité de lecture...

Administrativement, la production d'une étude d'évaluation économique est devenue une condition pour l'accès au marché des biens de santé dans beaucoup de pays développés (Royaume-Uni, Pays-Bas, Suède, Écosse, Australie, Canada, Italie, France...). Souvent produites par les producteurs de soins, ces études dont le statut varie selon les pays, sont analysées de manière critique par les agences d'évaluation des technologies médicales (« Heath Technology Assessement ») et peuvent servir de base à des décisions publiques.

Cette dimension administrative a entrainé un travail de normalisation et de standardisation des méthodologies. La HAS a ainsi publié plusieurs guides méthodologiques aussi bien pour les études coût-efficacité que pour les études d'impact budgétaire. L'objectif de cette normalisation est d'assurer un bon niveau de qualité et de comparabilité entre les différents travaux.

Les trois usages administratifs de l'EME

Si les travaux menés dans une perspective scientifique n'ont généralement d'autre but que d'apporter des informations et des connaissances nouvelles, ceux qui sont soumis aux agences d'évaluation, soit dans le cadre d'une démarche spontanée, soit dans celui d'une obligation administrative, obéissent à des objectifs bien précis mais variables selon les pays.

Très schématiquement, l'EME obéit en pratique à trois usages principaux, de rationnement, de tarification, et d'optimisation.

Le premier consiste à conditionner l'usage d'un bien ou service médical, voire même son accès au marché, à une condition d'efficience minimale. On sait que le NICE britannique a édicté une norme de coût minimal par année de vie sauvée (pondérée par la qualité) de l'ordre de £ 30.000. Le financement d'un bien ou d'un service médical dont le ratio cout-efficacité dépasse nettement cette norme est considéré comme un gaspillage de

ressources publiques. Cet usage très radical de l'EME est toutefois assez peu répandu dans le monde en dehors du cas emblématique du Royaume-Uni et de pays comme l'Australie ou l'Écosse.

Dans beaucoup de pays, dont la France et dans une certaine mesure l'Allemagne, l'EME est utilisée dans la négociation des prix régulés qui intervient entre les producteurs et les autorités publiques soit au moment de la mise sur le marché des biens et services concernés (en France) soit quelques mois plus tard (en Allemagne). Le but des autorités publiques est alors de s'assurer que le prix octroyé est proche « d'un prix d'efficience » jugé acceptable au regard des pratiques existantes, sans que soit défini un seuil d'efficience, sans non plus que la condition d'efficience ne soit impérative. Cela entre dans le cadre du principe de « value based pricing » qui veut que le prix d'un bien ou d'un service reflète sa valeur thérapeutique.

Enfin, il existe un troisième usage administratif de l'EME consistant à l'utiliser pour définir les conditions d'emploi des produits, par exemple à définir les populations cibles pour lesquels le produit concerné apportera à la collectivité un gain d'efficience.

Chapitre 2
Les principes de base de l'évaluation économique en santé

L'évaluation économique des technologies de santé est spécifique du fait qu'elle porte sur des interventions dont les effets s'exercent sur la santé des individus qui, outre les aspects humains et éthiques qui peuvent leur être associés, ne sont pas faciles à mesurer. Pour autant, du fait du financement largement socialisé des dépenses de santé, cette démarche trouve ses racines dans l'économie publique qui a développé un cadre d'analyse pour l'évaluation économique des politiques publiques dans la lignée du calcul économique appliqué[1]. Le principe fondamental est ainsi d'utiliser la rationalité économique pour éclairer les choix publics.

L'évaluation économique en santé : une évaluation bi-dimensionnelle

La rationalité économique s'exprime par la recherche d'une adéquation entre les objectifs ou les effets d'une action et les moyens qu'elle mobilise, dans le

1 Pour une présentation extensive des bases du calcul économique et de leur application à la santé, on peut se reporter à différents ouvrages, notamment :
–Brent R. *Cost-Benefit analysis and health care evaluations*. Edward Elgar, 2003.
–McIntosh E, Clarke P, Frew E, Louviere J (editors). *Applied Methods of Cost-Benefit Analysis in Health Care.* Oxford University Press, 2010.
Il existe également quelques manuels de référence sur l'évaluation économique des stratégies de santé. On citera en particulier :
–Drummond M, Sculpher M, Claxton K, Stoddart G, Torrance G . *Methods for the economic evaluation of health care programmes.* 4th edition, Oxford University Press, 2015.
–Drummond M, McGuire A. (Eds). *Economic evaluation in health care*. Oxford University Press, 2001.
–Annemans L. *L'économie de la santé pour non économistes – une introduction aux notions, aux méthodes, et aux écueils de l'évaluation économique en santé.* Academia Press, 2008.
–Kobelt G. *Health economics: an introduction to economic evaluation*. 3rd edition, Office of Health Economics, 2013.

but de faire le meilleur usage possible des ressources disponibles, par hypothèse limitées. Pour s'assurer de l'usage optimal de ces ressources, on met donc en relation le montant des ressources utilisées et les effets obtenus de façon à mesurer le degré d'efficience d'une intervention. La finalité de cette démarche est d'éclairer un choix possible entre plusieurs actions qui mobiliseraient les mêmes ressources dans des proportions variables, en identifiant celle qui, à coût donné, permet d'obtenir la plus grande quantité d'effet ou, alternativement, celle qui permet d'atteindre un niveau d'efficacité donné à moindre coût.

Bien évidemment, l'efficience n'est que l'un des critères à partir desquels on peut évaluer une action de santé :

- Dans une perspective sociétale, le critère de l'équité s'intéresse aux éventuelles disparités entre les patients, qu'il s'agisse de l'accès aux soins, de la couverture assurantielle et de la participation au financement par le reste à charge, ou de l'état de santé par exemple.
- Dans une perspective clinique, l'accent est plutôt mis sur l'efficacité qui se focalise sur les effets obtenus par différentes interventions médicales.
- Dans une perspective économique, l'efficience (*efficiency*) considère le coût nécessaire pour atteindre une certaine efficacité. Cette combinaison des aspects de coût et d'efficacité donne une nature bi-dimensionnelle à l'évaluation économique qui est donc du type « coût-efficacité ».

L'efficience de toute intervention de santé doit, de ce fait, être appréhendée en rapportant ses effets au coût des ressources qu'elle utilise. L'évaluation médico-économique consiste bien à développer une analyse de type « coût-efficacité » des différentes actions de santé possibles pour en estimer le degré d'efficience respectif, voire effectuer un classement en fonction de la performance des interventions possibles sur ce critère de l'efficience. Une définition formelle pourrait en être que l'évaluation économique des programmes de santé est une analyse comparative des coûts et des conséquences de stratégies thérapeutiques alternatives.

Cette définition souligne les deux conditions *sine qua non* de l'évaluation économique :

- D'une part, il est nécessaire d'organiser une comparaison entre des stratégies concurrentes dans une perspective de choix et d'aide à la décision, puisque décider c'est choisir. Il peut s'agir soit d'interventions alternatives pour traiter ou prévenir la même maladie ou facteur de risque qui sont donc exclusives les unes des autres, soit de programmes de santé indépendants car concernant des affections et des patients différents mais qui peuvent entrer en compétition sur les mêmes ressources.

- D'autre part, il faut prendre en compte simultanément les coûts et les conséquences des traitements pour caractériser l'efficience économique. Cette mise en relation entre les effets obtenus et les moyens employés par les différents usages des ressources disponibles est une façon de mesurer le rendement social de différents investissements possibles d'un budget collectif. La mesure des coûts donne le montant de l'investissement à réaliser et l'estimation des effets fournit une indication du bénéfice sanitaire obtenu en contrepartie.

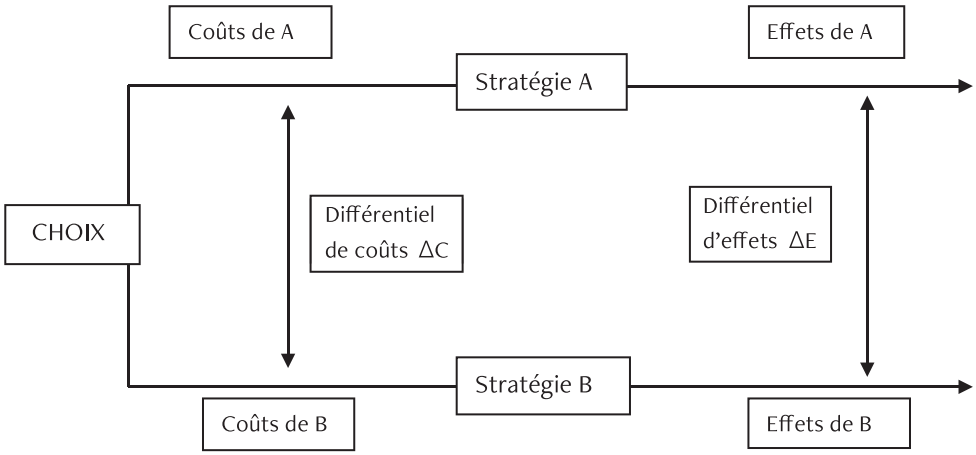


Figure 1. Structure type de l'évaluation médico-économique.

Une représentation graphique permet d'illustrer les aspects essentiels de cette définition (*figure 1*). La plupart du temps, cette approche est mise en œuvre pour déterminer l'efficience relative d'une innovation thérapeutique, la stratégie B ici, par rapport à la pratique courante, la stratégie A. Pour instrumentaliser cette notion d'efficience relative, il faut mesurer dans un premier temps les coûts attachés respectivement à la stratégie A et à la stratégie B, soit C_A et C_B, ainsi que leurs effets respectifs E_A et E_B. On peut dans un second temps former le ratio différentiel coût-résultat (RDCR) à partir du différentiel de coût ΔC (soit $C_B - C_A$) et du différentiel d'effet ΔE (soit $E_B - E_A$) :

$$\text{RDCR} = \frac{\Delta C}{\Delta E}$$

Il reste à dire que la formation de ce ratio suppose qu'on spécifie d'une part les modalités de mesure des coûts, et d'autre part le choix d'une

mesure des bénéfices thérapeutiques compte tenu de la diversité des mesures de résultats, ce que nous ferons un peu plus loin.

Précisons que, si cette représentation graphique permet d'illustrer la structure fondamentale de l'évaluation médico-économique, elle pourrait laisser croire que la comparaison doit nécessairement être menée deux à deux. Il n'en est rien et, comme nous le verrons plus en détails au chapitre 7, cette démarche peut tout à fait être mise en œuvre pour comparer, de façon généralisée, un nombre quelconque de stratégies alternatives.

Le champ d'application de l'évaluation économique

La démarche de l'évaluation médico-économique s'applique à de très nombreux types d'actions dans le domaine de la santé, et on utilise de façon synonyme les termes de programmes de santé, stratégies médicales, interventions sanitaires, technologies de santé, ou d'autres, sans que cela en modifie les caractéristiques. Elle peut ainsi être mobilisée dans les cas suivants :

- Toute intervention de santé, qu'elle soit préventive ou curative, médicale ou chirurgicale. Et il est tout à fait possible de comparer une stratégie de prévention à une stratégie purement curative, une approche médicale à une approche chirurgicale.
- Tous les produits de santé peuvent être candidats à cette approche, en particulier les médicaments mais aussi les dispositifs médicaux. À cet égard, il convient de souligner que si des produits de santé peuvent être réglementairement soumis à une évaluation économique pour leur accès ou leur maintien sur le marché, on n'évalue jamais des produits en tant que tels. Il s'agit en fait de stratégies thérapeutiques qui utilisent des produits de santé selon certaines modalités de telle sorte qu'on peut souvent définir plusieurs stratégies recourant au même produit[2].
- Toute technique médicale, qu'il s'agisse d'imagerie médicale (IRM), d'un plateau technique de greffe, ou de différents usages de la télémédecine.

2 On peut par exemple évaluer une stratégie médicamenteuse consistant à ajouter un médicament adjuvant à la thérapie standard (dans l'insuffisance cardiaque, l'ajout d'un bêta-bloquant au traitement standard, soit IEC, diurétique ± digitalique), ou différentes stratégies vaccinales utilisant le même vaccin selon que l'on injecte trois doses aux nourrissons ou une seule dose aux jeunes enfants. De façon similaire, on peut définir pour certains anti-cancéreux différentes stratégies selon qu'on les administre d'emblée à tous les patients ou qu'on teste ces patients par certains marqueurs pour ne traiter que les répondeurs compte tenu d'une signature génétique particulière.

- Toute action d'information, de dépistage, de prévention, de traitement ou de prise en charge organisée dans un parcours de soins peuvent faire l'objet d'une évaluation économique.

Dans les faits, l'évaluation médico-économique a plus fréquemment concerné les produits pharmaceutiques, conduisant à parler de « pharmaco-économie » pour désigner l'évaluation économique des médicaments. Mais il n'y a pas de différences fondamentales avec les autres sujets d'application et l'usage de ce terme peut être étendu sans dommage à tous les programmes de santé.

Le choix du comparateur

Dans tous les cas, il s'agit d'évaluer l'efficience d'une ou plusieurs stratégies particulières en les comparant entre elles ou par rapport à une stratégie de référence. Et le caractère comparatif de la démarche n'est pas limitatif car il existe toujours au moins un comparateur à toute innovation thérapeutique qui est la stratégie « ne rien faire » correspondant en réalité aux soins de support. En revanche, lorsqu'il existe déjà des alternatives thérapeutiques à une stratégie qu'on veut évaluer, l'un des problèmes à résoudre concerne le choix du comparateur, dans la mesure où celui-ci influence bien sûr les résultats de l'évaluation. Par rapport à une intervention coûteuse et peu efficace, il sera plus facile d'obtenir des résultats favorables à une innovation thérapeutique.

On pourrait de la sorte introduire un biais dans la comparaison et ceci explique qu'il existe assez souvent des recommandations quant au choix des stratégies à comparer. Il est souvent recommandé d'inclure parmi les stratégies évaluées une stratégie de référence, correspondant si possible à la pratique courante. En effet, si l'on est dans une logique d'accès au marché d'une technologie innovante, il est pertinent, notamment du point de vue du payeur, d'estimer si la généralisation de cette innovation serait un choix efficient ou pas. Nous verrons plus en détail au chapitre 7 comment on peut passer de l'évaluation à la décision.

La diversité des types d'études

Si, sur le plan des principes, l'évaluation médico-économique se ramène de façon assez simple à rapporter le différentiel de coûts entre deux stratégies médicales alternatives à leur différentiel d'effets, cette démarche peut se décliner de multiples manières compte tenu des choix méthodologiques qui doivent être faits pour réaliser une telle étude. Sans chercher ici à faire la liste exhaustive des options possibles, on peut néanmoins souligner quelques aspects essentiels de cette diversification de la démarche.

Nous présenterons en détail au chapitre 4 toutes les considérations utiles à la mesure des coûts dans une étude d'évaluation médico-économique. Disons simplement ici que la comparaison de différentes études entre elles peut être rendue délicate si elles ont été établies sur la base de choix différents quant aux coûts considérés.

En effet, on peut, par exemple, faire le choix de retenir des types de coût de nature différente, en ne considérant que des coûts médicaux ou en incluant des coûts correspondant à une consommation de ressources non médicales (aide sociale, transport). De même, le périmètre des coûts peut varier selon l'ampleur choisie en restreignant l'évaluation aux coûts hospitaliers pendant la phase aigüe d'une maladie ou en tenant compte des coûts ambulatoires subséquents. Enfin, comme on le précisera plus loin, la perspective adoptée pour mesurer les coûts en modifie le montant, en particulier selon qu'on ne considère que le point du vue du « payeur » (l'assurance maladie) ou qu'on opte pour une approche plus collective, incluant le reste à charge des patients.

La diversité de la mesure des effets

Si la mesure des coûts laisse ouverte plusieurs options introduisant une certaine variabilité, la diversification de l'évaluation économique en santé tient pour l'essentiel aux choix possibles pour mesurer les effets des stratégies prises en compte. De ce point de vue, il est usuel de distinguer trois types d'évaluation qui ont chacun leurs spécificités :

- L'approche coût-efficacité (ACE) prend en compte un indicateur de résultat qui exprime dans une métrique naturelle l'effet physique de chaque intervention. Il peut s'agir d'un indicateur pharmacologique, biologique ou d'évènements cliniques relevant de la morbi-mortalité (la fréquence de complications précisées ou la durée de la survie par exemple). Souvent utilisée dans le passé car pouvant puiser la mesure de l'efficacité aux sources de l'*Evidence-Based Medicine*, elle est considérée dorénavant comme limitée par le fait qu'elle ne peut prendre en compte qu'une seule dimension d'effet.
- L'approche coût-utilité (ACU) est plus sophistiquée que la précédente en ce qu'elle cherche à mesurer l'effet des interventions sur le bien-être des patients, à travers un indicateur composite qui appréhende à la fois la quantité de la survie et la qualité de cette survie. On peut dire que l'ACU va au-delà de l'ACE du fait qu'elle pondère la survie gagnée grâce à une innovation thérapeutique par la qualité de vie associée. Alternativement, elle permet de tenir compte, à travers la même métrique, d'un

éventuel différentiel de survie et/ou de qualité de vie entre stratégies alternatives, combinant ainsi une mesure objective et une mesure subjective de résultat. Le QALY (*Quality-Adjusted Life-Year*), qui est un équivalent en années de vie en bonne santé, est l'indicateur de référence en la matière. Et si le coût par année de vie gagnée est l'expression type d'une ACE, le coût par QALY gagné est l'expression de l'ACU.

- L'approche coût-bénéfice (ACB) vise également à appréhender dans une seule métrique l'ensemble des effets d'actions de santé alternatives. Mais ici, il s'agit de transcrire ces effets en équivalent monétaire, autrement dit de ramener des bénéfices sanitaires à des bénéfices pécuniaires. Comme les coûts et les effets sont mesurés dans la même métrique monétaire, le résultat de l'évaluation peut être exprimé dans le même format de ratio que l'ACE et l'ACU, mais aussi de différentes façons alternatives. Par exemple en inversant le ratio pour rapporter le différentiel de bénéfices monétisés au différentiel de coût, ce qui fournit une sorte de taux de rendement de l'investissement public. Mais, plutôt qu'un ratio, on peut aussi former une différence entre le différentiel de coût et le différentiel de bénéfices, de façon à obtenir le coût net ou le bénéfice net de l'intervention envisagée. L'ACB est peu utilisée dans le domaine de la santé du fait de difficultés réelles ou présumées à pouvoir valoriser en termes monétaires des modifications touchant à la santé des individus. Pour autant, il faut dire qu'elle est fréquemment mise en œuvre dans d'autres domaines, comme l'économie de l'environnement ou de la sécurité routière, pour orienter des choix, en recourant pour cela à des notions comme la disposition à payer (*willingness to pay*) pour valoriser l'impact de différentes mesures possibles[3].

On peut ajouter que, contrairement à l'ACE qui prend en compte l'efficacité telle qu'elle résulte d'indicateurs physiques, l'ACU et l'ACB procèdent, chacune à leur manière, à une valorisation de ces effets, la première en termes de bien-être ou d'utilité, la seconde en termes monétaires. Elles sont donc à la fois plus sophistiquées et moins parlantes pour les non-économistes, d'où la nécessité de préciser davantage la démarche sous-jacente[4].

---

3 Par ailleurs, l'ACB est normalement le cadre de référence du point de vue de la théorie économique, lorsqu'il s'agit de procéder à l'évaluation des politiques publiques ou des choix collectifs. Et un débat existe pour déterminer si, ou à quelles conditions, l'ACU peut être considérée comme fondée sur la théorie économique et, à ce titre, acceptable comme un substitut à l'ACB. Mais nous ne détaillerons pas davantage ces controverses académiques.

4 La question de la mesure des bénéfices de santé est traitée de façon complète dans l'ouvrage de Brazier J, Ratcliffe J, Salomon J et Tsuchiya A. *Measuring and valuing health benefits for economic evaluation.* Oxford University Press, 2007.

Les études coût-efficacité

L'approche coût-efficacité se présente donc comme la plus simple à mettre en œuvre du fait qu'elle se focalise sur une mesure des effets des actions de santé selon un indicateur en unités physiques sans transformation dans une métrique non naturelle. Ceci explique en partie qu'elle a été la plus fréquemment utilisée dans le passé, d'autant plus qu'elle est en fait un point de passage obligé lorsqu'on veut développer une ACU ou une ACB.

Plusieurs raisons expliquent son usage fréquent, à commencer par le fait qu'elle peut être documentée pour la mesure des effets à partir de la bibliothèque des indicateurs élaborés dans le cadre de l'*Evidence-Based Medicine* (l'EBM). Le développement de l'EBM a permis d'élargir le choix des mesures possibles qui sont des indicateurs quantifiés, objectivement évalués. En outre, ils sont pertinents du point de vue clinique, faciles à comprendre pour le décideur, aisés à mettre en œuvre car documentés par la littérature clinique. Enfin, ils permettent d'éviter la question controversée de la valorisation de l'efficacité clinique en termes de bien-être pour l'ACU ou de bénéfices monétaires pour l'ACB.

L'ACE est cependant limitée du fait qu'elle contraint l'évaluation à ne prendre en compte qu'une seule dimension d'efficacité alors que des stratégies alternatives peuvent avoir des effets différenciés sur plusieurs dimensions, par exemple en prolongeant la survie mais en dégradant la qualité de vie. Le choix de l'indicateur d'efficacité est donc primordial pour assurer la pertinence de l'évaluation.

Dans l'éventail des options ouvertes, on peut distinguer différentes dimensions sur lesquelles faire porter la mesure de l'efficacité :

- On peut vouloir mettre l'accent sur un paramètre pharmacologique ou biologique, comme l'efficacité microbiologique, la densité minérale osseuse, le taux de LDL-cholestérol, le niveau de l'hémoglobine glyquée, etc.
- Alternativement, on peut choisir une mesure faisant référence à une variable clinique, qu'il s'agisse du nombre de jours sans symptômes (d'asthme ou d'épilepsie), de la capacité de marche (dans l'insuffisance cardiaque), du délai avant retour à activité normale (après intervention chirurgicale selon différentes techniques), du nombre de personnes immunisées (suivant la stratégie vaccinale).
- Enfin, une dernière solution consiste à choisir une mesure de morbidité ou de mortalité, par exemple les complications de l'hyperlipidémie, la fréquence des décès, ou le nombre d'années de vie associées à chaque stratégie.

Ce choix dépend bien sûr d'un certain nombre de paramètres, notamment la disponibilité de données de qualité. Mais il recouvre aussi d'autres enjeux.

Les différentes notions d'efficacité

Il faut mentionner ici au moins deux différenciations possibles de la notion générale d'efficacité. On peut distinguer d'une part les indicateurs d'efficacité intermédiaire des indicateurs d'efficacité finale. D'autre part, il faut souligner l'écart existant entre l'efficacité théorique (*efficacy*) et l'efficacité pratique (*effectiveness*).

Ainsi, une possibilité consiste à recourir à un indicateur spécifique d'une certaine maladie pour comparer l'efficacité de traitements alternatifs. Ce peut être par exemple, dans le cas de la prise en charge de l'hyperlipidémie, le taux de cholestérol total (ou de LDL-cholestérol) atteint avec différents traitements hypolipémiants. Dans le cas de l'épilepsie, ce pourrait être le nombre de journées sans crise ou la proportion de patients obtenant une réduction d'au moins 50 % de la fréquence des crises. Ou le délai avant progression de la maladie dans plusieurs affections (cancers, insuffisance rénale). Il s'agit dans tous ces cas d'un indicateur d'*efficacité intermédiaire*.

Ils se différencient des indicateurs génériques qui consistent à recourir à la mesure d'un évènement qui peut survenir dans la plupart des aires thérapeutiques. Il peut s'agir du nombre de cas de guérison par exemple ou, préférentiellement, d'une mesure de la mortalité exprimée soit par la fréquence des décès, la survie globale ou l'espérance de vie. Dans le cas de l'hyperlipidémie, on peut ainsi opter pour une mesure de la mortalité toutes causes, ou de la mortalité cardiovasculaire, pour comparer l'efficacité de traitements alternatifs. Dans tous les cas, ces indicateurs expriment l'*efficacité finale* des stratégies thérapeutiques.

La question du choix entre un indicateur d'efficacité intermédiaire ou d'efficacité finale renvoie à un arbitrage entre plusieurs critères. D'un côté, la pertinence de la mesure pour une pathologie donnée et sa sensibilité pour faire apparaître des différences entre stratégies alternatives militent en faveur de l'efficacité intermédiaire. À cela s'ajoute le fait que ces indicateurs sont la plupart du temps bien plus faciles à documenter sur la base d'essais cliniques qui recourent à de telles mesures de résultats[5]. À l'inverse, il est souvent

5 Et ceci est d'autant plus souvent le cas que les industriels ont tendance à vouloir lancer leurs produits de plus en plus précocement sur la base d'une démonstration reposant sur des données non matures.

difficile, voire périlleux, de vouloir recourir à une mesure d'efficacité finale portant sur la mortalité lorsque le décès des patients a lieu longtemps après les interventions concernées. Dans le cas de maladies chroniques par exemple, comme l'hyperlipidémie ou le diabète, la mesure de l'efficacité de différents médicaments sur l'espérance de vie repose la plupart du temps sur le recours à différents outils de bio-statistiques pour développer une modélisation et une extrapolation permettant de passer d'un indicateur d'efficacité intermédiaire (le taux de LDL-cholestérol ou le taux d'hémoglobine glyquée) à l'efficacité finale. L'exercice reposant sur un certain nombre d'hypothèses (sur l'évolution à long terme de l'efficacité des produits, l'apparition d'effets secondaires, la fréquence des arrêts de traitements, etc.), le problème est celui du niveau de preuve et de la vraisemblance des résultats pris en compte.

Néanmoins, sauf s'il s'agit d'éclairer des choix de prescription dans une maladie donnée ou de gérer un budget par pathologie, on voit bien les limites d'une approche fondée sur un indicateur spécifique qui ne permet de faire de comparaisons que sur une aire thérapeutique restreinte. C'est pourquoi les recommandations et la pratique la plus courante mettent l'accent sur le choix d'un indicateur d'efficacité finale qui permet d'assurer une comparabilité des résultats au-delà d'une pathologie. Si le *Gold standard* en la matière est la mesure des années de vie gagnées (*Life-Years Gained*) c'est bien parce qu'elle autorise une comparaison sur l'ensemble des stratégies thérapeutiques (pour autant que le mortalité soit en jeu) et que, combinée à la mesure des coûts, elle permet d'exprimer l'analyse coût-efficacité sous la forme d'un ratio du coût par année de vie gagnée qui peut être parlant même au non-spécialiste.

Efficacy *versus* effectiveness

Par ailleurs, il est usuel de distinguer l'efficacité théorique (*efficacy* en anglais) qui donne une estimation des bénéfices thérapeutiques en environnement contrôlé de façon à neutraliser tous les facteurs confondants (typiquement un essai clinique randomisé), et l'efficacité pratique (*effectiveness* en anglais) qui en donne la mesure en situation réelle[6].

6 En général, l'efficacité théorique est supérieure à l'efficacité pratique du fait des multiples imperfections de la vie réelle qui font que le traitement est mal prescrit par les médecins (dosage, durée, indication) ou mal suivi par les patients (c'est le problème de l'adhésion des patients). On peut noter néanmoins que, dans le cas de certains vaccins, l'*effectiveness* est supérieure à l'*efficacy* lorsque l'effet de groupe (la *herd immunity*) permet de protéger indirectement les individus non vaccinés.

Parce qu'elle fournit une mesure de résultat en situation idéale ou expérimentale à travers les données issues d'un essai clinique, l'efficacité théorique donne un niveau de preuve élevé. Mais, dans la mesure où les essais cliniques sont généralement menés dans des conditions éloignées de la pratique clinique usuelle, leurs résultats ont une pertinence réduite par rapport à la mise en œuvre de ces stratégies en pratique courante. À l'inverse, la mesure en environnement réel que fournit l'efficacité pratique est plus pertinente parce qu'elle permet de prendre en compte l'effet de phénomènes non observables dans le cadre d'essais cliniques, que ce soit l'effet du produit à des doses prescrites différentes des recommandations, ou de la prescription à des patients ayant un profil différent de celui des essais cliniques, ou de l'observance imparfaite des patients. Mais une telle mesure suppose le recours à des types d'études, observationnelles ou autre, de moindre qualité méthodologique que les essais cliniques notamment parce qu'elles ne peuvent neutraliser le biais éventuel de facteurs confondants. Au total, le choix entre l'*efficacy* et l'*effectiveness* est délicat car fondé sur un arbitrage entre qualité et pertinence de la preuve.

Suivant l'environnement dans lequel on mesure les bénéfices thérapeutiques des actions de santé, l'analyse sera donc de type « *cost-efficacy* » ou « *cost-effectiveness* », étant entendu que cette dernière est généralement préférée lorsqu'elle est possible car plus représentative de la réalité. À cet égard, on peut souligner l'importance accordée aux données de vie réelle pour renseigner à la fois sur les coûts et les effets des stratégies thérapeutiques étudiées, comme on le verra plus en détail au chapitre 10. Mais bien évidemment, au lancement d'un nouveau produit de santé, on ne peut en général se référer qu'à son efficacité théorique pour autant qu'il n'a pas encore été prescrit à des patients.

Les études coût-utilité

Si les études coût-efficacité ont constitué le cadre d'analyse prédominant pendant un certain temps, elles sont dorénavant supplantées par les études coût-utilité qui correspondent aujourd'hui au format préférentiel, quoique non exclusif, de l'évaluation économique des technologies de santé. En effet, si plusieurs raisons expliquent la préférence initiale pour l'approche coût-efficacité, notamment la possibilité de recourir à des indicateurs de résultats validés et documentés par l'EBM, elle présente des limites de différents ordres.

Sur le plan pragmatique, l'une des principales limites de l'ACE tient au fait qu'elle conduit à focaliser la comparaison des stratégies sur un résultat

unidimensionnel d'efficacité, qu'il s'agisse du taux de pression artérielle ou de l'espérance de vie. Or, il se peut que les résultats des traitements ne se limitent pas à un seul effet commun. En comparaison de la stratégie thérapeutique courante, un traitement innovant peut, par exemple, améliorer la survie mais accroître la douleur ou réduire la mobilité des patients, ou freiner la progression de la maladie au prix d'effets secondaires importants. Selon qu'on choisisse comme indicateur d'efficacité la survie ou la qualité de vie, on voit bien que l'efficacité comparative de l'innovation thérapeutique sera positive ou négative.

À travers cet exemple, on voit que, bien souvent, une évaluation de type coût-efficacité risque de biaiser la comparaison en ne prenant en compte qu'une partie des effets des traitements alternatifs. Il est en effet très fréquent que les options étudiées affectent différemment la morbidité et la mortalité des patients, l'effet portant alors à la fois sur la qualité de vie et sur la durée (la quantité) de vie.

On peut ajouter que, d'un point de vue théorique, l'approche coût-efficacité n'est pas clairement reliée à la théorie économique du Bien-Être qui voudrait que les actions publiques soient évaluées dans le cadre de l'approche coût-bénéfice. Elle devrait donc être rejetée car ne reposant pas sur une mesure adéquate des résultats des actions entreprises.

Les QALYs et autres HALYs

Si pour les théoriciens, l'approche coût-utilité doit elle-même faire la démonstration de son enracinement dans la théorie économique du Bien-Être, elle peut y prétendre en mettant l'accent sur une mesure de résultat composite qui veut appréhender simultanément l'effet des stratégies thérapeutiques sur la qualité de vie (dont il reste à prouver qu'elle correspond au concept d'utilité de la théorie économique) et sur la durée de vie.

D'une manière générale, il s'agit donc de capturer dans une seule et même mesure l'espérance de vie, mesurée en années de vie, et la qualité de cette survie. Ceci nous renvoie au concept fondamental de HALYs, l'acronyme de *Health-Adjusted Life Years* en anglais, c'est-à-dire à la notion d'années de vie ajustées sur le niveau de santé. Si ce concept de HALYs peut être décliné de différentes façons, l'idée de base est bien de combiner un indicateur subjectif de qualité de vie associé à la santé (HRQoL ou *Health-Related Quality of Life* en anglais) avec une mesure objective de durée de la survie.

Il existe au moins deux façons de mettre en œuvre le concept de HALYs : les QALYs (*Quality-Adjusted Life Years*) et les DALYs (*Disability-Adjusted Life Years*). Sans entrer dans le détail ici puisque ces aspects seront traités dans les chapitres 5 et 6, on peut dire que, historiquement, les DALYs

ont été développés dans une perspective de santé publique par l'OMS pour donner une mesure du fardeau que font peser les maladies et les facteurs de risque sur la population de tous les pays de la planète. Il s'agit ainsi de capturer l'effet de ces affections à la fois sur la mortalité prématurée et sur le handicap qu'elles peuvent engendrer. En mesurant les années de vie sans incapacité perdues du fait de la maladie, il est donc question de donner une estimation des effets du « fardeau global de la maladie » (*global burden of disease*) au niveau populationnel par rapport à une situation idéale sans maladie.

Sur des bases différentes, les QALYs ont été développés par les économistes pour mesurer l'équivalent des années de vie en bonne santé qui pouvaient être gagnées grâce à des interventions sanitaires. Si de ce point de vue, on pourrait penser à une certaine complémentarité entre QALYs (gagnés) et DALYs (perdus), le facteur d'ajustement n'est ni de même nature (la qualité de vie dans le premier cas, le handicap ou l'incapacité dans le second), ni obtenus par les mêmes méthodes, même si l'OMS a récemment modifié la méthodologie d'estimation des DALYs.

Chapitre 3
Modèles coût-efficacité – Modèles d'impact budgétaire

Efficience et Solvabilité : deux notions distinctes

La recherche de l'*efficience*, c'est-à-dire de la « rentabilité » médico-économique d'un bien ou d'une stratégie de santé mesurée par son ratio incrémental coût-efficacité (ICER) ou son ratio différentiel coût résultat (RDCR) dans la terminologie de la HAS, a constitué historiquement – et continue de constituer – l'objet principal de l'évaluation économique des biens et stratégies de santé.

Mais cette efficience ne préjuge en rien de la capacité à payer des systèmes de santé. Des innovations jugées « efficientes » en théorie peuvent néanmoins être difficiles à financer en pratique.

Ainsi, dans le fameux cas du sofosbuvir (Sovaldi®), un médicament innovant dans l'hépatite C chronique revendiquant une guérison des patients après 12 semaines de traitement (pour certains génotypes), les études médico-économiques soumises par le laboratoire à la HAS ont fait ressortir, au prix revendiqué par la firme, un RDCR de 17 000 Euros/QALY par rapport au traitement par interféron/ribavirine[1], ce qui est un montant considéré comme acceptable (et même plutôt bas !). Le coût total du traitement n'en est pas moins apparu comme très élevé, compte tenu de la taille de la population cible – estimée à environ 130 000 personnes – et du coût d'acquisition du produit[2].

1 Chez les patients genotype 1. Cf. HAS (CEESP) : Avis d'efficience du Sovaldi® (sofosbuvir) en date du 15 avril 2014, en ligne sur le site de la HAS. Le modèle n'a suscité aucune « *réserve méthodologique majeure* » susceptible « *d'invalider tout ou partie de l'étude économique* ». Le sofosbuvir en association avec l'interféron et la ribavirine a obtenu une ASMR de niveau 2 dans la prise en charge de l'hépatite C (sauf génotype 3 pour lequel l'ASMR était de niveau 3). Avis de la Commission de la transparence en date du 14 mai 2014.

2 Le prix revendiqué par la firme est masqué dans l'avis d'efficience. Mais le prix HT pour 12 semaines de traitement était de 56 000 € durant la période ATU et il a été fixé à 41 000 €

Le cas du sofosbuvir a illustré la différence entre les notions d'« efficience » et de « solvabilité ». Les systèmes de santé peuvent éprouver des difficultés à financer des innovations jugées techniquement efficientes. Les modèles coût-efficacité (MCE) ou coût-utilité (MCU) font en effet abstraction des conditions de financement des biens qu'ils soient ou non considérés comme « efficients ».

Comme la plupart des ratios, le RCDR est un nombre « sans dimension » qui s'applique à une cohorte fictive de patients représentatifs, sans considération du nombre réel de personnes concernées par le traitement dans la « vraie vie » et, le plus souvent, sans limite de temps (la vie entière de la cohorte). Deux interventions de coûts (et d'efficacité) très différents peuvent avoir le même RDCR alors même que le coût d'une d'entre elles peut être difficile à supporter pour le système de santé (*figure 1*).

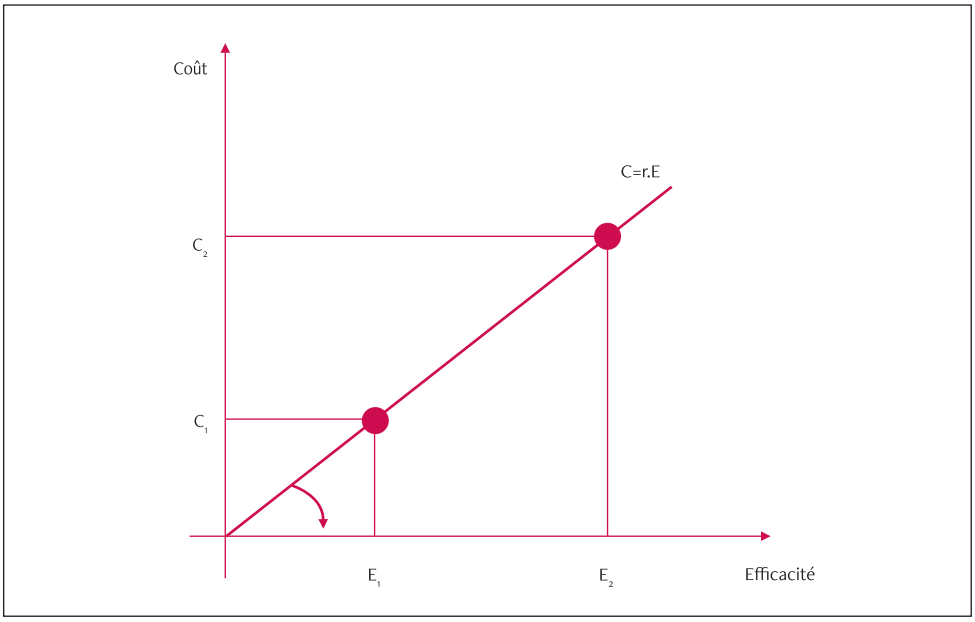


Figure 1. Coût, Efficacité et RDCR.
Deux interventions de coût (et d'efficacité) très différentes peuvent avoir le même RDCR figuré ici par la pente de la droite passant par l'origine du graphe (C, E).

par le Comité Economique des Produits de Santé (CEPS) le 20 novembre 2014 avant d'être abaissé à 28 700 € le 1er avril 2017.

Par ailleurs, la perspective des études est le plus souvent une perspective assez générale, sociétale ou payeur en dernier ressort, qui néglige des perspectives plus « locales », par exemple celle des établissements de santé concernés par une innovation qui, pour être rentable au niveau « macro-économique », peut néanmoins soulever des problèmes de financement au niveau « micro-économique ».

Enfin, l'actualisation des coûts (et des bénéfices), qui est requise dans les modèles coût-efficacité pour comparer des profils temporels de coûts différents, gomment l'échelonnement dans le temps des coûts engagés et des coûts évités alors que cet aspect peut être important dans la « vraie vie » en termes de planification budgétaire.

Pour ces raisons a été développée une classe de modèles dits « *modèles d'impact budgétaire* » (MIB) dont le but est précisément de documenter les conséquences budgétaires pour une institution donnée du choix d'une stratégie préventive ou thérapeutique jugée « coût-efficace ».

De nombreuses agences d'évaluation des technologies médicales ont ainsi inclus ces modèles dans leur processus d'évaluation et des guides méthodologiques ont été rédigés tant en France, par le Collège des Économistes de la Santé[3] et la HAS[4], qu'au niveau international à travers notamment une « task-force » spécialisée de l'ISPOR[5].

Ce chapitre est destiné à présenter les caractéristiques principales de ces modèles et d'en souligner l'intérêt et les limites.

Caractéristiques et méthodologie des MIB

La plupart des guides méthodologiques insistent sur les différences conceptuelles et pratiques qui séparent les modèles coût-efficacité ou coût-utilité (MCE/MCU) et les MIB.

Les premiers produisent une pure mesure de l'efficience d'une innovation dans l'absolu en considérant les seules données d'efficacité et de coût, dans une perspective assez générale, celle de la société ou celle du payeur. Les seconds au contraire ambitionnent de mesurer et de prédire l'impact de l'innovation sur le budget de chacune des institutions concernées par son introduction.

---

3 CES : *Guide méthodologique pour la mise en place d'une analyse d'impact budgétaire*, 2008.
4 HAS (Service évaluation économique et santé publique). *L'Analyse d'impact budgétaire, Guide méthodologique*, 2016. En ligne sur le site de la HAS.
5 Sullivan SD, Mauskopf JA, Augustovski F, *et al.* Principles of Good Practice : Report of the ISPOR 2012 Budget Impact Analysis Good Practice II Task Force. *Value in Health* 2014 ; 17 : 5-14.

Les premiers sont *normatifs* dans la perspective de la recherche d'une solution optimale, les seconds sont *positifs* et ne renseignent que sur les conséquences budgétaires d'un choix quel qu'il soit (*tableau 1*). L'un vient en appui de l'autre : en principe, la question de l'impact budgétaire se pose une fois résolue celle de l'efficience.

Tableau 1. Différences entre les modèles coût-efficacité (MCE) ou coût utilité (MCU) et les modèles d'impact budgétaire (MIB)

	MCE/MCU	**MIB**
Objectif	Efficience coût-résultat	Capacité à payer des institutions
Indicateur d'efficacité	Indicateurs cliniques QALYs	À travers la balance des coûts engagés et évités
Horizon temporel	Dépendant du contexte médical (espérance de vie des patients)	Dépendant de l'horizon budgétaire des institutions (3 à 5 ans)
Perspective	Sociétale ou payeur public en dernier ressort	Payeur public et institutions concernées (hôpitaux, institutions médico-sociales), etc.
Nature des coûts	En principe coûts directs et indirects réels hors transferts	Coûts budgétaires supportés par les institutions (tarifs)
Actualisation des coûts	Oui	Non
Prise en compte du contexte épidémiologique	Non (point de vue d'un patient représentatif)	Oui (impact de la diffusion de la technique)
Caractéristiques de la population d'intérêt	Représentative au plan national	Représentative de la population traitée par les institutions concernées (sous-populations)
Traitement de l'incertitude – Analyses de sensibilité	Déterministes et/ou stochastiques	Scénarios

Outre l'objectif qui n'est pas le même, les principales d'entre elles portent sur la perspective, les populations, les coûts, le temps et les analyses de sensibilité.

1. **La perspective** d'un MIB est celle d'une institution dotée d'une autonomie budgétaire et appelée à supporter tout ou partie du coût d'une innovation médicale.

Cette institution peut être locale (établissement de soins, cabinet, plateaux techniques, etc.) ou nationale.

Elle n'est pas nécessairement différente de celle d'un MCE/MCU, si ce n'est qu'elle exclut en principe une perspective « sociétale » globale qui agrègerait les effets sur les différents acteurs et institutions du système de santé et ferait ainsi disparaitre leurs spécificités.

Mais il est possible de conduire les deux types d'analyse selon une même perspective, par exemple, celle de l'assurance-maladie obligatoire, comme le recommande la HAS.

Les perspectives ne sont pas non plus exclusives et peuvent se décliner : effets sur l'assurance-maladie obligatoire, sur les assurances complémentaires, sur les établissements de santé…

Cette question de la perspective est exigeante et nécessite une description précise des circuits de financement des institutions.

Au niveau d'un établissement de soins, l'impact budgétaire d'un nouveau médicament peut être différent selon que celui-ci est financé au titre des forfaits soins de la T2A ou au titre de la « liste en sus », c'est-à-dire la liste des médicaments pris en charge par l'assurance maladie « en sus » des prestations d'hospitalisation. Ou selon qu'il est inscrit sur la liste des médicaments « rétrocédés » qui sont refacturés par les établissements de santé à l'assurance-maladie au titre des soins de ville.

2. La question de la **population traitée** est essentielle. Comme mentionné ci-dessus la détermination des ratios coût-efficacité ou coût-utilité est indépendante de la taille des populations cibles, ce qui n'est évidemment pas le cas de l'impact budgétaire. La construction des MIB suppose une estimation, à un moment donné du temps, de la taille de la population cible et de celle la « population rejointe », c'est-à-dire de la population effectivement traitée par le produit analysé, ainsi que de leurs évolutions dynamiques dans l'horizon de temps du modèle. Ces populations peuvent évoluer sous le triple effet des « cas incidents », c'est-à-dire des nouveaux patients, des cas guéris qui cessent leur traitement et des cas induits qui, antérieurement traités ou pas, sont orientés vers le nouveau traitement.

Elle suppose également que soient vérifiées que les caractéristiques d'âge, de sexe, de sévérité de la population traitée sont bien conformes à celles des données expérimentales et que les ajustements éventuellement nécessaires ont été effectués. Cela est particulièrement important si l'analyse est menée du point de vue d'une institution « locale », comme un hôpital, dont la patientèle n'est pas nécessairement représentative du patient moyen dans l'indication.

3. **Les coûts** associés à un nouveau produit de santé incluent les coûts d'acquisition, de préparation et d'administration, les coûts de suivi des patients et ceux des éventuels effets secondaires. Les coûts des examens complémentaires parfois nécessaires pour décider de l'administration du produit doivent également être pris en compte. L'évolution de ces coûts sur l'horizon temporel devrait également être considérée.

Comme les modèles MCE/MCU, les MIB prennent également en compte les coûts économisés du fait du traitement, par exemple les coûts des traitements auxquels le nouveau produit se substitue – si c'est le cas – ou encore celui des éventuelles complications évitées dans l'horizon de temps du modèle.

Contrairement à ce que pourrait laisser penser leur dénomination, les MIB prennent en compte l'efficacité des produits, mais uniquement à travers l'effet sur les coûts de traitements, majorés ou économisés.

Ces coûts sont mesurés dans une perspective « comptable », c'est-à-dire tels qu'ils sont enregistrés dans le budget des institutions. En prenant en compte les tarifs pratiqués effectivement, la part supportée par l'institution analysée, les transferts et subventions dont peuvent bénéficier les institutions, etc.

La logique est celle du scénario contrefactuel : quels sont les coûts induits et économisés par rapport à une situation de référence sans le produit analysé ?

Cette identification précise des coûts réels n'est pas toujours facile compte tenu de la confidentialité de certaines informations, par exemple les prix négociés des médicaments par les établissements de santé qui peuvent être inférieurs aux prix nominaux qui sont rendus publics.

4. Alors que **l'horizon temporel** des MCE/MCU dépend de la durée de la pathologie et peut donc aller jusqu'à la vie entière pour les pathologies chroniques, celui des MIB est limité à la période courante de planification budgétaire des institutions de santé, généralement de 3 à 5 ans.

L'échelonnement dans le temps des dépenses est pris en compte par le MIB et les impacts budgétaires sont calculés pour chaque année prise séparément dans les limites de l'horizon de temps du MIB. Ceci constitue une différence majeure avec les MCE/MCU où les coûts sont actualisés et agrégés dans une mesure unique sur tout l'horizon temporel du modèle.

Il en résulte que les MIB doivent – ou devraient – s'efforcer de refléter l'évolution des prix et des tarifs si ceux-ci sont réajustés en fonction du temps, de la diffusion de l'innovation, de l'apparition attendue d'options concurrentes, etc.

5. Enfin, les MIB peuvent – doivent – faire l'objet **d'analyse de sensibilité** comme les MCE/MCU de manière à tester la robustesse des résultats vis-à-vis des hypothèses. L'impact budgétaire sera en effet différent selon les hypothèses adoptées pour différents paramètres, par exemple le prix des biens ou services évalués, la rapidité de leur diffusion, le taux de substitutions aux procédures existantes, le nombre de patients induits, etc.

Les analyses de sensibilité en évaluation économique sont traitées au chapitre 9 du présent ouvrage, plutôt dans la perspective dominante des MCE/MCU. Les techniques classiques d'analyse de sensibilité, stochastique ou déterministe, ne sont pas toujours applicables dans le cas des MIB.

Beaucoup de paramètres utiles pour la simulation de l'impact budgétaire comme l'évolution des populations-cibles ou rejointes, l'effet des substitutions de traitement, l'apparition de solutions alternatives, l'évolution des prix et des tarifs constituent des prédictions qui ne sont pas nécessairement documentées par des études statistiques.

La méthode privilégiée de traitement de l'incertitude dans les MIB est plutôt celle des **scénarios**, faisant varier simultanément plusieurs paramètres, notamment ceux qui sont sous le contrôle de l'institution considérée[6]. Des instruments interactifs de simulation budgétaire, par exemple sous la forme de feuilles de calcul, peuvent être élaborés en ce sens et confiés aux institutions de manière à permettre un ajustement des paramètres à leurs caractéristiques propres.

6 Nuijten MJC, Mittendorf T, Persson U. Practical issues in handling data input and uncertainty in a budget impact analysis. *Eur J Health Econ* 2010. Publié en ligne sur springerlink.com.

Diversité et usage des MIB

Conséquence d'un formalisme méthodologique moins poussé que celui des analyses de type MCE/MCU, les MIB apparaissent très divers. C'est ce que confirme une revue récente de la littérature portant sur 79 MIB américains. Les auteurs expliquent cette variabilité par le caractère encore récent de ce type d'analyse et par le fait que les guides méthodologiques restent assez généraux[7] et rappellent plus des grands principes qu'ils ne fournissent de recommandations techniques précises.

De fait, la question se pose de l'usage d'un MIB, notamment dans le cas où l'impact budgétaire d'un traitement jugé efficient au prix revendiqué par le producteur entrainerait une dépense non « soutenable » pour le payeur.

Contrairement à l'efficience, qui peut être appréciée par des référentiels explicites (comme dans le cas du NICE au Royaume-Uni) ou implicites (par comparaison à des biens ou procédures couramment utilisés ou pratiqués), la notion de « solvabilité » est difficile à définir a priori. Elle varie d'institutions à institutions, de pays à pays, en fonction des budgets alloués aux organisations de financement de la santé, assureurs ou acheteurs, de leurs missions et des priorités qu'elles se donnent.

Par ailleurs, comme rappelé plus haut, un MIB n'est pas normatif et ne conduit pas à une recommandation formelle de prise en charge ou de tarification. Ainsi, l'institution concernée peut-elle tirer différentes conclusions pratiques des résultats d'un MIB.

Face à des biens ou services considérés comme efficients mais posant des problèmes de solvabilité trois solutions s'offrent en général à un payeur ou un acheteur :

- Augmenter le budget alloué au financement du bien ou du service concerné et de ses conséquences.
- Renoncer à des produits et donc à l'efficience potentielle induite.
- Alléger l'impact budgétaire en négociant un prix d'acquisition inférieur et/ou en limitant la population en ciblant par exemple aux cas les plus graves.

La première solution est toujours possible, ne serait-ce que par redéploiement du financement au détriment d'interventions ou de stratégies moins « efficientes ». Dans ce cas, à coût constant, l'institution et la collectivité enregistrent un gain net d'efficience. C'est l'inverse avec la seconde solution qui, toujours à coût constant, conduit à une perte nette d'efficience locale et globale.

---

7 Mauskopf JA, Earnshaw S. A Methodological review of US budget-impact models for new drugs. *Pharmaco-Economics* 2016 ; 34 : 1111-31.

La troisième, qui est combinable avec la première, est plus empirique. C'est celle qui a été conduite en France dans le cas emblématique du sofosbuvir rappelé en introduction : les autorités de santé ont restreint dans un premier temps les indications du produit aux patients les plus sévères (stades de fibrose F2 à F4)[8], et négocié un prix très inférieur au « prix d'efficience » du modèle coût-utilité. De surcroit la Loi de financement de la sécurité sociale (LFSS) pour 2015 a prévu un mécanisme de sauvegarde dit « enveloppe W » destiné à limiter l'impact budgétaire sur l'assurance-maladie de l'ensemble des nouveaux produits de l'hépatite C (les « antiviraux d'action directe »). Ce mécanisme instaurait une taxation spécifique et temporaire des entreprises productrices de ces médicaments, dès lors que le chiffre d'affaires de ces médicaments dépassait un montant W fixé par la Loi et qu'il croissait de plus de 10 %[9].

Les limites des MIB

Utiles pour la décision publique en complément d'une analyse coût-efficacité ou coût-utilité, les MIB présentent toutefois un certain nombre de limites qui tiennent à leur objet même, à savoir l'approche budgétaire de l'innovation médicale.

En premier lieu, la notion de solvabilité ne saurait se substituer à celle d'efficience. Il serait absurde de recommander la prise en charge d'un bien ou d'un service uniquement parce qu'il est aisément finançable. Les notions d'efficacité et d'efficience précèdent logiquement celle de solvabilité dans la hiérarchie des critères d'accès au marché.

En second lieu, un accent mis unilatéralement sur les MIB risque, faute de précaution, de conduire à la substitution de l'intérêt d'une institution à l'intérêt général. Autant l'efficience peut être appréciée de manière assez générale, autant celle de solvabilité n'a de sens que dans un contexte institutionnel précis. La notion de solvabilité met en question l'organisation du système de santé, l'origine des ressources, les parts respectives des financements publics et privés... Face à la question de l'insolvabilité d'une institution en

8 En application des recommandations du rapport du Pr. D. Dhumeaux : *Rapport sur la prise en charge des personnes infectées par les virus de l'hépatite B ou de l'hépatite C*, 2014. Consultable en ligne.

9 Art. 138-19-1 du Code de la sécurité sociale, créé par l'art. 3 de la LFSS pour 2015. Initialement prévu pour une durée de 3 ans de 2014 à 2016, le mécanisme a été prolongé d'un an par la LFSS pour 2017. Le montant W a été fixé à 450 M€ pour 2014, à 700 M€ pour 2015 et 2016 et à 600 M€ pour 2017. Les entreprises ont été redevables de 282 M€ en 2014 et de 11 M€ en 2015 au titre de ce mécanisme.

charge du financement d'une innovation médicale « efficiente », il importe de questionner le contexte organisationnel pour apprécier les éventuelles contradictions entre intérêt institutionnel (public ou privé) et intérêt général.

Se pose enfin la question du temps : la logique budgétaire est en effet plus « court-termiste » que la logique médicale sur laquelle se fondent les MCE/MCU. Ainsi, du point de vue des MIB, il est *a priori* possible que les impacts d'une pathologie aiguë et d'une pathologie chronique soit évalués sur une même échelle de temps – par exemple 3 ans – ce qui peut n'avoir guère de sens par rapport au temps médical. Cela peut constituer un obstacle à l'adoption d'une innovation dont les coûts se situent dans l'horizon temporel considéré mais dont les effets bénéfiques, médicaux et budgétaires, se situent au-delà. Le cas de l'hépatite C où les événements morbides évités, cirrhose, hépato-carcinomes, greffes hépatiques se situent plusieurs années après le temps du traitement, illustre à nouveau ce décalage. Il importe donc de ne pas considérer les MIB indépendamment des MCE/MCU qui présentent en principe cette dimension du long terme.

Chapitre 4
Identification et mesure des coûts

Tous les types d'évaluation ont en commun d'appréhender les coûts liés aux stratégies thérapeutiques comparées, quels que soient les choix faits par ailleurs sur la mesure des résultats. Fondamentalement, pour l'économiste, l'existence d'un coût n'est concevable que comme contrepartie de l'utilisation d'une ressource rare, car sans rareté, il n'y a pas de valeur. Ce principe de base qui consiste à définir le coût associé à une stratégie thérapeutique par les ressources qu'elle consomme ne permet cependant pas de procéder directement à l'estimation des coûts (le *costing*) qui nécessite de faire des choix, pertinents et cohérents entre eux, sur différents aspects qui relèvent de l'identification, de la mesure et de la valorisation des ressources consommées[1].

Quelle typologie pour l'identification des coûts ?

Il convient en premier lieu de distinguer les différentes catégories de coût possibles avant de s'interroger sur ceux qui sont pertinents pour l'étude à mener. La démarche de l'évaluation médico-économique (EME) fait traditionnellement référence à une typologie des coûts en croisant deux critères binaires indépendants permettant de définir quatre catégories différentes. Mais une classification un peu différente s'impose désormais.

La référence traditionnelle à deux critères a un sens moins évident qu'il n'y paraît. D'un côté, il est usuel de distinguer les coûts selon la nature des ressources consommées par les technologies de santé évaluées, en se limitant à une distinction simple entre les ressources qui sont utilisées dans

1 Pour un examen complet de toutes les questions relatives au *costing*, voir : « The main methodological issues in costing health care services – A literature review ». Zsolt Mogyorosy, Peter Smith. CHE Research Paper 7 ; Centre for Health Economics, York, 2005.

le secteur de la santé et celles qui sont issues d'autres secteurs d'activité. Ce premier critère pose ainsi la distinction entre coûts médicaux et coûts non médicaux de façon assez claire : l'utilisation d'un bien ou d'un service médical engendre un coût médical tandis que si ce bien ou service relève d'un autre secteur de l'économie (transport, aménagement de logement ou de véhicule, services sociaux à la personne), il s'agit d'un coût non médical. D'un autre côté, et de façon moins évidente, on distingue les coûts directs des coûts indirects selon que la consommation de ressources est imputable directement ou non à l'intervention évaluée.

En croisant ces deux critères, on obtient quatre catégories mutuellement exclusives (tableau 1). Les coûts directs médicaux et non médicaux ne posent pas de problème quant à leur interprétation, étant entendu qu'ils incluent les coûts éventuels des futures complications ou effets secondaires imputables au traitement considéré. En revanche, la notion de coût indirect pose différentes questions.

Tableau 1. Les différents types de coûts

	Coûts médicaux	**Coûts non médicaux**
Coûts directs	Ressources (biens et services) du secteur sanitaire utilisées pour le traitement consultations, examens, médicaments, hospitalisation...)	Ressources non sanitaires (autres secteurs) utilisées pour le traitement (transports, aménagement du logement, aide domestique, services sociaux...)
Coûts indirects	Coût des pathologies futures non liées au traitement	Coût du temps perdu du fait de la maladie Perte de productivité (absentéisme, présentéisme)

S'agissant des coûts indirects médicaux, ils désignent les coûts potentiels qui devraient être supportés dans le futur pour la prise en charge de maladies qui pourraient se déclarer, sans lien aux traitements considérés dans l'évaluation. Dans la mesure où l'EME est une démarche comparative, il est de toute façon inutile de chercher à mesurer ces coûts si on les retrouve de façon équivalente dans toutes les stratégies évaluées. En revanche, si une stratégie B permet de prolonger la survie par rapport à une stratégie A, il se peut en effet que les patients nécessitent des soins pendant ces années de vie

gagnées et la question qui se pose est de savoir s'il faut imputer ces coûts à la stratégie B. Cette question est très controversée et l'on retiendra simplement qu'outre les difficultés de mesure en cas de prise en compte, il faut s'en remettre aux recommandations institutionnelles qui les excluent en général[2].

Pour ce qui est des coûts indirects non médicaux, ils concernent en fait une ressource particulière, le temps, qui peut être mobilisé par la maladie et ses traitements de telle sorte qu'il n'est plus disponible pour la réalisation des activités courantes. Nous préciserons plus loin différents éléments relatifs à ce coût du temps perdu (le temps des patients ou de l'entourage, le temps travaillé ou le temps de loisir). On peut simplement dire ici d'abord que, dans le langage traditionnel, on désigne par le terme « coûts indirects » les seuls coûts indirects non médicaux du fait de l'exclusion générale des coûts indirects médicaux. Il faut ajouter que, à la suite d'un groupe d'experts américains chargés d'élaborer des recommandations[3], on considère dorénavant préférable de ne pas agréger n'importe quelle perte de temps dans la catégorie des coûts indirects de façon homogène. La proposition est de différencier le temps mobilisé pour la réalisation du traitement de la maladie, et le temps qui est perdu du fait de la maladie elle-même. Dans le premier cas, il s'agit d'éléments à inclure dans le coût direct des traitements, dans le second cas, il s'agit du coût indirect à proprement parler.

La typologie de la HAS

Sur cette base, La HAS a émis des recommandations faisant référence à une classification des coûts beaucoup plus claire[4]. Partant de l'idée qu'il faut mobiliser des ressources pour produire les stratégies thérapeutiques évaluées, la HAS préconise de distinguer les coûts de production pour chacune des alternatives, qu'elle appelle les « coûts directs », et le coût des autres ressources n'entrant pas dans le processus de production qui constituent les coûts indirects. Cette nouvelle définition distinguant les coûts directs et indirects sur la base du seul critère du processus productif des soins recoupe très largement la typologie traditionnelle à une exception près. Là où la tradition

2 Sur ce point, voir « Costing in economic evaluation », W. Brouwer, F. Rutten et M. Koopmanschap, in *Economic evaluation in health care: merging thepry with practice*, M. Drummond & A. McGuire eds. Oxford University Press, 2001.
3 Voir en particulier Luce BR, *et al.* (1996). Estimating costs in Cost-Effectiveness Analysis. Chapter 6 in: Gold *et al.* (eds) *Cost effectiveness in health and medicine*. Oxford University Press.
4 *Choix méthodologiques pour l'évaluation économique à la HAS*. Haute Autorité de Santé, octobre 2011.

agrégeait tout le temps perdu sous le terme unique de coût indirect, la nouvelle approche distingue le temps (des patients et des aidants) mobilisé pour la production des soins qui doit être considéré comme une composante des coûts de production, donc des coûts directs, et le temps perdu du fait de la morbi-mortalité qui est constitutif des coûts indirects car n'entrant pas dans la production des interventions étudiées. Les coûts directs se trouvent donc augmentés du coût du temps des patients dévolu aux soins.

Cette redéfinition des catégories de coût n'est pas sans conséquence sur l'évaluation dans la mesure où les recommandations stipulent assez souvent, et en France en particulier, de n'inclure que les coûts directs (nouvelle manière) mais de ne pas considérer, sauf exception, les coûts indirects puisqu'ils ne correspondent pas à des ressources mobilisées pour mettre en œuvre les stratégies thérapeutiques évaluées.

Le choix d'une perspective pour la mesure des coûts

Après avoir identifié les postes de coût à retenir dans l'évaluation, il faut préciser selon quelle perspective on se place pour en mesurer l'ampleur, c'est-à-dire expliciter la ou les catégories d'agents du point de vue desquels on considère le coût des ressources mobilisées. Les coûts d'un séjour hospitalier, d'un examen biologique ou d'un médicament ne sont en effet pas les mêmes pour l'assurance maladie, les patients, l'État ou d'autres parties prenantes.

Plusieurs perspectives peuvent être pertinentes. La perspective sociétale qui englobe toutes les parties prenantes est toujours intéressante puisqu'elle correspond au point de vue global et considère donc le coût total de chaque ressource utilisée sans préciser comment se répartit la charge de ces coûts entre les différentes catégories d'agents économiques. Mais s'il s'agit d'éclairer un choix particulier, il peut être plus pertinent de se situer dans la perspective spécifique du cadre de décision concerné. En particulier, on peut être amené à se placer dans une perspective « payeur », c'est-à-dire en général du point de vue du système d'assurance maladie s'il s'agit de faire admettre au remboursement un produit de santé ou la prise en charge d'une stratégie thérapeutique. Cette approche est néanmoins limitative puisqu'elle ignore la partie des coûts non pris en charge par le système assurantiel, en particulier le reste à charge des patients. Entre ces deux approches extrêmes, on peut choisir d'adopter une perspective dite collective qui prend en compte à la fois le point de vue de l'assurance maladie obligatoire et celui des patients

(y compris la part remboursée par l'assurance maladie complémentaire), voire d'autres parties prenantes lorsque c'est pertinent.

On peut ajouter que le périmètre des coûts à considérer est défini à la fois par les types de coût retenus (directs, voire indirects) et par la perspective adoptée. Mais il est aussi conditionné par la nature comparative de l'évaluation économique qui limite l'ampleur des coûts à mesurer aux seuls éléments pouvant faire apparaître une différence entre les stratégies alternatives. Ainsi, si un même bilan médical est nécessaire à l'initiation de traitements substituables ou à intervalle régulier, il n'est pas utile de les prendre en compte.

La valorisation des ressources

Une fois le périmètre des coûts à prendre en compte défini, leur mesure passe par l'articulation de deux types d'informations. Il faut en premier lieu connaître ou estimer le volume des ressources utilisées (le nombre de séjours hospitaliers, de consultations, d'examens biologiques, etc.), ce qui peut se faire à partir de différentes sources de données que nous préciserons au chapitre 10. Si l'information peut être issue d'un essai clinique contrôlé randomisé, on sait que ceci conduit pour une part à délaisser certains coûts (ambulatoires en particulier) et pour une autre part à biaiser la consommation de ressources du fait d'un protocole qui correspond rarement à la pratique courante. De ce fait, il est préférable de rechercher des sources permettant d'obtenir des données correspondant à la pratique courante : études observationnelles, registres ou fichiers de patients, bases de données de remboursement.

En second lieu, il faut procéder à la valorisation des coûts à proprement parler, c'est-à-dire associer une valeur spécifique à chacune des ressources consommées. Comme la conception économique des coûts n'est pas une conception comptable ou financière qui conduirait à mesurer le montant de la dépense associée à l'utilisation d'une ressource, à un débours pour un agent quelconque en considérant son prix, cette valorisation peut être complexe.

D'un point de vue économique, ce que coûte réellement la mobilisation d'une ressource pour un usage particulier, c'est la valeur des avantages perdus et qu'on aurait pu avoir par le meilleur usage alternatif de cette ressource. C'est la notion de coût d'opportunité : en affectant une ressource à un usage, on renonce aux avantages que procurerait le meilleur usage alternatif. Et si ce coût d'opportunité est en général mesuré par le prix d'une ressource, ce n'est vrai que lorsqu'il s'agit d'un prix de marché résultant du

libre jeu de l'offre et de la demande. Ce qui est rarement le cas dans le domaine de la santé. S'agissant bien souvent de biens et services non marchands, il n'existe pas nécessairement de prix unitaire permettant de procéder de façon simple à cette valorisation. Les prix administrés ou les tarifs peuvent ne pas donner une valorisation correcte des coûts d'opportunité.

Compte tenu de la difficulté de mise en œuvre de la notion de coût d'opportunité, on accepte l'idée d'une mesure qui s'en approche le plus possible en se référant à la notion de coût de production de la ressource considérée. Lorsque c'est possible, on doit donc valoriser les ressources utilisées sur une telle base. À défaut, on accepte de prendre comme base d'estimation les tarifs des biens et services consommés s'ils représentent une approximation raisonnable des coûts, quitte à les augmenter s'il existe un prix réel différent du tarif (par exemple dans le cas des honoraires libres des professionnels de santé ou de certains produits). Cette valorisation par les tarifs, éventuellement ajustés, est notamment utilisée pour mesurer les coûts ambulatoires. En revanche, l'approche par les tarifs ou prix est moins acceptable pour les coûts hospitaliers dont les coûts de production peuvent différer soit parce que les tarifs les sous-estiment soit parce que, dans le cas d'un secteur hospitalier privé, le prix peut être augmenté d'une marge de profit qui les surestiment.

Les coûts hospitaliers

Lorsque les stratégies thérapeutiques comparées nécessitent l'utilisation de ressources hospitalières, il est important d'estimer les coûts correspondants de façon correcte et précise dans la mesure où, compte tenu du fait qu'ils sont généralement bien plus élevés que les coûts ambulatoires, leur valorisation peut avoir un effet majeur sur le différentiel de coût. Deux méthodes bien différentes peuvent être utilisées, chacune ayant ses avantages et ses inconvénients, l'une procédant d'une logique de forfaitisation (le *gross-costing*), l'autre d'une logique de décomposition fine des éléments de coût (le *micro-costing*).

Dans les pays comme la France où les séjours hospitaliers sont financés de façon forfaitaire, on peut se référer au coût estimé pour chaque type de séjour. On part dans ce cas d'une classification des séjours sur la base d'une similitude de leur nature médico-économique, en fonction du diagnostic, de la nature des actes effectués et de la consommation de ressources durant le séjour. Ce système, initialement développé aux États-Unis sur la base de DRGs (*Diagnosis-Related Groups*), a été progressivement mis en œuvre en France à partir de 1992 dans le cadre du PMSI (Programme

de Médicalisation des Systèmes d'Information) permettant de définir une liste de GHM (Groupes Homogènes de Malades) régulièrement modifiée depuis. À chaque GHM, à libellé médical, est associé en général un seul, mais parfois plusieurs, GHS (Groupe Homogène de Séjours) qui est un tarif sur la base duquel les établissements hospitaliers sont remboursés par l'assurance maladie (la T2A). Mais ce tarif n'est en général pas adapté pour estimer le coût réel du séjour, plusieurs éléments concourant à une dissociation entre tarif et coût[5]. De ce fait, il est recommandé de se référer au coût de production des séjours tel qu'il est estimé par l'ENCC (Échelle Nationale des Coûts à méthodologie Commune) sur la base d'un travail de comptabilité analytique qui permet d'estimer le coût moyen par GHS sur la base d'un échantillon d'établissements volontaires.

Cette approche par coût forfaitaire de GHS est intéressante par plusieurs aspects. D'une part, elle permet d'obtenir directement le coût complet d'un séjour hospitalier en fonction de sa nature et des actes réalisés, à partir de postes de coût regroupés en sous-ensembles cohérents (par exemple personnel médical, personnel soignant, actes médico-techniques, restauration, blanchisserie, structure). D'autre part, ce coût est estimé sur la base d'un échantillon suffisamment large pour être représentatif de l'ensemble des établissements en gommant d'éventuelles spécificités locales. Enfin, elle est transparente et facilement contrôlable. Mais l'approche par *gross-costing* n'est pas exempte de limites. Elle n'est en effet facilement mobilisable que lorsqu'il est facile d'identifier les GHS pertinents, ce qui n'est pas toujours le cas par exemple lorsqu'on veut estimer le coût d'un acte technique (la coronarographie) qui peut renvoyer à un grand nombre de GHS. De même, lorsque la différence entre deux stratégies tient à une réduction marginale de la durée de séjour ou à une modification de l'utilisation des ressources durant l'hospitalisation, l'estimation des coûts par forfait ne permet souvent pas d'en tenir compte du fait d'un manque de précision.

On peut alors utiliser la méthode alternative de *micro-costing* qui consiste à procéder au cas par cas au travail de comptabilité analytique de mesure et de valorisation de chacune des ressources mobilisées durant un séjour hospitalier en observant de façon précise le déroulement de séjours représentatifs. Cette technique permet de tenir compte d'une modification

5 D'une part, le tarif est inférieur au coût complet du séjour parce que celui-ci est financé éventuellement par plusieurs tarifs qui se superposent (comme pour les produits onéreux de la liste en sus payés aux établissements en supplément du tarif du séjour). D'autre part, le coût complet tel qu'il est estimé par l'ENCC est inférieur au coût réel parce qu'il ne tient pas compte des coûts de structure. Mais ces derniers sont néanmoins estimés.

de la structure ou du niveau des ressources utilisées à un niveau très fin puisqu'on va pouvoir détailler davantage les postes de coût que dans l'approche du *gross-costing* qui s'en tient à des sous-ensembles. Par ailleurs, lorsqu'on veut estimer le coût d'une stratégie thérapeutique innovante pour laquelle il n'y a pas de GHS correspondant, le *micro-costing* s'avère la seule possibilité. Pour autant, cette approche s'avère lourde à mettre en œuvre pour obtenir une estimation de coûts hospitaliers dont la représentativité sera nécessairement limitée du fait qu'elle sera fondée sur un petit nombre de cas, avec le risque d'un biais lié à des particularismes locaux en fonction des établissements choisis.

De façon pragmatique, la solution se trouve parfois dans la combinaison des deux approches. Si l'on prend l'exemple d'une innovation thérapeutique dont l'un des effets est de réduire le temps infirmier nécessaire durant le séjour hospitalier du fait d'une moindre surveillance. On peut partir de valeurs issues des forfaits hospitaliers pour l'ensemble des autres ressources en retranchant le poste « coût infirmier » du forfait et réaliser un micro-costing uniquement sur cette composante.

Les coûts indirects

Selon la perspective utilisée et le type d'EME mise en œuvre, on peut estimer nécessaire d'inclure tout ou partie des coûts indirects puisqu'on a déjà évoqué plusieurs grandes composantes constituant le coût du temps perdu (voir *supra*) :

❶ le coût engendré par le temps passé pour les soins ;
❷ le coût du temps passé par l'entourage à l'aide informelle ;
❸ le coût lié à la perte de productivité (par absentéisme ou présentéisme[6]) du fait de la maladie ;
❹ le coût relatif à la perte de loisir.

Mais l'intégration de ces éléments pose différents problèmes. D'une part, on peut se demander, d'un strict point de vue de l'efficience économique, s'il faut les inclure. En effet, la question est de savoir s'il s'agit « d'inputs », de ressources, dont il faut prendre la mesure économique, ou s'il s'agit d'effets non sanitaires liés à la détérioration de l'état de santé. Pour les composantes

6 Le présentéisme renvoie à la perte de productivité partielle qui peut se manifester lorsqu'un individu effectue son travail alors qu'il est malade et ne peut alors avoir qu'une productivité réduite.

❶ et ❷, il s'agit bien de ressources mobilisées pour la mise en œuvre de la stratégie visée, et il est normal de les prendre en compte. En revanche, pour les composantes ❸ et ❹, il y a des chances qu'ils soient appréhendés par la mesure de résultat, surtout s'il s'agit de QALYs. Pour ces composantes, c'est donc un problème de cohérence qui se pose et il faut faire attention au problème de double comptabilisation : il n'est pas légitime de compter une année de vie sauvée à la fois comme coût économisé (par la production ainsi épargnée) au numérateur du ratio formé, et comme résultat amélioré au dénominateur sous la forme d'un QALY.

À ce problème possible de cohérence s'ajoute un problème de pertinence de leur inclusion en fonction de la question à éclairer et de la perspective adoptée. S'il s'agit d'aider à la décision un payeur par exemple, ces coûts indirects peuvent ne pas faire partie de son budget et ne pas le concerner, comme c'est le cas pour le NHS au Royaume-Uni. À l'inverse, du point de vue sociétal, ces coûts sont pertinents et peuvent être pris en compte si les effets sous-jacents ne sont pas intégrés dans la mesure de résultat.

D'autre part, l'évaluation des coûts indirects peut poser des problèmes d'équité. Si l'on considère les pertes de productivité au niveau individuel, cela peut conduire à valoriser différemment le temps des individus en fonction de leur statut. Dans ce cas, on serait conduit à privilégier les interventions qui concernent prioritairement les actifs par rapport aux inactifs, les jeunes par rapport aux vieux, les cadres par rapport aux ouvriers. Si cet argument peut conduire à exclure les coûts indirects de l'évaluation, une autre solution consiste à valoriser au même prix le temps perdu quelle que soit la personne.

Au-delà de ces problèmes, l'estimation des coûts indirects qu'on veut prendre en compte suppose qu'on réponde à deux questions comme pour les autres ressources utilisées : comment mesurer le temps relatif à l'une ou l'autre des quatre composantes identifiées ? Quelle valeur attribuer à ce temps ? Si la première question a généré moins de développements que la seconde, elle n'en est pas moins centrale. Disons simplement ici que de nombreux questionnaires utilisables de façon prospective ou rétrospective ont été mis au point pour obtenir une estimation du temps perdu pour chacune des quatre composantes des coûts indirects. Il peut s'agir de questionnaires génériques (le WPAI – Work Productivity and Activity Impairment – pour les pertes de productivité par exemple) ou spécifiques à certaines pathologies (par exemple MIDAS – migraine disability assessment – pour le temps de loisir) qui se distinguent par le type de répercussion pris en compte et la période considérée.

La valorisation des coûts indirects

La valorisation des coûts indirects est un sujet controversé qui renvoie à de nombreuses considérations de différents ordres qu'on ne détaillera pas ici[7]. Si l'on ne considère pas que ces éléments sont déjà pris en compte dans les QALYs et ainsi valorisés de façon non monétaire, il existe principalement deux méthodes pour estimer la valeur monétaire du temps perdu par la maladie ou son traitement. L'approche de référence est celle dite du « *capital humain* » qui est fondée sur le principe du coût d'opportunité. D'un point de vue collectif, la valeur d'une journée de travail perdue se mesurerait par la richesse non produite pendant cette absence au travail[8]. En s'appuyant sur les bases de la microéconomie, la théorie du capital humain préconise de valoriser le temps de travail perdu par le salaire brut correspondant tandis que le temps de loisir perdu devrait être valorisé par le salaire net. Si cette approche a le mérite de la facilité d'application, elle pose un certain nombre de problèmes du fait des hypothèses sur lesquelles elle a été élaborée et qui conduisent potentiellement à surestimer les coûts indirects, en particulier l'hypothèse de plein emploi.

Ainsi, un groupe d'économistes hollandais a critiqué cette approche et a proposé comme alternative celle dite des coûts de friction pour valoriser les coûts liés aux pertes de production. En fait leur critique porte non seulement sur la valorisation elle-même mais aussi, et peut-être surtout, sur la mesure du temps perdu. Partant du constat qu'il existe généralement un certain volant de chômage et donc de travailleurs disponibles, ces auteurs estiment que l'absence au travail des malades fait jouer différents mécanismes de compensation, allant du rattrapage par les collègues à l'emploi de travailleurs intérimaires lorsqu'il s'agit d'une absence programmée, voire à l'embauche d'un nouveau salarié en cas de décès prématuré. Il ne conviendrait alors de mesurer comme temps perdu que celui non compensé pendant la phase temporaire d'adaptation, ou période de friction. Par ailleurs, si cette approche accepte le principe de base d'une valorisation du temps perdu par le salaire correspondant, il faudrait ajouter certains coûts additionnels liés par exemple au recrutement et à la formation des remplaçants. Si l'approche par les coûts de friction a pour elle un degré de réalisme supérieur, elle

7 Pour plus de détails, voir « Costing in economic evaluation », W. Brouwer, *et al. Op. cit.*
8 Ainsi, on ne se place pas du point de vue de l'individu malade qui peut ou non bénéficier d'un système d'indemnisation en cas de perte de revenu. Si l'on se place du point de vue restrictif de l'assurance maladie, on pourrait certes mesurer ces coûts indirects par les indemnités journalières (IJ) versées. Mais au niveau collectif, ces IJ sont des revenus de transfert entre bien-portants et malades et cette redistribution est un jeu à somme nulle.

s'avère assez lourde à mettre en œuvre pour estimer à la fois le temps réellement perdu et l'ensemble des coûts de friction.

Il n'en reste pas moins que cette controverse entre les deux approches est importante dans la mesure où la comparaison des coûts indirects estimés par l'une et l'autre qui a été faite sur plusieurs exemples montre une très grande différence d'estimation, ce qui ne milite pas en faveur de leur prise en compte[9].

Horizon temporel et actualisation des coûts

Un des points essentiels dans la conception d'une EME tient au choix de l'horizon temporel, c'est-à-dire de la période de temps que l'on retient pour mesurer les coûts et les effets de santé des technologies de santé concernées. Étant entendu que cette période doit être la même pour toutes les options envisagées, le principe de base est de retenir une durée qui permet de prendre en compte toute la séquence des coûts engendrés et de faire ressortir tous les effets de santé générés par chacune d'elles. Il s'agit en cela de ne pas biaiser la comparaison en faveur de l'une ou l'autre des stratégies. Si l'on compare par exemple dans le cas d'une maladie chronique un traitement médicamenteux à prendre à vie à une intervention chirurgicale dont le coût est limité au moment de sa réalisation mais dont les bénéfices thérapeutiques se manifestent pendant toute la vie des patients, il ne faut pas faire l'évaluation sur une période d'un an au risque de défavoriser l'option chirurgicale, mais raisonner sur la vie entière des patients. Par ailleurs si au moins l'une des stratégies a pour effet de modifier l'espérance de vie des patients, il faut là aussi choisir un horizon temporel « vie entière » de façon à pouvoir prendre en compte ce bénéfice sanitaire. Même si la détermination de l'horizon temporel pertinent doit être faite au cas par cas, le choix d'une période de référence « vie entière » permet également de tenir compte d'éventuelles complications ou séquelles à long terme et se révèle souvent nécessaire.

Dans le cas très fréquent où l'horizon temporel est supérieur à un an, il est nécessaire de procéder à ce qu'on appelle l'actualisation des coûts[10].

---

9 Ainsi, une étude canadienne portant sur des femmes atteintes de lupus érythémateux disséminé a montré que selon la façon de valoriser le coût lié aux pertes de production, voire le temps de loisir perdu, on aboutissait à un coût annuel moyen par patiente allant de 1 424 dollars canadiens à 22 604, soit un écart de 1 à 15 (Clarke AE, *et al. J Rheumatol* 2000 Nov ; 27 (11) : 2597-604).

10 L'actualisation s'applique aussi potentiellement aux effets de santé, mais il existe une importante controverse à ce sujet, notamment lorsqu'il s'agit des QALYs. Voir le chapitre 6 à ce sujet.

L'actualisation est nécessaire dès lors que l'on veut comparer des valeurs monétaires relatives à des dates différentes car une somme de 100 € dans un an n'est pas égale à la même somme dans 10 ans et, en général, les individus ne sont pas indifférents à la date de survenue d'un événement, en particulier la perception d'un revenu ou le financement d'une dépense. On préfère généralement toucher 100 € aujourd'hui que dans un an. On parle ainsi en microéconomie d'une « préférence pour le présent », qui peut être facilement mesurée pour chaque individu par la somme X qui le conduirait à être indifférent entre la perception d'une somme de 100 € aujourd'hui et une somme 100 + X € dans un an. Cette prime demandée de X € est la contrepartie d'un facteur d'impatience, du refus d'une abstinence aujourd'hui : on préfère consommer aujourd'hui plutôt que demain. C'est cet effet « temps » que cherche à capturer le principe de l'actualisation qui, en revanche, ne vise à tenir compte ni d'une éventuelle évolution des prix qui modifierait le pouvoir d'achat de ces 100 € (un effet d'inflation), ni de l'incertitude sur le futur qui peut faire craindre la réalité d'un remboursement futur (« un tiens vaut mieux que deux tu l'auras ») en fonction du degré d'aversion au risque (un effet risque).

Au niveau collectif, cette préférence pour le présent se matérialise par un taux de substitution entre le futur et le présent qui peut être compris comme un taux d'intérêt (sans risque et hors inflation) lorsqu'on emprunte aujourd'hui pour rembourser demain, mais aussi comme un taux d'actualisation lorsqu'on veut savoir quelle est la valeur présente d'une valeur future. Dans une perspective de préférence sociale pour le temps, on peut montrer que le niveau du taux d'actualisation à prendre en compte doit être fondé sur le taux d'intérêt (hors inflation) existant sur des actifs sans risque, par exemple les obligations publiques, c'est-à-dire de l'ordre de 3 % à 4 % dans la période actuelle.

Ainsi, l'actualisation des coûts permet de ramener n'importe quelle séquence de coûts supportés à des dates différentes à leur valeur actuelle selon un principe de calcul simple. Si l'on considère une valeur future G relative à la période t. La valeur actualisée VA à la date présente (période 0) de cette somme future est donnée par le montant $\frac{G}{(1+r)^t}$ où r est le taux d'actualisation. Si par exemple G vaut 1000 € et que t = 1, on en déduit que VA = 990 € si r = 1 % (soit 1000/1,01), mais VA = 952 € si r = 5 % (1000/1,05).

Mais s'il existe de bons arguments pour justifier le recours à l'actualisation, il faut reconnaître que cette pratique n'est pas sans effet sur la comparaison des coûts entre stratégies alternatives. Si l'on compare par

exemple sur une période de 6 années une stratégie chirurgicale A dont le coût initial est de 10 000 € auquel s'ajoute un coût de suivi annuel de 1 000 € à une stratégie médicale dont le coût annuel est constant à 2 500 € (tableau 2), la comparaison des coûts totaux est sensiblement modifiée par le taux d'actualisation choisi. Les calculs montrent que si on n'actualise pas, les deux traitements ont le même coût cumulé de 15 000 €. En revanche, si on actualise à 3 % ou *a fortiori* à 5 %, le coût total actualisé diffère entre les deux options, et ce d'autant plus que le taux d'actualisation est important.

Tableau 2. Comparaison des coûts de 2 stratégies, A chirurgicale et B, médicale, à l'horizon de six ans et influence du taux d'actualisation

	année 0	**année 1**	**année 2**	**année 3**	**année 4**	**année 5**
Traitement A	10 000 €	1 000 €	1 000 €	1 000 €	1 000 €	1 000 €
Traitement B	2 500 €	2 500 €	2 500 €	2 500 €	2 500 €	2 500 €

	taux d'actualisation		
	0 %	3 %	5 %
Traitement A	15 000 €	14 580 €	14 329 €
Traitement B	15 000 €	13 949 €	13 324 €

Dans cet exemple, on voit que l'actualisation, qui réduit la valeur actualisée des coûts futurs, avantage le traitement B qui reporte une partie des coûts dans le futur par rapport au traitement A qui est défavorisé parce qu'il fait supporter l'essentiel des coûts à la première période. L'actualisation n'est donc pas neutre, et ceci est particulièrement important lorsqu'on considère des stratégies préventives, notamment vaccinales, dont la caractéristique est d'engendrer des coûts immédiats pour éviter des coûts futurs dont la valeur actualisée est réduite. Ceci peut conduire à faire des variantes en considérant plusieurs valeurs du taux d'actualisation.

Chapitre 5
La qualité de vie et sa mesure

Le thème de la *qualité de vie* s'est progressivement imposé dans le monde de la santé à partir des années 70 tant en santé publique qu'en recherche clinique et il a marqué la reconnaissance de la légitimité du point de vue des patients sur leur propre prise en charge, potentiellement différent de celui des professionnels de santé.

Ce thème a suscité un important effort de réflexion et de recherche, chaque discipline médicale, chaque société savante, mettant au point et validant ses propres instruments de mesure de la qualité de vie dans son domaine. Les agences d'enregistrement – un temps réticentes en raison de doutes sur la validité et la fiabilité de ces instruments – ont fini par les reconnaitre comme des mesures légitimes du bénéfice pour la délivrance des AMM.

Ce chapitre est spécifiquement consacré à la qualité de vie reliée à la santé (« *Health Related Quality of Life* », HRQoL[1]) dans une perspective médico-économique. On s'intéressera tout particulièrement au concept d'« utilité » qui, combiné à l'espérance de vie, a donné naissances aux QALYs (« *Quality-Adjusted Life Years* »), dont on traitera au chapitre suivant, et aux études coût-utilité.

On ne considérera donc pas les instruments et les méthodes de mesure de la qualité de vie en recherche clinique ou en santé publique et, en dehors d'un bref rappel des définitions et de la typologie des instruments, nous renverrons aux excellentes synthèses existantes[2].

---

1 Sauf mention contraire, l'expression « qualité de vie » dans ce chapitre s'entend comme « qualité de vie reliée la santé ».

2 Par exemple A. Leplège. *Les Mesures de la qualité de vie*. Que sais-je, Octobre 1999, ou A. Leplège et J. Coste (direction) : *Mesure de la santé perceptuelle et de la qualité de vie : méthodes et applications*. ESTEM, 2002.

On ne considérera pas non plus la qualité de vie *en général* qui a fait l'objet de nombreux travaux comme mesure de développement socio-économique alternative au traditionnel PIB par tête qui a tendance à mettre l'accent sur la seule production de biens matériels indépendamment de leurs conditions de production et de leurs usages sociaux[3].

La « Qualité de vie reliée à la santé » : mesure du bonheur ou score de capacité fonctionnelle ?

L'OMS a donné dès 1993 une célèbre définition de la qualité de vie comme « *la perception qu'a un individu de sa place dans l'existence, dans le contexte de la culture et du système de valeurs dans lesquels il vit en relation avec ses objectifs, ses attentes, ses normes et ses inquiétudes* ». Et elle ajoute que la qualité de vie « *est influencée de manière complexe par la santé physique du sujet, son état psychologique, son niveau d'indépendance, ses relations sociales ainsi que sa relation aux éléments essentiels de son environnement* ».

L'OMS n'a sans doute pas voulu s'enfermer dans une définition normative qui, donnant une définition universelle du bonheur et du bien-être, aurait risqué de refléter l'effet d'une culture dominante, mais la contrepartie de cette louable préoccupation est le très faible degré d'opérationnalité de sa définition.

En pratique, les études et mesures de la qualité de vie reliée à la santé, ont contourné la difficulté de plusieurs manières : d'abord, en dressant une liste de capacités fonctionnelles physiques ou mentales « objectives », relativement transculturelles, en en évaluant la qualité de vie par le degré de limitation qu'impose l'état de santé à l'exercice de ces capacités ; ensuite en opérant la fameuse distinction entre les instruments[4] de mesure dits « spécifiques » à un contexte physiopathologique donné et les instruments « génériques » qui ambitionnent de mesurer la qualité de vie « en général » indépendamment du contexte médical ; enfin, en accordant une attention toute particulière à la transposition et à la validation des instruments de mesure d'une culture ou d'une langue à l'autre.

---

3 Des économistes comme Amartya Sen ou Robert Stiglitz ont ainsi proposé des « indicateurs de développement humain » qu'on peut assimiler à des mesures de qualité de vie en général, intégrant, outre la dimension économique, des aspects éducation, santé, environnement, etc.

4 Un instrument de mesure de la qualité de vie se compose d'un questionnaire descriptif des états de santé et de valeurs numériques attachées à chacun de ces états, elles-mêmes déterminées à partir d'études psychométriques. Des algorithmes de calcul de scores permettent d'agréger les différentes questions en score par dimension (douleur, anxiété, relations sociales, etc.) ou score global.

Instruments génériques et Instruments spécifiques

De très nombreux instruments spécifiques ont été développés dans le cadre de l'évaluation clinique des traitements. Ce sont souvent des scores symptômes, par définition très différents d'une pathologie à l'autre, qui ont parfois anticipé la problématique de la qualité de vie en santé. C'est ainsi que David A. Karnofsky a publié dès 1949 le score éponyme (le « *Karnofsky Performance Status* » ou KPS) pour classer les états d'autonomie des patients sous chimiothérapie anti-cancéreuse.

Reposant sur des instruments différents et visant des symptômes différents, les mesures de qualité de vie ainsi obtenues en sont pas comparables entre elles. D'où le développement d'instruments dits « génériques », valables pour différentes pathologies, conçus pour couvrir tout le champ médical avec des questions plus générales. L'évaluation économique se voulant universelle et indépendante du domaine d'application, n'utilise pratiquement que ce type d'instrument.

Il en existe de nombreux spécimens, dont le plus célèbre est sans doute l'échelle « courte » en 36 items dite « *SF-36* » (Short form-36) – et ses déclinaisons SF-12, SF-8 – issue des travaux menés au début des années 80 au sein du groupe Medical Outcome Study (MOS) sous la direction de John Ware, alors chercheur à la Rand Corporation en Californie[5].

La SF-36 possède trois caractéristiques que l'on trouve dans la plupart des autres échelles comme le « *Nottingham Health Profile* » (NHP)[6], le « *Sickness Impact Profile* » (SIP)[7], le « *Quality-of-Well Being* », le « *World Health Organisation Quality of Life* » (WHOQOL-100)[8], le « Duke Health profile »[9], etc.[10] :

---

5 Leplège A, Ecosse E, Verdier A, *et al.* The French SF-36 Health Survey: Translation, cultural adaptation and preliminary psychometric evaluation. *J Clin Epidemiol* 1998 ; 51(11) : 1013-23.

6 Bucquet D, Condon S, Ritchie K. The French version of the Nottingham Health Profile. A comparison of items weights with those of the source version. *Social Science & Medicine* 1990 ; 30(7).

7 Chwalow AJ, Lurie A, Bean K, *et al.* A French version of the Sickness Impact Profile (SIP): stages in the cross cultural validation of a generic quality of life scale. *Fundam Clin Pharmacol* 1992 ; 6(7) : 319-26.

8 http://www.who.int/healthinfo/survey/whoqol-qualityoflife/en/.

9 Parkerson GR, Broadhead WE, Tse CK. The Duke Health Profile. A 17-item measure of health and dysfunction *Med Care* 1990 ; 28(11) : 1056-72.

10 Coons SJ, Rao S, Keininger DL, *et al.* A comparative review of generic quality-of-life instruments. *Pharmacoeconomics* 2000 ; 17(1) : 13-35.

- Une orientation délibérément opérationnelle en direction de la recherche clinique et de l'évaluation des politiques de santé publique, avec la possibilité d'une auto-administration du questionnaire.
- Une appréciation de la qualité de vie sur différentes « *dimensions* », chacune reflétant une capacité fonctionnelle, physique, mentale, sociale.
- Une synthèse de ses capacités sous la forme d'une appréciation générale.

Profils et Scores

A. La plupart de ces instruments sont *multidimensionnels*, c'est à dire qu'ils sont conçus et structurés de manière à appréhender les différents aspects de la qualité de vie. Le nombre et la formulation de ces dimensions peuvent varier d'une échelle à l'autre mais on retrouve généralement une dimension physique (douleur, handicap), une dimension psychique ou mentale (stress, anxiété, etc.), une dimension relationnelle, une dimension cognitive, etc. Elles sont ainsi au nombre de 8 dans l'échelle SF-36[11], chaque dimension comportant elle-même plusieurs questions. La dimension « activité physique » (PF) compte 10 items, la dimension « relations avec les autres » (SF) 2, la dimension douleur (BD) 2, etc.

La multiplicité des dimensions définit un « profil » de qualité de vie qui peut être meilleure ou moins bonne selon les dimensions. La *figure 1* montre l'évolution du profil de qualité de vie de patientes atteintes de cancer du sein lors d'un programme de réhabilitation psychologique et physique nommé « Move Beyond Cancer »[12]. On notera que si le programme améliore la qualité de vie sur beaucoup de dimensions il ne change cependant pas l'appréciation générale de l'état de santé de patients qui continuent de se percevoir comme gravement malades.

B. Aux profils multidimensionnels de la qualité de vie, s'ajoutent donc – ou s'opposent – les mesures *unidimensionnelles*, les « scores » censés synthétiser en une mesure unique les différentes dimensions. Ces scores peuvent s'obtenir de plusieurs manières : en complément d'une analyse multidimensionnelle au moyen d'une question de synthèse, souvent posée en fin de questionnaire, par exemple sous la forme d'une « échelle visuelle analogique »

---

11 1. (PF) Fonctionnement physique – 2. (RP) Limitation physique – 3. (BP) Douleur physique – 4. (SF) Fonctionnement social – 5. (RE) Limitation émotionnelle – 6. (MH) Santé mentale générale – 7. (VT) Vitalité – 8. (GH) Etat de santé général. PF, RP, BP et GH donnent un score « physique » et VT, SF, RE et MH, un score « mental ».

12 Ganz PA, Kwan L, Stanton AL, *et al.* Physical and psychosocial recovery in the year after primary treatment of breast cancer. *J Clin Oncol* 2011 ; 29(9) : 1101-109.

(EVA) ; en effectuant une somme pondérée des différentes dimensions ; en utilisant des instruments ne comportant qu'une seule dimension.

C'est par exemple le cas de l'indice de Karnofsky dans le cancer[13]. L'échelle SF 36 offre la possibilité de calculer des scores synthétiques par la somme pondérée des différentes dimensions : PF, RP, BP et GH donnent un score « physique » et VT, SF, RE et MH, un score « mental ».

Les EVA sont des instruments très simples – largement utilisés pour évaluer la douleur – où les patients sont invités à situer leurs états de santé entre les deux extrêmes que sont le « meilleur » et le « pire » état de santé possible, auxquels sont conventionnellement affectés les notes 1 (ou 100) et 0.

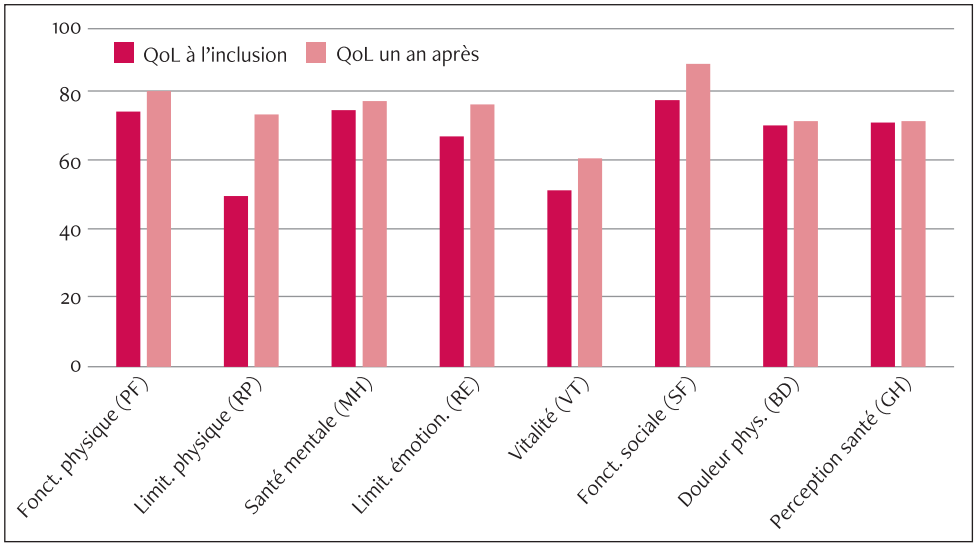


Figure 1. Profil de qualité de vie de patientes atteintes de cancer du sein au moment de leur entrée dans un programme de réhabilitation psychologique et physique et après un an de suivi[14].
QoL : quality of life.

13 L'indice de Karnofsky comporte un indicateur unique allant de 100 % pour l'état « normal, pas de signe de maladie » à 10 % pour « *moribond, processus fatal progressant rapidement* ». Entre 100 % et 80 % le sujet est « *capable de vivre une vie normale* ». Entre 70 % et 40 % il est « incapable de travailler, capable de vivre chez lui et d'assumer ses besoins personnels ». Une assistance variable est nécessaire ». Au-dessous de 40 %, il est « *incapable de s'occuper de lui-même et nécessite des soins hospitaliers ou l'équivalent* ».

14 Ganz PA, Kwan L, Stanton AL, *et al.* Physical and psychosocial recovery in the year after primary treatment of breast cancer. *J Clin Oncol* 2011 ; 29(9) : 1101-109.

Scores et Utilités

Ces scores unidimensionnels ne doivent pas être confondus avec des « utilités » qui sont également des mesures unidimensionnelles et synthétiques de l'état de santé mais qui reposent sur des fondements théoriques et pratiques très différents.

Une utilité est en théorie économique une mesure quantitative de la satisfaction ou du bien-être tiré de la consommation d'un bien choisi par un agent de préférence à un autre (cf. l'annexe à ce chapitre : l'utilité en théorie économique). Les concepts importants sont ceux de « choix » et de « préférence ». La qualité de vie est indirectement dérivée des choix – le plus souvent déclaratifs – entre les options thérapeutiques proposées aux personnes interrogées, chacune comportant des bénéfices et des risques. En effectuant un choix, éventuellement celui du *status quo*, le sujet révèle l'utilité qu'il attache à son état de santé.

La mesure résultant d'un choix de cette nature, même virtuel, n'est pas nécessairement la même que celle issue d'une notation directe sur une EVA.

Supposons qu'un patient, souffrant d'une pathologie chronique et questionné sur sa qualité de vie, réponde par un « score » de 0,75 ou 0,8 sur une EVA de type [0,1]. On peut concevoir que ce même patient, à qui on proposerait de choisir entre rester dans son état actuel ou subir une intervention chirurgicale qui le guérira mais au prix d'un risque opératoire élevé, estimé par exemple à 5 %, refuse cette option.

Dans ce cas, la théorie de l'utilité espérée (cf. annexe) montre que le patient valorise en fait son état de santé à un minimum de 0,95 (soit 95 % de chances de recouvrer une santé parfaite, estimée conventionnellement à 1,0). Il révèle ainsi par son choix une valeur de qualité de vie supérieure à celle spontanément cotée avec l'EVA.

L'utilité est en quelque sorte une évaluation « payante » de la qualité de vie, même s'il s'agit d'un prix virtuel issu d'une « *denk experiment* ». La qualité de vie s'apprécie alors par le sacrifice qu'il faut consentir pour l'améliorer significativement et elle est d'autant plus mauvaise que le sacrifice accepté est grand. Et on peut se plaindre de son état de santé sans pour autant être prêt à prendre un risque important pour l'améliorer[15]…

Ces mesures dites « *choice based* » sont donc en ligne avec l'objectif de l'évaluation économique en santé qui est la quête de l'efficience dans

15 L'inverse est sans doute moins vrai : on ne prendra pas de risques importants pour se débarrasser d'une maladie dont on ne se plaint pas !

l'allocation des ressources rares à des fins alternatives. Cela explique que les guidelines recommandent pour la mesure de la qualité de vie le recours à de méthodes « *choice based* » ou, dans la terminologie de la HAS, des « *scores de préférence* ».

Élicitation des préférences : Standard Gamble (SG) et Time Trade-Off (TTO)

La littérature propose deux approches pour calculer ces « *scores de préférence* » ou ces utilités, la méthode des loteries (« Standard Gamble ») et celle de l'arbitrage temps (« Time-Trade-Off »).

A) La première est directement issue de la théorie de l'utilité espérée (cf. annexe). Elle consiste à proposer au sujet un choix entre deux options. Celle de rester dans l'état de santé dont on cherche à mesurer l'utilité et celle de subir une intervention permettant de retrouver un état de pleine santé (dont l'utilité est conventionnellement fixée à 1) mais au risque de subir, avec la probabilité p, une complication fatale conduisant à l'« état mort » (d'utilité 0). L'arbre de décision (*figure 2*) illustre ce choix[16].

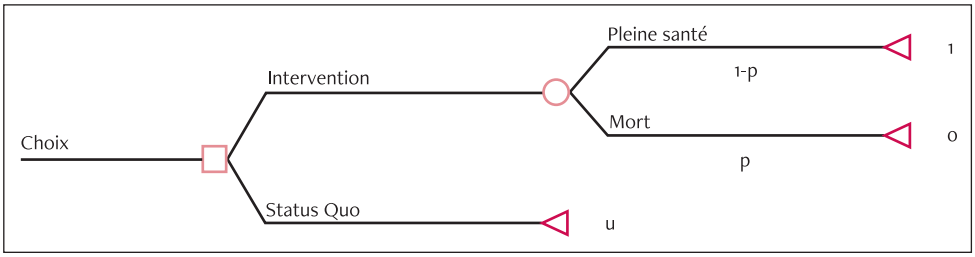


Figure 2. Le « Standard Gamble ».

Si le risque opératoire est nul, un patient rationnel et « conséquentialiste » (cf. Annexe) choisira l'option intervention. Si inversement la probabilité de décès est égale à 1, ce même patient choisira le *status quo*. D'après la théorie de l'utilité espérée il existe, sous certaines hypothèses de comportement rationnel, une probabilité d'équilibre p* pour laquelle le patient est indifférent entre les deux options. Cela signifie que pour lui l'utilité du statut quo est égale à celle de l'intervention qui, selon la même théorie, est elle-même

16 Cf. chapitre sur la modélisation.

égale à la somme pondérée des utilités de ses conséquences à savoir la bonne santé (1) ou le décès (0).

On en déduit donc immédiatement que : u = 1-p*.

L'utilité du *status quo* est égale au complément à 1 du risque maximal qu'on est prêt à prendre pour retrouver une bonne santé : plus le risque est élevé plus l'utilité du *status quo* est faible.

B) L'arbitrage temps (TTO)

La deuxième méthode, développée par les économistes de l'université McMaster à Toronto[17], propose aux sujets un choix entre rester dans l'état *status quo* durant toute leur durée de vie restante (soit E sur la *figure 3*) ou accéder immédiatement à un état parfait au prix d'une réduction de cette espérance de vie (de E à x sur la *figure 3*).

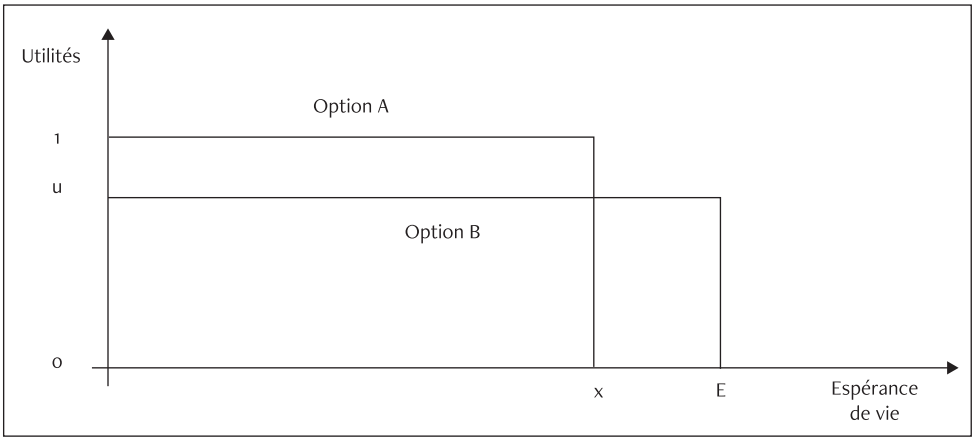


Figure 3. Le « Time-Trade-Off ».

On cherche par tâtonnement à identifier le sacrifice (E-x) pour lequel les deux options sont équivalentes. On en déduit alors que si une personne estime équivalent le fait de rester, disons, 20 ans dans son état de santé actuel ou de retrouver immédiatement un état de santé parfait au prix d'un sacrifice de 2ans d'espérance de vie, c'est qu'elle évalue l'utilité de l'état actuel à 18/20 soit 0,9 sur l'intervalle [0,1].

17 Torrance GW. Measurement of health state utilities for economic appraisal. *J Health Econ* 1986 ; 5(1) : 1-30.

C) L'une et l'autre de ces méthodes possèdent la caractéristique d'être issues d'un choix offert aux patients et de refléter leurs préférences pour différentes situations cliniques ; elles sont l'une et l'autre recommandées par les guidelines d'évaluation économique.

Elles sont cependant différentes à bien des égards. Elles reposent sur des postulats théoriques différents. Dans l'arbitrage temps, le sujet est confronté à une diminution certaine de son espérance de vie et non à un risque de décès probabilisable. Le traitement du temps est également différent : le choix de l'arbitrage temps porte sur des couples quantité/qualité de vie alors que la dimension temporelle ne figure pas explicitement dans les loteries.

Empiriquement les mesures ne sont pas nécessairement identiques : l'utilité associée à une baisse importante de l'acuité visuelle (moins de 20/40 pour le meilleur œil dans l'échelle de Snellen) chez des patients présentant une rétinopathie est ainsi de 0,770 mesurée en TTO et de 0,853 mesurée en SG[18].

La méthode d'élicitation des préférences n'est donc pas sans influence sur l'estimation du ratio coût-utilité et une analyse rétrospective des études coût-utilité remise au NICE britannique montre une grande hétérogénéité des choix méthodologiques ce qui reste problématique[19].

Finalement le choix entre SG et TTO relève surtout de considérations pratiques, la méthode TTO se révélant plus facile et plus intuitive à utiliser dans les enquêtes de terrain.

Les états de santé « pires que la mort »

Une des questions que posent ces méthodes est la définition conventionnelle des états de référence et notamment de celui auquel est affectée la valeur 0. En général, cette a valeur est associée à l'état « mort » considéré comme le pire possible. Mais des personnes peuvent considérer que certains états de santé, associant par exemple de grandes douleurs à une perte totale d'autonomie, sont « pires que la mort ».

---

18 Sharma S, Brown GC, Brown MM, *et al.* Validity of the time trade-off and standard gamble methods of utility assessment in retinal patients. *Br J Ophthalmol* 2002 ; 86(5) : 493-6. Les gradients sont cependant comparables et les mesures dútilité baissent continument quand l'acuité visuelle diminue.

19 Tosh JC, Longworth JL, Geirge E. Utility values in National Institute for Health and Clinical Excellence (NICE) Technology Appraisals. *Value in Health* 2011 ; 14 : 102-9.

Techniquement, cela nécessite d'adapter les méthodes d'interrogation[20]. Dans le cas du Time Trade-Off, une fois admis que l'état dont on cherche à mesurer l'utilité est pour la personne interrogée « *pire que la mort* », le choix offert devient celui entre une mort immédiate (0 sur tout la période) et une vie partagée entre x années de bonne santé suivies de (E-x) années dans l'état en question, E étant l'espérance de vie (*figure 4*). À l'équilibre (indifférence entre les deux options), on aura :

$$1.x + u.(E\text{-}x) = 0 \text{ et donc : } u = -\frac{x}{E - x.}$$

Si par exemple, une personne dont l'espérance de vie est de 35 ans, est indifférente entre une mort immédiate et la perspective de passer 10 ans en bonne santé suivie de 25 années dans un état de handicap majeur accompagné de douleurs sévères, l'utilité attachée à état est de – 0,40.

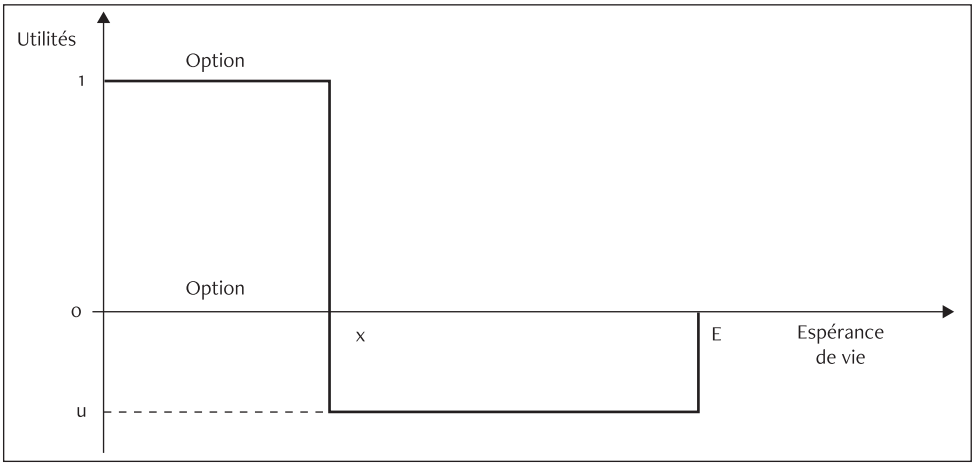


Figure 4. Le Time-Trade-Off pour les « états pires que la mort ».

Cette solution n'est pas sans soulever des problèmes philosophiques et techniques. Philosophiquement, comme la mort peut alors entrainer des gains d'efficience, elle pourrait déboucher sur des recommandations « d'euthanasie médico-économique », surtout si les utilités sont valorisées à partir d'enquêtes en population générale. De manière générale cette notion d'états

20 Patrick DL, Starks HE, Cain KC, *et al.* Measuring preferences for health states worse than death. *Med Decision Making* 1994 ; 14 : 9-14.

« pires que la mort » pose la question du statut de la mort dans les études médico-économiques. Elles ont, en effet, tendance à considérer la mort comme un « état de santé » ce qui est évidemment hautement contestable.

Mesures de l'utilité et « Mapping »

En pratique la mesure des préférences est un exercice délicat. Il faut en effet poser des questions très abstraites, très artificielles, appelant des réponses dont il est difficile de vérifier la plausibilité. Le dispositif pour mesurer des utilités avec un minimum de fiabilité est lourd : il nécessite un grand nombre de sujets, des interrogateurs formés, du temps pour expliquer aux sujets le sens de la démarche. Ces conditions ne sont pas toujours remplies en pratique et les plans de développement cliniques comportent rarement une mesure de l'utilité.

Beaucoup d'études ont recours à des techniques de « mapping » qui sont admises, sous réserve d'être exécutées correctement, par les agences d'évaluation[21]. La technique consiste à transposer les valeurs d'un indicateur clinique, d'un score symptômes ou même d'une mesure de qualité de vie couramment utilisée en recherche clinique dans une échelle dont on connait les scores d'utilité.

Cette transposition suppose :

- Qu'ait été préalablement mesuré simultanément, sur une même population, l'indicateur clinique et l'index d'utilité.
- Qu'ait été établie une relation statistique de type régression linéaire[22] entre l'indicateur clinique et l'utilité :

$$u = \alpha.i + \beta + \epsilon$$

ou u est l'utilité, i le score symptômes et α et β des coefficients de régression[23].

Une fois ce « mapping » réalisé, toutes les études utilisant l'indicateur clinique peuvent transformer ce dernier en utilités.

La littérature fournit de nombreux exemples d'études de « *mapping* », la plupart liant un indicateur clinique (ou de qualité de vie de type SF-36) aux index d'utilité de l'échelle 5Q-5D, la plus utilisée. En rhumatologie par

21 Longworth L, Rowen D. Mapping to obtainEQ-5D utility values for use in NICE health technology assessments. *Value in Health* 2013 ; 16 : 202-10.

22 D'autres formes fonctionnelles peuvent être utilisées, par exemple des régressions logistiques qui contraignent la variable expliquée (l'utilité) entre 0 et 1.

23 Et ε le terme d'erreur supposé d'espérance nulle et de variance constante

exemple des études de « mapping » ont lié le WOMAC, un score symptôme utilisé dans l'arthrose, aux utilités de l'EQ-5D[24]. Dès lors toute étude utilisant le WOMAC peut se référer à cet algorithme pour transformer l'indicateur d'état de santé en indice d'utilité.

La technique de « *mapping* » doit cependant être menée avec beaucoup de soin. Il convient de s'assurer que la population sur laquelle est mesuré le score symptôme que l'on veut transformer en utilité est comparable à celle qui a donné lieu à l'algorithme de transposition. Celui-ci doit, par ailleurs, être validé et se montrer sensible aux variations des indicateurs d'état de santé.

La qualité de vie de qui ? Patients ou population générale ?

Une question en débat est celle de savoir qui doit exprimer une appréciation ou des préférences sur les états de santé.

Selon un premier point de vue, les préférences doivent être celles des patients qui connaissent et vivent les états de santé qu'il s'agit d'évaluer. Cette option paraît raisonnable mais elle soulève l'objection que les patients pourraient se trouver en situation de conflit d'intérêt dans la mesure où l'appréciation de leur propre état pourrait entrainer des allocations de ressources en leur faveur.

Selon un second point de vue, le financement des soins étant très largement assuré, du moins en Europe, par des fonds publics, la valorisation des états de santé doit être faite du point de vue de la population générale. Sinon les ratios coût/utilité confronteraient un numérateur reflétant le point de vue des patients et un dénominateur mesurant la charge de la collectivité. L'objection consiste alors à se demander comment la population générale pourrait évaluer de manière pertinente des états de santé dont elle n'a pas l'expérience ?

Il n'y a pas de réponse définitive à ce dilemme. Si certaines agences, comme le NICE, recommandent que la valorisation des états de santé reflète les préférences de la population générale, en pratique la plupart des études d'évaluation de la qualité de vie se réfère au point de vue des patients.

Un outil privilégié en évaluation économique : l'EQ-5D[25]

Les trois types de mesure de la qualité de vie en santé, profil, score et utilités se retrouvent dans un même instrument, sans doute l'un des plus utilisés en évaluation économique, l'EuroQol 5 Dimensions (EQ-5D).

---

24 Kiadaliri AA, Englund M. Assessing the external validity of algorithms to estimate EQ-5D-3L from the WOMAC Health Qual Life Outcomes. 2016 ; 14(1) : 141.

25 Cf. Kind P, Brooks R, Rabin R. *EQ-5D, Concepts and Methods*, Springer, 2005.

L'EuroQol Group, un groupe de chercheurs des Pays-Bas, de Norvège, de Suède, de Finlande et du Royaume-Uni, a conçu dans les années 80 un instrument générique destiné à l'évaluation médico-économique (EQ-5D) qui a connu un très large succès et auquel est consacrée une importante littérature théorique et pratique. L'EuroQol Group s'est ouvert à d'autres chercheurs du monde entier et s'est institutionnalisé, avec la création d'une fondation, d'un site internet, etc.

L'EQ-5D se compose de deux parties :

1. Une partie « profil » à 5 dimensions.

2. Une partie EVA sous la forme d'une échelle verticale graduée de 0 à 100, sur laquelle le répondant est invité à situer « *son état de santé aujourd'hui* ».

Dans sa version française, les 5 dimensions sont les suivantes : i) Mobilité, ii) Autonomie de la personne, iii) Activités courantes, iv) Douleur ou gêne, v) Anxiété ou dépression.

À chacune des dimensions étaient associées à l'origine 3 modalités de réponse : 1. Absence de problèmes – 2. Présence de problèmes – 3. Problèmes importants.

L'échelle permettait ainsi de coter 253 états de santé (5^3) allant de l'état « 11111 » (aucun problème sur toutes les dimensions) à « 33333 » : problèmes sévères sur toutes les dimensions.

Un profil « 21221 » correspondait ainsi à un état de santé avec restriction de mobilité, sans entrave pour l'autonomie mais pénalisant les activités courantes, avec des douleurs mais sans effet sur le moral de la personne questionnée.

Pour répondre à la critique du caractère schématique et peu sensible de la version 3 niveaux, une version à 5 niveaux de réponses a été récemment développée : 1. Absence de problèmes (pour se déplacer, pour effectuer des activités courantes, etc.) – 2. Problèmes légers – 3. Problèmes modérés – 4. Problèmes sévères – 5. Incapacité (ou douleurs extrêmes et anxiété extrême pour les deux dernières dimensions).

La version EQ-5D(5L) décrit ainsi 3.125 (5^5) états de santé entre le meilleur (« 111111 ») et le pire (« 55555 »).

Ce qui fait l'originalité de l'EQ-5D et justifie son large usage en évaluation économique c'est, qu'à chaque profil descriptif de l'état de santé, a été attaché un coefficient d'utilité déterminé par une méthode « *choice based* », généralement le TTO.

En France le travail a été effectué de manière rigoureuse sur l'EQ-5D (3L)[26]. Un travail d'enquête menée en population générale par des enquêteurs formés spécialement a permis d'obtenir des « scores de préférences » pour certains états pivots, à partir desquels ont été obtenus par estimation statistique, l'ensemble des scores des 543 états.

La *figure 5* donne les scores des états pivots. À l'état « 11111 » a ainsi été normalement affecté le coefficient 1. Le score « 33333 » est considéré comme « *pire que la mort* » avec un score négatif de -0,60. L'état de santé n'entrainant que des problèmes légers d'autonomie (toilette, habillement) pour (« 12111 ») est côté 0,78, tandis qu'un état caractérisé par des douleurs sévères (« 11131 ») est côté 0,47. À chaque descriptif d'état de santé obtenu directement ou par « mapping » peut ainsi être attaché un coefficient d'utilité.

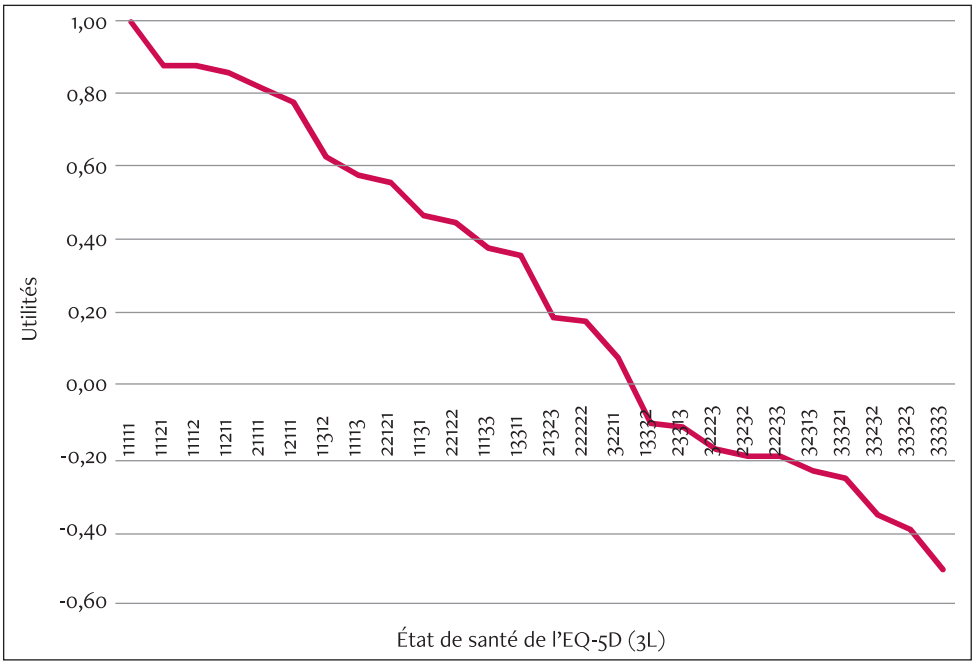


Figure 5. Utilités attachées à certains états de santé de l'échelle 5Q-5D (3L) pour la France.
Source : Chevalier J, *op. cit.*

26 Chevalier J. *Mesure de l'utilité attachée aux états de santé. Valorisation de l'index d'utilité EQ-5D et évolution de l'échelle actuelle en France*, Thèse pour le Doctorat en Sciences Économiques, Université Paris-Dauphine, 2010.

Annexe : l'utilité en économie

L'utilité (utility) est un concept clé de la théorie économique développé à la fin du XIXe siècle dans le cadre des théories de la valeur et du marché : comment s'établit sur le marché la valeur des biens soumis à transactions ?

Du côté offre, les choses sont (relativement) simples : les industriels veulent couvrir leurs coûts de production et dégager le plus large profit possible. Tout cela se calcule dans différentes situations, concurrence, monopole, oligopole, etc. Du côté de la demande c'est plus compliqué : les consommateurs sont supposés choisir leurs consommations en cherchant à maximiser leur « utilité » (leur bien-être ou leur satisfaction) sous contrainte de budget.

À la fin du XIXe siècle, certains auteurs, en ligne avec la philosophie utilitariste de Jeremy Bentham, ont avancé l'idée qu'il était possible de mesurer le bien-être et la satisfaction des consommateurs, dans une unité ad-hoc, « l'util », assimilable à une grandeur physique, mesurable, additionnable[27].

Mais cette voie, très marquée philosophiquement, a vite été abandonnée au profit d'une approche plus objective, dénuée de connotations hédonistes, fondée sur le concept de « préférence ». Une préférence est observable et ne nécessite pas d'interprétation psychologisante : peu importe les raisons qui sous-tendent les préférences du moment que celles-ci sont clairement exprimées par les agents.

Toutefois ces préférences sont qualitatives et se prêtent mal au calcul économique. Les économistes ont donc cherché à *quantifier ces préférences*, à les exprimer sous forme numérique à travers ce qu'on appelle précisément les utilités.

La tentative a d'abord été faite en univers certain. Elle a été ensuite renouvelée dans un cadre plus sophistiqué avec la prise en compte de l'incertitude.

1. La théorie classique en univers « certain »

Supposons qu'un individu « représentatif » soit confronté aux choix entre des éléments d'un ensemble (x_1, x_2, x_3, ..., x_n) qu'il connait parfaitement et entre lesquels il exprime des préférences (il préfère par exemple x_2 à x_1, etc.).

27 L'auteur le plus célèbre est sans conteste l'économiste britannique Francis Ysidro Edgeworth (1845-1926) qui, dans son livre *Mathematical Psychics* paru en 1888, allait jusqu'à imaginer un modèle d'instrument « psychophysique », « l'hédonimétre », qui mesurerait l'intensité du plaisir comme un ampèremètre mesure celle du courant électrique.

La théorie montre alors que, sous certaines conditions, on peut attacher à chaque élément x_i de l'ensemble choix un nombre $u(x_i)$ tel que $u(x_i)$ est supérieur (ou égal) à $u(x_j)$ si et seulement si l'agent préfère x_i à x_j (ou est indifférent entre x_i et x_j). Les coefficients d'utilité donnent ainsi une représentation quantitative des préférences de l'agent. CQFD.

Deux conditions pour obtenir ce résultat sont essentielles et définissent le « comportement rationnel » d'un agent parfaitement informé :

i) la condition de *totalité* selon laquelle l'agent est capable d'exprimer une préférence (ou une indifférence) pour tout couple d'éléments de l'ensemble de choix (il n'est jamais en position de ne pas pouvoir choisir).

ii) la condition de transitivité selon laquelle, si l'agent préférence x_i à x_j et x_j à x_l, il préfère x_i à x_l.

Si ces conditions (plus quelques autres plus techniques) sont vérifiées, on démontre que les préférences peuvent être représentées par des nombres, les utilités, qui quantifient les préférences.

Ce résultat, mathématiquement élégant, a un intérêt théorique : il jette un pont entre l'univers qualitatif des préférences et celui quantitatif des utilités ; il permet de formaliser des « fonctions d'utilité » d'où on peut déduire des fonctions de demande ; il met en évidence les conditions du bon fonctionnement d'un marché, etc.

Mais il n'en a aucun sur le plan pratique. On montre en effet qu'il existe une infinité d'ensemble de coefficients ordonnés de la même manière (« à une transformation monotone près ») qui peuvent valablement représenter le même système de préférences. Comme on n'a aucune raison d'en choisir un plutôt que l'autre, la portée pratique du résultat est très limitée. On parle alors d'utilités « ordinales », ce qui ne nous avance guère du point de vue du calcul empirique par rapport à la représentation de préférences elles-mêmes ordinales.

2. L'utilité espérée de von Neuman-Morgerstern (vNM)

Les choses changent si on introduit l'incertitude dans le modèle comme l'ont montré dès 1944 les mathématiciens John von Neumann et Oskar Morgenstern, les pères de la théorie des jeux[28].

Dans ce cas le consommateur est confronté non pas à un choix entre des biens mais à un choix entre des « *perspectives aléatoires* » ou « loteries », c'est-à-dire des biens auxquels sont affectés des probabilités de réalisation.

28 Dans leur célèbre livre : « Theory of Games and Economic Behavior », paru en 1944 aux éditions Princeton University Press. La théorie a été développée par Leonard Savage dans son ouvrage « The Foundations of Statistics » en 1953.

Cette situation se retrouve très largement dans beaucoup de domaines : la santé où les traitements sont affectés de probabilité de réussite ou d'échec, la finance où les choix d'investissement sont soumis à des aléas conjoncturels, l'assurance, le marché du travail, etc.

À chaque élément d'un choix (par exemple un traitement chirurgical vs un traitement médical) est attachée la liste de ses conséquences possibles ainsi que leur probabilité de survenue : par exemple l'intervention a 98 % de chances de réussir et 2 % de « chances » d'entraîner une complication grave.

On suppose que les agents ne s'intéressent pas à l'événement en tant que tel (l'intervention chirurgicale) mais *uniquement à ses conséquences* (guérir ou souffrir de la complication). On postule en outre qu'il est parfaitement informé et qu'il exprime des préférences sur les conséquences (il préfère par exemple guérir à subir une complication) et qu'il en connaît les probabilités. Il peut en revanche avoir plus ou moins peur du risque, être de nature plutôt prudente ou plutôt intrépide.

On démontre alors – sous un ensemble d'hypothèses dont certaines définissent comme précédemment la rationalité de l'agent – que, non seulement les préférences entre les perspectives peuvent être représentées par des nombres, les utilités, mais que l'utilité attachée à un choix est égale à la somme des utilités des conséquences pondérées par leur probabilité d'occurrence (théorème de l'utilité espérée).

En fixant arbitrairement des utilités à des états de référence, on peut calculer les utilités de n'importe quel état de santé, qui deviennent alors des « utilités cardinales » définies à une transformation linéaire près.

L'exemple simple illustrera ces points :

- Supposons qu'un patient, dans un état de santé donné dont on cherche à évaluer la qualité de vie, a le choix entre un traitement chirurgical et un traitement médical.

Le premier a trois conséquences possibles : succès (le patient est guéri), échec sans complication (status quo) et échec avec complications péri-opératoires graves (voire décès). Les probabilités attachées à ces conséquences sont respectivement de 78 %, 20 % et 2 %.

- Le traitement médical offre une probabilité de succès de 60 % et d'échec (status quo) de 40 % mais est dépourvu de risques graves.
- Notons u(S), u(G) et u(D), les utilités respectives des conséquences : status quo, guérison, décès.
- D'après le théorème l'utilité espérée, l'utilité attachée au choix chirurgical s'écrit :

$$U(C) = 0{,}20 \times u(S) + 0{,}78 \times u(G) + 0{,}02 \times u(D)$$

Et celle de l'option médicale :

$$u(M) = 0{,}40 \times u(S) + 0{,}60 \times u(G)$$

- Fixons enfin par convention la valeur des utilités des deux états extrêmes, par exemple 1 pour la guérison et 0 pour le décès.

Les utilités respectives de la chirurgie et du traitement médical deviennent :

$$u(C)=0{,}20.u(S)+0{,}78 \text{ et } u(M)=0{,}40.u(S)+0{,}60$$

Maintenant il reste à demander ses préférences à notre patient rationnel, après l'avoir informé de toutes les données du problème.

S'il préfère le traitement chirurgical, c'est que ce dernier lui rapporte plus d'utilité donc :

$u(C) > u(M)$ d'où on en déduit que $u(S) < 0{,}90$

S'il préfère le traitement médical c'est que

$u(C) < u(M)$ d'où : $u(S) > 0{,}90$

S'il est indifférent entre la chirurgie et le traitement médical

$u(C) = u(M)$ d'où : $u(S) = 0{,}90$

En exprimant un choix entre deux traitements aux conséquences risquées, l'un chirurgical l'autre médical, le patient révèle finalement *une mesure cardinale de l'utilité de son état de santé* qui peut être intégrée au calcul économique (sous la forme d'études « coût-utilité »).

Cette valeur sera déterminée en recherchant par tâtonnement la situation dans laquelle le patient est indifférent entre les différentes options qui lui sont offertes.

Attention : la valeur des utilités dérivées de cette manière dépend de celles attachées aux états de référence. Deux utilités attachées à deux états de santé différents ne peuvent se comparer que si les mêmes valeurs conventionnelles sont attachées aux mêmes états de référence. C'est une précaution méthodologique qui n'est pas toujours respectée en pratique.

La théorie de l'utilité espérée est très puissante et elle est largement utilisée, notamment en finance ou pour l'analyse stratégique.

Elle a toutefois été critiquée en raison de l'axiomatique très lourde et complexe qu'il faut imposer à l'agent pour obtenir le résultat. Des auteurs comme Maurice Allais en France ou Daniel Ellsberg aux États-Unis ont montré que des individus rationnels pouvaient ne pas suivre l'axiomatique de vNM[29].

---

29 D'où les « paradoxes » d'Allais ou d'Ellsberg. Cf. Allais M : « Le Comportement de l'homme rationnel devant le risque : critique des postulats et axiomes de l'école américaine ». *Econometrica* 1953 ; 21 : 503-46. Ellsberg D : Risk, Ambiguity, and the Savage Axioms. *Quarterly Journal of Economics* 1961 ; 75 (4) : 643-69.

La nature des probabilités a été questionnée (probabilités « objectives » ou « subjectives »). Des modèles de choix dans l'incertain reposant sur d'autres principes ont été développés sans toutefois atteindre le même niveau de simplicité et d'opérationnalité.

Des critiques plus philosophiques ont également été formulées à l'encontre d'une théorie qui, tout en rompant avec un utilitarisme hédoniste, reste néanmoins « *conséquentialiste* » en ce que les agents sont supposer ne s'intéresser qu'aux conséquences des choix et non à leurs modalités. C'est une hypothèse que l'on peut trouver forte par exemple dans le cas de la chirurgie.

En dépit de ces limites, la théorie de l'utilité espérée reste dominante et c'est à elle que font explicitement allusion les « guidelines » qui mettent l'accent sur des mesures « *choice based* », comme le NICE britannique ou recommandent l'utilisation de « scores de préférences » comme la HAS.

Chapitre 6
Les QALYs : Définition, calcul, interprétation

Qu'est-ce qu'un QALY ?

Le QALY (« *quality adjusted life year* ») est défini par le NICE britannique comme « *une mesure de l'état de santé d'une personne ou d'un groupe pour lesquels les bénéfices en termes de durée de vie sont ajustés pour refléter la qualité de vie* »[1]. Il constitue l'indicateur de résultat recommandé pour l'évaluation économique des technologies médicales.

Le guide de la HAS donne une définition proche : « *une unité de mesure de la durée de vie pondérée par la qualité de vie liée à la santé, cette dernière étant valorisée par un score de préférence* ». La recommandation 10 du Guide méthodologique pour l'évaluation économique de la HAS[2] le désigne comme le critère à privilégier quand l'analyse de référence est de type « coût-utilité », c'est-à-dire quand des mesures de qualité de vie fondées sur des « scores de préférence » sont disponibles (cf. chapitre sur la qualité de vie). En l'absence de telles mesures, le guide recommande de recourir à des analyses coût-efficacité dont le critère privilégié est la survie brute (recommandation 9).

Les QALYs se calculent à partir de l'espérance de vie d'un patient dans une condition donnée chaque année étant pondérée par un coefficient compris entre 0 et 1 qui en décrit la qualité.

Si durant une année, la qualité de vie est « normale » (pour une personne de même âge et sexe), la qualité de vie sera notée 1. La qualité des années au-delà de l'espérance de vie est logiquement notée 0.

1 « A measure of the state of health of a person or group in which the benefits, in terms of length of life, are adjusted to reflect the quality of life », https://www.nice.org.uk/glossary
2 HAS : *Choix méthodologique pour l'évaluation économique à la HAS*, 82 p., Octobre 2011.

Formellement on a donc pour un état de santé i un nombre de QALYs Q(i) égal à :

$$Q(i) = \sum_{t=1}^{n_i} q_{it} \qquad [1]$$

avec n_i : l'espérance de vie brute dans l'état *i*
et $0 \leq q_{it} \leq 1$

La survie « brute » est un cas particulier où l'on fait l'hypothèse que toutes les années se valent sans considération de leur qualité. Dans ce cas, on aura $q_{it} = 1$ sur toute la période et donc $Q(i) = n_i$ (le nombre de QALYs est égal à l'espérance de vie).

Graphiquement, le nombre de QALYs attaché à une intervention se représente par l'aire sous la courbe dans un graphique quantité/qualité de vie . Le gain en années de vie brute se mesure sur l'axe des espérances de vie (E1-E0), tandis que le gain en QALYs correspond à la différence entre les aires sous la courbe de l'intervention et de la référence, produit des gains en quantité et en qualité de vie (*figure 1*).

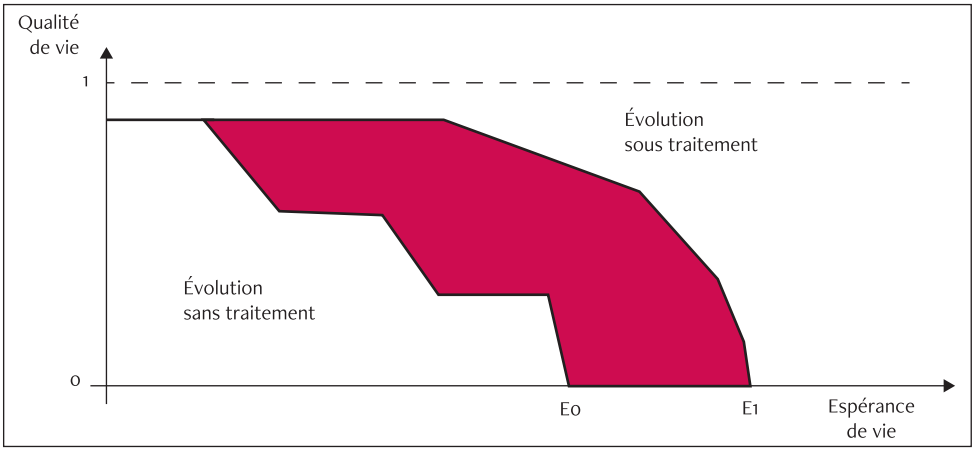
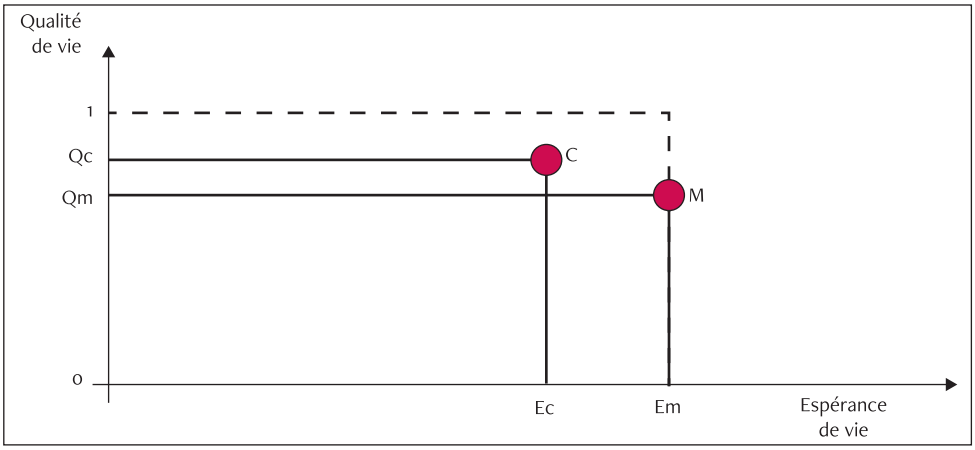


Figure 1. Graphique illustrant le calcul du nombre de QALYs attaché à une intervention ; il est représenté par l'aire sous la courbe dans un graphique quantité/qualité de vie. Le gain en QALYs se mesure par la différence des aires sous la courbe.

Pour déterminer le nombre de QALYs associé à une intervention de santé, il faut donc pouvoir disposer, pour cette intervention et pour la situation de référence :

- De l'espérance de vie des patients ;
- D'un descriptif de leur état de santé année après année ;
- D'une expression de cet état de santé par un coefficient de qualité de vie déterminé par exemple à l'aide d'une échelle de type EQ-5D (cf. chapitre sur la qualité de vie).

La figure 2 illustre l'utilisation des QALYs en vue d'un choix entre un traitement médical et un traitement chirurgical lors d'une pathologie chronique.



Avec :
C : traitement chirurgical et M : traitement médical
Ec et Qc : espérance de vie et qualité de vie du traitement chirurgical
Em et Qm : espérance de vie et qualité de vie du traitement médical

Figure 2. On suppose le choix entre un traitement chirurgical qui améliore un état chronique par rapport à un traitement médical chez des patients ayant 20 ans d'espérance de vie mais leur fait subir le risque d'une mortalité péri-opératoire non nulle. Il suffit d'une petite variation de qualité de vie (par exemple de 0,80 à 0,90) pour compenser une mortalité opératoire importante de 5 % : dans ce cas l'option chirurgicale « rapportera » 17,1 QALYs (19 × 0,9) contre 16,0 à l'option médicale (20 × 0,8).

L'origine des poids

L'étape de définition et de quantification des états de santé est certainement la plus délicate. De fait on observe une grande hétérogénéité dans la nature des coefficients de qualité et dans la manière dont ils sont élaborés.

En théorie, pour mériter la qualification d'« utilités », les poids devraient être des « scores de préférences » dérivés d'études « *choice-based* », comme les « standard gamble » (SG) ou le Time-Trade-Off (TTO) (cf. chapitre sur la qualité de vie).

Mais le recours à des « scores de préférences » originaux est loin d'être général. Selon une revue de littérature australienne portant sur toutes les études soumises au « *Pharmaceutical Benefits Advisory Committee* » (PBAC) entre 2002 et 2004, 39 % d'entre elles utilisaient des poids originaux dérivés d'échelles de type EQ-5D, 36 % empruntaient les « scores de préférences » à des études déjà publiées et 24 % avaient recours à des méthodes non fondées sur les préférences (« *rating scale* »).

La même étude rapporte que dans 58 % des cas seulement les coefficients provenaient de patients. Dans 24 % des cas ils provenaient de professionnels de santé et dans 9 % des investigateurs eux-mêmes[3].

Cette dépendance par rapport à des mesures de qualité de vie hétérogènes et souvent imprécises, constitue une faiblesse de la méthode.

Dans certains cas, le cancer par exemple[4], les échelles peuvent manquer de sensibilité de sorte que le calcul en QALYs ne se distinguera guère de celui en années de vie brutes. Dans d'autres, un excès de sensibilité peut conduire à une grande volatilité des résultats.

---

3 Scuffham PA, Whitty JA, Mitchell A, and Viney R. The use of QALY weights for QALY calculations: a review of industry submissions requesting listing on the Australian Pharmaceutical Benefits Scheme 2002-4. *Pharmacoeconomics* 2008 ; 26(4) : 297-310.

4 Tengs a analysé 110 interventions de prévention, de détection précoce et de traitement du cancer ayant fait l'objet d'analyse en termes de survie brutes et de QALYs. Il conclut : « *For interventions aimed at cancer, the outcome measures of cost/life-year and cost/QALY are highly correlated with one another. Although adjusting for quality of life can make an important difference in the evaluation of alternative approaches to cancer prevention and control, it often does not* ». Cf. Tengs TO: Cost-Effectiveness versus Cost-Utility Analysis of Interventions for Cancer: Does Adjusting for Health-Related Quality of Life Really Matter? *Value in Health* 2004 ; 7 (1) : 70-78.

Pourquoi les QALYs ?

Le QALY, dont la paternité est généralement partagée entre Milton C. Weinstein[5], professeur de politique et gestion de la santé à la Harvard Medical School et les économistes de l'Université McMaster à Toronto[6], a connu un grand succès au point de s'imposer dans la littérature comme l'indicateur de référence dès les années 80, avec notamment l'article pionnier d'Alan Williams sur le pontage coronarien dans le prestigieux British Medical Journal [7] ! Cela s'explique aisément :

- Le QALY est un concept *simple et* intuitif qui peut apparaitre comme une extension « naturelle » de la notion d'espérance de vie.
- Le nombre de QALYs est une *mesure unidimensionnelle* du bénéfice médical qui peut donc être facilement rapporté au coût qui est également une mesure unidimensionnelle de l'usage des ressources. La question de savoir s'il est efficient d'engager des ressources pour financer un traitement diminuant la quantité de vie mais augmentant sa qualité serait insoluble si ces deux dimensions n'étaient pas fusionnées dans un indicateur unique.
- Le nombre de QALY est un *indicateur ordonné du bénéfice médical* en ce sens que plus de QALYs est toujours préférable à moins de QALYs. Ce n'est pas le cas de la plupart des mesures biologiques ou cliniques dont beaucoup doivent se situer entre deux bornes de normalité. Il est donc symétrique du coût qui est lui aussi un indicateur ordonné de l'usage des ressources (moins de ressources utilisées est préférable à plus de ressources).
- Il est *universel* en ce sens que tout traitement médical a vocation à améliorer soit la quantité de vie soit la qualité de vie soit les deux. Le QALY permet ainsi des comparaisons inter-pathologies qui seraient impossibles autrement. Comment comparer les efficiences relatives d'un traitement

---

5 Weinstein MC, Stason WB. Foundations of cost-effectiveness analysis for health and medical practices. *N Engl J Med* 1977 ; 296 : 716-21.

6 Torrance GW, Thomas WH, Sackett DL. A utility maximization model for evaluation of health care programs. *Health Serv Res* 1972 ; 7 : 118-33.

7 Williams A. The Economics of Coronary Artery Bypass Grafting. *BMJ* 1985, 291 : 326-329. Il montrait, dans le cas de patients présentant une angine de poitrine sévère et une atteinte coronaire gauche, que le pontage coronarien permettait un gain de 6 ans d'espérance de vie par rapport au traitement médical (11 ans *versus* 5) et un gain moyen en qualité de vie valorisé à 0,37 par année (0,67 *vs* 0,30) soit un gain total de 5,87 QALYs (7,37 *vs* 1,5). Allan Williams concluait ainsi son article : « *les données sur lesquelles sont basées cette évaluation sont brutes et ont besoin d'être raffinées mais la méthodologie est puissante, prometteuse et ouverte à discussion* »

indiqué en deuxième ligne dans le cancer du sein métastatique et d'un médicament indiqué dans le diabète de type 2 après échec d'un traitement de première intention ? Le QALY donne une métrique commune à des interventions très différentes quel que soit leur but (préventif, curatif, symptomatique, etc.), quelle que soit la technique analysée (médicament, dispositif médical, acte chirurgical, etc.) et quelle que soit l'indication.

- On peut enfin ajouter que c'est un concept *résilient* qui, grâce à ses côtés pratiques et empiriques, a résisté jusqu'à présent aux critiques nombreuses dont il a fait l'objet dès son apparition, critiques qui, pour être généralement pertinentes, restent le plus souvent très théoriques et n'offrent guère d'alternatives aussi simples et aussi évocatrices.

Les Critiques des QALYs

Nous les rangerons dans trois catégories, les critiques tenant aux domaines d'application, les critiques purement techniques et les critiques philosophiques et éthiques.

a) Sur le premier point, nombre d'auteurs ont souligné le manque de pertinence de la métrique des QALYs dans des contextes médicaux particuliers qui posent pourtant de réels problèmes de santé publique.

C'est le cas des soins pédiatriques avec la difficulté des enfants à exprimer un point de vue sur leur qualité de vie[8], des maladies rares ou ultra rares où les données disponibles peuvent ne pas être très fiables, des pathologies mentales et psychiatriques[9], des soins palliatifs pour lesquels la notion de « qualité de vie » n'a pas toujours de sens[10] et plus généralement des cancers à des stades avancés[11].

Dans les pathologies chroniques, on a reproché aux QALYs de manquer de sensibilité, de ne pas envisager les retombées sur la qualité de vie des personnes constituant l'entourage de patients, les « care givers »[12], de ne

8 Griebsch I, Coast J, Brown J. Quality-adjusted life-years lack quality in pediatric care: a critical review of published cost-utility studies in child health. *Pediatrics* 2005 ; 115(5) : 600-14.

9 Mavranezouli I. A review and critique of studies reporting utility values for schizophrenia-related health states. *PharmacoEconomics* 2010 ; 28(12) : 1109-21.

10 Normand C. Measuring Outcomes in Palliative Care: Limitations of QALYs and the Road to PalYs. *Journal of Pain and Symptom Management* 2012 ; 43(3) : e2-e6.

11 Garau M, Shah KK, Mason R *et al.* Using QALYs in cancer: a review of the methodological limitations. *PharmacoEconomics* 2011 ; 29 (8) : 673-85.

12 Haydock A. QALYs: a threat to our quality of life. *J Applied Philosophy* 1992 ; 9(2) : 183-8.

pas valoriser des éléments importants pour les patients au-delà de l'effet purement thérapeutique du traitement, par exemple l'espoir d'une amélioration de leur sort[13].

Dans les pathologies aigues, certains auteurs ont soutenu que les patients pouvaient préférer un état très douloureux mais très court à un état prolongé de douleur modérée même si ce dernier « produisait » un nombre plus élevé de QALYs[14]. De manière générale, le QALY semble mieux approprié aux maladies chroniques qu'aux états morbides aigus[15].

b) Par ailleurs, techniquement, les capacités métrologiques de l'indicateur sont parfois mises en cause. La métrique peut être biaisée par l'hétérogénéité des états de référence (les « *anchor values* »), auxquels sont associées par convention les valeurs 0 (état « mort ») et 1 (état de bonne santé).

Cette notion de « bonne santé » est problématique : s'agit-il d'une notion absolue définie par un niveau maximal de capacité physique, mentale et sociale ou d'un état de bonne santé relatif définie par référence à l'appartenance à un groupe d'âge, de sexe et de milieu culturel bien défini ?

Dans le premier cas la mesure risque d'être irréaliste et manquer de sensibilité ; dans le second elle perd sa prétention à l'universalité.

Une variation significative de l'état de santé pour une personne âgée ou pour un patient en phase terminale peut ne pas être reflétée par le QALY si l'état de référence est un état de bonne santé absolu et inatteignable[16].

Un bon usage des QALYs pourrait nécessiter d'en limiter l'usage à des comparaisons intra-pathologies ou intra-indications après avoir vérifié la concordance des états de référence. Mais de ce fait, le QALY ne serait plus l'outil privilégié d'allocation des ressources à l'échelle du système de santé que vantent ses promoteurs.

c) Enfin, philosophiquement, la critique porte sur le caractère utilitariste de la mesure qui n'envisage le bénéfice d'une intervention qu'à travers la production de QALYs. Une vie vaut-elle la somme des années de vie qui lui reste à vivre ? Ou au contraire chaque vie a-t-elle une égale valeur sans considération de « reste à vivre » ? C'est la critique fondamentale adressée par John Harris, philosophe spécialiste d'éthique médicale, qui, dès 1987, a

13 Devlin NJ, Lorgell PK. QALYs as a measure of value in cancer. *J Cancer Policy* 2017 ; 11 : 19-25.
14 Mehrez A, Gafni A. Quality adjusted life years, utility theory and healthy-years equivalents. *Med Decis Making* 1989 ; 9 : 142-9.
15 Bala MV, Zarkin GA. Are QALYs an appropriate measure for valuing morbidity in acute diseases? *Health Econ* 2000 ; 9 : 177-80.
16 Cf. Garau, *et al. Op. cit.*

accusé les QALYs d'être « *ageists* » c'est-à-dire discriminatoires vis-à-vis des personnes âgées[17].

Moralement, argumente Harris, une vie « *compte pour une* » et non pour le nombre de QALYs potentielles qu'elle contient. La vie d'une personne dont l'espérance de vie en bonne santé est de 50 ans ne vaut pas le double d'une personne qui, en raison de son âge, n'a que 25 ans d'espérance de vie.

C'est pourtant ce que suppose l'usage des QALYs. Le même traitement produira plus de QALYs s'il est administré à une personne de 40 ans qu'à un vieillard de 80 ans. À coût égal, le recours aux QALYs conduira à des allocations de ressources inférieures, pour les personnes âgées, les handicapés, les malades les plus sévères, etc. Il serait ainsi contraire à l'éthique de « *qualifier* » la vie humaine.

Importante au plan philosophique, cette critique n'a peut-être pas la même portée au plan pratique : comme mentionné plus haut, il existe une forte corrélation entre la valeur par vie sauvée (sans considération du « reste à vivre ») et la valeur par année de vie sauvée, notamment dans les pathologies graves[18], de sorte que la perspective adoptée pourrait n'avoir que peu d'effet empirique sur le processus décisionnel. Le manque de sensibilité des QALYs dans les cas les plus graves compenserait en quelque sorte leur tendance discriminatoire.

QALYs et Pondération

Un principe de base de la philosophie des QALYs est explicité par le fameux aphorisme de Williams : « *A QALY is a QALY is a QALY no matter to whom it accrues* »[19].

Ce principe fondamental, présenté comme une impératif éthique, stipule qu'une année de vie supplémentaire en bonne santé doit avoir une valeur égale quelle qu'en soit le récipiendaire, jeune ou vieux, malade ou bien portant (dans le cas d'un traitement préventif), patient atteint d'une pathologie rare ou souffrant d'une maladie commune, riche ou pauvre, etc. La valeur d'un gain ou d'une perte d'un QALY ne *doit* pas dépendre des caractéristiques médicales, sociales ou économiques des individus concernés.

---

17 Harris J. QALYfing the value of life. *J Med Ethics* 1987 ; 13(3) : 117-23.
18 Cf. Tengs, *op. cit.*
19 Weinstein MC. A QALY is a QALY is a QALY – or is it. *J Health Econ* 1988 ; 7 : 289-90. Williams A. Cost-effectiveness, is it ethical? *J Med Ethics* 1992 ; 18 : 7-11.

Ce principe normatif – qui ne nie pas que des individus différents peuvent avoir une perception différente de leur état de santé – fait l'objet de jugements contradictoires, s'agissant de patients qui pourraient mériter, au nom de valeurs collectives, d'être favorisés par rapport à d'autres.

C'est par exemple le cas des maladies rares. Les membres d'un jury citoyen réuni par le NICE en 2004 ont jugé logique et acceptable que la prise en charge d'un traitement par le NHS *ne dépende pas* du nombre de patients concernés. Mais d'un autre côté, ils admettent que la rareté rend « spéciale » une maladie dont les patients méritent un effort de solidarité additionnel au nom de la « double peine » d'être à la fois malades et victimes de cette « injustice » d'être frappé d'une maladie génétique rare (ou d'être atteint jeune d'une maladie mortelle)[20].

Et de fait, en violation du « principe de Williams », les « médicaments orphelins », destinés à des cibles thérapeutiques très étroites, bénéficient de conditions d'accès au marché plus favorables que les médicaments ordinaires. De même le ratio coût efficacité des anti-cancéreux est aux États-Unis le double de ceux des médicaments dans d'autres indications[21].

Le NICE lui-même a adopté un seuil d'acceptation de l'ordre de £ 100 000 par QALY pour le ratio coût-efficacité des médicaments dits « ultra-orphelins » au lieu des £ 30 000 habituels. L'année de vie en bonne santé d'un patient atteint d'une maladie rare aurait ainsi trois fois plus de valeur que celle d'un individu ne souffrant que de pathologies « normales ». Cela peut être choquant au nom de l'égalité des patients mais compréhensible au nom d'un impératif compassionnel[22].

Principe de Williams : une justification théorique ou pratique ?

Le « principe de Williams » repose pourtant sur de solides fondations théoriques.

Pour la théorie économique, un seul instrument ne peut satisfaire deux objectifs différents de politique économique comme l'efficience et l'équité. Le processus d'allocation des ressources constitue un dispositif technique visant à au meilleur rapport coût/avantages. L'objectif de distribution

20 NICE Citizens Council Report – Ultra Orphan Drugs, Londres, Novembre 2004.
21 Bae YH, Mullions CD. Do value thresholds for oncology drugs differ from nononcology drugs? *J Managed Care* 2014 ; 20(11) : 1086-92.
22 Enfin Alan Williams lui-même a tardivement émis des doutes sur la validité de son principe. Cf. Alan Williams, Discovering the QALY or how Rachel Rosser changed my life, Université de York.

dépend d'un jugement de valeur sur l'équité de la répartition des avantages optimisés entre les individus ou les groupes d'individus.

Si, par exemple, la concurrence en économie de marché garantit l'efficience dans l'allocation des ressources, il faut utiliser d'autres instruments, par exemple des mesures fiscales appropriées, pour corriger la répartition des avantages dans le sens d'un objectif d'équité sociale.

C'est exactement la position des tenants du « principe de Williams » : « *Les questions d'équité et de justice sociale ne sont pas incluses quantitativement dans l'approche par les QALYs, au-delà de l'hypothèse que chaque QALY a le même poids pour toutes les individus* » écrivent trois éminents promoteurs de cette approche. Et ils ajoutent : « *Non pas parce que ces questions ne sont pas importantes mais parce qu'elles doivent être appréciées par le décideur en tant que considérations additionnelles à côté du gain collectif en QALYs* »[23].

Le QALY n'a pas vocation à tout résoudre. C'est un instrument visant à l'efficience pas à l'équité des choix en santé et les questions de justice sociale doivent être traités indépendamment de l'approche QALYs. Le QALY d'un patient atteint d'une maladie rare a la même valeur que celui d'un patient « ordinaire », mais la société peut accepter de compenser par ailleurs (au moyen d'allocations forfaitaires par exemple) l'« injustice » ressentie du fait de la rareté de la maladie.

Cette position orthodoxe serait totalement admissible si les systèmes de santé fonctionnaient tout entier selon ces principes. Mais c'est loin d'être le cas et l'équité dans l'accès aux soins (consultations, visites, examens, hospitalisations, etc.) est largement assurée par des prix administrés qui sont également des instruments d'allocation des ressources.

De sorte que l'adhésion au principe de Williams relève finalement d'une logique plus empirique que théorique.

- Il serait pratiquement impossible – même si c'était souhaitable – d'affecter un poids au QALY qui reflèterait l'ensemble des caractéristiques médico-socio-économiques des patients. Qui serait habilité à affecter ces poids et à assumer des choix difficiles entre patients ou groupes de patients ?
- Il serait difficile voire impossible d'agréger les QALYs individuels pour estimer un gain collectif si chaque QALY reflétait une réalité différente, les uns reflétant la gravité de la pathologie, les autres des caractéristiques socio-économiques.

23 Weinstein MC, Torrance G, McGuire A. QALYs: The Basics. *Value in Health* 2009 ; 12 (sup.1) : 5-9.

QALY et agrégation des préférences

L'agrégation des QALYs individuels en une mesure collective constitue un autre point délicat et souvent négligé. Dans les travaux empiriques, le résultat d'un traitement est présenté le plus souvent sous la forme d'un gain total de QALYs sur une population de patients ou, ce qui revient au même, sous la forme d'un gain moyen par patient.

Peut-on légitimement additionner les QALYs de différents individus ? Si les QALYs ne mesuraient que de simples durées de vie la réponse serait positive : deux individus survivant l'un 5 ans et l'autre 10 ans « produisent » 15 années de survie, soit 7,5 ans en moyenne. Mais le QALY incorpore une mesure subjective de la qualité de vie et les subjectivités individuelles ne sont pas aisément additives.

En pratique, l'addition des QALYs se légitime en faisant l'hypothèse (souvent implicite) d'un « *individu représentatif* » dont les préférences individuelles seront prises comme celles de la collectivité. Cet individu représentatif peut être l'individu moyen ou médian de la population analysée. En toute rigueur, il serait nécessaire de s'assurer que la population est suffisamment homogène quant à ses préférences vis-à-vis des états de santé pour que l'individu représentatif le soit réellement.

QALYs et actualisation

La question de l'actualisation des QALYs est l'une des plus débattues dans la littérature sans qu'une solution définitive n'ait émergé à ce jour.

Notons que ce débat est spécifique aux QALYs. L'actualisation est admise sans contestation quand les bénéfices sont monétarisés dans le cadre d'une étude coût-bénéfice. Elle n'a pas de sens dans le cadre d'étude coût-efficacité où le bénéfice est un indicateur clinique, par exemple un taux d'hémoglobine glyquée chez des patients diabétiques...

Par analogie avec l'actualisation des coûts, les guides méthodologiques du NICE comme de la HAS recommandent : a) d'actualiser à la fois les coûts et les résultats (donc les QALYs) dès que l'horizon temporel d'une intervention est supérieur à 12 mois et : b) d'appliquer le même taux d'actualisation aux coûts et aux QALYs[24].

24 HAS : Choix méthodologiques pour l'évaluation économique à la HAS, Octobre 2011 (recommandations 6).

Avec actualisation la formule [1] des QALYs devient :

$$Q(i) = \sum_{t=1}^{n_i} \frac{q_{it}}{(1+r)^t} \qquad [2]$$

avec n_i : l'espérance de vie brute dans l'état i
r : le taux d'actualisation
et $0 \leq q_{it} \leq 1$

Actualiser les bénéfices signifie qu'entre deux traitements aux effets identiques on favorisera celui qui produit ses résultats favorables le plus rapidement (ou ses effets négatifs le plus tardivement). Cela peut paraître raisonnable mais pose de redoutables questions à la fois en faveur et en défaveur de l'actualisation.

Deux arguments sont évoqués à son appui. D'abord la cohérence entre la mesure des résultats et celle des coûts. Citons à nouveau Williams : « *parce qu'il est possible à tout moment de transformer à la marge la santé en argent, et vice versa, et puisque l'argent est (idéalement) alloué dans le temps par référence à un taux social de préférence inter-temporel*, [le taux d'actualisation] *il serait incohérent d'appliquer à la "santé" un taux d'actualisation différent de celui appliqué à la "richesse"* ».

Ensuite, la question de l'équité horizontale. L'actualisation permet de traiter également des patients recevant des traitements de même efficacité et de même coût dans une même indication, mais dont le bénéfice intervient à des moments différents du temps. L'actualisation des coûts et des bénéfices (au même taux) annule en quelque sorte l'effet du temps en ramenant coûts et bénéfices à leur valeur présente.

En revanche, l'argument de l'aversion au risque qui devrait conduire à une dépréciation des bénéfices futurs parce que leur réalisation est nécessairement soumise à une dose d'incertitude n'est pas recevable. Il confond les traitements analytiques du temps et du risque qui sont de nature différente. L'actualisation reflète une *préférence pure* pour le présent indépendante de la question du risque. Un gain futur doit être déprécié même à risque nul.

Actualiser… ou ne pas actualiser ?

En face de ces arguments « pour », il existe de non moins solides arguments « contre ».

D'abord d'ordre méthodologique. Si les coefficients de qualité de vie sont estimés par la méthode du « time trade-off » (TTO), comme le

recommandent le NICE par exemple, ils incorporent déjà une préférence pour le présent à travers le taux d'échange d'une amélioration de la qualité de vie aujourd'hui contre une diminution future de la quantité de vie. L'actualisation de QALYs obtenus dans ces conditions constituerait un double compte.

Par ailleurs, l'actualisation à taux positif, en dépréciant les bénéfices futurs, constitue une incitation contraire à la prévention dont la logique voudrait qu'au contraire les bénéfices futurs soient suffisamment valorisés pour compenser des sacrifices immédiats. À la limite qu'elle fasse l'objet d'un taux d'actualisation négatif exprimant une « préférence pour le futur ». De même, dans les maladies chroniques, l'actuation défavorise le choix d'un traitement de fond à long terme par rapport à un traitement symptomatique à courte échéance.

L'analogie monétaire n'est pas très convaincante. Alors qu'un agent économique peut facilement transférer de l'argent dans le temps grâce aux marchés financiers, cela est impossible s'agissant d'années de vie humaines : pour vivre une année dans le futur il faut en avoir vécu une dans le présent ! Et contrairement à la citation d'Alan Williams, la santé ne peut pas être continument échangée contre de l'argent.

Une question épineuse – et rarement abordée – est de savoir à quel moment intervient le « gain » en QALYs et quelle est donc la date de départ de l'actualisation. Prenons le cas d'un patient qui gagne un an d'espérance de vie grâce à une intervention cardiaque et qui meure dix ans plus tard d'un cancer. Faut-il ajouter le gain à la fin de la vie ? Faut-il l'ajouter au moment de l'intervention à l'origine du gain ? Faut-il l'ajouter à la fin de la période d'observation de l'essai clinique qui a prouvé le gain ?

Le paradoxe de Keeler et Cretin

Finalement l'argument le plus fort en faveur de l'actualisation est purement logique. Il a été donné il y a plus de trente ans par Keeler et Cretin[25] et il s'applique chaque fois que l'on rapporte un coût monétaire à un bénéfice non monétaire.

Il s'exprime ainsi : en l'absence d'actualisation des coûts et des gains en QALYs *au même taux*, il est toujours possible d'améliorer l'efficience d'un programme de santé en le décalant dans le temps. Ce processus étant

25 Keeler EB, Cretin S. Discounting of life-saving and other non-monetary effects. *Manage Sci* 1983; 29: 300-6.

répétable le programme est repoussé à l'infini et la recherche de l'efficience conduit à une absence de décision !

Un exemple simple suffit à le démontrer. Supposons un temps divisé en deux périodes, « maintenant » (période 1) et « plus tard » (période 2) et supposons qu'un programme de santé, développé sur une seule période (par exemple la première), coûte 1000 unités de compte (uc) et rapporte 10 QALYs.

Comparons maintenant ce programme à lui-même en le décalant d'une période. Le ratio coût/utilité du programme initial en période 1 est de 100 uc/QALY comme celui du même programme différé en période 2. Mais puisque les programmes se déroulent sur deux périodes différentes, il faut les actualiser pour que la comparaison ait un sens.

Si on actualise, disons à 10 %, les coûts et les bénéfices du programme futur, son coût actualité devient 909 uc et le bénéfice actualisé 9,09 QALYs[26]. Le ratio coût utilité reste à 10 uc/QALY. C'est l'effet équité horizontale mentionné plus haut.

Si maintenant on n'actualise que les coûts et pas les QALYs, on aura un coût actualisé de 909 uc face à un gain non actualisé de 10 QALYs. Le ratio coût-utilité du programme différé baisse à 90.9 uc/QALY et il devient plus efficient que le programme non différé.

Conclusion : le fait de ne pas actualiser le gain en QALYs *améliore l'efficience du programme différé*. D'où la régression à l'infini qui empêche le programme de voir le jour.

À noter que le paradoxe joue encore quand on actualise les coûts *à un taux positif mais inférieur à celui des coûts*. Si face aux coûts actualisés à 10 %, on actualisait les gains au taux de 5 %, le ratio coût-utilité du programme différé passerait à 95,48 uc/QALY et il resterait plus efficient que le programme non différé.

Ce paradoxe, très discuté dans la littérature, a toutefois davantage une portée logique qu'une importance empirique. Des agences – belges ou hollandaises – ont recommandé des taux différenciés pour les coûts et les bénéfices sans que les programmes de santé de ces pays n'aient été repoussés à l'infini…

Quels taux d'actualisation ?

En théorie, le taux d'actualisation décrit une préférence sociétale pure pour le présent. Le taux d'intérêt servi à long terme par les marchés financiers sur

26 En application de la formule de l'actualisation : S1=S0/(1+r), soit ici 1000/1,10.

des classes d'actifs sans risque (par exemple des obligations d'État bien cotées par les agences de notation) pourrait refléter cette préférence. Mais ces taux varient selon la conjoncture (généralement entre 1 et 5 %) et les gouvernements peuvent avoir leur propre préférence pour le présent, indépendamment des marchés financiers.

La plupart des « *guidelines* » font référence à des « *taux tutélaires* » conventionnels utilisés pour les états pour les choix d'investissements publics.

Ainsi le NICE a longtemps appliqué aux coûts et aux bénéfices le taux unique de 3,5 %, recommandé par le « Green Book » du Trésor Britannique[27]. De même la HAS « *considère que le taux d'actualisation retenu dans l'analyse de référence doit être identique au taux commun à tous les secteurs d'investissement public tel que défini par le Centre d'analyse stratégique* » soit 4 % pour les horizons temporels inférieurs à 30 ans avec une décroissance jusqu'à 2 % au-delà (à 500 ans !)[28]. Cette nouvelle tendance, en France et en Grande-Bretagne, de choisir des taux déclinant dans le très long terme s'explique essentiellement par le souci de ne pas déconsidérer des choix ayant des impacts environnementaux à des échelles de temps « planétaire ». De manière générale, les taux tutélaires pour un long terme de l'ordre de 30 ans varient entre 3 et 5 %. Cela correspond à peu près à la durée maximale des emprunts obligataires.

Un taux ou deux taux ?

La question qui revient fréquemment en débat est celle de l'unicité des taux pour les coûts et les QALYs. Beaucoup d'économistes plaident pour l'application d'un taux d'actualisation plus faible aux effets de santé qu'aux coûts[29], au motif que la préférence collective pour le présent en matière monétaire n'est pas nécessairement la même que pour les états de santé.

Les choix de santé sont supposés être moins « court-termistes » que les choix financiers, ils doivent valoriser les effets des politiques de prévention

27 HM Treasury. *The Green Book, Appraisal and Evaluation in Central Government*, 2003, rev. 2011. Annexe 6.

28 Commissariat Général du Plan : *Révision du taux d'actualisation des investissements publics*. Rapport du groupe d'experts présidé par D. Lebègue, 2005.

29 van Hout BA. Discounting costs and effects : a reconsideration. *Health Economics* 1998 ; 7 : 581-94. Gravelle H, Smith D. Discounting for health effects in cost-benefit and cost-effectiveness analysis. *Health Econ* 2001 ; 10 : 587-99. Brouwer WB, Niessen LW, Postma MJ, Rutten FH. Need for differential discounting of costs and health effects in cost effectiveness analyses. *BMJ* 2005 ; 331 : 446-8.

et, dans certains cas, incorporer les retombées sur les générations futures. L'utilisation d'un taux unique ramène les choix de santé à des choix de consommation « ordinaire », négligeant leur potentiel de création de richesses, humaines et matérielles, dans le futur[30]. Un autre argument se fonde sur l'augmentation de l'espérance de vie qui conduirait les générations actuelles à élargir leur horizon temporel et à prêter plus d'attention aux effets futurs des choix présents.

Si la HAS reste pour le moment sur sa position de taux unique, le NICE britannique a récemment assoupli sa position et accepté la pratique d'un double taux de 3,5 % pour les coûts et de 1,5 % pour les « *health effects* » quand ceux-ci portent leurs effets à très long terme (30 ans et plus). Elle rejoint ainsi les Pays-Bas où les taux respectifs sont de 4 % et de 1,5 % respectivement pour les coûts et les effets de santé[31].

Mais dans ce cas, la question de la valeur du second taux reste posée et les chiffres retenus, variables d'un pays à l'autre, reposent davantage sur des « dires » d'experts et des affichages politiques que sur une analyse scientifique.

Sur ce point comme sur beaucoup d'autres, on conclura comme Weinstein, Torrance et McGuire : *« This continues to be a dilemma that remains to be fully resolved »*[32].

30 Cf. van Hout, *op. cit.*
31 College voor zorgverzekeringen (CVZ). Guidelines for pharmacoeconomic research, updated version. Netherlands : CVZ, 2006.
32 Weinstein MC, Torrance G, McGuire A, *op. cit.*

Annexe 1
Le fondement théorique des QALYs comme fonction de bien-être collectif

Les QALYs sont généralement regardés comme une construction empirique étendant la notion d'espérance de vie pour y inclure une dimension qualitative.

Toutefois cette position ne saurait satisfaire les économistes (ou une majorité d'entre eux) pour qui les concepts analytiques doivent être fondés en théorie, c'est-à-dire rapportés aux comportements d'optimisation sous contrainte des agents qui constituent le cœur de l'analyse micro-économique.

Dès les années 80, des économistes ont cherché à donner aux QALYs un tel fondement en partant d'une situation où les agents expriment des préférences plus globales sur des états de santé définis par des couples qualité-quantité de vie[33].

Ils les ont ainsi dotés d'une fonction d'utilité en situation d'incertitude (dite utilité multi-attribut de von Neuman-Morgensten) du type :

$U = U(e, q)$

Ou « U » représente l'utilité, « e » l'espérance de vie et « q » la qualité de vie (supposée uniforme) sur le reste à vivre. La situation d'incertitude se traduit par le fait que les états de santé sont affectés d'une probabilité dont les agents tiennent compte dans l'expression de leur préférence.

On peut alors montrer, sous des conditions très précises et très techniques (les préférences sur les loteries définies sur les états de santé sont « neumanniennes », il existe une « mutuelle indépendance en utilité » entre les deux arguments, le taux de substitution durée-qualité est constant et l'agent est neutre au risque par rapport à la longévité) que la fonction U = U(e, q) peut s'écrite sous la forme :

$U(e,q) = e.u(q)$

qui rappelle celle des QALYs « ordinaires » où u(q) représenterait le « score de préférence ».

Le QALY empirique peut donc être déduit d'un modèle de comportement des agents en situation d'incertitude. CQFD.

Cette analyse, sur laquelle on ne s'attardera pas, a deux conséquences :

33 Pliskin, JS, Shepard DS, Weinstein MC. Utility Functions for Life Years and Health Status. *Operations Research* 1980 ; 28 : 206-24.

- La première c'est que, sous ces conditions, le QALY n'est pas la multiplication d'une espérance de vie par un « score de préférence » mais il est *lui-même tout entier une utilité* ou un « score de préférence ». Autrement dit entre deux états de santé A et B, on est fondé à dire, après agrégation, que la collectivité préfère A à B (ou est indifférente entre A et B) si *et seulement si* QALYs(A) $\geq$ QALYs(B).
- La seconde c'est que le QALY devient de ce fait une *mesure de bien-être relatif à la santé* et pas seulement un outil d'évaluation des technologies médicales. On peut donc légitiment affecter aux systèmes de santé un objectif de maximisation du nombre total de QALYs à l'échelle de la société. Le calcul coût-utilité devient un outil de construction du bien-être social.

Les axiomes supportant cette promotion des QALYs au rang d'indicateur de bien-être collectif sont cependant très restrictifs et on a pu monter que des agents supposés rationnels n'y obéissaient pas nécessairement[34].

34 Le Pen C. Théorie de l'utilité et mesure des états de santé. *Analyse et Prévision* 1997 ; 129-30 : 37-54.

Annexe 2 : QALYs et DALYs

Les QALYs ont été développés par des économistes de la santé en liaison plus ou moins forte avec la théorie micro-économique dont sont issus beaucoup de concepts (notion d'utilité, de bien-être, agrégation des préférences, allocation des ressources, actualisation, etc.).

Dans une autre perspective et avec une autre référence théorique, les épidémiologistes et les spécialistes de santé publique ont développé un instrument voisin, les DALYs (Disability Adjusted Life Years).

Le concept est similaire. Les DALYs comptent des années de vie en les pondérant d'un score de handicap. Mais le but est différent. Il s'agit de mesurer l'état de santé d'une population et d'évaluer l'écart entre un état de santé parfait (où toute la population atteint l'espérance de vie maximale, libre de toute maladie et de tout handicap) et l'état de santé réel.

Cet écart entre l'idéal et la réalité est mesurée par la somme des années de vie perdues du fait d'une mortalité précoce (years of life lost, YLL) et des années de vie passées avec un handicap (Years Lost due to Disabilty, YLD). Les DALYs sont la somme de ces deux composantes.

Cet indicateur est utilisé par l'OMS notamment pour évaluer et comparer entre pays le « poids ou le fardeau de la maladie » (« *burden of illness* ») c'est-à-dire la perte de ressources humaines liées à la mortalité précoce et au handicap.

Le but n'est donc pas d'évaluer une technologie médicale mais l'état sanitaire d'une population ainsi que la performance globale de son système de santé. Le concept est empirique et démographique et ne fait référence à aucune hypothèse comportementale.

Les DALYs n'échappent cependant pas à certaines règles et difficultés rencontrées avec les QALYs, notamment en matière de pondération pour refléter les préférences sociétales en faveur de tel ou tel groupe social, et l'actualisation pour comparer des états sanitaires à deux moments du temps.

Chapitre 7
La mise en œuvre du critère de l'efficience

Les chapitres précédents ont permis de voir comment utiliser la démarche de l'évaluation économique pour concevoir et mettre en œuvre une étude visant à estimer le ratio différentiel coût-résultat (RDCR) de stratégies thérapeutiques alternatives. Dans la mesure où la valeur numérique du RDCR ne permet pas de conclusion directe, il reste à interpréter les résultats obtenus pour compléter la démarche en se situant dans une perspective d'aide à la décision. Pour cela, il convient d'abord d'identifier les stratégies les plus efficientes en mobilisant le concept de *dominance*, afin de déterminer ensuite la *frontière d'efficience* qui permet de restreindre l'univers des choix aux seules options qui surclassent les autres. Nous verrons enfin que si cette approche fondée sur la logique de l'efficience permet en théorie de résoudre le problème du choix optimal, elle bute en pratique sur des problèmes qu'on peut résoudre par la prise en compte de la disposition à payer du décideur, définissant ainsi une valeur-seuil qui permet aussi de raisonner en termes de bénéfice monétaire net.

La sélection des stratégies efficientes

La recherche de l'efficience consiste à optimiser l'usage des ressources disponibles, par hypothèse rares. Le principe de base repose ainsi sur la notion de non-gaspillage entendu comme processus d'optimisation (des effets de santé) sous contrainte budgétaire. Dans des cas simples, la mise en œuvre de ce principe est assez immédiate et s'exprime comme la recherche de l'option la plus efficace pour un coût donné, ou alternativement la moins coûteuse à efficacité donnée.

Mais en général, la comparaison est plus complexe parce qu'il existe plus de deux possibilités concurrentes et que les alternatives à envisager n'ont pas le même coût ou la même efficacité.

On peut s'appuyer sur un raisonnement graphique pour mener à bien cette analyse comparative, ce qui facilite aussi l'identification des stratégies dites dominées. La représentation graphique de référence passe par la construction du plan efficacité-coût en mettant l'efficacité en abscisse (l'axe des x) et le coût en ordonnée (l'axe des y). Chaque stratégie est identifiée par un point (x,y) selon son efficacité x et son coût associé y. En général, on obtient des effets de plus en plus grands en supportant des coûts de plus en plus élevés, mais on essaie de « payer » les gains d'efficacité le moins cher possible. Graphiquement, la recherche de l'efficience revient donc à « coller » le plus à l'axe des abscisses quand on se déplace vers la droite. Dans la *figure 1*, on voit que le plan efficacité-coût permet de positionner les différentes stratégies considérées comme autant de points. Le problème qui reste à résoudre est celui du choix de la stratégie la meilleure.

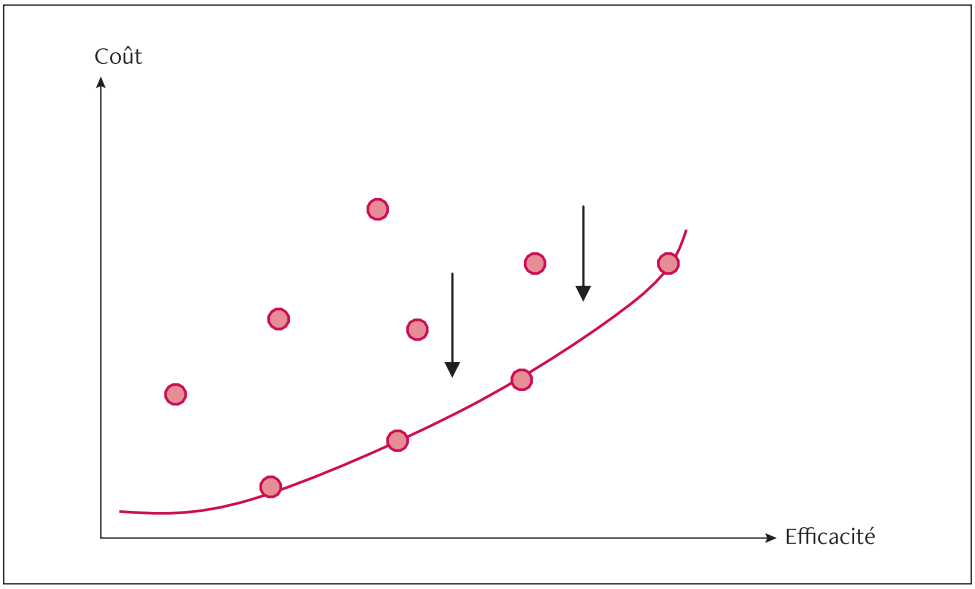


Figure 1. Représentation schématique du plan efficacité-coût.

Le graphique à 4 quadrants et la notion de stratégie dominée

Pour répondre à cette question, on peut tout d'abord considérer le cas simple où il n'existe que deux options, la pratique courante A et une innovation thérapeutique B dont l'efficience est indiquée par son RDCR, c'est-à-dire le rapport ΔC / ΔE du différentiel de coût ΔC (soit $C_B - C_A$) au différentiel d'effet ΔE (soit $E_B - E_A$) quelle que soit la métrique utilisée (en QALYs ou autres). On peut donc utilement représenter graphiquement la comparaison dans le plan (ΔE ; ΔC), le différentiel d'effets étant sur l'axe horizontal et le différentiel de coût sur l'axe vertical (*figure 2*) puisque le RDCR correspond alors à la pente de la droite qui relie deux stratégies[1].

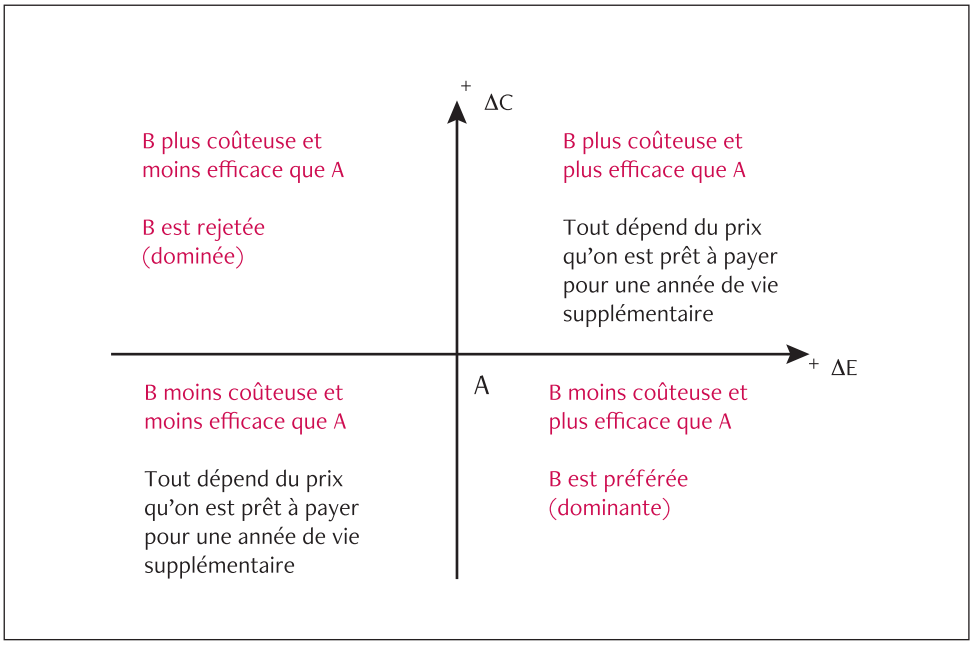


Figure 2. Le graphique à 4 quadrants.

La stratégie A étant prise comme référence, la stratégie B se trouvera à gauche de A sur l'axe horizontal si elle est moins efficace, ou à droite si elle est plus

1 Le plan (ΔE ; ΔC) correspond donc à un changement de repère par rapport au plan efficacité-coût : au lieu de partir de l'origine et de positionner chaque stratégie par son efficacité et son coût, on prend pour origine la stratégie de référence A et on positionne une stratégie alternative B par son différentiel de coût et d'efficacité par rapport à A.

efficace. Et la stratégie B sera au-dessus de A si elle est plus coûteuse ou en-dessous si elle est moins coûteuse. Il en résulte un graphique dit à 4 quadrants selon le signe de chacune de ces deux différences et 4 situations possibles :

- Dans le quadrant Sud-Est, la stratégie B est à la fois plus efficace et moins coûteuse que la stratégie A. cette configuration révèle une situation dite de dominance forte. On peut en conclure directement que l'option B doit être préférée à A et, suivant le type de décision considérée, qu'elle doit par exemple être prise en charge et remboursée alors que la stratégie A, dominée, doit être rejetée.
- Dans le quadrant Nord-Ouest, symétriquement, l'option B est à la fois moins efficace et plus coûteuse que l'option A. La stratégie B est alors fortement dominée et doit être rejetée. Ici aussi, la décision peut être facilement éclairée.
- Dans le quadrant Nord-Est en revanche, la situation est moins simple alors qu'elle est la plus fréquente : c'est celle où l'innovation thérapeutique B est certes plus efficace que la pratique courante A mais aussi plus coûteuse. Il n'y a pas de relation de dominance forte et il n'est pas possible de conclure en l'état. Tout dépend du prix que le décideur est prêt à payer pour obtenir une année de vie ou un QALY supplémentaire (selon la métrique de résultat choisie) en fonction de sa disposition à payer. Dans ce cas, on ne peut éclairer la décision que si l'on dispose d'un critère de décision supplémentaire, comme nous le verrons plus loin.
- Dans le quadrant Sud-Ouest, la situation est caractérisée par le fait que la stratégie B est moins efficace mais aussi moins coûteuse et le même type de question apparaît : quelle est la réduction des coûts qui rend acceptable pour le décideur de perdre des années de vie ou des QALYs en substituant la stratégie B à la stratégie A ? La réponse dépend de sa disposition à recevoir qui peut être égale ou non à sa disposition à payer lorsqu'il s'agit de gagner des années de vie ou des QALYs.

De la dominance forte à la dominance faible

Nous venons de voir que le critère de dominance forte ne permet d'exclure que certaines stratégies qui, dans une comparaison deux à deux, sont moins efficaces et plus coûteuses qu'une alternative[2]. Ceci ne veut pas dire que

2 D'un point de vue instrumental, cela signifie qu'une relation de dominance forte apparaît lorsque le ratio incrémental mesuré est négatif puisque le différentiel de coût au numérateur

toutes les stratégies non fortement dominées (celles qui sont dans le quadrant Nord-Est par rapport à une autre) s'avèrent de bons choix potentiels. Certaines peuvent en effet être faiblement dominées mais cette notion de dominance faible ou étendue ne peut apparaître que s'il existe au moins trois options alternatives[3]. Supposons par exemple que, pour une indication donnée, il existe trois stratégies alternatives A, B et C dont on a estimé les coûts et QALYs associés comme indiqué dans le *tableau 1*.

Tableau 1. L'identification d'une stratégie dominée faiblement

	Coût moyen par patient	**QALYs moyens par patient**
Stratégie A	40 000 €	4
Stratégie B	70 000 €	5
Stratégie C	60 000 €	4,5

Lorsque le choix s'étend sur plusieurs stratégies, il faut raisonner de manière incrémentale, en classant les stratégies par niveau croissant d'efficacité et en mesurant le RDCR de chaque stratégie par rapport à celle qui la précède. Ici, le RDCR de A par rapport à l'abstention de traitement est de 10 000 €/QALY, le RDCR de C par rapport à A est de 40 000 €/QALY et celui de B par rapport à C est de 20 000 €/QALY[4].

Comme on peut le voir sur la *figure 3*, il est possible d'atteindre le niveau d'efficacité moyen de la stratégie C en utilisant une stratégie mixte consistant à traiter certains patients avec la stratégie A et d'autres avec la stratégie B pour un moindre coût. Dans notre exemple, on voit facilement que pour obtenir 4,5 QALYs en moyenne comme avec la stratégie C, il suffit de traiter un patient sur deux avec la stratégie A (4 QALYs) et l'autre avec

est positif et le différentiel d'effet au dénominateur est négatif. Dans la présentation des résultats, on n'exprime jamais cela en présentant le ratio négatif (qui n'a pas de sens) mais en disant que la stratégie est fortement dominée.

3 Le terme de dominance faible est une traduction du terme anglais *weak dominance* emprunté à la théorie des jeux, par opposition à *strong dominance* pour la dominance forte. Mais on peut également parler de dominance généralisée (*extended dominance*) au lieu de dominance faible.

4 Le calcul montre que pour le $RDCR_{A/O}$, $\Delta C = 40\,000$ et $\Delta E_{A/O} = 1$; pour le $RDCR_{C/A}$, $\Delta C = 20\,000$ et $\Delta E = 0{,}5$; et pour le $RDCR_{B/C}$, $\Delta C = 10\,000$ et $\Delta E = 0{,}5$.

la stratégie B (5 QALYs) pour un coût moyen de 55 000 € (110 000/2). En cela, on dit que la stratégie C est faiblement dominée par une stratégie mixte, ici composée pour moitié de chacune des deux autres disponibles. Et même s'il ne s'agit pas, pour des raisons éthiques, de mettre en œuvre cette stratégie mixte qui désavantagerait certains patients, la stratégie C doit être exclue de l'univers des choix du fait qu'elle n'est pas efficiente[5].

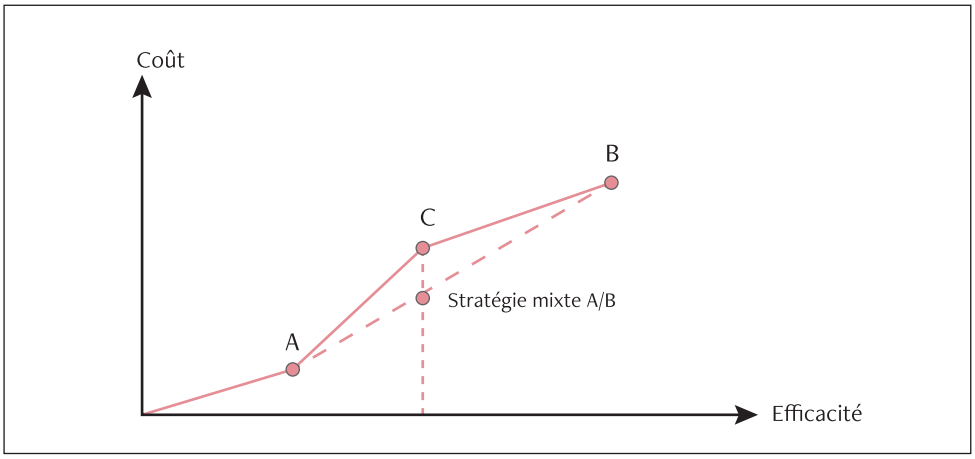


Figure 3. Représentation graphique d'une situation de dominance faible.

La définition de la dominance faible explique la nécessité de considérer simultanément l'ensemble des stratégies pertinentes dans une analyse comparative généralisée plutôt que des comparaisons deux-à-deux qui ne permettraient pas de prendre en compte cette situation. En considérant toutes les options et les critères de dominance pour les discriminer, on peut obtenir la frontière d'efficience qui identifie les meilleurs choix possibles.

La frontière d'efficience comme univers des choix optimaux

Lorsqu'il y a plusieurs stratégies thérapeutiques alternatives, la recherche de l'efficience consiste à éliminer les stratégies fortement et faiblement dominées qui seraient associées à un gaspillage des ressources. Il ne subsiste alors

5 Sur le plan instrumental, il est facile de repérer une situation de dominance faible par le fait qu'après avoir augmenté en considérant des stratégies de plus en plus efficaces, le RDCR (qui correspond à la pente de la droite allant d'une stratégie à l'autre) diminue, comme ici en passant du $RDCR_{C/A}$ au $RDCR_{B/C}$.

que les stratégies non dominées, réputées efficientes entre lesquelles il reste à choisir, comme nous le verrons plus loin. Mais à ce stade, l'identification des stratégies non dominées permet de définir la frontière d'efficience qui est la courbe passant dans le plan efficacité-coût par tous les points représentant ces stratégies, comme dans la *figure 1*. Pour y arriver, il faut généraliser la démarche précédente en commençant par classer les stratégies par niveau de résultat croissant, calculer les RDCR incrémentaux, exclure les stratégies fortement ou faiblement dominées et recalculer les RDCR entre les stratégies non dominées.

Tableau 2. Coût et QALYs de 7 stratégies alternatives

	Coût de traitement par patient	QALYs par patient
Stratégie A	10 000 €	2
Stratégie B	70 000 €	5
Stratégie C	27 000 €	3,2
Stratégie D	20 000 €	3
Stratégie E	40 000 €	4
Stratégie F	60 000 €	4,5
Stratégie G	25 000 €	3,4

Supposons par exemple qu'il existe 7 stratégies alternatives dans le traitement d'une maladie, pour lesquelles on a pu estimer le coût moyen de traitement par patient et les QALYs associés, reportés dans le *tableau 2*.

Si on reclasse ces stratégies par niveau de QALY croissant et qu'on calcule les RDCR successifs (*tableau 3*), on voit que la stratégie C est fortement dominée par la stratégie G qui est à la fois plus efficace et moins coûteuse, de telle sorte que son RDCR est de -10 000. De même, la stratégie F est dominée faiblement par une stratégie mixte composée des options E et B (comme dans l'exemple précédent) et l'on voit que le $RDCR_{B/F}$ diminue alors qu'il a augmenté avec toutes les options précédentes. Après élimination des stratégies C et F, on peut donc savoir d'une part quelles sont les options non dominées qui peuvent faire l'objet d'un choix efficient, et d'autre part calculer les RDCR associés et tracer ainsi la frontière d'efficience passant par ces options selon des segments de droite ayant pour pente la valeur du RDCR correspondant (*figure 4*).

Tableau 3. Calcul des RDCR des 7 stratégies alternatives reclassées

	Coût de traitement par patient	QALYs par patient	RDCR	RDCR recalculés
Stratégie A	10 000 €	2	5 000	5 000
Stratégie D	20 000 €	3	10 000	10 000
Stratégie C	27 000 €	3,2	35 000	//
Stratégie G	25 000 €	3,4	– 10 000	12 500
Stratégie E	40 000 €	4	25 000	25 000
Stratégie F	60 000 €	4,5	40 000	//
Stratégie B	70 000 €	5	20 000	30 000

Dans cet exemple, le choix doit donc être fait entre les stratégies A, D, G, E et B sans qu'il soit possible ici de conclure de façon plus précise. Pour lever l'indétermination de la décision, il faut un critère de choix supplémentaire.

En toute rigueur, la décision devrait être prise en fonction d'un budget prédéfini qui contraint les choix possibles. Dans l'exemple précédent, puisqu'il s'agit de choisir entre des traitements alternatifs, cela supposerait qu'on ait défini un budget pour la pathologie sous-jacente et qu'on connaisse le nombre de patients à traiter. S'il y a 1000 patients, on choisira l'option G si l'on dispose d'un budget de 25 M€, ou l'option E si on a 40 M€. Outre que cette solution suppose qu'on définisse une enveloppe budgétaire stricte et qu'on connaisse la taille de la population à traiter, le problème est souvent plus complexe parce qu'on dispose d'un budget global pour toutes les aires thérapeutiques et qu'il faut faire des choix entre des stratégies indépendantes, portant sur des maladies différentes, et pas seulement des stratégies alternatives, voire parce que l'on ne fixe pas de budget au départ. Les solutions au problème de la décision sont alors moins directes[6].

6 Pour un traitement plus complet de cette question, voir par exemple le rapport de la HAS : *Valeurs de référence pour l'évaluation économique – revue de la littérature*. HAS, décembre 2014.

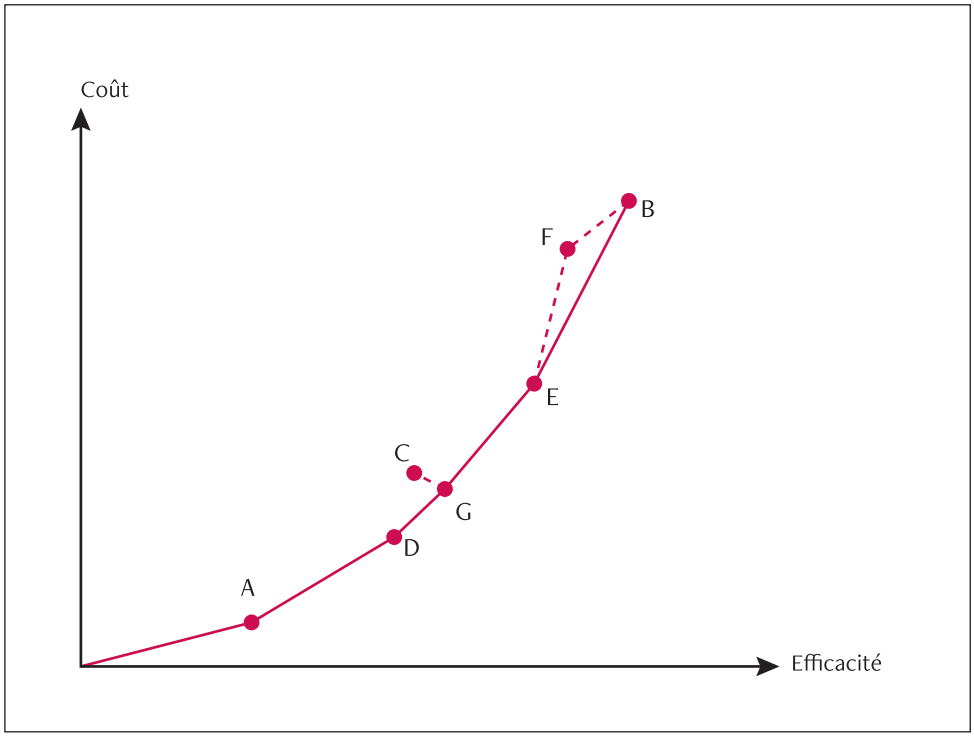


Figure 4. La frontière d'efficience.

Le respect de la contrainte budgétaire : une possibilité théorique

Dans le cas où les choix collectifs conduisent à définir un budget pour les dépenses de santé, ou une partie d'entre elles, il suffit en théorie de recourir aux principes de l'optimisation *sous contrainte* pour éclairer les choix selon le critère de l'efficience. Pour cela, il convient en premier lieu d'établir ce qu'il est convenu d'appeler une *League Table*, c'est-à-dire un tableau de classement des stratégies en fonction de leurs performances respectives mesurées ici par leur RDCR, la meilleure option ayant le coût/QALY le plus faible. Si l'on considère par exemple des stratégies indépendantes car concernant des pathologies différentes, il faut les classer par niveau de RDCR croissant comme dans le *tableau 4*[7].

7 Les données de ce tableau sont librement inspirées de la *League Table* présentée par M. Drummond *et al.* dans *Methods for the economic evaluation of health care programmes*, 3rd edition, Oxford University Press, 2005.

Pour savoir quelles interventions doivent être prises en charge, il faut également connaître le coût de mise en œuvre de chacune d'elles, en fonction du coût moyen par patient et de la taille de la population à traiter, ainsi que l'enveloppe budgétaire disponible, que nous supposerons être un budget global de 6 M€.

Tableau 4. *League Table* et budget global

Stratégie	Coût/ QALY	Coût par patient	Population à traiter	Coût total de la stratégie	Coût cumulé des stratégies
Contrôle du cholestérol et régime alimentaire	220 €	150 €	3 000 000	450 000 000 €	450 000 000 €
Neuro-chirurgie du cerveau	240 €	180 000 €	300	54 000 000 €	504 000 000 €
Conseil du généraliste pour l'arrêt du tabac	250 €	50 €	4 500 000	225 000 000 €	729 000 000 €
Traitement anti-hypertenseur pour la prévention des AVC	940 €	350 €	4 000 000	1 400 000 000 €	2 129 000 000 €
Pose d'une prothèse de hanche	1 180 €	45 000 €	35 000	1 575 000 000 €	3 704 000 000 €
Dépistage du cancer du sein	5 780 €	100 €	5 000 000	500 000 000 €	4 204 000 000 €
Transplantation cardiaque	7 840 €	360 000 €	800	288 000 000 €	4 492 000 000 €
Hémodialyse à l'hôpital	21 970 €	170 000 €	5 000	850 000 000 €	5 342 000 000 €
Traitement de la polyarthrite rhumatoïde par biothérapies	25 000 €	32 500 €	15 400	500 500 000 €	5 842 500 000 €
Traitement par immunomodulateur du myélome multiple	32 500 €	29 000 €	18 000	522 000 000 €	6 364 500 000 €

Le principe simple à utiliser consiste à choisir, c'est-à-dire à prendre en charge, les stratégies par niveau d'efficience décroissant (de RDCR croissant) et, compte tenu du coût total de chacune d'elles qui vient augmenter le coût cumulé, arrêter le choix sur la dernière intervention qui rentre dans le budget disponible. Dans notre exemple, on voit qu'un budget de 6 Md€ permet de financer toutes les stratégies sauf la dernière qui est la moins efficiente et dont le coût total de 522 M€ ne rentre pas dans l'enveloppe disponible du fait que les stratégies plus efficientes mobilisent déjà le budget à hauteur de 5 842,5 M€[8].

Le point le plus important tient à ce que la mise en œuvre stricte de cette démarche d'optimisation sous contrainte budgétaire conduit à une détermination endogène de ce qu'il est convenu d'appeler la valeur-seuil, c'est-à-dire le C/QALY maximum associé à la décision, soit ici 25 000 €. Cette valeur correspond à ce que la collectivité, compte tenu du budget dédié à la santé, accepte de payer au maximum pour obtenir des effets de santé ici mesurés en QALYs. Cette valeur-seuil dépend du budget alloué à la santé d'une part et de l'ensemble des stratégies disponibles d'autre part. Ainsi, l'arrivée d'une innovation thérapeutique dont le C/QALY serait inférieur à 25 000 € devrait être financée par priorité avant le traitement de la polyarthrite rhumatoïde par une biothérapie qui pourrait être évincé du panier de soins remboursés si le coût total de l'innovation est important et accaparerait le budget restant. La valeur-seuil serait alors réduite à celle de l'innovation en question.

Outre que cette démarche suppose la définition d'un enveloppe budgétaire fermée à respecter strictement, sa mise en œuvre suppose, en pratique, une information complète pour toutes les stratégies thérapeutiques sur les paramètres exposés dans le *tableau 4*. Or, l'impossibilité pratique par défaut d'information de disposer d'un tel tableau pour toutes les interventions envisageables interdit de procéder selon ce raisonnement et nous oriente vers la recherche d'une autre approche.

Le recours à une valeur seuil exogène

Que ce soit uniquement pour des raisons pratiques ou parce qu'on ne souhaite pas limiter les programmes de santé pris en charge en fonction d'un budget fermé, la solution pratique consiste à se donner une valeur-seuil de façon exogène, c'est-à-dire un montant maximum de coût par QALY qu'on

8 Nous n'entrerons pas ici dans le détail des problèmes éventuels de l'indivisibilité des programmes qui ne permet pas d'épuiser exactement le budget disponible.

accepte de payer de sorte à prendre en charge toutes les stratégies dont le RDCR est inférieur à cette norme de référence. Cette valeur-seuil peut être déterminée sur la base de différents principes dont nous retiendrons simplement qu'elle exprime finalement la disposition à payer de la collectivité pour certains effets de santé, qu'il s'agisse de QALYs ou d'une autre métrique de résultat.

Il est usuel de citer à titre d'exemple la proposition de l'OMS de définir deux seuils, celui des interventions très efficientes correspondant à une année de revenu par habitant (environ 30 000 € en Europe) et celui des interventions efficientes à trois années de revenu par habitant (soit environ 90 000 € en Europe)[9]. On sait également que le Royaume Uni utilise explicitement une fourchette allant de 20 000 £ à 30 000 £ comme valeur-seuil par QALY. Si elle est définie explicitement, elle permet de lever l'indétermination qui persiste pour les stratégies se situant dans le quadrant NE et SO de la figure à 4 quadrants : elle correspond en effet à la pente d'une droite qui divise ces quadrants en deux et permet de savoir, selon la situation d'une stratégie dans le plan, si elle est réputée efficiente ou non au regard de cette norme (*figure 5*).

Cette notion de valeur-seuil et son utilisation ont soulevé de nombreuses discussions et analyses dont l'essentiel est présenté dans le rapport précité de la HAS. Disons ici simplement que sa mise en œuvre peut être rendue compliquée si l'on décide de ne pas utiliser de valeur unique mais des seuils de référence différents, par exemple tenir compte de considérations éthiques, qu'il s'agisse de médicaments orphelins, de traitements pédiatriques ou pour d'autres raisons. Dans le cas simple d'une valeur-seuil unique, le problème qui demeure tient au fait que le choix d'une prise en charge collective (du coût des stratégies dont le RDCR n'excède pas cette référence) signifie implicitement l'acceptation de disposer d'un budget ouvert et d'une escalade de la dépense. En effet, au fur et à mesure de l'arrivée d'innovations thérapeutiques dont le RDCR sera sous cette norme, on acceptera d'augmenter le panier de soins remboursés sans aucun contrôle direct sur l'évolution des dépenses associées.

9 Il faut préciser que cette norme de l'OMS porte normalement sur la valeur d'un DALY (*Disability-Adjusted Life Year*) et non d'un QALY. Pour une discussion en référence au cas américain, voir Neuman PJ, Cohen JT, Weinstein MC : Updating cost-effectiveness – the curious resilience of the $50,000 per QALY threshold, *N Engl J Med*, 2014 ; 371(9) : 796-7.

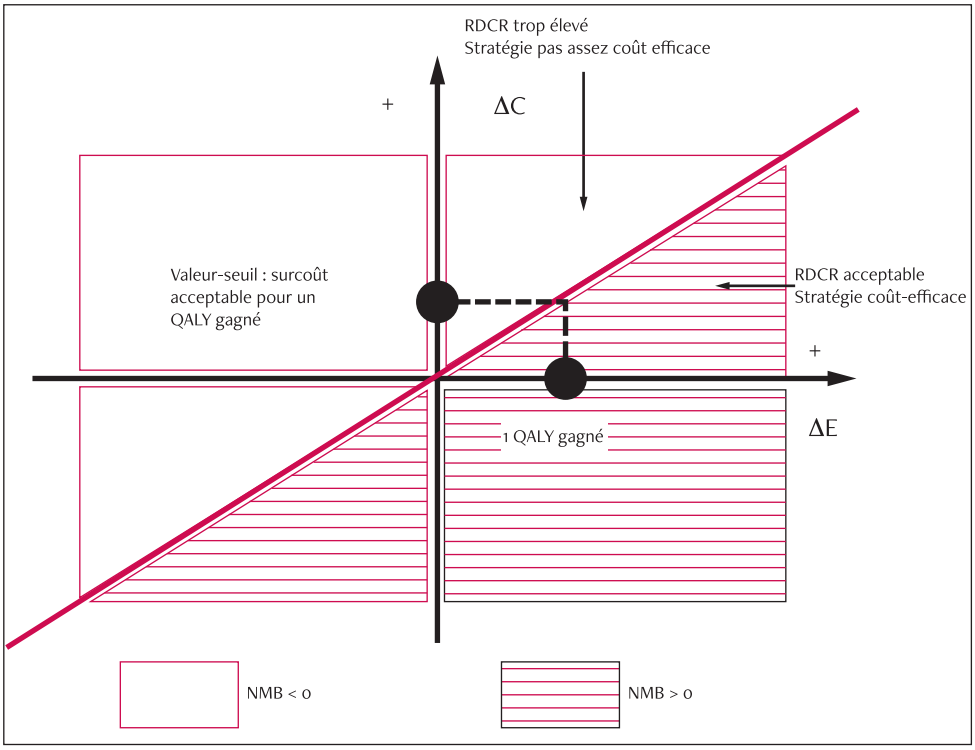


Figure 5. Graphique à 4 quadrants illustrant les notions de valeur seuil et de bénéfice monétaire net ou Net Monetary Benefit (NMB).

La mise en œuvre de cette démarche fondée sur la recherche de l'efficience peut donc être limitée par les possibilités réelles qu'elle autorise, notamment en termes d'information disponible, et par ses conséquences en termes budgétaires si l'efficience est le seul critère de décision. Ceci explique pourquoi les politiques de santé utilisent différents moyens de régulation des dépenses, non seulement au moment de la mise sur le marché d'innovations thérapeutiques mais aussi dans la phase ultérieure. En particulier, cette limite pratique de la démarche justifie le recours à *l'analyse d'impact budgétaire* à titre complémentaire de façon à mieux appréhender l'effet sur le budget d'un acteur du système de santé, en particulier l'assurance maladie, de la prise en charge collective de chaque nouvelle stratégie thérapeutique (voir le chapitre 3 à ce sujet).

De la valeur seuil au Bénéfice Monétaire Net

Si l'on choisit d'expliciter une valeur-seuil, une de ses utilisations intéressantes est de permettre de transcrire le résultat de l'évaluation économique sous un autre format que le RDCR. Si l'on représente la valeur-seuil par λ, on est amené à conclure qu'une stratégie est réputée efficiente si son RDCR es inférieur à cette valeur, soit $\Delta C/ \Delta E < \lambda$. Cette expression peut être réarrangée sous la forme $(\lambda. \Delta E) - \Delta C > 0$, conduisant à définir le bénéfice monétaire net (ou NMB en anglais pour *Net Monetary Benefit*) par $(\lambda. \Delta E) - \Delta C$ dans laquelle $\lambda. \Delta E$ représente la valeur monétaire que la société donne au gain en santé ΔE, c'est-à-dire ce qu'elle est prête à payer au maximum pour l'obtenir et ΔC correspond à ce qui est payé réellement. Donc, si $(\lambda. \Delta E) - \Delta C > 0$, l'intervention coûte moins que ce qu'on est prêt à payer, et le bénéfice monétaire net mesure le surplus correspondant à cette différence.

Outre l'intérêt du NMB par rapport au RDCR sur le plan de la technique statistique, on voit donc qu'il nous fournit une information supplémentaire en quantifiant l'ampleur du « bénéfice » réalisable par rapport à la disposition à payer exprimée par la valeur-seuil. Plus généralement, on peut établir un lien entre la valeur du NMB et le graphique à 4 quadrants puisque le NMB est positif dans le quadrant SE (stratégie dominante) et dans les quadrants NE et SO en-dessous de la valeur-seuil, tandis qu'il est négatif dans le quadrant NO (stratégie dominée) et dans les quadrants NE et SO au-dessus de la valeur-seuil (*figure 5*).

Si aucune valeur de référence n'a été définie, il est possible d'exprimer le montant du NMB en fonction de ce paramètre et de faire la représentation graphique sous la forme d'une droite, le NMB augmentant d'autant plus que la valeur-seuil est élevée.

Enfin, notons qu'on peut aussi définir le bénéfice de santé net (ou *Net Health Benefit*) à partir de l'expression $\Delta E - (\Delta C/ \lambda)$, qui permet de voir si le gain en santé effectif ΔE de la stratégie évaluée est supérieur à celui qui serait obtenu si on utilisait les fonds correspondants à la stratégie prise en charge la moins efficiente, donc si on gagne des effets de santé supplémentaire en réattribuant une partie du budget à la nouvelle stratégie à la place de l'ancienne.

Chapitre 8
Évaluation médico-économique et modélisation

Introduction

L'évaluation médico-économique fait extensivement appel aux techniques de modélisation pour l'aide à la décision.

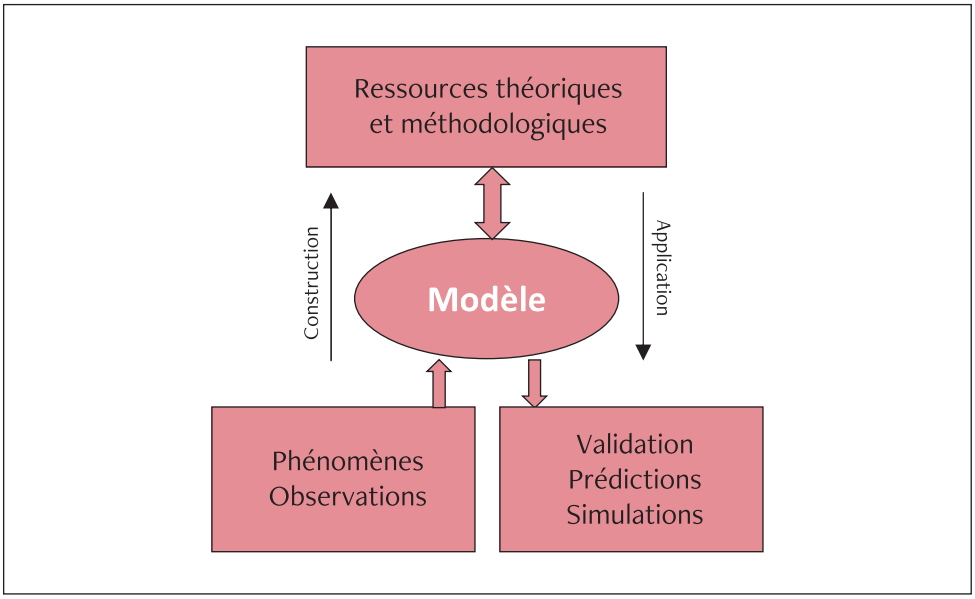
L'intérêt d'un modèle est de rassembler dans un cadre cohérent et structuré des informations de sources diverses, essais cliniques, usage de ressources, grille tarifaire, coefficients de qualité de vie, etc. qui toutes doivent concourir à la décision, mais qui ne sont disponibles qu'indépendamment les unes des autres. Le modèle les remet en situation d'interdépendance selon un schéma interprétatif explicite, adéquat aux objectifs du décideur (remboursement, tarification, agrément, définition de conditions d'usage, etc.).

La modélisation est un exercice complexe qui requiert des compétences économiques, épidémiologiques, statistiques, mathématiques. Ce chapitre n'a pas vocation à couvrir l'intégralité du champ de la modélisation, notamment dans ces aspects les plus techniques qui demandent des savoirs spécialisés dont certains relèvent encore de la recherche. L'objectif est d'expliquer et de justifier le recours à la modélisation et de donner un aperçu des modèles les plus utilisés dans la littérature, de leurs avantages et de leurs limites.

La démarche de la modélisation

Dans une définition large du terme on peut dire que la modélisation est *la* manière dont la science appréhende le monde en mélangeant des données concrètes issues de l'observation et des idées ou spéculations théoriques

issues de la réflexion intellectuelle. La modélisation est ainsi au cœur de la démarche scientifique avec une phase inductive qui conduit de l'observation à la théorie et une phase déductive qui conduit de la théorie à la validation des hypothèses et à la prédiction. Le schéma ci-dessous inspiré de B. Walliser[1] décrit cette démarche.



Dans la phase de construction, le modélisateur élabore une représentation hypothétique des faits observés en mobilisant des ressources théoriques ou méthodologiques préexistantes qui constituent en quelque sorte sa « boite à outils ». Celle-ci peut contenir des hypothèses théoriques qu'on cherchera par exemple à vérifier ou à invalider et des méthodes statistiques, mathématiques ou informatiques qui serviront à étayer la démonstration. Dans la phase descendante, le modélisateur retourne aux faits pour valider sa construction (le modèle explique-t-il bien les faits connus ?), fournir des prédictions (que se passera-t-il si le modèle est « vrai » ?) et procéder à des simulations (que se passera-t-il si telle ou telle hypothèse est modifiée).

1 Walliser B. *Systèmes et Modèles*. Le Seuil, 1977.

Qu'est-ce qu'un modèle ?

De manière générale, on définira un modèle comme une représentation *simplifiée*, *raisonnée* et *quantifiée* de la réalité.

Simplifiée car le modèle a comme fonction de réduire la réalité à ses articulations essentielles, c'est-à-dire celles qui sont utiles à la démonstration pour laquelle il est construit. Le degré optimal de simplification est une question difficile. Trop simple le modèle risque de caricaturer la réalité et de manquer de représentativité. Trop complexe, il perd sa raison d'être et son utilité pour la décision. Le principe de parcimonie méthodologique voudrait qu'il n'y ait pas de complexification sans valeur ajoutée pour le décideur[2]. Ce n'est pas toujours le cas...

Raisonnée car tout modèle repose sur un schéma interprétatif qui lui sert à organiser et structurer les observations. C'est en fonction de ce schéma que seront sélectionnées les variables et les relations entre variables qui entreront dans le modèle. L'adéquation de la structure technique du modèle à son schéma interprétatif constitue un des principes de sa validation interne.

Quantifiée enfin, car un des principaux objectifs de la modélisation est de donner une expression chiffrée à des impressions ou intuitions subjectives, quitte à ce que ces informations soient entachées d'une certaine imprécision, notamment quand il s'agit de mesurer quantitativement des effets qualitatifs. Par exemple s'agissant de la valeur de la vie humaine dans les modèles coût-bénéfice ou des « utilités » attachées à tel ou tel état de santé dans les études coût-utilité.

Un modèle est d'un certain point de vue toujours un compromis. Compromis entre exhaustivité et opérationnalité, entre quantification et précision, entre empirisme et théorie.

Modèles explicatifs – Modèles de simulation

Schématiquement, on peut définir deux types de modèles, les modèles à vocation explicative et les modèles de simulation. Les premiers ont plutôt un but heuristique, les seconds plutôt un objectif d'aide à la décision.

Les premiers cherchent à corréler une variable dite « expliquée » (par exemple la durée moyenne de survie d'une cohorte de patients chez qui un cancer a été diagnostiqué) à des variables dites « explicatives » comme l'âge, la taille et la localisation de la tumeur, l'ancienneté de la maladie, le nombre

2 Principe connu dans la littérature sous le nom du « rasoir d'Ockham » : « *Pluralitas non est ponenda sine necessitate* » (*Le multiple ne doit pas être utilisé sans nécessité*).

de sites métastatiques, la présence d'un biomarqueur, la nature du traitement administré, etc. Des analyses statistiques utilisant des méthodes reconnues comme la régression linéaire, la régression de Cox à risque proportionnel ou les estimateurs de Kaplan-Meyer, permettent de mesurer la force et la significativité statistique de la corrélation. On vérifiera ainsi si la nature du traitement a une influence jugée « significative » sur la survie des patients toutes causes confondues, ce qui pourrait conduire à la délivrance d'une AMM ou à une recommandation de pratique clinique.

Au-delà des modèles statistiques utilisés pour la mesure des corrélations, *les essais cliniques constituent eux-mêmes des « modèles » expérimentaux à vocation explicative*. Dans la mesure où leurs caractéristiques (nombre de patients, critères d'inclusion et d'exclusion, choix du groupe contrôle, randomisation, nature du suivi, etc.) sont déterminées pour vérifier une hypothèse d'efficacité d'un traitement, ils ne peuvent prétendre à être une simple description empirique de la réalité et ils répondent aux trois critères ci-dessus[3].

Les modèles de simulation obéissent à une autre logique. Il ne s'agit pas de démontrer la véracité d'une hypothèse *mais d'en explorer les conséquences*.

À supposer qu'un traitement ait démontré son efficacité dans un essai clinique ayant inclus quelques centaines de patients, que se passera-t-il s'il est administré à toute la population éligible ? Quel en sera le coût annuel pour l'assurance-maladie en fonction de différentes hypothèses de prise en charge ? Quel en sera le bénéfice en termes d'année de vie sauvées ? En termes de QALYs ? Quels gains et coûts supplémentaires peuvent être espérés si l'on suppose que les effets du traitement se prolongent au-delà de la durée de l'essai ? Les modèles de simulation offrent au décideur des scénarios alternatifs d'évolution lui permettant d'apprécier les conséquences des options qui lui sont offertes.

Ils lui fournissent des *prédictions* plus que des *prévisions*. Idéalement ils répondent à la question : « *que peut-on attendre de telle ou telle décision sur telle ou telle variable, toutes choses égales par ailleurs ?* » sans pouvoir autant garantir que les effets attendus seront les effets réels dans la mesure où la clause « *toutes choses égales par ailleurs* » n'est jamais totalement vérifiée dans la vie réelle.

---

3 Voir sur ces aspects le rapport de la DGS « L'Évaluation de l'impact d'un médicament sur la santé des populations et la santé publique » établi par Bernard Bégaud, Catherine Le Galès et Jacques Massol, DGS, Janvier 2002.

Ils se situent en général en aval des modèles explicatifs dans une logique d'extrapolation : extrapolation dans le temps avec la mesure des conséquences à long terme ; extrapolation à d'autres schémas thérapeutiques ; extrapolation à d'autres populations.

L'apport des modèles en évaluation économique

Plus spécifiquement, les questions qui sont le plus souvent abordées par le biais de la modélisation en évaluation économique sont les suivantes :

- L'introduction ou la valorisation de variables de consommation de ressources dans des modèles purement médicaux.
- La mesure des effets à long terme d'un traitement, coûts et bénéfices, au-delà de la durée des essais cliniques requis pour la mise sur le marché.
- La transformation de « end-points » intermédiaires (par exemple des taux de cholestérol ou de glycémie) en « end points » terminaux (taux de décès toutes causes) permettant des comparaisons inter-traitements et le calcul de coût par année de vie sauvée ou de coût par QALY.
- L'extrapolation, via des comparaisons indirectes, des résultats d'un essai à d'autres comparateurs que celui sélectionné par les investigateurs de manière à refléter différentes stratégies thérapeutiques.
- La prise en compte d'options qu'il ne serait pas éthique de tester dans un cadre d'essai clinique (par exemple déterminer l'efficience d'un programme de dépistage systématique d'un cancer versus une absence de programme).

Les modèles peuvent également servir à tester ou à définir des seuils, par exemple des seuils d'efficacité ou des seuils de coût, en-deçà ou au-delà desquels un traitement n'est pas « efficient ». Ils peuvent grâce aux analyses de sensibilité tester la robustesse d'un résultat ou d'une conclusion.

La validation des modèles

La validation est une étape importante dont dépend la crédibilité du modèle et de ses résultats. Il est d'usage de distinguer au moins deux niveaux de validation.

La *validation interne* consiste à s'assurer de la cohérence du modèle, de sa conformité aux principes méthodologiques choisis, de son adéquation au problème posé et de sa capacité à répondre aux questions d'intérêt. Il s'agit d'une *validation a priori*, portant sur la structure du modèle plus que sur ces résultats. La fiabilité de ces derniers dépend alors de la pertinence

des choix méthodologiques, par exemple ceux portant sur l'horizon temporel (long ou court), sur la sélection des alternatives offertes au décideur, sur la définition des variables, sur l'origine des données.

La *validation externe* consiste à confronter les prédictions du modèle à des données connues. C'est une *validation a posteriori* reposant sur l'adéquation des modèles à la réalité. Dans cette perspective empirique un bon modèle est un modèle qui décrit bien la réalité, indépendamment de sa structure théorique ou méthodologique.

Ce type de validation est très utilisé pour les modèles « explicatifs » où la validation consiste essentiellement à mesurer la qualité des ajustements (« *goodness of fit* ») à l'aide de tests statistiques comme les tests du Chi-2, de Kolmogorov-Smirnov, etc.

Il est plus délicat pour les modèles de simulation, sauf à disposer d'un *scénario de référence* que le modèle doit pouvoir retracer avec prévision.

Une variante consiste à retrouver des résultats obtenus par d'autres modèles appliqués à des problèmes similaires (parfois appelée « *validation croisée* »). Cela peut être le cas d'un modèle développé pour évaluer un nouveau médicament dans une indication donnée, qui, appliqué aux données concernant un médicament plus ancien, donnera des résultats proches de ceux des modèles développés spécifiquement pour ce dernier.

Il est de bonne pratique de modélisation d'expliciter les conditions de validation internes ou externes des modèles produits.

Analyses de sensibilité

Les analyses de sensibilité qui visent à mesurer l'influence des hypothèses d'un modèle sur ses résultats font partie intégrante de la démarche de modélisation. Elles permettent en particulier de déterminer les variables « stratégiques » dont dépendent principalement les recommandations faites au décideur. Le chapitre suivant est consacré à cette question.

Une classe de modèles simples : les arbres de décision

Les arbres de décision constituent les modèles les plus simples et les plus intuitifs de modélisation médico-économique. Un examen détaillé s'impose néanmoins car beaucoup des concepts se retrouvent dans des approches plus sophistiquées (modèle de Markov ou modèle d'évènements discrets) qui ont été développées pour dépasser les limites des arbres de décision.

Un arbre de décision décrit la structure logique d'un parcours patient en partant d'une décision de traitement, en envisageant les principaux

évènements aléatoires, mais probabilisables, qui peuvent en découler et en déterminant un résultat final sous la double forme d'un bénéfice et d'un coût attachés à chacune des options du choix initial. À partir de ces éléments on peut déduire les ratios coût-efficacité et coût-efficacité incrémental du traitement d'intérêt.

Un arbre de décision comporte deux types d'éléments : des « nœuds » et des « branches ».

On distingue trois principaux types de nœuds :

- Les « nœuds de décision » (« *decision nodes* », conventionnellement figurés par des carrés, généralement placés à la racine des arbres qui figurent les différentes options offertes au décideur (traiter ou ne pas traiter, utiliser le produit A ou le produit B, etc.).
- Les « nœuds aléatoires » (« *chance nodes* »), conventionnellement figurés par des ronds, qui décrivent les différents événements aléatoires qui suivent la décision initiale et dont la survenue éventuelle détermine le résultat de l'analyse (les patients répondent au traitement ou pas, un évènement indésirable grave survient ou pas, etc.).
- Les « nœuds terminaux » (« *terminal nodes* »), conventionnellement figurés par des triangles, qui marquent la fin du temps logique de l'analyse et auxquels sont collectés les différents résultats cliniques et économiques, coûts des traitements, indicateurs d'efficacité clinique, années de vie sauvées, QALYs, etc.

Les branches assurent la liaison entre les événements. Les branches issues d'un « nœud aléatoire » portent les probabilités de survenue des événements aléatoires décrits par le nœud. Ceux-ci doivent être exhaustifs (c'est-à-dire décrire toutes les éventualités) et mutuellement exclusifs. Les probabilités attachées aux branches issues d'un même « *chance node* » somment à 1. Ces probabilités sont des *probabilités conditionnelles* qui dépendent des événements survenus antérieurement dans la logique de l'arbre.

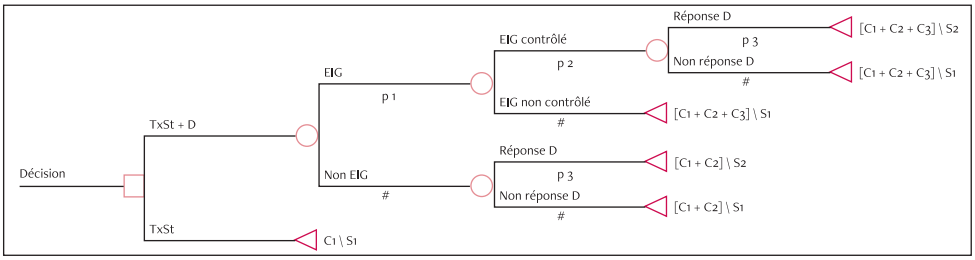
La démarche de modélisation comprend quatre étapes illustrées par l'exemple suivant : i) la construction de l'arbre proprement dit ; ii) la collecte et l'entrée des données ; iii) la résolution du modèle et l'obtention des résultats ; iv) les analyses de sensibilité.

Un exemple d'arbre de décision

Cet exemple hypothétique porte sur l'évaluation médico-économique d'un médicament D administré en complément (*add-on)* d'un traitement standard (TxSt) d'une pathologie grave. Ce médicament augmente significativement

la durée de survie des patients mais entraine chez certains d'entre eux des événements indésirables graves (EIG) qui peuvent conduire à l'arrêt du traitement. La *figure 1* montre la structure de l'arbre.

Structure



Variables

Variable	Libellé	Valeur	Variable	Libellé	Valeur
S1	Moyenne de survie des patients sous traitement standard	12 mois	S2	Moyenne de survie des patients répondants à D	24 mois
P1	Probabilité EIG	10 %	C1	Coût annuel du traitement standard	60 000 uc
P2	% EIG contrôlés. Patients maintenus sous traitement	60 %	C2	Cout additionnel du traitement d'intérêt	40 000 uc
P3	Taux de répondeurs au traitement	80 %	C3	Coût de traitement des EIG	20 000 uc

Résultats

	Traitement standard	**Traitement standard + D**	Δ
Durée moyenne de survie	12 mois	21,2 mois	9,2 mois
Coût total de traitement	60 000 uc	105 600 uc	45 600 uc
Cout par année de vie sauvée	60 000 uc	59 773 uc	59 478 uc*

(*) Coût incrémental par année de vie sauvée

Figure 1. Exemple d'arbre de décision comparant un traitement standard et un traitement combiné (add on D). Voir le texte pour les abréviations.

a) À sa racine on trouve la décision d'ajouter ou non le médicament au traitement standard. Les caractéristiques de ce dernier sont bien connues en termes de survie (variable S1) et de coût (C1 mesuré en unités de compte uc). C'est à ce « temps » que seront rapportées les conséquences du choix en termes de coûts et de bénéfices cliniques.

Dans le cas où le traitement D est ajouté, le premier événement logique est la survenue possible d'un EIG avec une probabilité connue (p1). Si un EIG survient, ce dernier peut ou pas être contrôlé par une médication spécifique dont le coût s'ajoute à celui de D. S'il est contrôlé, le patient est maintenu sous le traitement D (probabilité p2) et bénéficie du surcroit de la survie que ce dernier entraine, pour autant que le patient réponde au traitement (probabilité p3). Dans le cas inverse le produit D est arrêté et le patient ne bénéficie que du traitement standard.

Si aucun EIG ne survient, le patient bénéficie également du traitement D sous réserve de réponse au traitement. La branche issue du nœud « Non EIG » est ainsi le « clone » de celle suivant le double évènement EIG et traitement maintenu. Si le patient ne répond pas à D ou si le traitement D est interrompu, il bénéficiera néanmoins des effets du traitement standard.

b) Cet arbre simple comporte 8 variables. Trois d'entre elles sont des probabilités : probabilité de survenue d'un EIG, probabilité que le traitement soit ou non interrompu (au prix de l'adjonction d'une médication spécifique), probabilité de réponse au traitement (si pas d'EIG ou si traitement maintenu).

Deux sont des variables de résultats cliniques, survie moyenne sous traitement standard et survie avec l'adjonction de D (conditionnellement à

la réponse au traitement). Trois enfin sont des variables de coût de traitement, coût du traitement standard, coût additionnel du traitement d'intérêt D et coût de la médication anti-EIG.

Ces probabilités peuvent être documentées à partir d'essais cliniques, de méta-analyses, de données médico-administratives, etc.

c) La résolution du modèle (« roll back ») se fait en remontant les branches de leurs extrémités à la racine (où est prise la décision). À chaque étape on calcule la valeur espérée du « end point » choisi (survie ou coût).

Dans cet exemple, une fois l'arbre « remonté », les valeurs espérées à la racine de la survie et du coût sont respectivement de 12 mois et de 60 000 uc dans la branche traitement standard et de 21 mois et 102 000 uc dans la branche avec l'adjonction de D, soit un avantage de 9 mois de survie au bénéfice de D compensé par un surcout de 42 000 uc Cela donne un ratio incrémental coût efficacité de 56 000 uc par année de vie sauvée.

d) La dernière étape est l'analyse de sensibilité qui peut être menée sur une variable ou sur des combinaisons de variables, de manière déterministe ou stochastique. Ce point est traité au chapitre suivant.

Avantages et limites des arbres de décision

Cette technique des arbres de décision est particulièrement bien adaptée à des situations médicales relativement simples, avec un nombre limité d'états de santé, dans un horizon de temps relativement court. Un domaine de prédilection est par exemple l'analyse des tests diagnostiques ou des stratégies de dépistage.

Un test biologique se déroule en effet dans un temps court, possède des caractéristiques techniques bien définies (sensibilité, spécificité, valeur prédictive), offre des alternatives souvent binaires et bien identifiées (présence ou non d'un marqueur génétique par exemple) avec des conséquences elles-mêmes bien caractérisées (traiter ou pas par exemple).

Leur limite principale est la gestion du temps. Ils se déploient en effet dans un temps logique qui est mal adapté à des pathologies de long terme où les états de santé évoluent, naturellement et sous l'effet des traitements, dans un temps chronologique, avec des états récurrents, des phases de progression et de régression, comme en connaissent les pathologies chroniques.

Il est évidement possible de complexifier les arbres pour modéliser de telles situations mais ils perdent alors leur principal attrait que constitue leur coté intuitif. D'autres techniques, comme les modèles de Markov ou les

modèles d'évènements discrets permettent de traiter plus efficacement ces situations complexes.

Les Modèles de Markov

Introduits dans la littérature médicale par Beck et Pauker en 1983[4], les modèles de Markov[5] sont largement aussi intuitifs et transparents que les arbres de décision tout en permettant de modéliser des situations cliniques plus complexes et plus évolutives. C'est ce qui explique qu'ils soient devenus une des méthodes de modélisation les plus utilisées, notamment pour les pathologies chroniques.

Un modèle de Markov décrit l'évolution dans le temps de l'état de santé d'un patient type. Les concepts de base sont ceux de temps, d'état de santé et de probabilités de transition qui se définissent conjointement.

a) Le temps est divisé en périodes fixes ou cycles (« temps discret »). Partant d'un état de santé initial à l'instant 1, le patient peut, à chaque cycle, soit rester dans le même état soit opérer une transition vers une autre état en fonction de probabilités prédéfinies.

L'état de santé étant supposé rester stable pendant toute la durée d'un cycle, le choix de la périodicité du modèle dépend de leur évolutivité. En général pour les pathologies chroniques évolutives, la plupart des modèles utilisent une périodicité annuelle, mais des temps plus courts sont possibles (mensuels, trimestriels, semestriels), en fonction, par exemple, des recommandations de pratiques cliniques ou des données cliniques disponibles.

b) Un état de santé est une description clinique de l'état d'un patient à un moment donné qui dépend du problème analysé. Le modèle doit inclure tous les états possibles en relation avec le problème analysé.

À chaque état, décrit médicalement, est assigné deux variables, un coefficient de la qualité de la vie vécue dans l'état (si la mesure d'efficience est le coût par QALY) et le coût des soins reçus pendant une période. De cycle en cycle, les patients accumulent ainsi des années de vie (ou des QALYs) et des coûts en fonction des états qu'ils parcourent sur l'horizon temporel du modèle.

---

4 Beck JR, Pauker SG. The Markov process in medical prognosis. *Med Decis Making* 1983 ; 3 : 419-58.

5 D'après Andreï Markov (1856-1922) mathématicien russe considéré comme le père de l'analyse des processus stochastiques, décrivant l'évolution dans un temps discret ou continu des variables aléatoires. Les modèles de Markov utilisés en médico-économie diffèrent souvent des modèles « purs » en ce qu'ils intègrent souvent des hypothèses non-markoviennes comme des probabilités de transition variables.

c) D'une période à l'autre le patient peut (ou non) effectuer une ***transition vers un autre état***. Le fait de rester dans un même état est considéré comme une transition de l'état vers lui-même. Les transitions interviennent selon des « *probabilités de transition* » définies a priori. Par définition, toutes les probabilités issues d'un nœud somment à 1.

Deux états sont particuliers : l'état initial auquel sont affectés tous les patients et l'état « mort » à partir duquel aucune transition n'est possible. Un tel état est dit « état absorbant » et tous les patients entrés à l'état initial ont vocation à terminer dans un état absorbant si la simulation va à son terme.

d) La résolution du modèle donne la répartition des patients entre les états après un certain nombre de cycles, le nombre de périodes passées dans chaque état ainsi que le coût cumulé par état et globalement.

Un exemple de modèle de Markov

Dans l'exemple très simplifié ci-dessous on utilise un modèle de Markov pour calculer le ratio coût-efficacité incrémental d'un médicament qui ralentit la progression d'un cancer d'un stade local vers un stade avancé et métastasique dans une cohorte de patients d'une soixantaine d'années.

a) Les patients entrent dans le modèle à un stade précancéreux mais on les suppose à haut risque de développer la maladie compte tenu d'un certain nombre de facteurs de susceptibilité (âge, antécédents, marqueurs biologiques, etc.). Ces patients, qui sont asymptomatiques et dont la qualité de vie est proche de la normale (q = 0,90), font l'objet d'un suivi régulier pour un coût annuel (consultations, examens complémentaires, traitements, etc.) relativement faible, fixé dans notre modèle à 10 unités de compte (uc). La mortalité dans cet état asymptomatique est proche de la mortalité générale pour des patients de cet âge (environ 1 %).

b) Chaque année, environ 8 % des patients se voient diagnostiquer une tumeur qui reste localement stable la plupart du temps, en partie sous l'effet d'un traitement plus intensif, avec éventuellement des phases d'hospitalisation, et plus coûteux (50 uc/an). La qualité de vie est dégradée (q = 0,70) et la mortalité est plus forte (10 %).

c) Une fois la tumeur diagnostiquée, environ 40 % de ces patients progressent annuellement vers un stade avancé de la maladie. La qualité de vie est alors très dégradée (q = 0,30) et la mortalité très forte (30 % par an), en dépit d'un traitement intense avec une alternance d'hospitalisations et de séjours à domicile. On estimera à 200 uc le coût annuel de traitement des patients dans cet état.

d) Le traitement d'intérêt a comme effet de ralentir la progression de la maladie une fois celle-ci déclarée. Le taux de patients progressant du stade tumeur stable au stade avancé est ainsi réduit à 10 %. Le traitement n'affecte pas la qualité de vie des patients au stade local mais le coût annuel de traitement passe de 50 à 75 uc.

La *figure 2* représente le modèle avec les 4 états (asymptomatique, tumeur stable, état avancé, mort) qui sont figurés par des cercles dans lesquels sont indiqués le coût annuel de traitement et le coefficient de pondération d'une année de vie pour le calcul des QALYs.

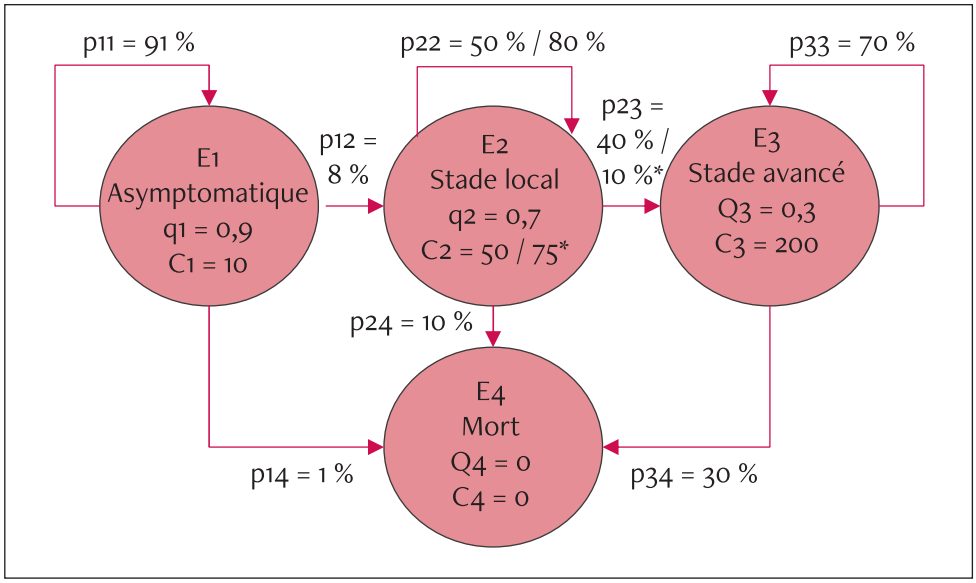


Figure 2. Modèle de Markov avec ses 4 états (asymptomatique, tumeur stable, état avancé, mort). Le modèle ne comporte ni saut (un patient asymptomatique doit passer au moins un cycle dans l'état tumeur stable avant de progresser vers l'état avancé) ni retour en arrière, un patient pouvant tout au plus être stabilisé dans un état donné.Les probabilités de transition sont figurées par des flèches. Les variables marquées d'un astérisque sont modifiées par le traitement d'intérêt (État 2).

La *figure 3* montre la répartition dynamique de la population entre les quatre états après 40 cycles. Au cycle 1 toute la population est dans l'état initial. Au cycle 41, elle est à près de 99 % dans l'état « mort ». Elle transite par les

états « localement stable » et « avancé » avec un décalage temporel (il faut être passé au moins une fois par l'état « stable » pour parvenir à l'état « avancé » et une perte d'effectifs (tous les patients ne progressent pas et certains décèdent dans l'état stable).

Ces profils sont caractéristiques des modèles de Markov avec des états initiaux qui décroissent tendanciellement, des états finaux qui croissent tendanciellement et des états intermédiaires qui alternent phases de croissance et de décroissance.

Figure 3. Répartition de la population entre les états en fonction des cycles (sans traitement).

Le *tableau 1* indique les **résultats coût-efficacité** comparés des deux options. Le traitement, dont le seul effet est de ralentir la progression de la maladie, permet d'accroitre l'espérance de vie des patients de 1,60 QALYs en moyenne mais augmente le coût de 66 uc par patient, ce qui donne un ratio coût-utilité incrémental de 41,55 uc/QALY.

Les résultats pourraient être actualisés en les pondérant à chaque période par un coefficient $\frac{1}{1 + r}$ où r est le taux d'actualisation correspondant à une période.

Comme dans tous les modèles, des analyses de sensibilité mono- ou multi-variables permettent de rester la robustesse des hypothèses et leur influence sur le résultat final. Elles sont présentées au chapitre suivant.

Tableau 1. Ratios coût-efficacité dans les deux options thérapeutiques

	Sans traitement P23 = 40 % - C2 = 50		**Avec traitement P23 = 10 % - C2 = 75**		Δ	
	QALYs	**Coût (uc)**	**QALYs**	**Coût (uc)**	**QALYs**	**Coût (uc)**
E1	9,55	106	9,55	106	0	0
E2	1,24	89	3,11	333	1,87	244
E3	0,71	474	0,44	296	– 0,27	– 178
Total	**11,51**	**669**	**13,10**	**735**	**1,59**	**66**
Ratio C/U	**58,12**		**56,10**		**41,55**	

Avantages et limites des modèles de Markov

Les modèles de Markov sont des outils puissants permettant de décrire des trajectoires de patients relativement complexes tout en restant intuitifs et transparents à l'utilisateur. Ils offrent une gestion d'un temps chronologique avec des états de santé qui varient d'un cycle à l'autre et des variables de coût et d'efficacité qui se cumulent dans le temps. Ils sont particulièrement bien adaptés à la modélisation des pathologies chroniques évolutives au long cours.

Ils sont cependant soumis à un certain nombre de restrictions et limites qui peuvent la plupart du temps être surmontés mais au prix d'une complexification des modèles.

a) Le caractère « sans mémoire » des modèles de Markov est leur principale limite. Les probabilités de transition d'un état à l'autre ne dépendent que de l'état de départ sans considération des états antérieurs et donc sans prise en compte de l'histoire des patients. Or des patients hétérogènes par leur caractéristiques et leurs histoires peuvent se trouver à un moment donné dans un même « état » markovien et les probabilités d'évolution vers un autre état peuvent être déterminées par leur histoire antérieure.

Une manière de résoudre ce problème est de dupliquer les états de santé de manière à regrouper des populations avec un même historique et des probabilités identiques d'évolution. Mais cela complexifie les modèles.

b) De même, les modèles de Markov ne tiennent pas compte du fait que la transition d'un état vers un autre peut dépendre de l'ancienneté des

patients dans l'état de départ. Dans le modèle ci-dessus, la probabilité de transition entre l'état « stable » et l'état « avancé » augmente vraisemblablement avec le temps passé dans le premier. Ce défaut peut toutefois être contourné par la création d'« *états tunnels* » qui constituent une succession d'états temporaires que les patients doivent parcourir successivement d'une période à l'autre, chacun d'entre eux étant affecté d'une probabilité de transition différente vers l'état « avancé ».

c) Une hypothèse souvent contestée est la *fixité des probabilités de transition*. Cette hypothèse est difficilement acceptable pour certaines variables comme le taux de mortalité en population générale, par exemple, qui varie fortement avec l'âge. L'obstacle peut être surmonté en adoptant des « *probabilités dépendantes du temps* », par exemple en indexant le taux de décès sur le nombre de cycles en suivant une loi de mortalité du type loi de Gompertz.

d) Enfin, les modèles de Markov gèrent mal la question, bien connue des biostatisticiens, des « risques compétitifs » affectant les probabilités de transition. Dans notre exemple, les patients qui décèdent dans un état (E1 ou E2) ne sont plus susceptibles de passer dans un autre état (E2 ou E3). Les modèles de Markov gèrent comme des risques indépendants des risques qui sont en réalité interdépendants.

Les modèles d'évènements discrets (DES)

Les modèles d'évènements discrets (« *Discrete Event Simulations* », DES) ont été proposés par certains auteurs pour dépasser les limites des modèles de Markov[6]. Issus de de la recherche opérationnelle et conçus pour gérer les files d'attente, les défaillances d'équipements, la maintenance et plus généralement tous les phénomènes de durée, ils reposent sur le concept « d'évènement » par opposition aux modèles de Markov dont l'élément de base est l'état de santé.

Alors que ces derniers affectent aux patients des probabilités de passage d'un état à l'autre à intervalles fixes (temps discret), les DES affectent des probabilités au *délai de survenue d'un* évènement, par exemple le délai séparant deux stades d'une maladie comme le délai entre l'aggravation d'une maladie et le décès. Ces durées sont modélisées en temps continu à partir de lois exponentielles ou de lois plus générales comme les lois de Weilbull qui, grâce à leur flexibilité, peuvent représenter de nombreuses distributions de probabilités.

6 Caro J, Möller J, Getsios D. Discrete Event Simulation: The Preferred Technique for Health Economic Evaluations? *Value in health*, 2010 ; 13(8) : 1056-60. Voir aussi le rapport de la task-force « Modeling Good Research Practices » consacré au DES. Karnon J, Stah Jl, Brennan A *et al.*, Modeling using Discrete Event Simulation, *Value in Health* 2012 ; 15(6) : 821-7.

Alors que les modèles de Markov sont fondamentalement des *modèles de cohorte* déterminant la valeur espérée des gains et des coûts d'un choix médical, les modèles DES simulent des trajectoires individuelles en utilisant des méthodes de micro-simulation stochastique de type Monte Carlo (cf. chapitre suivant sur les analyses de sensibilité).

Un nombre n de patients parcourent successivement le modèle et, à chaque évènement, capitalisent rétrospectivement les coûts et les utilités accumulés durant la période (variable) qui l'a séparée de l'évènement précédent. À la fin de l'analyse les résultats individuels sont moyennés sur l'ensemble de la cohorte.

Le *tableau 2* résume les différences entre les deux approches.

Tableau 2. Comparaison des modèles de Markov et des modèles d'évènements discrets (DES)

	Modèles de Markov	**DES**
Nature des population	Cohortes	Individus
Concept de base	**États** (évènement : passage d'un état à un autre)	**Évènement** (État : situation des individus entre deux évènements)
Hypothèse de base	Trajectoire « sans mémoire » (l'évolution d'un état à l'autre ne dépend pas des états antérieurs	Processus « avec mémoire ». Le temps de survenue d'un évènement dépend de la trajectoire antérieure
Mode de calcul	Macro-simulation	Micro-simulation
Gestion du temps	Durée fixe de séjour dans un état	Durée variable avant la survenue d'un évènement
Gains et coûts	Attachés à un état prédéfini	Attaché à chaque unité de temps entre deux évènements
Mise en œuvre	Relativement simple	Relativement complexe
Données requises	Peu nombreuses (probabilité de transition, qualité de vie dans chaque état)	En général plus nombreuses (trajectoires des patients)
Logiciels statistiques	Spécifique, peu coûteux	Généraux / Généralement chers

Un exemple de DES

L'exemple précédent de modèle de Markov peut être converti en DES en gardant les mêmes états de santé mais en substituant aux probabilités de transition d'un état à l'autre, la durée moyenne passée entre deux états. Le modèle comporte trois « évènements » principaux,

- La déclaration de la maladie, correspondant au passage de l'état asymptomatique à l'état de maladie localement stable,
- L'aggravation de la maladie, correspondant au passage de l'état stable à l'état avancé, et
- Le décès dans l'état avancé.

Deux de ces évènements (déclaration et aggravation de la maladie) sont en « compétition » avec un autre risque, le risque de décès avant de parvenir à l'évènement, et chaque individu simulé tirera au sort le risque qui survient en premier.

Supposons, à titre illustratif, que la durée moyenne dans l'état asymptomatique avant la déclaration de la maladie est de 5 ans, que la durée au stade local avant progression de la maladie vers un stade avancé est de 3 ans et que la durée moyenne avant décès des patients aggravés est de 1,5 ans.

Si on conserve les mêmes valeurs d'utilité et de coût que précédemment mais en les attachant non plus aux unités de temps passés dans un état mais à chacune des années passées entre deux évènements, la durée moyenne de vie des patients sera de 7 QALYs pour un coût moyen de 800 uc. Si un traitement administré au stade local permet de doubler le temps avant aggravation pour un coût annuel passant de 50 uc à 75 uc, l'espérance de vie des patients passe à 9,15 QALYs pour un coût additionnel de 300 uc soit un ratio coût-utilité incrémental de 140 uc/QALY environ.

Avantages et limites des DES

Les DES possèdent de solides attraits comme le fait de traiter les trajectoires des patients ou leur mode de résolution par micro-simulation permettant d'obtenir des résultats sous forme de distribution de probabilités et non seulement des valeurs moyennes. Ils sont particulièrement bien adaptés aux problèmes médicaux où les gains s'expriment en durée, avec des durées médianes de survie, des survies sans progression, des durés totale de survie.

En revanche, ils ont plus délicats à utiliser et nécessitent un nombre plus élevé de données. Ils ne sont pas facilement représentables sous forme de diagramme et, partant, moins intuitifs et moins transparents à l'utilisateur.

Les « package » logiciels disponibles sont souvent des logiciels de recherche opérationnelle complexes, peu ou pas adaptés aux questions de santé et coûteux[7].

Comme on l'a rappelé plus haut, les limites des modèles markoviens « purs » peuvent être contournées par des pratiques de modélisation adaptées (création d'états spécifiques pour mieux décrire les trajectoires, indexation des probabilités de transition sur le temps passé dans un état, états tunnels, intégration de méthodes de Monte Carlo pour passer d'une approche cohorte à une approche patients, etc.), tout en restant plus simples à mettre en œuvre que les DES

En pratique les comparaisons effectuées par certains auteurs montrent que les différences sont plus conceptuelles et esthétiques qu'opérationnelles : les résultats des analyses sont rarement inversés quand on confronte les deux approches sur le même problème avec les mêmes données[8].

7 On notera toutefois que TreeAge Pro 2017, le logiciel le plus utilisé en évaluation économique pour les arbres de décision et les modèles de Markov inclut désormais un module DES simplifié.

8 Standfield L, Comans T, Scuffham P. Markov modeling and Discrete Event Simulation in health care: a systematic comparison. *Int J Technology Assessment in Health Care* 2014 ; 30(2) : 165-172.

Annexe : Un modèle simple de conversion taux/durée

Les modèles de Markov utilisent des taux de survenue d'un évènement par période (par exemple le décès). Les modèles DES modélisent des durées avant survenue d'un nement.

Il existe une manière de convertir des taux en durées et réciproquement si on suppose que les courbes de survie des patients obéissent à des modèles paramétriques en temps continu dont le plus simple est le modèle exponentiel déclinant (modèle DEALE : « *Declining Exponential Approximation of Life Expectancy* »)[9].

Il repose sur l'hypothèse d'un taux de mortalité constant r qui n'est acceptable que si la mortalité spécifique due à la pathologie est très supérieure à la mortalité générale liée à l'âge. Par exemple dans le cas de cancers avancés chez des sujets relativement jeunes.

Dans ce cas, on peut montrer que le taux de survie des patients (ou d'autres indicateurs de durée comme le taux de survie sans progression de la maladie) est distribué dans le temps selon la loi exponentielle. Si r désigne le taux de mortalité, il s'écrit :

$$S(t) = e^{-rt}$$
$$\textit{avec } S(0) = 1 \text{ et } \lim_{t \to \infty} S(t) = 0$$

L'espérance de vie des patients est alors égale à l'inverse du taux de mortalité :

$$E = \int_0^{\infty} e^{-rt} dt = \frac{1}{r}$$

La variance et l'écart-type de la moyenne sont égaux respectivement à $\frac{1}{r^2}$ et $\frac{1}{r}$ et la médiane de survie à $\frac{1}{r}\ln(2)$.

Si x% des patients sont décédés après T périodes et si donc la proportion de survivants est de (1 – x), le taux de mortalité par période sera égal à : $\frac{1}{T}.\ln\left(\frac{1}{1-x}\right)$

Pour un taux de mortalité spécifique de 8 %, l'espérance de vie des patients est ainsi de 12,5 ans et la médiane de survie de 8,66 ans. Inversement

9 Cf. Beck JR, Kassirer JP, Pauker SG. A convenient approximation of life expectancy (the « DEALE ») – I. Validation of the method. *Am J Med* 1982 Dec ; 73(6) : 883-8, et Beck JR, Pauker SG, Gottlieb JE, *et al.* A: A convenient approximation of life expectancy (the « DEALE ») – II. Use in medical decision-making. *Am J Med* 1982 ; 73(6) : 889-97.

si la médiane de survie est de 5 ans le taux annuel de mortalité spécifique sera de 13,9 % et l'espérance de vie de 7,12 ans.

La *figure 4* montre la relation entre le taux spécifique de mortalité, l'espérance de vie et la médiane de survie.

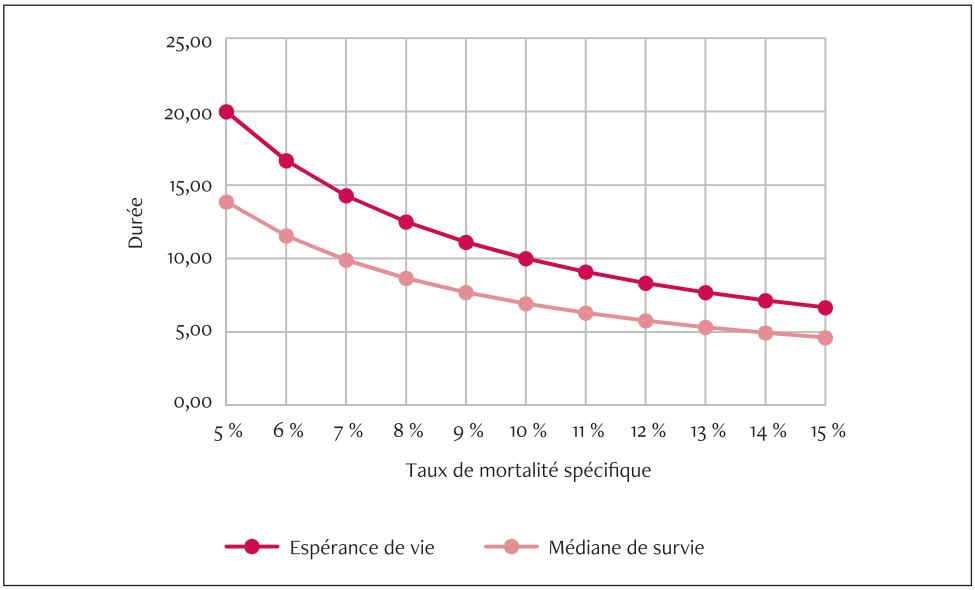


Figure 4. Modèle DEALE – relation entre le taux spécifique de mortalité, l'espérance de vie et la médiane de survie (exprimées dans la même unité de temps).

L'usage du modèle DEALE suppose toutefois la vérification de ses conditions de validité (un important excès de mortalité par rapport à la mortalité générale).

La loi exponentielle constitue la base de la plupart des modèles paramétriques de durée. La loi de Weibull, par exemple, est l'une de ses généralisations les plus utilisées et permet de modéliser toutes sortes d'évolution des taux de mortalité, notamment quand ceux-ci augmentent avec le temps.

La fonction de survie de la loi à deux paramètres s'écrit :

$$S(t; k, \lambda) = e^{-(\lambda.t)k}$$

où k, dit « paramètre de forme » et λ, « paramètre d'échelle » s'estiment à partir des données de la distribution.

L'espérance de vie est égale à :
$\frac{1}{\lambda}\Gamma\left(1+\frac{1}{k}\right)$ la variance à : $\left(\frac{1}{\lambda}\right)^2$.

$$\left[\Gamma\left(1+\frac{2}{k}\right) - \Gamma^2\left(1+\frac{1}{k}\right)\right],$$

ou $\Gamma(.)$ dénote la fonction gamma, et la médiane de survie à : $\frac{1}{\lambda}.\ln(2)^{\frac{1}{k}}$.

On retrouve la distribution exponentielle et son taux de mortalité constant pour $k = 1$[10]. Pour $k > 1$, on aura des taux de mortalité croissants (phénomène de vieillissement) et, pour $k < 1$, des taux de mortalité décroissants (cas de la mortalité infantile par exemple).

10 On a en effet : $\Gamma(2) = 1$ et $\Gamma(3) = 2$.

Chapitre 9
Incertitude et analyses de sensibilité

L'incertitude dans les études médico-économiques

« *Uncertainty in economic evaluation is pervasive, entering the evaluative process at every stage.* »[1]

Dans la mesure où l'évaluation économique prétend fournir une aide à la décision publique elle ne peut échapper à l'interrogation sur la pertinence et la fiabilité de ses recommandations et donc sur le degré d'incertitude qui affecte ses résultats.

Une grande partie de ces interrogations ne sont pas propres à l'évaluation économique et touche également la recherche clinique : la question d'intérêt a-t-elle été bien formulée ? Les comparateurs adéquats bien été identifiés ? Les données utilisées sont-elles fiables ? Les indicateurs de résultats sont-ils pertinents ? Robustes ?

Les essais cliniques de phase III randomisés en double aveugle et leurs techniques statistiques, menés conformément aux règles des agences d'enregistrement, ont été développés précisément pour réduire (sans l'annuler !) la marge d'incertitude concernant les avantages médicaux d'un traitement par rapport à une référence. Une attention particulière a été portée à la construction d'« intervalles de confiance » pour les indicateurs de résultats qui informent le décideur sur le risque (dit de première espèce) qu'il prend en acceptant la conclusion de supériorité (ou de non infériorité) d'un traitement par rapport à un autre[2].

---

1 « *L'incertitude dans l'évaluation économique est envahissante et affecte le processus d'évaluation à chaque étape* ». Briggs A : Handling uncertainty in economic evaluation, *BMJ* 1999 ; 319 : 120.

2 Ou plus exactement le risque de rejeter l'hypothèse nulle d'égalité des traitements alors qu'elle est « vraie ».

L'évaluation économique, si elle fait extensivement appel aux résultats de la recherche clinique, ne peut prétendre aux mêmes processus de validation dans la mesure où ses méthodologies sont rarement expérimentales et font le plus souvent appel à la modélisation.

L'incertitude liée aux modèles est double et il faut distinguer l'incertitude de structure de l'incertitude de paramètre.

Le première est relative au modèle lui-même. Les questions posées sont d'ordre qualitatif. Elles portent sur le choix du type de modèle (statique, dynamique, Markov, DES, etc.), sur l'horizon temporel, sur la définition des états de santé et sur les possibilités de transition d'un état à l'autre, sur le choix des indicateurs de résultats cliniques (QALYS, années de vie, années de vie sans symptômes, etc.), sur la définition des coûts (directs, indirects), sur l'origine des données médicales et économiques.

Ces questions n'appellent pas de réponses techniques et elles relèvent des méthodes traditionnelles de l'évaluation scientifique : revue par les pairs, communication dans les colloques spécialisés, publication dans des revues à comité de lecture, appréciation par des commissions d'évaluation ad-hoc, comme celle de la HAS en France, du NICE ou de l'IQWIG au Royaume-Uni ou en Allemagne. Ces dernières ont d'ailleurs publié des « guidelines » pour réduire cette incertitude de structure et rendre les analyses aussi comparables entre elles que possible.

L'incertitude de paramètre est d'une autre nature. Elle tient à la valeur des données d'entrée dans le modèle, qui, pour une structure donnée, en détermine les résultats. La question est de savoir dans quelle mesure tel ou tel paramètre d'entrée (par exemple un coût d'hospitalisation ou un coefficient de qualité de vie) a – ou non – une influence déterminante sur le ratio coût-résultat produit par le modèle. Il s'agit donc de tester la robustesse des résultats et d'identifier les variables stratégiques pour la décision.

Cette incertitude de paramètre est classiquement traitée par le recours aux analyses de sensibilité. Ce chapitre est principalement consacré à ces analyses. Dans une dernière partie, plus technique, nous aborderons rapidement la question de la construction d'un intervalle de confiance d'un rapport de deux variables aléatoires, les coûts et les efficacités, selon que l'on connait, ou pas, leurs distributions de probabilité.

Qu'est-ce qu'une analyse de sensibilité ?

Le principe d'une analyse de sensibilité consiste à mesurer l'effet de la variation d'une variable d'entrée (input) sur une variable de sortie (output) d'un modèle de structure donnée.

L'objectif est de tester la robustesse d'un résultat aux variations du paramètre choisi. Un résultat sera d'autant plus « robuste » qu'il varie peu en réponse à un changement de valeur du paramètre.

Il existe plusieurs approches des analyses de sensibilité. Briggs par exemple en distingue trois sortes :

a) Les *analyses unidimensionnelles* (« *one way* ») étudient l'effet sur le résultat final de la variation d'un paramètre pris isolément, toutes choses égales par ailleurs. Cela suppose de définir des « bornes de variation » raisonnables et de calculer le résultat pour des valeurs du paramètre comprises entre les bornes. Les extrêmes sont censés représenter les valeurs les plus favorables et les plus défavorables aux objectifs du décideur.

Comme la réponse du résultat n'est pas identique d'un paramètre à l'autre, cette analyse de sensibilité permet également de les hiérarchiser selon leur degré d'influence sur le résultat final et de représenter visuellement cette hiérarchie sous la forme d'un graphique de type « tornado »[3] (voir plus loin).

L'analyse « one way » peut être répétée pour différentes valeurs d'un autre paramètre. On obtient ainsi une analyse « two way » évaluant l'effet des variations conjointes de 2 paramètres, toutes les autres étant maintenus constantes. L'analyse peut être étendue à un troisième paramètre (« *three way* ») mais elle devient vite difficilement lisible. D'autres techniques apparaissent mieux adaptées aux variations conjointes de plusieurs paramètres.

b) *Les analyses de scénarios meilleur et pire* (« *best case-worst case* ») consistent à traiter simultanément tous les paramètres en leur affectant successivement les valeurs extrêmes de leurs bornes de variation de manière à dégager pour les décideurs le scénario le plus pessimiste, où tous les paramètres prennent les valeurs les plus défavorables, et le scénario contraire le plus optimiste.

c) *Les analyses de sensibilité stochastiques ou probabilistes* enfin qui utilisent les techniques de Monte Carlo pour réitérer le calcul des résultats un très grand nombre de fois en faisant varier les paramètres selon des lois statistiques prédéfinies entre les bornes elles-mêmes prédéfinies. Par opposition à cette méthode stochastique, les analyses précédentes sont qualifiées de « déterministes ».

On s'intéressera ici essentiellement aux analyses déterministes de type « one way » et aux analyses stochastiques.

3 Il existe quelques traductions en français de « tornado graph ». Aucune n'est vraiment satisfaisante et nous en resterons à l'expression anglaise qui est bien établie.

Quel type d'analyse pour quels paramètres ?

Schématiquement, un modèle médico-économique fait intervenir trois types de paramètres : des paramètres certains, non susceptibles de varier dans l'horizon temporel de l'étude ou dont les éventuelles variations sont sans rapport avec l'objet de l'analyse ; des paramètres qui conditionnent les résultats mais qui sont contrôlés par les acteurs ; des paramètres aléatoires qui échappent au contrôle des acteurs.

a) Appartiennent à la première catégorie des paramètres démographiques (par exemple les taux de mortalité générale), des paramètres économiques (par exemple des tarifs hospitaliers), des paramètres institutionnels (par exemple des modes de prise en charge, des taux de remboursement), des paramètres médicaux (par exemple l'état des connaissances sur les pathologies).

Les études médico-économiques cherchent à établir l'efficience d'un technique *dans un environnement institutionnel donné et stable* et dans la mesure où ces paramètres décrivent cet environnement, aucune analyse de sensibilité n'est requise.

Dans le long ou très long terme, ces facteurs environnementaux sont susceptibles de varier. Mais les analyses à long ou très long terme ne sont qu'indicatives et, compte tenu du rythme du progrès des sciences et des techniques, le traitement d'intérêt lui-même a peu de chance de survivre à de longues périodes. Par ailleurs, compte tenu de l'impossibilité de prévoir l'état médico-socio-économique des systèmes de santé à 20/30 ans, les analyses à long terme ne sont que des extensions des analyses à terme rapproché sous une hypothèse – peu réaliste mais difficile à éviter – d'un environnement institutionnel, médical, économique à peu près stable.

b) La deuxième catégorie regroupe ce qu'on pourrait appeler les « paramètres de contrôle » dont le plus évident est le prix et/ou le coût du traitement d'intérêt. Dans la mesure où celui-ci résulte d'une stratégie industrielle ou commerciale, il ne peut être considéré comme une variable aléatoire. Mais bien évidemment les résultats en termes de ratio coût-efficacité en dépendent fortement. C'est typiquement ces types de paramètres qui peuvent (doivent) être soumis à des analyses de type « one way » pour envisager l'effet de différentes hypothèses de tarification.

c) La dernière catégorie de paramètres correspond à des *variables aléatoires* issues de sources diverses, essais cliniques, études ad-hoc, avis d'experts, etc. Les valeurs de référence entrées dans le modèle sont par définition entachées d'incertitude et connues à une distribution de probabilité près. Les exemples les plus fréquents sont les variables d'efficacité, les

coefficients de qualité de vie, et plus généralement les paramètres qui résultent d'une étude en amont du modèle. L'analyse de sensibilité concernant ces variables peut être du type « one way » entre des valeurs extrêmes, mais elles sont candidates aux analyses stochastiques qui utilisent les méthodes de Monte Carlo pour en déduire leur distribution empirique.

Les analyses de type « one way »

Le processus d'une analyse de sensibilité « one way » suppose :

- de choisir les variables à analyser ;
- de fixer les valeurs extrêmes raisonnables que peuvent prendre ces variables ;
- de recalculer pour chacune d'entre elles le résultat final en termes de ratio coût-efficacité.

La *figure 1* montre une analyse de sensibilité « one way » effectuée à partir du modèle d'arbre de décision présenté au chapitre précédent.
Ce modèle, qui portait sur l'évaluation d'un traitement D ajouté en « add-on » à un traitement existant, comportait 8 variables dont une « variable de contrôle », la coût du traitement (C2), dont la valeur de référence était de 40 000 uc. Le coût par année de vie sauvée était alors de 57 828 uc. la *figure 1* montre l'effet d'une variation du coût du traitement entre 0 et 100 Kuc sur le coût par année de vie sauvée qui varie alors de manière linéaire entre 0 et 135 000 uc. Un objectif de 50 000 uc/année de vie sauvée pourrait être atteint avec coût du traitement D de 35.000 uc toutes choses égales par ailleurs.

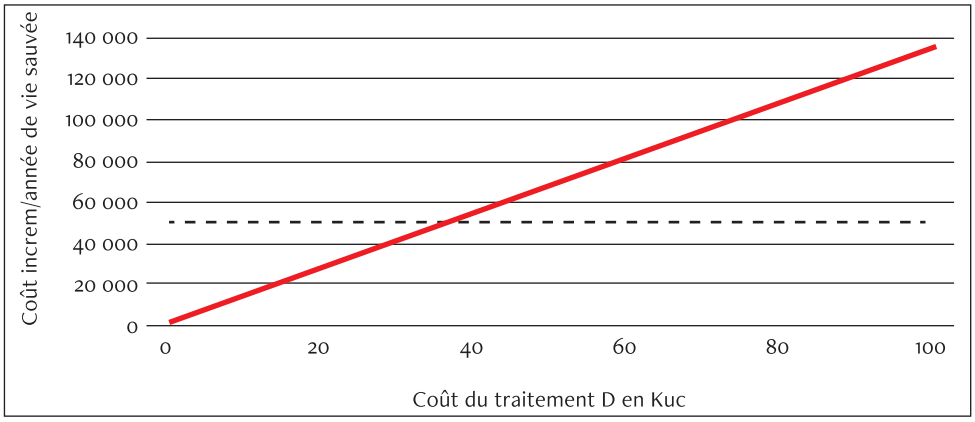


Figure 1. Analyse de sensibilité « one way » sur la variable coût du traitement.
Modèle d'arbre de décision (voir chapitre 8).

La *figure 2* montre deux analyses « two way » correspondant à ce même modèle d'arbre de décision et au modèle de Markov du chapitre 8.

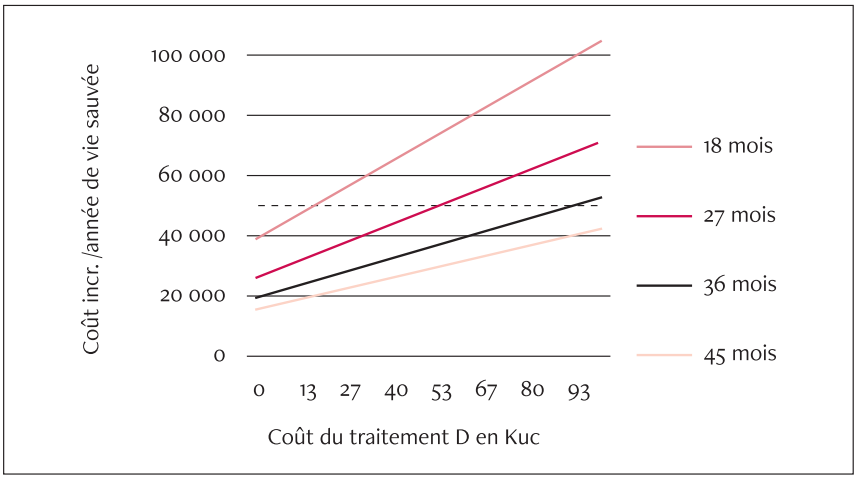


Figure 2a. Analyse de sensibilité « two way » (variables S2 et C2).
Modèles d'arbre de décision (chapitre 8).

		Coût du traitement état "stable" (en uc)				
		50	62,5	75	87,5	100
Efficacité (p23)	0,0 %	T	T	T	T	NT
	5,0 %	T	T	T	NT	NT
	10,0 %	T	T	T	NT	NT
	15,0 %	T	T	NT	NT	NT
	20,0 %	T	T	NT	NT	NT

T : supériorité de l'option traitement - NT supériorité de l'option non traitement

Figure 2b. Analyse de sensibilité « two way » (variables p23 et C2).
Modèles de Markov (chapitre 8).

Dans le premier cas (2a), on mesure l'effet d'une variation conjointe de la durée de survie des patients répondeurs (S2) et du coût du traitement D (C2). L'analyse « one way » sur le coût de traitement est en fait répétée pour différentes valeurs de la durée de survie, donnant ainsi une analyse à 3 dimensions.

L'objectif de 50 000 uc/année de vie sauvée peut être atteint avec différentes combinaisons cout/efficacité : 17 000 uc pour une survie de 18 mois, 53 000 pour une survie de 27 mois, etc. jusqu'à plus de 100 000 uc pour une survie de 45 mois.

Dans le deuxième cas, on envisage une variation conjointe de l'efficacité du traitement (variable p23 qui dénote la probabilité annuelle de passer du stade local au stade avancé et de son coût (variable C2). Le tableau montre les combinaisons p23/C2 pour lesquelles le ratio cout-utilité du traitement d'intérêt est supérieur à celui du non traitement (cases notées T) et les combinaisons aboutissent à un résultat inverse (cas NT).

Les graphiques « tornado »

Les graphiques « tornado » constituent une représentation graphique usuelle des analyses « one way » effectuées sur plusieurs variables en permettant de visualiser les influences comparées de chacune d'entre elles.

La *figure 3*, relative au modèle d'arbre de décision présenté au chapitre précédent, montre l'effet sur le ratio coût-efficacité incrémental des variations « one way » de quatre paramètres relatifs au traitement d'intérêt, le taux d'effets indésirables graves (p1), le taux de réponse au traitement (p3), le coût du traitement (C2) et le bénéfice en termes de survie pour les patients répondants (S2).

Le graphique est centré sur le ratio coût efficacité correspondant aux valeurs de référence (ici 56 000 uc/année de vie) et chaque barre montre les variations de ce ratio en fonction des variations ordonnées de chaque paramètre. Le ratio C/E passe ainsi de 36 000 uc/année de vie à 109 000 quand la survie moyenne des patients répondants décline de 30 à 18 mois. Cette représentation permet de hiérarchiser les facteurs d'influence.

Ces analyses « one way », « two way », etc., bien que très utilisées et faciles à mettre en œuvre, ont cependant deux limites :

- Les bornes de l'intervalle de confiance sont arbitraires.
- Il n'existe pas de définition objective de ce qu'est un résultat « robuste » et de ses marges acceptables de variations.

En toute logique il devrait revenir au décideur de fixer les bornes des intervalles, notamment pour les « variables de commande » et d'édicter ses « règles d'acceptabilité » des résultats.

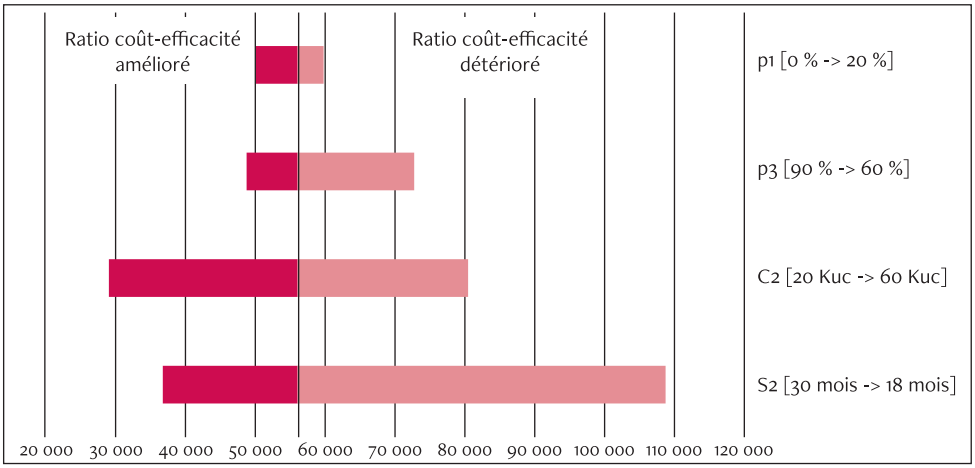


Figure 3. Graphe « Tornado » sur des variables caractéristiques du traitement.
Modèle d'arbre de décision (chapitre 8).
Nota : Le sens de variation des paramètres est important. En augmentant, certains dégradent le rapport cout-efficacité (p1, taux d'effets indésirables ou C2, coût du traitement) tandis que d'autres l'améliorent (p3, le taux de réponse au traitement ou S2 la survie des patients répondants).

Les analyses de Monte Carlo et l'analyse stochastique

Les analyses de sensibilité stochastiques consistent à utiliser les méthodes de Monte Carlo ou de micro-simulation pour engendrer une distribution empirique de la variable de résultat, permettant de construire des tests statistiques de significativité.

Ces méthodes sont très utiles quand on ne connait pas les lois de distribution des variables de résultats, notamment dans les « grands » modèles avec de nombreux états, et quand les approximations par des lois normales ne semblent pas réalistes.

Elles sont par ailleurs les seules à pouvoir être utilisées dans le cas des DES où il n'existe pas de valeurs espérées calculables comme dans les modèles markoviens.

Si on note E la valeur espérée de l'efficacité, mesurée par exemple en QALYs, obtenue par un arbre de décision ou par un modèle de Markov, l'idée est d'en produire un estimateur sans biais qui sera la somme des ratios empiriques obtenus en réitérant un grand nombre de fois l'analyse (en parcourant l'arbre ou le modèle de Markov). Chaque itération sera légèrement

différente de la précédente car, à chaque nœud de chance, un tirage aléatoire orientera différemment le modèle donnant un résultat différent. L'estimateur Ê converge en probabilité vers E et plus le nombre d'itérations est grand plus la valeur empirique est proche de la valeur théorique.

Il est utile de penser l'analyse de Monte Carlo sous la forme d'une cohorte d'individus fictifs qui parcourent l'arbre et qui, à chaque nœud de chance, tirent au sort le chemin qu'ils suivront. Si par exemple un nœud de chance offre deux possibilités A et B affectées respectivement des probabilités 0,20 et 0,80, l'individu tire au sort une valeur uniformément distribuée sur l'intervalle [0,1] et prend le chemin A si le chiffre tiré au sort est inférieur à 0,20 et le chemin B dans le cas inverse. Au terme des simulations on dispose d'une distribution empirique des individus permettant des calculs d'inférence statistique, comme pour une « vraie » cohorte. D'où la dénomination de « micro-simulation » souvent donnée aux analyses de Monte Carlo.

Le *tableau 1* montre ainsi les statistiques relatives aux coûts et aux QALYs associés aux stratégies « traitement » et « non traitement » du modèle de Markov présenté au chapitre 8. L'analyse de Monte Carlo permet d'obtenir des estimations de la moyenne, de l'écart-type de certains percentiles de la distribution des résultats. À partir de la moyenne et de l'écart-type on peut utiliser les formules classiques de détermination des bornes d'un intervalle de confiance à 95 % si on accepte une approximation normale des distributions[4].

Dans ce cas, représenté *figure 4* et *tableau 2*, les différences seraient significatives pour le gain QALYs (les intervalles de confiance sont disjoints) mais pas pour les coûts, ce qui conduirait à une conclusion de dominance pour la stratégie « traitement ». Mais l'hypothèse de normalité est assez fortement rejetée dans ce modèle et la construction d'intervalles de confiance par la méthode des percentiles (2,5 % et 97,5 % pour un seuil de 5 %) conduit à une conclusion de non significativité aussi bien pour les différences de survie que pour les coûts.

4 Soit $[\overline{X} - 2.\frac{\sigma(x)}{\sqrt{n}}\ ;\ (\overline{X}) + 2.\frac{\sigma(x)}{\sqrt{n}}]$ où $\overline{X}$ et σ (x) sont la moyenne empirique et l'écart-type d'un échantillon de taille n (2 est la valeur quantile d'ordre $1 - \alpha/2$ de la loi de Student à $n - 1$ degrés de liberté quand $\alpha = 5$ %).

Tableau 1. Coût et efficacité de chacune des stratégies évaluées par micro-simulation (n = 10 000)
Modèle de Markov (Chapitre 8)

	Non Traitement		**Traitement**	
	Coût	**QALYs**	**Coût**	**QALYs**
Moyenne*	**667 (669)**	**11,55 (11,51)**	**742 (753)**	**13,18 (13,10)**
Écart-type	585	9,69	693	10,15
Médiane	525	8,65	575	10,35
Minimum	5	0,45	5	0,45
Maximum	5 875	83,75	6 225	94,05
2,5 %	25	1,35	25	1,35
97,5 %	2 175	38,05	2 390	40,35
Int. de conf. (5 %)**	[630-704]	[11,30-11,75]	[698-786]	[12,90-13,40]

* Entre parenthèses les valeurs espérées calculées par le modèle. ** Approximation normale.

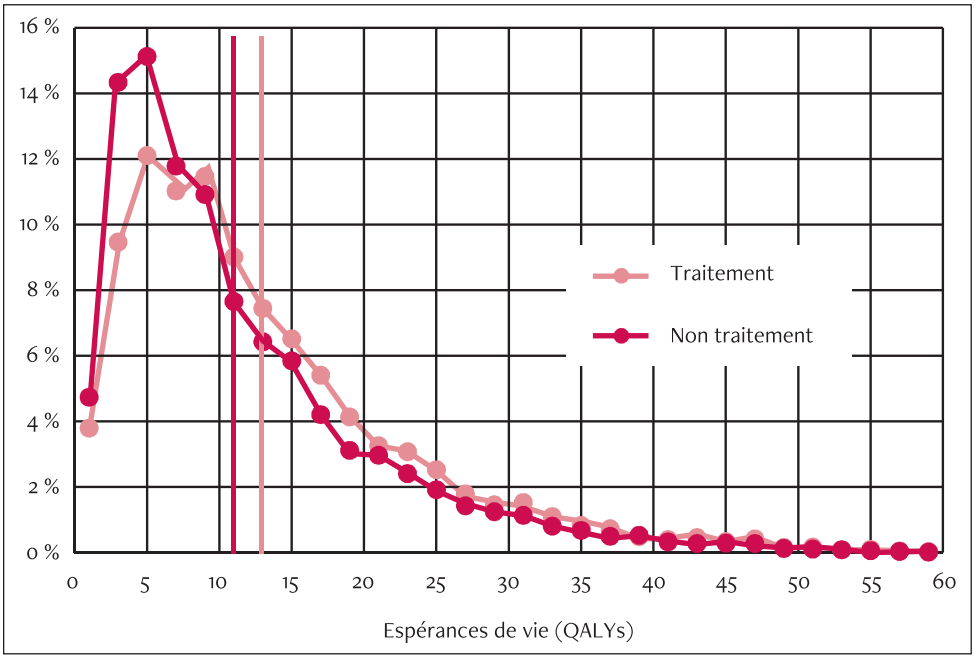


Figure 4. Distributions empiriques des QALYs avec ou sans traitement par simulation de Monte Carlo.
Modèle de Markov (chapitre 8).

La courbe d'acceptabilité

Une manière courante de présenter les résultats d'une analyse de Monte Carlo est la courbe d'acceptabilité du ratio coût efficacité (« *Cost-Effectiveness Acceptability Curve* », CEAC).

Partant d'une distribution empirique des efficacités incrémentales et de coûts incrémentaux, la question est de savoir quelle est la probabilité qu'une stratégie soit jugée « acceptable » pour différentes valeurs de la « *disposition à payer* » (« *willingness-to-pay* ») soit différentes valeurs seuil du ratio coût efficacité incrémental (ou RDCR).

La figure suivante représente la distribution des couples bénéfice incrémental (en QALYs) et coût incrémental (en uc) associés à une stratégie

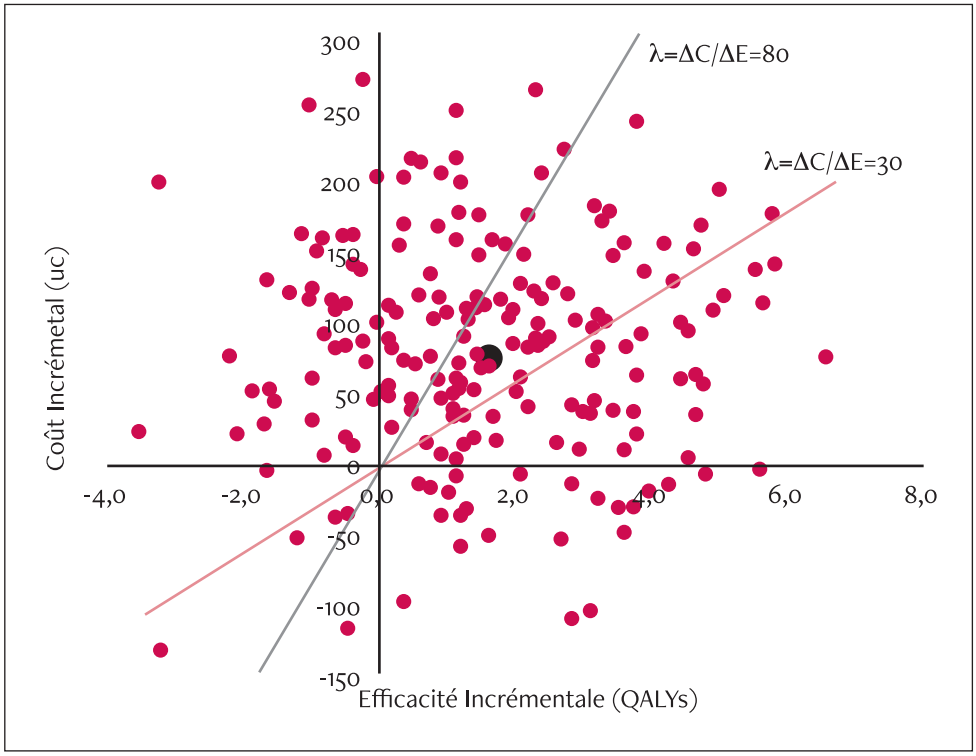


Figure 5. Résultat d'une simulation stochastique (Monte Carlo) dans le plan coût et efficacité. Chaque point représente un « individu » de la cohorte fictive. Le ratio seuil coût efficacité est mesuré par la pente d'une droite passant par l'origine.

thérapeutique comme celle du chapitre 8. Ces données sont issues d'une analyse de sensibilité stochastique. Le point noir donne l'estimation ponctuelle (« point estimate ») des efficacité et cout incrémentaux ($\Delta C = 75$ et $\Delta E = 1{,}63$) et les autres points représentent la distribution des individus de la cohorte fictive de l'analyse de Monte Carlo.

Dans 19 % des cas, les couples efficacité-coût se situent dans quadrant Nord-Ouest des stratégies dominées (différentiel d'efficacité négatif et différentiel de coût positif). C'est l'inverse dans le cadrant Sud-Est (12 %) des stratégies dominantes (différentiel d'efficacité et différentiel de coût négatif. Dans ces deux cas, le seuil d'acceptation du ratio-coût efficacité ou coût-utilité n'a pas d'influence sur la décision d'acceptation ou de rejet.

Le quadrant Sud-Ouest concerne le cas particulier où les différentiels des coûts et des résultats sont tous négatifs (3 %). De manière conservatrice on assimilera ce cas à celui des stratégies dominées : on n'acceptera pas un traitement médicament moins efficace au seul prétexte qu'il est moins cher.

Reste le cas général (67 %) du quadrant Nord-Est ou les différentiels d'efficacité et de coûts sont tous deux positifs. Dans ce cas, le RDCR est représenté par la pente d'une droite passant par l'origine ($\Delta C = \lambda.\Delta E$). Plus la valeur de cette pente est élevée plus le ratio acceptable est lui-même élevé. Dés lors, tous les points se situant sous cette droite ont un ratio coût-efficaité inférieur à la valeur seuil et sont donc jugés « acceptables ». Plus la pente est élevée, plus le nombre de points « acceptables » est élevé. Sur le graphique 5, on observe que le nombre de points acceptables augmente quand la valeur seuil passe de 30 uc/QALYs à 80 uc/QALYs.

La courbe d'acceptabilité se déduit du graphique précédent en faisant varier la pente de la droite représentant la valeur-seuil et en comptant la proportion de points « *acceptables* » pour chaque valeur. Cette proportion augmente logiquement quand le seuil d'acceptabilité s'élève. Elle devient égale à 100 % pour des valeurs extrêmes du seuil (*figure 6*).

Un point délicat : l'intervalle de confiance d'un rapport coût-efficacité

Le calcul d'intervalles de confiance pour des ratios coût-efficacité (moyens ou incrémentaux) est une question difficile pour plusieurs raisons dont la plus importante est qu'un ratio coût-efficacité est un rapport de variables aléatoires, exprimées dans des unités différentes,et que la loi de distribution d'un tel rapport n'a généralement rien à voir avec la loi de distribution du numérateur (coût) et du dénominateur (efficacité). À supposer même que le coût et l'efficacité soient distribués normalement, leur rapport ne l'est pas.

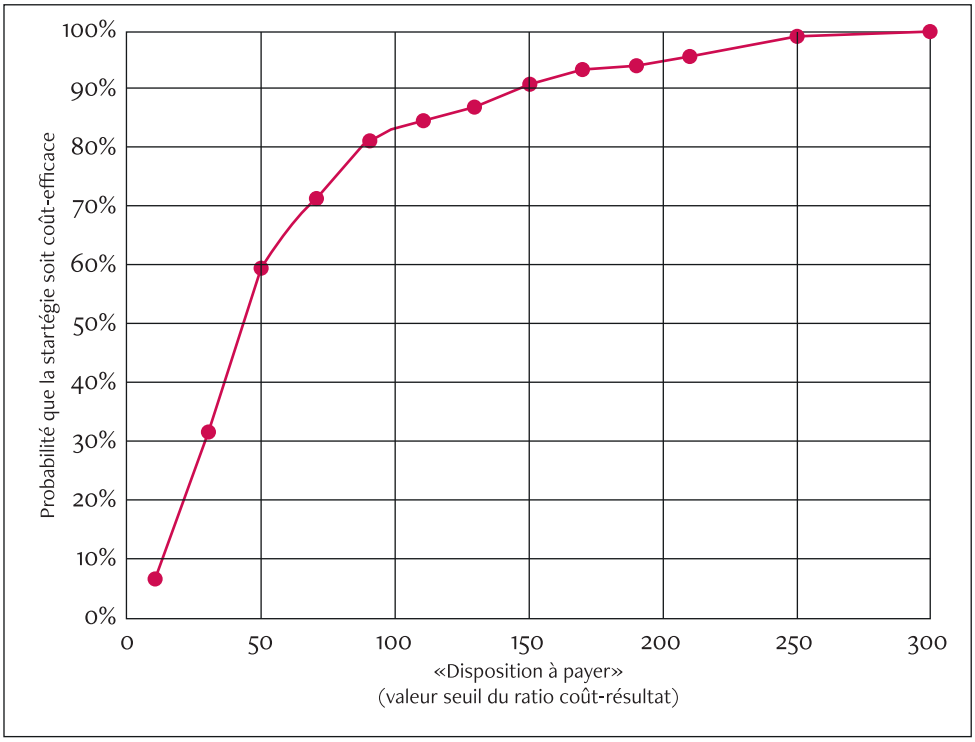


Figure 6. Courbe d'acceptabilité issue de l'analyse stochastique (*figure 5*).
Pour une valeur de 100 uc/QALYs, on note que la stratégie analysée a 85 % de chances d'être coût-efficace. Cette probabilité ne concerne que les 67 % de cas du cadran Nord-Est du plan coût-efficacité. En ajoutant les 12 % de stratégies dominantes, qui sont acceptables quel que soit le ratio seuil, la probabilité d'acceptation sous l'hypothèse d'une valeur seuil de 100 uc/QALYs est de 70 % environ. Elle passerait à 76 % si le ratio seuil passait à 200 uc/QALYs.

Par ailleurs, si l'intervalle de confiance inclut la possibilité d'une efficacité (ou d'un delta d'efficacité) nulle, *le ratio coût-efficacité devient infini à l'approche de l'efficacité 0 et aucun intervalle de confiance ne peut être calculé*.

On considérera dans les exemples qui suivent qu'on se situe dans le 1er quadrant du plan Coût-Efficacité c'est-à-dire dans le cas très général d'un traitement à la fois (strictement) plus efficace et plus cher que la référence qui justifie le recours à l'analyse médico-économique.

Différentes approches sont cependant possibles, toutes relativement techniques et toutes présentant diverses limites. Nous nous contenterons ici de les évoquer rapidement[5].

La « boîte de confiance »

Le concept de « boite de confiance » (*confidence box*) constitue une manière relativement simple d'aborder le problème. Elle consiste à dessiner dans le plan coût-efficacité – dans lequel le ratio coût efficacité r = C/E est la pente d'une droite passant par l'origine (C = r.E) – la « boîte » représentée par les valeurs extrêmes des intervalles de confiance à 95 % des deux variables, efficacité et coût.

Par définition il y a 95 % de chances que la boite contienne les « vraies » valeur de C et de E. Il est alors assez logique de prendre comme bornes de l'intervalle de confiance du ratio coût efficacité, r_m et r_M, la pente des droites encadrant la boite, c'est-à-dire passant respectivement par ses coins « nord-ouest » et « sud-est ».

La *figure 7* donne un exemple de « boîte de confiance » pour la branche « traitement » du modèle de Markov du chapitre 8 en acceptant l'approximation normale des intervalles de confiance. Sous ces hypothèses, l'intervalle de confiance du ratio coût-utilité est égal à [51,1-60,6] uc par QALYs.

Cette méthode simple a toutefois plusieurs limites. Elle suppose une distribution normale des coûts et des efficacités ; elle suppose que coûts et efficacité (ou deltas de coûts et d'efficacité) se situent préférentiellement dans le quadrant nord-est du plan coût-efficacité ; elle reste indéterminée si l'intervalle de confiance de l'efficacité inclut 0 ; elle suppose enfin que les distributions des coûts et des efficacités sont indépendantes.

Les « ellipses de confiance »

Cette dernière restriction peut être levée par l'utilisation « d'ellipse de confiance » (*figure 8*) Si les variables de coût et d'efficacité sont des variables aléatoires suivant des distributions normales et corrélées entre elles positivement (exemple d'un traitement qui augmente la durée de vie et donc le coût) ou négativement (exemple d'un traitement qui diminue le recours à l'hospitalisation), il y a 95 %

5 Sur ces questions, on se réfèrera par exemple à AH Briggs : « Statistical issues in economic evaluation », in *Encyclopedia of Health Economics*, édité par AJ Culyer, Elsevier 2014, Vol. 3 pp. 352-61 ou à O'Brien BJ et Briggs AH : « Analysis of uncertainty in health care cost-effectiveness studies : an introduction », *Stat Meth In Medical Research* 2012 ; 11(6) : 455-68.

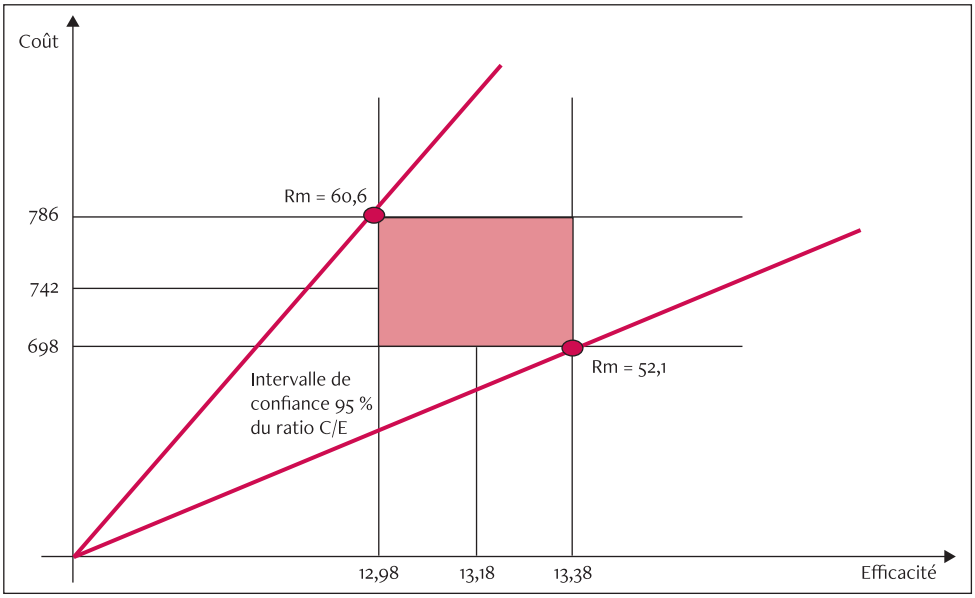


Figure 7. La « boîte de confiance ».
Modèle de Markov du chapitre 8.

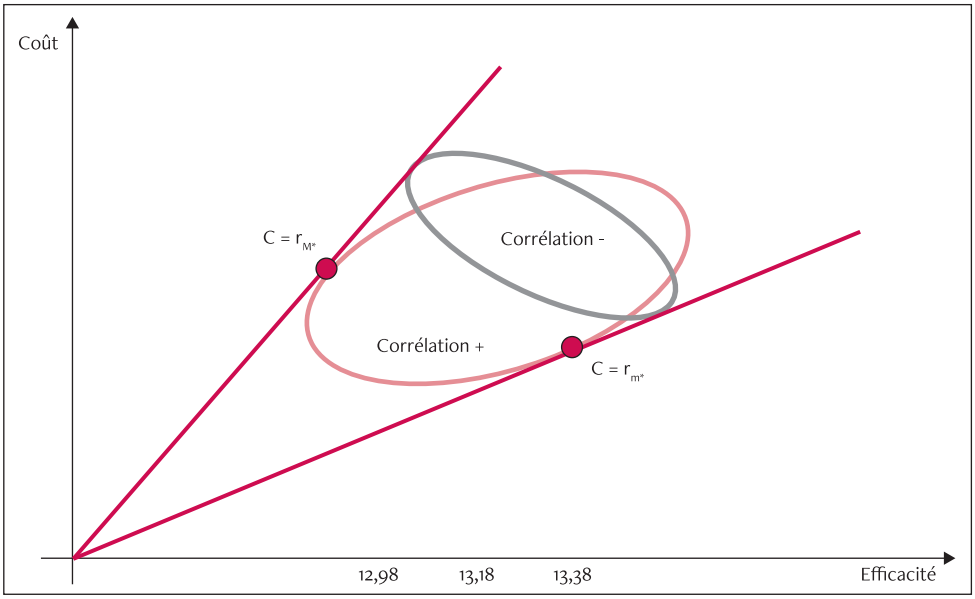


Figure 8. La méthode des ellipses de confiance.

de chances que les « vraies » valeurs d'efficacité et de coût se situent dans une ellipse centrée sur les valeurs moyennes dans le plan C,E.

Comme précédemment, on prendra comme bonne de l'intervalle de confiance pour le ratio coût-efficacité r, les pentes r_m et r_M des tangentes à l'ellipse.

La forme elliptique du domaine de confiance s'explique par les caractéristiques des lois normales bi-variées, de type $\phi(C,E)$ avec les moyennes, m(E) et m(C), les variances, $s(E)^2$ et $s(C)^2$, la covariance Cov(C,E) des deux variables et leur coefficient de corrélation ρ égal à Cov(C,E)/s(E).s(R).

On montre que les points (C,E) sont distribués selon des courbes d'iso-densité de probabilité d'équation :

$$\frac{1}{1-\rho^2}\left[\left(\frac{E-m(E)}{s(E)}\right)^2+\left(\frac{C-m(C)}{s(C)}\right)^2-2.\rho.\left(\frac{E-m(E)}{s(E)}\right).\left(\frac{C-m(C)}{s(C)}\right)\right]\sim\chi^2_\alpha$$

où χ^2_α est la valeur de la loi du chi-2 à deux degrés de liberté pour le risque α[6].

Cette expression est l'équation d'une ellipse en coordonnées cartésiennes dans le plan (C,E) et définit la frontière du domaine de confiance, pour un risque donné. La taille et la forme des ellipses dépendent des paramètres de la distribution conjointe des coûts et des efficacités. On peut montrer que le grand axe suit la droite de régression de C sur E et passe par le point moyen du nuage de points. Les ellipses sont d'autant plus allongées que la corrélation est forte. Elles sont orientées nord-ouest si la corrélation est positive, sud-est si elle est négative et est-ouest si elle est nulle (grand axe parallèle à l'axe des efficacités).

Les pentes des tangentes aux ellipses qui définissent les bornes de l'intervalle de confiance du ratio coût/efficacité s'obtiennent en remplaçant C par son expression C = r.E dans l'expression ci-dessus. On obtient une équation du second degré en X paramétrée par la pente r. La condition pour que cette équation n'admette qu'une racine double est une nouvelle équation quadratique en r dont la solution donne les pentes des deux tangentes. Le calcul est fastidieux mais simple.

À titre d'exemple, la *figure 9* montre les ellipses correspondant à deux variables coût et efficacité indépendantes distribuées normalement pour deux niveaux du seuil de confiance, 90 % et 95 %. Les ellipses ont alors une forme circulaire et leur grand axe est parallèle à l'axe des efficacités.

6 Soit 4,605 pour un risque de 10 % (90 % des points dans l'ellipse) ou 5,991pour un risque de 5 % (95 % de points dans l'ellipse).

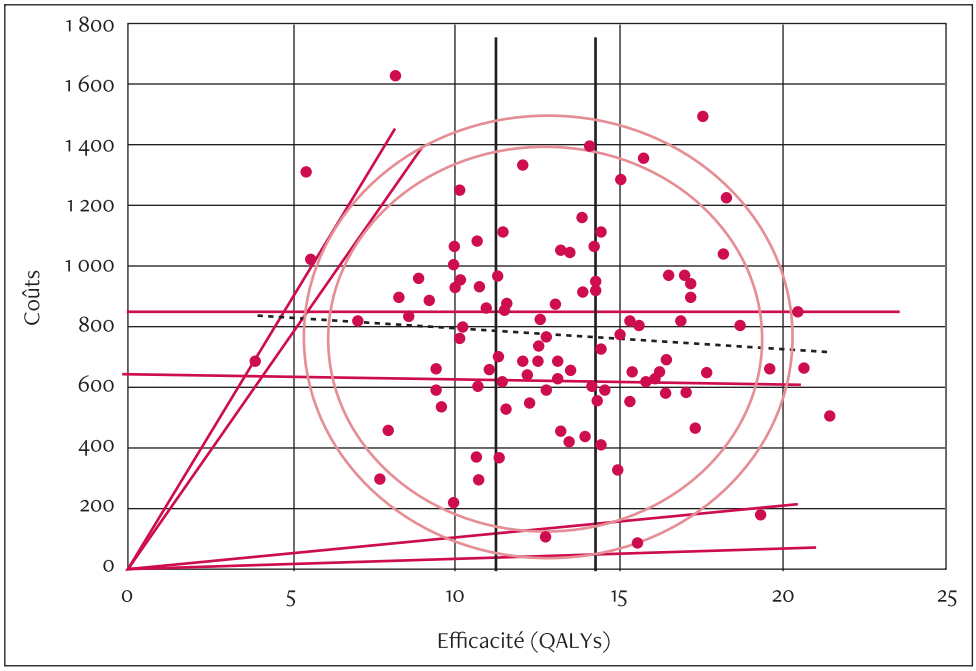


Figure 9. Un exemple d'ellipse de confiance pour deux variables indépendantes normalement distribuées.
Nota : Les deux variables suivent des lois normales indépendantes dont les moyennes et écart types sont respectivement de 13,2 et 3,2 pour l'efficacité (QALYs) et de 740 et 300 uc pour les coûts.

Le théorème de Fieller

Le statisticien britannique Edgar C. Fieller a proposé dès les années 50 une méthode relativement simple et élégante pour résoudre le problème des intervalles de confiance d'un rapport de deux variables aléatoires[7]. Cette approche, également paramétrique, est très discutée dans la littérature spécialisée.

Considérant que le ratio r = C/E des deux variables aléatoires Coût (C) et Efficacité (E), que l'on suppose normalement distribuées et corrélées entre elles (ou celui des variables ΔC et ΔE dans le cas d'un ratio incrémental), peut s'écrire sous la forme, r.E-C = 0, il s'intéresse à la

7 Fieller EC. « Some problems in interval estimation ». *Journal of the Royal Statistical Society* 1954, Series B. 16 (2) : 175-185.

distribution du terme (r.E-C) dont la variance V s'écrit par définition : r^2.var(E) + var(C) – 2.r.cov(C,E). Il peut alors construire une variable aléatoire distribuée selon une loi normale standard en divisant l'expression précédente par son écart-type

$$\frac{r.E - C}{\sqrt{V}} \sim N(0,1)$$

Il suffit alors de rapporter cette expression au quantile d'ordre 1-α/2 de la loi $N(0,1)$, noté z_α pour obtenir une équation du second degré en r dont les racines (si elles existent) donnent les bornes de l'intervalle de confiance du ratio r.

L'équation qui est assez simple (en dépit des apparences) s'écrit :

$$r^2\ [E^2 - z_\alpha^2.var(E)] - 2.r\ [E.C - z_\alpha^2.cov(E,C)] - C^2 - z_\alpha^2.var(C)] = 0$$

Deux remarques. On montre que l'approche de Fieller donne des résultats proches de la méthode des ellipses. En revanche, dans le cas de variables indépendantes avec donc cov(C,E) = 0, on ne retrouve pas nécessairement les résultats de la « boite » de confiance.

Les approches empiriques : Monte Carlo et « bootstrap » non paramétrique

Toutes les voies précédentes approchent la construction de l'intervalle de confiance du ratio cout-efficacité à partir des distributions supposées normales du numérateur et du dénominateur prises séparément.

Une voie différente consiste à considérer le rapport C/E comme *une variable en tant que telle* et à en retracer la distribution empirique à partir d'une simulation de Monte Carlo.

Le *tableau 3* donne les résultats de 500 itérations de Monte Carlo effectuées sur les ratios coût-efficacité issus du modèle de Markov du chapitre 8.

Les intervalles de confiance construits à partir des percentiles 2,5 % et 97,5 % (pour un risque alpha de 5 %) sont très larges et ne donnent pas de différences significatives.

Appliquer la méthode de Monte Carlo aux résultats de modèles de type Markov ou DES est relativement simple dans la mesure où ils se prêtent aux micro-simulations.

Tableau 2. Simulations de Monte Carlo des ratios coût-efficacité – Modèle de Markov (chapitre 8)

Statistique	Traitement (n = 500)	Non traitement (n = 500)
Moyenne	74,16	84,79
Écart-type	70,97	84,07
Min	11,11	11,11
Max	443,71	466,20
Médiane	51,68	54,01
Percentile 2,5 %	11,11	11,11
Percentile 97,5 %	274,55	329,24
IC (approx. normale)	67,8-80,6	77,2-92,4

Mais cela est plus difficile dans le cas des études médico-économiques observationnelles menée « en vie réelle » pour lesquelles on ne dispose en général que d'un échantillon unique de mesures (par exemple du ratio coût-efficacité) dont on ne connaît pas la loi de distribution et à partir duquel il n'est donc pas possible de calculer des intervalles de confiance. D'autant que la taille de l'échantillon est souvent insuffisante pour justifier une approximation normale.

Le bio-statisticien américain Bradley Ephron a proposé dans les années 70 une méthode originale (dite du « *bootstrap* » non paramétrique) pour approcher la distribution empirique et déterminer les intervalles de confiance[8].

La méthode – dont on ne donnera ici qu'un aperçu très succinct – consiste à produire B échantillons « bootstrappés » de même taille que l'échantillon initial, chacun d'entre eux étant composé de valeurs tirées au sort (avec remise) au sein de l'échantillon initial avec des probabilités égales.

On montre que les échantillons « bootstrappés » permettent d'approcher les caractéristiques de la distribution empirique de l'échantillon de départ et de déterminer alors empiriquement l'intervalle de confiance de la moyenne par la méthode des quantiles (2,5 % et 97,5 % pour un risque α de 5 %)

8 L'ouvrage de référence est celui de Efron B et Tibshirani R, *An Introduction to the Bootstrap*, 1994, Chapman & Hall/CRC. Il existe de nombreuses présentations techniques de la méthode du « bootstrap ». Certains « packages » statistiques informatiques incluent des fonctions « bootstrap ». Le terme fait métaphoriquement référence à une situation difficile dont on se sort miraculeusement comme le baron de Münchhausen qui prétendait s'être extrait de sables mouvants avec son cheval en tirant sur les attaches (« straps ») de ses bottes.

Supposons que l'on dispose d'une série de ratios coût-efficacité mesurés sur une population de 20 patients dans une étude médico-économique observationnelle (colonne E0 du *tableau 3*). On construit B échantillons de 20 variables en tirant chaque fois au sort une valeur du tableau initial. Les colonnes E1, E2, ..., EB du *tableau 3* donnent des exemples de ce ré-échantillonnage. On constate que la composition des échantillons « boostrappés » varie, certaines variables de l'échantillon initial se retrouvant plusieurs fois dans l'un des échantillons, d'autres au contraire n'y figurant pas.

Tableau 3. Un exemple d'échantillons « bootstrappés » (1,2, ..., B) à partir d'un échantillon initial observé (0)

	E0	E1	E2	.....	EB
1	135,01	29,65	**29,65**		**64,31**
2	62,09	**18,68**	135,01		28,36
3	108,87	28,36	**108,87**		**64,31**
4	69,27	**154,82**	**108,87**		**154,82**
5	18,68	64,31	88,94		29,65
6	64,31	**62,09**	**62,09**		98,12
7	47,59	**135,01**	**29,65**		64,31
8	44,06	98,12	**98,12**		**154,82**
9	134,47	**99,48**	47,59		62,09
10	134,66	**154,82**	**24,50**		**113,28**
11	29,65	**62,09**	154,82		135,01
12	28,36	**62,09**	98,12		**154,82**
13	154,82	**18,68**	**134,47**		**154,82**
14	37,46	47,59	**134,47**		**113,28**
15	115,63	**62,09**	69,27		134,66
16	99,48	115,63	**64,31**		**24,50**
17	88,94	69,27	**24,50**		**24,50**
18	98,12	**99,48**	**62,09**		**113,28**
19	113,28	**135,01**	**98,12**		115,63
20	24,50	134,66	**64,31**		62,09

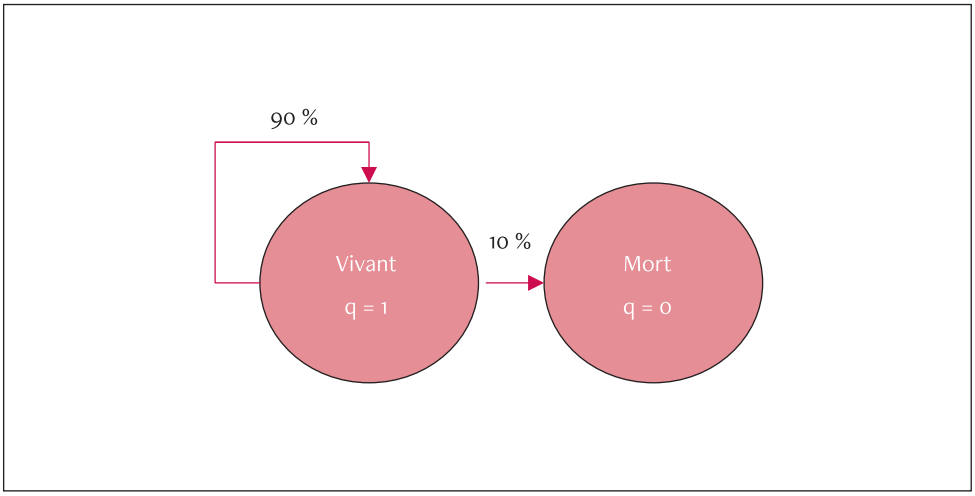
Le *tableau 4* donne le ratio coût efficacité moyen de l'échantillon de départ la moyenne et l'intervalle de confiance de ce ratio à partir des échantillons « bootstrappées » quand B = 100.

Tableau 4. Caractéristiques comparées de la distribution initiale et des distributions « bootstrappées »

	Échantillon Original	Bootstrap (B = 100)
Moyenne	80,46	80,53
IC 95 % borne basse		60,13
IC 95 % borne haute		96,10

Annexe
Comparaison des résultats d'un modèle paramétrique et des estimations de Monte Carlo

Pour comparer les résultats obtenus par des analyses de Monte Carlo avec les « vraies » valeurs issues d'un modèle paramétrique connu, on supposera un modèle de Markov d'espérance de vie très simple, comportant deux états « vivant » ou « mort » avec un taux de mortalité spécifique constant de 10 % par an.



La durée moyenne de survie est alors distribuée selon loi paramétrique de type DEALE (cf. annexe Chapitre 8.) qui permet d'en connaitre les caractéristiques « vraies », moyenne, écart-type, médiane, percentiles, etc. Des simulations numériques de type Monte Carlo permettent d'obtenir de très bonnes estimations des valeurs « paramétriques », d'autant meilleures que le nombre des itérations est grand (*tableau 5*).

Dans ce cas simple, la loi de distribution étant parfaitement connue, les indicateurs statistiques peuvent être calculés théoriquement rendant inutile une analyse de Monte Carlo. En revanche, pour tout modèle un peu plus complexe, produisant des résultats dont on ne connait pas les lois de distribution – ce qui est le cas assez général des analyses médico-économiques – la méthode de Monte Carlo fournit des distributions empiriques à partir de laquelle il est possible de déterminer des estimateurs sans biais de caractéristiques de la distribution inconnue.

Tableau 5. Comparaisons des caractéristiques statistiques du modèle paramétrique DEALE et de trois simulations de Monte Carlo

	Modèle paramétrique DEALE	**Simulations de Monte Carlo**		
		Sim. 1 (n = 100)	**Sim. 2 (n = 1 000)**	**Sim. 3 (n = 10 000)**
Moyenne	10,00	10,31	10,33	10,04
Écart-Type	10,00	9,92	9,72	9,66
Variance	100,00	98,43	94,42	93,26
9e décile*	1,05	1,00	2,00	1,00
Médiane	6,93	7,00	7,00	7,00
1er décile*	23,03	25,00	23,00	23,00

* : 90 % de la population a vécu plus de 1,05 ans et 10 % plus de 23,03 ans.

Chapitre 10
Quelles données pour l'analyse médico-économique ?

Source unique ou sources combinées ?

La réalisation d'une étude requiert l'utilisation de données permettant de chiffrer les différents paramètres constitutifs de l'évaluation économique envisagée. Plusieurs aspects interfèrent pour déterminer les données à utiliser. En premier lieu, il faut considérer la nature des données nécessaires qui dépend bien évidemment du type d'approche prévu, les études coût-utilité nécessitant des données de qualité de vie qu'il n'y a pas lieu de considérer pour une étude coût-efficacité. En second lieu, la méthode employée pour réaliser l'évaluation économique conditionne, au moins en partie, les sources mobilisables. Si l'on recourt en général à une modélisation, comme détaillé au chapitre 8, on peut dans certains cas opter pour une étude prospective en parallèle d'un essai clinique, c'est-à-dire utiliser une étude économique superposée à un essai clinique, dénommée en anglais des « *piggyback study* »[1]. Ces deux configurations conduisent à se référer à des données d'origines différentes.

En effet, dans le cadre d'une modélisation, il faut accepter de mobiliser plusieurs sources de données car chacune ne permet en général de documenter qu'une partie des paramètres à estimer. C'est en particulier le cas en France, dans la mesure où il existe un problème d'articulation entre les bases médico-administratives qui permettent de suivre les consommations de ressources d'un côté et les sources d'information clinique de l'autre, notamment pour des raisons d'interconnexion des systèmes d'information. Il n'existe

1 Le mot « piggyback », difficile à traduire, exprime l'idée que l'étude médico-économique s'appuie sur un sous-jacent clinique.

donc pas de base de données générale qui puisse être utilisée pour obtenir l'ensemble des données nécessaires pour réaliser une étude d'évaluation économique sur la base d'un modèle[2]. En revanche, la réalisation d'une *piggy-back study* permet de se référer à une seule et même source de données pour estimer la consommation de ressources en volume et les effets de santé, mais il faut en général recourir à une autre base pour valoriser le montant des ressources utilisées.

Données de coût et données d'efficacité

Il est utile de commencer par distinguer les données à considérer selon leur nature dans la mesure où les sources de données sont souvent différentes suivant le type d'information que l'on recherche. La première distinction à faire renvoie de façon évidente aux deux parties constitutives de toute évaluation économique en santé, entre les éléments qui permettent de documenter les coûts d'une part, et ceux qui renseignent les résultats de santé d'autre part – autrement dit, ce qui va être pris en compte dans le numérateur et ce qui va alimenter le dénominateur du ratio à estimer.

Du côté des coûts, deux types d'information sont nécessaires pour estimer le montant associé à chacune des stratégies prises en compte. Il convient d'abord de documenter l'utilisation des ressources associée à chacune d'entre elles, de façon à identifier tous les postes de coût pertinents, avant de pouvoir en donner une traduction monétaire par une valorisation qui suppose l'obtention des coûts unitaires correspondants. Comme on l'a vu au chapitre 4, les données relatives à la consommation de ressources peuvent être ordonnées suivant les différentes catégories de coût, de même que celles qui vont en permettre la valorisation dépendent des choix méthodologiques faits sur la perspective retenue.

En regard de ces données de coût, il convient d'obtenir les données permettant d'estimer la mesure du résultat associé à chaque option thérapeutique retenue. Dans le cadre d'une étude coût-efficacité, il s'agit de l'information relative à la mesure d'efficacité retenue, qu'il s'agisse d'un indicateur d'efficacité intermédiaire (l'effet sur le taux de cholestérol des différentes interventions) ou d'un indicateur d'efficacité finale sur la morbi-mortalité (en particulier l'espérance de vie).

---

2 Pour un aperçu du champ couvert par les bases de données existantes en France et à l'étranger, voir en particulier le rapport réalisé par B. Bégaud, D. Polton et F. von Lennep à la demande de la ministre de la santé : « Les données de vie réelle, un enjeu majeur pour la qualité des soins et la régulation du système de santé », mai 2017.

Mais en général, on se réfère à une étude coût-utilité, et il convient de valoriser ce résultat d'efficacité en lui attribuant une valeur d'utilité traduisant la qualité de vie, de façon à pouvoir en déduire une mesure en termes de QALYs. Comme pour les coûts, ce sont donc deux types d'information différents qui sont nécessaires pour estimer la mesure d'effet.

Intérêt et limites des « piggyback studies »

Lorsqu'on veut réaliser une *piggyback study* en parallèle d'un essai clinique, il faut mettre au point un cahier d'observation économique qui doit être rempli pour chaque patient de l'essai en même temps que le cahier d'observation clinique (le CRF – *case report form*), sans oublier de collecter les données de qualité de vie permettant de calculer des QALYs si on souhaite conduire une étude coût-utilité. Cette approche est intéressante car elle permet d'obtenir de façon homogène des données pour les bénéfices cliniques et la consommation de ressources sur les patients de l'essai. Elle présente néanmoins un certain nombre de limites dont on se contentera ici d'évoquer les principales[3].

En premier lieu, la question de la collecte des données et de leur pertinence est souvent problématique dans la mesure où d'une part la consommation de ressources est la plupart du temps collectée à partir de « *case report forms* » spécifiques qui doivent avoir été bien conçus, et où, d'autre part, il faut s'assurer de disposer d'un processus de collecte performant pour éviter de constituer une base contenant trop de données manquantes. En second lieu, il existe un important problème de transposabilité et de généralisabilité des résultats qui dépasse certes le seul cas des *piggyback studies*, mais qui se pose avec plus d'intensité dans ce type d'études. Le problème de transposabilité concerne d'abord la consommation de ressources. Outre qu'elle dépend des centres investigateurs, et peut ne pas correspondre aux pratiques d'autres pays ou contextes, cette composante de l'évaluation des coûts comporte un biais lié au protocole de l'essai qui conduit en général à une sur-consommation par rapport à la pratique courante. De plus, la question du choix des prix unitaires à considérer pour valoriser ces ressources peut ne pas être simple. Par ailleurs, les résultats d'efficacité peuvent ne pas

3 Pour une discussion des avantages et limites de cette approche, on peut consulter l'article de O'Sullivan AK, Thompson D, Drummond MF : « Collection of health-economic data alongside clinical trials : is there a future for piggyback evaluations ? » *Value in Health* 2005 Jan-Feb ; 8(1) : 67-79. Ou pour une discussion plus longue l'ouvrage de H. Glick *et al.* « *Economic evaluation in clinical trials* », Oxford University Press, 2007.

être valables dans d'autres populations que celles de l'essai du fait par exemple de différences épidémiologiques, sans parler du problème plus général de manque de représentativité par rapport aux résultats en vie réelle. Enfin, la question des comparateurs pertinents est cruciale : ils peuvent différer d'un pays à l'autre et celui ou ceux inclus dans l'essai clinique peuvent ne pas être pertinents dans un contexte spécifique. Au total, ces limites peuvent être jugées rédhibitoires et expliquent en partie pourquoi l'approche par la modélisation est préférée la plupart du temps.

Des données pour la modélisation

Dans l'approche par modélisation, on peut en effet élaborer la comparaison souhaitée par construction et par conséquent ne pas être soumis aux limites de la comparaison aux seules stratégies considérées dans un essai clinique. Mais la multiplicité des options thérapeutiques qu'on peut ainsi envisager ne fait que renforcer le problème des données nécessaires pour faire tourner le modèle. Dans ce cas, l'hétérogénéité des sources de données est la règle mais il convient de chercher les meilleures possibles, plusieurs cas de figure pouvant se présenter. Lorsqu'on ne dispose d'aucune information pour un ou plusieurs paramètres, il faut alors s'en remettre à des dires d'experts ou de pures spéculations, étant entendu que le caractère hypothétique des valeurs retenues en réduit la validité. A l'inverse, lorsqu'on dispose de plusieurs sources d'information, il convient d'utiliser la meilleure pour avoir une estimation qui constitue alors le cas de base, et d'utiliser les autres pour effectuer des études de sensibilité.

Qualité et fiabilité des données

La question de la qualité des sources de données est un sujet complexe. En ce qui concerne les données cliniques, il est usuel de distinguer leur niveau de preuve scientifique en fonction de leur source, les essais comparatifs randomisés de forte puissance et les méta-analyses étant jugés supérieurs aux essais de faible puissance, et surtout aux études cas-témoins et aux études rétrospectives et aux séries de cas[4]. Plus largement, on oppose bien souvent les données issues d'études conduites selon un plan expérimental à celles provenant d'études observationnelles, les premières étant considérées

4 Sur ce point, voir par exemple le rapport de la HAS : « Niveau de preuve et gradation des recommandations de bonne pratique », avril 2013.

comme fournissant un meilleur niveau de preuve[5]. Si l'approche expérimentale, en neutralisant les facteurs confondants grâce à la randomisation des individus, permet normalement de conclure qu'une différence sur les variables mesurées est imputable à la différence des stratégies comparées, elle pose le problème de ne pas être représentative de la mise en œuvre de ces stratégies dans la vie réelle. A l'inverse, si elles correspondent mieux à la vie réelle, les données issues d'études observationnelles sont souvent soumises à différents biais. Le biais de sélection et le biais de confusion, tenant au fait que les patients recevant un traitement plutôt qu'un autre ne sont pas choisis au hasard mais en fonction de leurs caractéristiques, faussent en particulier la comparaison même si certains d'entre eux peuvent être corrigés.

Pour les données cliniques, une possibilité est de rechercher les informations nécessaires dans l'*Evidence-Based Medicine*, c'est-à-dire dans les nombreux essais cliniques dont les résultats sont publiés, ce qui nécessite une recherche bibliographique adaptée. Un des problèmes méthodologiques liés à ce choix apparaît lorsqu'on doit considérer des stratégies alternatives qui n'ont pas fait l'objet d'une comparaison directe entre elles dans un même essai clinique. Il faut alors recourir à des méthodes statistiques permettant de faire des comparaisons indirectes, en particulier des méta-analyses en réseau, ce qui ne va pas sans poser de problèmes[6]. Tout ou partie des informations nécessaires pour décrire l'utilisation des ressources associée à chaque stratégie peut également être issu des essais cliniques, par exemple la consommation médicamenteuse ou la fréquence d'hospitalisation.

Mais on peut également se référer à des données provenant d'études observationnelles, soit lorsqu'on procède à la réévaluation de stratégies déjà mises en œuvre, dans le cadre d'études post-inscription, soit même dès l'évaluation initiale d'une nouvelle option thérapeutique, pour estimer certains paramètres épidémiologiques pertinents, identifier les comparateurs appropriés préexistants et leurs performances, ou mesurer les coûts associés. De telles données de vie réelle peuvent provenir par exemple de registres ou d'études de cohorte, que le portail épidémiologie-France répertorie de façon utile pour certaines d'entre elles[7].

---

5 Pour une discussion des aspects méthodologiques, voir l'annexe 3 du rapport pré-cité dans la note 2.

6 On peut renvoyer le lecteur intéressé à un article en français écrit par Y. Cabon et N. Molinari : « Incohérence dans le cadre des méta-analyses en réseau ». *Revue d'Épidémiologie et de Santé Publique*, 63 (Suppl 2), 2015, p S75-S76.

7 Il existe également différentes bases publiques ou privées permettant de recueillir des données sur les diagnostics et les prescriptions en médecine de ville. Voir le rapport cité en note 2.

Une tendance croissante : Les bases de données médico-administratives

Les bases de données médico-administratives constituent cependant une des principales sources pour les données en vie réelle, leur avantage provenant du fait que, outre qu'elles sont souvent à large échelle, elles fournissent des informations collectées antérieurement à la recherche menée qui n'exerce donc pas d'influence sur les pratiques et leurs résultats. Mais dans le même temps, il faut reconnaître que ceci est une limite à leur apport puisque les informations collectées en routine peuvent ne pas être en adéquation avec les données nécessaires. On sait en particulier qu'il existe un problème d'articulation entre les bases médico-administratives centrées en général sur la mesure de la consommation de ressources et les autres sources de données, registres ou cohortes par exemple, qui fournissent l'information sur les aspects cliniques.

La France dispose depuis avril 2017 d'un Système National des Données de Santé (SNDS), selon les principes de l'article 193 de la loi de modernisation du système de santé votée en janvier 2016. Le SNDS a vocation à intégrer plusieurs des bases médico-administratives et à en faciliter l'accès, en particulier le SNIIRAM (base de données de remboursement de l'assurance-maladie), le PMSI (base de données des établissements de santé), et CepiDC (base de données sur les causes médicales de décès), ainsi que d'autres éléments tels que les données des assurances complémentaires. Potentiellement, on dispose ainsi en France d'un formidable outil permettant de nombreuses analyses[8] dont l'utilisation requiert néanmoins un savoir-faire certain et un droit d'accès.

En effet ces bases de données – de volumétrie très importante – ont été constituées à des fins de gestion (tarification hospitalière, liquidation des dossiers de remboursement) et non d'évaluation médicale ou médico-économique. Elles permettent de faire des études historiques de suivi de patients plus adaptées à la description de la prise en charge standard d'une pathologie qu'à l'évaluation d'un nouveau traitement non encore sur le marché.

---

8 Pour avoir une idée des principales possibilités offertes par ces bases, voir en particulier Bezin J, Duong M, Lassalle R, et al. : « The national healthcare system claims databases in France, SNIIRAM and EGB : Powerful tools for pharmacoepidemiology ». *Pharmacoepidemiol Drug Saf* 2017. Les fiches techniques 3 et 4 du « Guide méthodologique pour l'évaluation économique à la HAS » publié en octobre 2011 fournissent des listes intéressantes des différentes bases utiles pour la mesure des ressources consommées et la valorisation de ces ressources.

Par ailleurs, elles sont souvent faiblement médicalisées en raison des règles imposées par le secret médical. La base SNIIR'AM pour les soins de ville en France ne comporte pas les diagnostics (en l'absence de codage des pathologies) qu'il faut donc reconstituer grâce à des algorithmes d'analyse des consommations de soins.

En revanche, elles sont exhaustives, objectives, longitudinales et elles permettent de suivre des épisodes de soins sur plusieurs années avec notamment l'articulation ville-hôpital (par exemple les séjours évités).

Quelle que soit leur richesse, elles ne peuvent prétendre se substituer aux autres sources de données que sont les essais cliniques, les études observationnelles, avec lesquelles elles peuvent être combinées.

L'accès aux bases publiques du SNDS enfin est conditionné au respect de différentes règles tenant à la finalité des études qui doivent servir, directement ou indirectement, l'« intérêt public », à l'exclusion de finalités marketing (par exemple le « ciblage » de prescripteurs), et au respect d'une procédure d'anonymisation stricte des individus qui ne doivent en aucun cas pouvoir être identifiés ou ré-identifiés à partir des données utilisées.

Une partie de ces données (à caractère non individuel) sont toutefois en accès libre, ouvrant ainsi la voie au développement de *l'open data* réclamé par certains.

D'une manière plus générale, il faut dire que ce domaine des données de santé est en constante évolution en partie du fait des capacités offertes par les technologies de l'information qui augmentent considérablement les possibilités en matière de génération de données, avec les *big data* et autres objets connectés, mais aussi d'analyse automatisée de ces données, avec le développement de l'algorithmique et à terme du *machine learning*.

Données internationales vs données françaises

Enfin, les données de qualité de vie sont souvent cruciales pour pouvoir mener à bien une étude coût-utilité. Idéalement, il faut pour cela disposer de données sur des patients français, ou à tout le moins européens. Et si l'on se réfère à l'instrument privilégié qu'est l'EQ-5D, il est nettement préférable d'utiliser la matrice française des valeurs associées aux différents états de santé. Un des meilleurs moyens d'obtenir de telles données est donc de faire passer le questionnaire de l'EQ-5D à des patients français de façon à pouvoir valoriser correctement les différents états de santé identifiés dans un modèle. Une alternative à cette génération de données est d'utiliser les données disponibles dans la littérature, ce qui peut se faire soit à partir d'une recherche bibliographique spécifique, soit en interrogeant la base de données du CEA

Registry qui permet d'obtenir directement les poids d'utilité attribués aux différents états de santé dans les études coût-utilité publiées sur le sujet d'intérêt[9]. Mais il n'est alors souvent pas possible de s'assurer de la pertinence de ces données dans le cadre français car il est parfois difficile de savoir dans quel contexte national elles ont été élaborées, voire de connaître l'instrument utilisé pour dériver ces poids d'utilité, EQ-5D ou autre.

9 Le CEA Registry est une base de données bibliographiques sur les études d'évaluation économique publiées. Elle propose d'une part un résumé d'information sur chacune des études identifiées, et d'autre part une information spécifique sur les poids d'utilité référencés dans ces études. Elle est accessible en ligne gratuitement sur le site http://healtheconomics.tuftsmedicalcenter.org/cear4/Home.aspx

Chapitre 11
Médico-économie, éthique médicale et philosophies sociales

« *L'analyse coût-efficacité est-elle éthique ?* » se demandait Allan Williams en 1992 dans le *Journal of Medical Ethics*[1] en distinguant le cas de l'évaluation économique en tant que telle de celui relatif à certaines hypothèses faites dans la mesure des coûts et, surtout, des résultats qui peuvent s'avérer discriminatoires dans la priorisation des traitements. Une autre question, un peu différente, mérite également d'être posée : comment justifier la prise en compte de l'efficience comme critère de prise en charge collective ? La première question, celle de Williams, vise surtout la compatibilité de l'analyse médico-économique avec l'éthique médicale. La seconde avec les valeurs de la protection sociale dans un pays donné.

Choix éthiques et mesure des résultats

« *C'est dans le domaine de la mesure des résultats*, écrit Williams qui a toujours milité en faveur des QALYs, *que j'ai rencontré les objections* [éthiques] *les plus intenses* ». Nous avons largement fait état dans le chapitre 6 des questions en débat, qui tournent essentiellement autour de la pondération des QALYs, et nous n'y reviendrons pas. Dans l'article de 1992, Allan Williams défend timidement son principe d'égalité (« *a QALY is a QALY is a QALY no matter who gets it* ») tout en reconnaissant que les personnes interrogées dans différentes enquêtes ont plutôt tendance à penser qu'un élément de discrimination en faveur de certaines populations ne serait pas nécessairement injuste, qu'elles soient définies par leur âge, leur comportement, leur situation de famille ou leur pathologie.

1 Williams A. Cost-Effectiveness Analysis: is it ethical? *Journal of Medical Ethics* 1992 ; 18(1) : 7-11. Repris dans Culyer AJ et Maynard A (eds) : *Being Reasonable about the Economics of Health, Selected Essays by Allan Williams*, Edward Elgar Pub, 1997.

Dans les choix tragiques de répartition d'une ressource médicale rare entre malades – pas nécessairement monétaire, cela peut être aussi le cas de greffons – les enquêtes se prononcent en général en faveur des jeunes contre les vieux, des personnes « vertueuses » contre celles qui prennent des risques avec leur santé, de celles avec charge de famille contre les personnes isolées, etc. Donc en faveur d'un principe de discrimination positive. Est-il pour autant admissible de surpondérer la vie d'un patient non-fumeur par rapport à un patient fumeur ? Le tabac diminue t'il la valeur d'une vie ? Celle-ci est-elle au contraire plus élevée si la pathologie est rare ? Autant de questions délicates qui renvoient à des jugements de valeur extra-économiques.

Le cas des maladies rares, que nous avons rapidement évoqué au chapitre 6, est exemplaire et a donné lieu à des débats passionnés en Angleterre où les tenants de l'orthodoxie médico-économique se sont élevés contre l'idée d'un statut particulier accordé aux médicaments orphelins, sans toutefois craindre la contradiction puisque cette même orthodoxie préconise que les états de santé soient valorisés du point de vue de la population générale... plutôt favorable à la pondération [2].

De toute manière, comme le note Williams à juste titre, il n'existe pas de mesure d'efficacité qui ne soit entachée de considérations éthiques, implicites ou explicites. Un end-point « technique » d'un essai clinique comme un taux de survie à 2 ans suppose par exemple : i) qu'une survie de moins de 2 ans n'a pas de valeur ; ii) que, pour ceux qui ont survécu, la prolongation de la survie au-delà de 2 ans n'a pas de valeur ; iii) que la qualité de la survie ne compte pas ; iv) que la personnalité du patient ne compte pas. Autant de considérations qu'il est possible de contester. L'introduction des QALYs a au moins le mérite de prendre en considération le point iii) par introduction explicite de la qualité de vie des patients et d'ouvrir la discussion sur le point iv) à travers le débat sur la pondération.

Portons au crédit des QALYs de rendre explicite les systèmes de valeurs qui sous-tendent les choix thérapeutiques les plus techniques qui resteraient sans doute implicites si la question économique n'était pas clairement posée [3].

---

2 Pour les partisans de l'orthodoxie, cf. McCabe C, Claxton K, Tsuchiya A. Orphan drugs and the NHS: should we value rarity ? *BMJ* 2005 ; 331(7523) : 1016-9. Pour eux la réponse est négative : La rareté de la pathologie n'implique aucun traitement de faveur et ne suspend pas la logique du coût d'opportunité.

3 Williams A. QALYs and Ethics: A health economist's perspective. *Social Science and Medicine* 1996 ; 43(12) : 1795-804. Repris dans Culyer AJ et Maynard A (eds), *op. cit.*

Choix éthiques et coûts indirects

En ce qui concerne les coûts, Williams soulève deux questions éthiques importantes, pas ou mal résolues dans la littérature, celle des coûts indirects et notamment de la valorisation des « *pertes de production* » associées aux arrêts de travail et celle la répartition de la charge financière.

On n'abordera pas ici cette dernière qui porte sur l'équité dans le financement des soins mais qui a finalement plus à voir avec le contexte institutionnel qu'avec les études coût-efficacité en tant que telles. Il est vrai que la répartition des coûts de la santé n'est pas sans influence sur la distribution des soins. Selon qu'un système est public ou privé et, s'il est public, selon qu'il est financé par l'impôt sur le revenu (comme au Royaume-Uni) ou par un système de cotisations et de taxes affectées (comme en France), les effets redistributifs sont différents. Ils sont plutôt « verticaux », du riche au pauvre, dans le premier et plutôt « horizontaux », du bien-portant au malade, dans le second. Mais cette caractéristique générale s'impose à toutes les modalités d'affectation des ressources dans un contexte institutionnel donné, qu'on utilise ou non les études coût-résultat pour sélectionner ou prioriser les options thérapeutiques.

La question des coûts indirects est en revanche spécifique. Faut-il affecter des « *pertes de production* » aux enfants, aux retraités, aux chômeurs et aux personnes effectuant un travail domestique non rémunéré ? En toute logique non. Mais du coup, le soin délivré à un actif qui est susceptible d'économiser des coûts indirects est priorisé par rapport à celui d'un inactif qui n'en économise pas ! D'où un effet discriminatoire qui heurte l'exigence éthique d'égalité dans l'accès aux soins.

Parmi les actifs, la question des coûts indirects peut également être discriminatoire si l'on considère la situation professionnelle des individus – et notamment du niveau de rémunération – ou l'état de la conjoncture économique : les « pertes de production » seraient ainsi plus faibles pour les ouvriers que pour les cadres supérieurs, plus faibles en situation de sous-emploi qu'en situation de plein-emploi, puisque le travailleur en arrêt de travail peut être plus aisément remplacé (aux coûts de transaction près !).

Il n'existe pas de solution « technique » à ces problèmes. On peut décider d'adopter le point de vue du payeur, comme le recommandent les « guidelines » de la HAS, qui exclut la valorisation des « *pertes de production* » mais, outre qu'il s'agit d'un choix politique, restrictif par rapport à la perspective sociétale, reste posée la question technique des « *indemnités journalières* » qui constituent un débours ne concernant que les salariés et donc discriminatoire à l'égard des inactifs.

On peut décider, pour éviter le problème, de ne prendre en compte que les coûts médicaux directs, en dehors des arrêts de travail indemnisés, mais on se prive des effets de la « resocialisation » des patients qui constitue une conséquence importante – voire même la plus importante – des traitements médicaux.

On peut affecter conventionnellement un coût indirect forfaitaire à tous les individus quel que soit leur statut professionnel, calculé par exemple à partir du salaire moyen ou du salaire médian, en s'appuyant sur la théorie des coûts d'opportunité, mais il s'agit d'un choix éthique (traiter ou pas tout le monde de la même manière) doublé d'une convention de calcul (valoriser le temps d'activité domestique par le salaire horaire moyen).

En absence de solution pleinement satisfaisante ce qu'on peut exiger des études c'est que ces questions et leurs conséquences soient clairement posées, surtout quand des choix méthodologiques alternatifs conduisent à des recommandations différentes, quand, par exemple, la prise en compte des coûts indirects inverse le sens de ces dernières.

Choix éthiques et analyses coût-résultat per se

Si les méthodes de mesures des résultats et des coûts peuvent entrainer des choix éthiques, le recours aux analyses coûts-résultats pour procéder à des recommandations d'utilisation et/ou de prise en charge, *constitue un choix éthique en tant que tel*, par rapport à d'autres approches, par exemple celles, longtemps majoritaires dans les milieux médicaux, consistant à ne considérer que l'effet médical des traitements indépendamment de leurs coûts, au motif que les questions économiques ne relevaient pas de la responsabilité des médecins mais de celle des pouvoirs publics. Cet argument n'est guère recevable car il constitue à l'égard de la société une injonction de payer potentiellement sans limite à l'initiative exclusive de la sphère médicale.

En fait la compatibilité de l'analyse médico-économique et de l'éthique médicale peut être défendue de plusieurs manières.

La première, peut-être la plus classique, réside dans l'argument du coût d'opportunité. Dans un contexte de ressources rares, le « vrai » coût d'un traitement n'est pas la somme d'argent qu'il est nécessaire de dépenser pour le mettre en œuvre, mais le *bénéfice perdu* du fait que cet argent n'a pas été dépensé ailleurs, pour un autre type de soins délivrés à d'autres patients. Toute dépense est un sacrifice, un renoncement à un choix alternatif. Et pour paraphraser Alan Williams, on ne voit pas sur quels fondements éthiques on serait fondé à ignorer l'effet de ses propres décisions sur les parties tierces. Non seulement il n'est pas contraire à l'éthique médicale d'intégrer

le coût des traitements dans les critères de décision, mais on pourrait soutenir qu'il ne serait pas éthique de ne pas le faire !

Un deuxième argument tient à l'équité dans *la répartition des ressources* entre patients ou entre indications thérapeutiques. Calculer le ratio coût-résultat des interventions diagnostiques ou thérapeutiques dans différentes indications contribue à révéler les possibles distorsions entre des situations ou une dépense considérable est engagée pour des résultats modestes par rapport à d'autres où c'est l'inverse. La prise en compte des ratios coûts-résultats tend à harmoniser les efforts entre les différents domaines thérapeutiques et éviter la concentration des dépenses dans certaines situations « privilégiées », pour une raison ou pour une autre. Elles constituent alors un levier d'équité dans l'affectation inter-pathologies des ressources.

Enfin, l'évaluation économique a essentiellement une *portée générale* destinée à formuler des recommandations d'usage des ressources collectives en santé à partir de mesures statistiques. Il ne s'agit nullement de décider le traitement à administrer à un patient en particulier qui relève de la responsabilité exclusive du clinicien. On peut certes objecter qu'une recommandation négative au plan collectif peut priver d'accès au traitement un patient individuel, mais c'est aussi le cas de l'évaluation clinique, quand une AMM est refusée à un médicament particulier pour manque de significativité statistique du résultat clinique[4], alors même que ce traitement peut avoir un effet bénéfique pour certains patients. Les études coût-efficacité ne se distinguent pas à cet égard des études cliniques. L'efficacité et l'efficience se démontrent statistiquement à partir de mesures collectives et ces démonstrations sont de fait opposables aux individus, pourvu qu'elles soient menées conformément aux règles de l'art, tant scientifiquement qu'éthiquement, même si la solidité méthodologique des études d'efficience est sans doute généralement inférieure à celle des études d'efficacité.

La question des seuils d'efficience

L'évaluation médico-économique est, nous le rappelons dans la préface, à la fois une discipline académique et une pratique administrative. Les études médico-économiques sont de plus en plus requises par les autorités de santé dans de nombreux pays pour l'« accès au marché » des produits de santé, pour en déterminer les conditions de prise en charge collective et/ou la tarification.

4 D'autant que la fixation d'un risque alpha à 5 % pour définir la supériorité d'un traitement par rapport à son comparateur est parfaitement arbitraire.

À nouveau, elle ne se distingue guère à cet égard de la recherche clinique : un essai de phase III d'un nouveau médicament constitue à la fois une production scientifique, susceptible de publication dans des revues à comité de lecture, et un élément d'un dossier administratif ouvrant la voie à la mise sur le marché du médicament.

À cette différence près que, contrairement à la recherche clinique, les conditions d'utilisation administrative des études médico-économiques *varient très fortement d'un pays à l'autre*, avec, très schématiquement, une opposition entre les pays où l'efficience, mesurée dans des études médico-économiques répondant à des normes méthodologiques précises, constituent le critère final d'accès au marché et ceux où elle ne constitue qu'un élément d'information porté à la connaissance des autorités publiques en tant que critère parmi d'autres d'aide à la décision[5].

Le Royaume-Uni, avec la création du NICE dès 1999, a été sans conteste le premier pays à ériger l'efficience médico-économique en critère final d'accès au marché, notamment dans le domaine du médicament. Cela est passé par une codification très stricte des méthodes d'évaluation et par la fixation d'un *seuil de référence*, exprimé en cout par QALY, pour définir l'efficience.

La codification est bien connue et peut être résumée comme suit[6] :

- Les études doivent être de type coût-utilité
- Le critère de résultat est le QALY
- Les indices d'utilités doivent être dérivées de méthodes fondées sur les choix (généralement des « time trade-off »)
- Les préférences sous-tendant les choix doivent être celles de la population générale (et non des patients ou des professionnels de la santé)
- Les QALYs ne doivent pas être pondérés
- Le calcul des coûts doit être mené dans la perspective du NHS
- Coûts et QALYs doivent être actualisés au même taux recommandé par le NICE.

Le critère d'accès au marché est alors un ratio incrémental coût-utilité fixé en 2004 entre £ 20.000 et £ 30.000/QALY. Il a été élargi entre £ 20.000 et £ 50.000/QALY à partir de 2014 avec la possibilité d'aller jusqu'à £ 100.000/QALY dans le cas des médicaments avancés (« *Highly specialized technologies* » ou HSTs dans la terminologue du NICE[7]) et des maladies ultra-rares.

---

5 Nous n'envisageons ici que le cas des systèmes publics de prise en charge. Des assureurs privés peuvent également utiliser les études médico-économiques pour constituer leur « *formulary* » de médicaments remboursables.

6 NICE : *Guide to the methods of technology appraisal*, April 2004.

7 Il n'existe pas de définition universelle des pathologies rares ou ultra rares. En Europe on se réfère au chiffre de 5 cas pour 10 000 habitants pour les pathologies rares et de 1 cas pour

Mais dans tous les cas, qu'il s'agisse des valeurs initiales ou des valeurs élargies, *aucun fondement théorique ou même explicite n'a jamais été donné à ces valeurs seuils*. Pas plus d'ailleurs que dans les autres pays ayant établi le même type de règles comme les Pays-Bas, l'Australie ou l'Écosse (qui, dans le cadre de la dévolution, a son propre système de santé).

Consciente qu'un même seuil ne pouvait s'appliquer à des pays de niveaux économiques très hétérogènes, l'OMS a suggéré un seuil d'efficience égal à 3 fois le PIB par habitant, mais sans davantage de justification[8]. Quant à la recherche académique, elle débouche sur une multiplicité de valeurs en fonction des modèles et des hypothèses utilisées par leurs auteurs[9].

Des pays comme la France avec la HAS en 2004 ou l'Allemagne avec l'IQWIG en 2003, ont également créé des agences d'évaluation des technologies médicales et leur ont également confié une mission d'évaluation économique des technologies innovantes[10]. Dans les deux cas toutefois, l'efficience qui ressort des études fournies aux agences ne constitue pas un critère exclusif d'accès au marché. Elle est considérée comme un élément d'information, utile voire indispensable, mais pas déterminant, pour la négociation ou la fixation des prix.

Conséquence : les « guidelines » méthodologiques de la HAS et de l'IQWIG sont plus souples que leurs homologues du NICE (par exemple sur la question des QALYs) et *surtout il n'existe, ni en France ni en Allemagne, de seuil d'efficience officiel*, même de type indicatif.

À partir donc des mêmes données et des mêmes études, on peut parvenir à des prises en charge différentes de patients qui se voient privés de traitements efficaces mais non efficients dans certains pays, alors que ces traitements sont accessibles dans des pays voisins, de même niveau socio-économique[11].

---

10 000 habitants pour les pathologies ultra-rares. Sur les HSTs, cf. la note du NICE : *Process for the evaluation of highly specialised technologies*, avril 2017, qui établit la possibilité d'une pondération des QALYs et fixe explicitement un seuil d'efficience à £ 100.000/QALY.

8 En termes de DALYs et non pas de QALYs. Cf. *The world health report 2002 – Reducing risks, promoting healthy life*, OMS. Cela représenterait pour la France un seuil de 94 000/QALY en 2016.

9 Cf. l'excellente revue de littérature de la HAS : *Valeurs de références pour l'évaluation économique en santé*, document de travail, Décembre 2014.

10 C'est la Loi de financement de la sécurité sociale pour 2008 qui a, en France, donné cette compétence à la HAS. Une commission spécialisée de la HAS, la Commission d'évaluation économique et de santé publique (CEESP), a été créée à cet effet.

11 On ne manque pas d'exemples, notamment en oncologie, des médicaments non pris en charge par le NHS au Royaume-Uni – car considérés comme non efficients par le NICE – pour lesquels la HAS en France a émis un avis positif pour le remboursement et l'usage hospitalier.

Ce qui ne peut manquer de soulever des interrogations quant à la portée pratique de l'évaluation économique. En tant que discipline académique, elle obéit à des règles méthodologiques universelles mais en tant que pratique administrative elle subit les contraintes des contextes institutionnels de prise de décision qui varient de pays à pays.

Médico-économie et valeurs sociétales

Ces différences ne manquent pas d'être troublantes et elles doivent être mises en relation avec les valeurs qui fondent les différents systèmes de protection sociale avec notamment la fameuse opposition entre systèmes « *beveridgiens* » et systèmes « *bismarckiens* »[12].

Le NHS britannique, créé en 1948, repose sur les principes formulés dès 1942 par Lord Beveridge dans son fameux rapport au parlement[13] où il préconisait la création d'un État Providence (« *Welfare State* ») destiné à mettre la population à l'abri des cinq grands maux sociaux qui avait frappé les démocraties occidentales au début du XXe siècle – la pauvreté, l'insalubrité, la maladie, l'ignorance et le chômage – et dans lesquels il voyait l'origine des dictatures et de la guerre.

Ces principes ont conduit à la création d'un système national de santé distribuant à toute la population des soins gratuits (« *au point de délivrance* ») financés par la taxation générale et géré directement par l'État. La perspective était celle d'une médecine sociale et frugale, mettant l'accent sur les soins primaires, avec des restrictions touchant les patients (astreints de s'inscrire chez un généraliste sana accès direct aux spécialistes) et des médecins qui sont des contractuels du NHS, exerçant en groupe et payés à la capitation. Destiné à couvrir gratuitement les soins primaires de l'ensemble de la population, le système était rationnant par nature, notamment pour les soins de spécialité non urgents délivrés exclusivement à l'hôpital, le plus souvent avec liste d'attente.

Dans un tel système, l'organisation et la planification des soins relèvent des missions de l'État et l'évaluation économique est apparue dans les années 90 comme un outil de priorisation de choix de santé qui venait théoriser et rationaliser, notamment à travers la notion de coût d'opportunité,

12 Le Pen C. Is there a « continental » view of health economics evaluation? *Eur J Health Econ* 2009 ; 10(2) : 121-3.
13 Beveridge W. *Report to the Parliament on Social Insurance and Allied Services, 1942*.

des pratiques de rationnement qui préexistaient en cohérence avec la logique globale du système[14].

Les ordonnances de 1945 créant la sécurité sociale en France se sont inspirées de principes très différents, issus de la doctrine sociale allemande élaborée par le chancelier Bismarck en 1883. La logique est celle d'un système assurantiel – plutôt qu'un système d'assistance – essentiellement destiné aux salariés dans les entreprises, financés par des cotisations sociales salariales et patronales, géré paritairement par les partenaires sociaux au sein de caisses autonomes vis-à-vis d'un État, qui, conçu comme un état garant plutôt qu'un état gérant, délègue aux partenaires sociaux la responsabilité des différentes branches de la protection sociale.

Contrairement au NHS, l'organisation des soins n'a pas été affectée par la création de la sécurité sociale : la médecine est demeurée libérale avec libre choix du praticien – sans distinction officielle entre soins primaires et soins secondaires[15] – et les médecins sont restés payés à l'acte avec des tarifs conventionnels nationaux qui peineront à voir le jour[16].

Contrairement au NHS, la logique assurantielle et le financement contributif ont créé un véritable *droit aux soins* sous la seule condition d'efficacité. Dans un tel système refuser un accès à des soins efficaces pour des raisons économiques, ou sous prétexte qu'ils ne sont pas « efficients », constitueraient une « perte de chance » et une violation du contrat social formalisée par la fameuse devise, systématiquement rappelée dans les débats parlementaires : « *de chacun selon ses moyens, à chacun selon ses besoins* ». Et ce n'est pas à l'État de définir les besoins…[17] L'évaluation économique ne peut alors être qu'un élément d'information, aidant par exemple à la tarification des biens de santé, mais sans pouvoir prétendre à être un critère final d'accès au patient[18].

---

14 Les médecins et médias français n'étaient pas les derniers à s'indigner d'un système où ne délivraient pas des soins lourds à des patients présentant des facteurs de risques trop importants, par exemple le tabagisme.

15 Il a fallu attendre la loi du 13 Août 2004 pour introduire la notion de « médecin traitant » et la loi HPST du 18 Juillet 2009 pour introduire celle de médecin de « premier recours ».

16 Les tarifs nationaux ne deviennent véritablement opposables qu'en 1971 avec la signature de la première convention nationale entre les syndicats médicaux et les casses d'assurance-maladie. Une exception nait toutefois dès 1980 avec la création du secteur II conventionnel à honoraires libres !

17 Depuis les années 90 un processus d'étatisation progressif du système de santé est en cours et il n'est pas surprenant que l'émergence en France du thème de l'évaluation économique soit contemporaine de cette tendance.

18 Pour autant bien sûr que les réformes en cours ne modifient pas les termes du contrat social de 1945.

L'Évaluation économique est-elle utilitariste ?

Cette affinité particulière de l'évaluation économique avec les systèmes beveridgiens ont conduit à s'interroger, au-delà des aspects institutionnels, sur les rapports entre l'évaluation économique et les philosophies sociales qui inspirent à leur tour ces systèmes.

L'évaluation économique est à ce titre souvent décrite comme ayant ses racines dans l'« utilitarisme », cette philosophie morale et politique – souvent caricaturée – qui prescrit qu'une bonne politique publique est celle qui augmente le bonheur collectif, conçu comme l'agrégation des bonheurs individuels de tous les être sensibles, et qui, selon la fameuse formule de Jeremy Bentham (1748-1832), vise au « *plus grand bonheur du plus grand nombre* » [19].

Sous la forme du « welfarisme » à la Beveridge, l'utilitarisme a certainement imprégné les politiques sociales de la seconde moitié du XXe siècle, notamment dans les pays anglo-saxons, et comme le dit John Rawls, l'adversaire le plus connu et le plus résolu de l'utilitarisme, celui-ci constitue le programme implicite de tous les états providence.

Et il est exact que l'affectation de ressources rares à des programmes de santé alternatifs, priorisés en fonction de leur efficience (donc de leur rapport coût-efficacité), a comme objectif explicite de maximiser le nombre agrégé de QALYs qu'autorise la contrainte budgétaire. On démontre aisément qu'aucune autre méthode d'affectation ne permet d'obtenir un plus grand nombre de QALYs agrégés, considérés comme une mesure du « bonheur » dans un sens utilitariste.

Si l'utilitarisme, sous la forme plus politique du « welfarisme » c'est-à-dire de politiques visant à élever systématiquement le bien-être des populations, a été dominante dans la deuxième moitié du XXe siècle, elle a été contestée par tout un courant de pensée, dominé notamment par John Rawls dans son monumental ouvrage de 1971 « A Theory of Justice »[20].

Que reproche Rawls à l'utilitarisme ? Sans entrer dans un débat qui dépasse à la fois nos compétences et le cadre de cet ouvrage, essentiellement de méconnaitre l'infinie pluralité des bonheurs – ce qui suffit à ses yeux pour discréditer la recherche d'un bonheur collectif unique comme action politique –, de concevoir les individus comme interchangeables sans les

---

19 La notion de bonheur dans la philosophie utilitariste ne relève pas d'un simple hédonisme sensuel. Pour de nombreux utilitaristes après Bentham, la qualité du « bonheur » compte autant sinon plus que la quantité. On connait la célèbre formule de John Stuart Mill : « *Il vaut mieux être un homme insatisfait qu'un porc satisfait. Il vaut mieux être Socrate insatisfait qu'un imbécile satisfait* ». Stuart Mill J, *L'Utilitarisme*, Flammarion, Col. Champs classiques, 1988 (p. 54).

20 Rawls J. *Théorie de la justice*, trad. Catherine Audard, Le Seuil, 1987.

considérer en tant que personnes libres dotées de droits fondamentaux, de sacrifier les droits de quelques-uns au « bonheur » des autres, de mépriser les droits de l'homme et les autres valeurs fondamentales que Bentham qualifie de « *non-sens sur des échasses* », de confondre le juste et le bien, etc.

Comment un utilitariste pur et dur peut-il être certain que les sommes économisées en ne traitant pas à grand frais un patient en phase terminale sont effectivement affectées à sauver un plus grand nombre de vies ? Comment est-il sûr qu'un desserrement de la contrainte budgétaire, permettant de traiter l'un et les autres, n'augmenterait pas le bien-être collectif ? Est-il assuré que l'ensemble des choix collectifs faits dans d'autres domaines, comme l'éducation, la police, la justice, la défense, etc., sont efficients ? Pourquoi demande-t-il en quelque sorte aux malades de financer d'autres malades ?

C'est cette dimension sacrificielle, cette arithmétique des vies propre à l'utilitarisme, qui heurte ceux pour qui l'acte de soigner et l'attention portée au malade sont plus importants que la mesure du résultat clinique. La morale rawlsienne, couramment qualifiée de néo-kantienne, se fonde sur des devoirs et des obligations à respecter, plutôt que sur des objectifs à atteindre. Et la distribution des ressources rares entre individus et pathologies – qui est inévitable car la morale n'abolit pas la rareté – doit obéir à des impératifs de justice et d'équité plus que d'efficience.

Ces principes, dans le détail desquels nous ne rentrerons pas, peuvent être très schématiquement résumés sous la forme d'un contrat social stipulant de manière hiérarchique :

1) « *Un égal accès à un système pleinement adéquat de libertés de base* ». Cet égalité d'accès à ce qu'il appelle les « biens premiers » est fondamental dans la pensée de Rawls.

2) Des inégalités sociales et économiques qui doivent être (i) au plus grand bénéfice des membres les plus désavantagés de la société ; (ii) attachées à des fonctions et à des positions ouvertes à tous, conformément au principe de l'égalité équitable des chances[21].

En mettant l'accent sur les valeurs et les droits fondamentaux plutôt que sur les résultats, sur l'individu plutôt que sur le groupe, sur l'équité plutôt que sur l'égalité, Rawls se sépare radicalement de l'utilitarisme.

Pour revenir à la question qui titre ce paragraphe, on nous permettra de citer intégralement la conclusion d'Allan Williams à son article de 1996

21 Pour un exposé rapide et pédagogique de la doctrine notamment de la notion de « biens premiers », cf. Rawl J : Unité sociale et bien premiers, *Raisons Politiques*, Presses de Sciences Po, 2009 ; 33 : 9-43.

dans « *Social science and Medicine* »[22], qui nous semble fournir une réponse raisonnable en faisant la différence entre l'outil et son usage :

« *Quoi que l'approche des QALYs pour la détermination des priorités en santé soit « conséquentialiste » (en ce sens qu'elle met l'accent sur l'importance de mesurer l'impact des soins médicaux sur la santé des personnes), elle n'est pas nécessairement « utilitariste » (en ce sens qu'elle n'exige pas que le but se résume à la maximisation des gains en santé). Les QALYs* [et plus généralement l'évaluation économique[23]] *ont un rôle important à jouer dans n'importe quel système de fixation de priorité pourvu que l'impact des soins sur la santé des personnes soit pris en considération* ».

Ainsi ce n'est ni l'évaluation économique, ni son outil emblématique le QALY qui sont par eux-mêmes utilitaristes, mais l'usage que pourrait en faire des pouvoirs publics qui planifieraient la production et la distribution des soins sous l'angle exclusif de l'efficience économique, oubliant le droit des personnes, les règles déontologiques, et la nature du contrat social qui, en France comme finalement dans une majorité de pays, fonde le rapport entre la nation et son système de protection sociale.

22 Williams A. QALYs and Ehtics: A heath economists's perspective. *Social Science and Medcine* 1996 ; 43(12) : 1795-804.
23 Ajouté par nous.

L'évaluation médico-économique s'est imposée ces dernières années, en France comme dans beaucoup d'autres pays, comme une dimension incontournable des politiques de santé. Il n'est guère de recommandations portant sur tel ou tel bien de santé ou, plus généralement, sur telle ou telle stratégie médicale à visée préventive ou thérapeutique, qui n'envisage sa dimension médico-économique, à des fins de décision ou d'information (...).

L'ajout de la dimension économique à l'évaluation médicale n'est guère fortuit. Il correspond à la nécessité de faire face à l'inéluctable croissance des dépenses de santé qui, dans beaucoup de pays développés, absorbent environ 10 % de leur richesse nationale, voire davantage. Rechercher des « gains d'efficience », c'est-à-dire trouver les stratégies et techniques médicales les plus économiques pour atteindre un objectif sanitaire et social donné, est devenu un impératif quasi universel. Dans les pays où existent des budgets fermés – ou quasi fermés – des dépenses de santé, la recherche de l'efficience permet d'optimiser l'usage de ces ressources et d'obtenir, sous cette contrainte, les meilleurs résultats collectifs possibles en termes de santé publique.

Au-delà des aspects de politique de santé, l'émergence et le succès des approches médico-économiques correspondent à une vraie révolution conceptuelle, philosophique, éthique dans un monde de la santé qui n'a jamais fait bon ménage avec l'argent et le marché et où le terme de « médico-économique » est presque un oxymore.

Or, cette nouvelle discipline académique qu'est devenue l'analyse médico-économique, avec ses méthodes, ses références, ses congrès, ses revues scientifiques et ses enseignements, reste relativement mal connue des professionnels de santé en dehors des cercles spécialisés, au demeurant assez restreints.

L'objectif poursuivi dans cet ouvrage est de donner au non-spécialiste – professionnel de la santé ou non, économiste ou non – un aperçu des concepts et méthodes qu'elle utilise.