

L'évaluation des technologies de santé : les protocoles de deuxième génération

R. LAUNOIS¹, S. PAYET¹, L. RIOU-FRANÇA¹, K. LE LAY¹, C. FRANCESCONI¹,
M. DEVAUX¹, A. QUÉMÉNER¹, E. LOOTVOET¹

¹REES France, 28, rue d'Assas, 75006 Paris, France

RÉSUMÉ

On oppose souvent la rigueur et la bonne « validité interne » des essais randomisés au pragmatisme et à la transposabilité des résultats obtenus en pratique réelle dans le cadre d'études observationnelles. A priori il est indispensable de s'entendre sur le contenu et la portée des termes utilisés. Si les essais permettent d'estimer la quantité d'effet produite par une action de santé au regard d'une autre, il est fondamental de replacer les résultats qu'ils fournissent, par rapport à l'ensemble des programmes qui pourraient être mis en place pour atteindre un objectif de santé publique déterminé. Instruments incontournables de la connaissance scientifique, ils ne peuvent à eux seul éclairer la décision politique et ne sauraient être le vecteur de l'évaluation des technologies de santé. Les bases médico-administratives et les études observationnelles apportent des informations différentes. Une enquête est dite d'observationnelle quand on peut affirmer que tout se serait passé de la même façon si l'étude n'avait pas eu lieu. Le véritable enjeu est alors de veiller à ce que ces études portent sur des populations comparables et lorsqu'elles ne le sont pas, de spécifier les techniques d'ajustement qui doivent être utilisées. Deux types d'outils semblent aujourd'hui à même de pallier les éventuelles insuffisances des schémas précédents : l'analyse de décision et les études quasi expérimentales. Cet article se propose d'étudier ces outils de deuxième génération et de les codifier pour faciliter leur mise en œuvre dans des domaines où il existe de véritables enjeux de santé publique : prévention et dépistage du cancer, maladie d'Alzheimer, troubles de l'obésité etc.

Mots-clés : Synthèse quantitative, Mélange de comparaison de traitement, Analyse de Monte Carlo, Bootstrap, Courbe d'acceptabilité.

SUMMARY

Title

Health technology assessments propose to study the differential impact of health interventions in a complex care system which is characterised by the multitude of individual behaviours and the diverse nature of the institutions involved. Current systems for data collection lend themselves poorly to this rigorous analysis of efficacy of treatments in the actual situations where they are used. Randomised trials endeavour to neutralise any parasitic interference which could compromise testing for a causal relationship between the treatment administered and the result obtained. Their methodology which establishes the term *ceteris paribus* in the principle of good practice lends itself poorly to an analysis of individual behaviour. Observational studies are start from actual treatment situations to describe them as reliably as possible. By definition, however, these assume that the natural course of events is not deviated by any intervention. The absence of an experimental plan increases the likelihood of bias and makes it more difficult to test for causal relationships. They lend themselves poorly to testing for actual effectiveness. The two instruments to be recommended are : decision analysis and quasi-experimental studies. Decisional analysis help to avoid the problems of external validity associated with randomised clinical trials by associating parameters which are extracted from data obtained from everyday practice. Quasi-experimental studies are based on the reality of behaviour of the prescriber and his/her patients; their impact on efficacy, quality of life social costs of the disease and of treatments may be identified under normal conditions of use.

Key-words: Quantitative synthesis, Mix treatment comparisons, Monte Carlo analysis, Bootstrap, Acceptability curve.

1. INTRODUCTION

Pour évaluer les conséquences d'une action de santé publique, différents designs d'étude sont possibles. Les essais randomisés, souvent considérés comme le Gold Standard, s'efforcent de neutraliser toute interférence parasitaire qui pourrait compromettre la recherche d'un lien de causalité entre la stratégie étudiée et le résultat obtenu. Toutefois, leur méthodologie qui érige la clause « *ceteris paribus* » en principe de bonnes pratiques se prête mal à une analyse des comportements et des structures, et les effets des comportements sur l'efficacité et l'adhésion aux programmes de prévention ne sont pas intégrés. A ce type d'études, s'opposent les enquêtes dites observationnelles.

Elles partent des réalités de terrain pour les décrire de la manière la plus fidèle possible, et supposent par définition que le cours naturel des choses ne soit infléchi par aucune intervention. L'inconvénient de ce design est que l'absence de plan expérimental augmente la possibilité de biais, et rend plus difficile la recherche des causalités. Une étude observationnelle, si elle permet de rendre compte des pratiques médicales « dans la vraie vie », reste tout de même sujette aux biais de recrutement. L'attribution d'un traitement plutôt qu'un autre ne se fait pas aléatoirement mais sur la conviction d'indication du prescripteur, ceci ne garantit pas a priori que deux groupes de traitement à comparer partagent les mêmes caractéristiques initiales.

Comment peut-on alors combler l'écart entre l'expérimentation et la vie réelle ? Comment établir un lien entre le contexte artificiel des essais cliniques et la réalité d'une prise en charge au jour le jour ? Aujourd'hui, des techniques spécialisées, de plus en plus précises et sophistiquées offrent les réponses attendues. Ces techniques permettent (i) de synthétiser au mieux toutes les données disponibles sur l'efficacité relative des programmes ou des traitements quelle que soit la modalité choisie pour exprimer le critère de jugement : différence absolue de risque, risque relatif, ou rapport de cotes. (ii) d'estimer les coûts à partir de données observationnelles prospectives ou rétrospectives issues de populations rendues comparables par l'utilisation des méthodes appropriées (iii) d'intégrer les données d'efficacité et de coût associées aux différentes stratégies dans le cadre d'une analyse décisionnelle renouée d'inspiration bayésienne. Cet article se propose d'étudier ces outils de deuxième génération et de les codifier pour faciliter leur mise en œuvre dans des domaines où il existe de véritables enjeux de santé publique.

2. LA SYNTHÈSE DE LA LITTÉRATURE EXISTANTE

Pour choisir les valeurs centrales d'une analyse décisionnelle quelle qu'en soit la forme : arbre de décision ou modèle de markov [1, 2, 3], l'analyste est amené à faire des choix entre des essais multiples dont les designs, les populations, les critères de jugement et les comparateurs sont bien souvent différents. Jusqu'à présent la synthèse des informations pertinentes relevait d'une « analyse statistique raisonnée » dont la démarche n'était pas sans rappeler le rôle et l'importance du jugement clinique dans la décision médicale. L'analyste cherchait à réaliser les combinaisons optimales à partir des données médicales disponibles pour prévoir les trajectoires de vie concevables et chiffrer leurs conséquences humaines, sociales et

financières. Les modèles s'appuyaient sur une analyse qualitative de la littérature et sur le rapprochement des bras extraits des différents essais qui étaient disponibles, ce qui rompait la randomisation. Dans ce contexte, les estimations fournies par un modèle sur l'efficacité, les coûts et les conséquences de la pathologie étaient plus ou moins entachés d'arbitraire, puisque leurs conclusions dépendaient des hypothèses adoptées. Ainsi, une synthèse purement qualitative de la littérature n'est pas suffisante pour avoir une idée précise de l'efficacité d'une stratégie (programme de santé, traitement, survenue d'effets indésirables ou de complications, etc.). Aujourd'hui, de nouvelles techniques statistiques sont disponibles pour contourner ces biais, synthétiser au mieux l'information. Les modèles de mélange de comparaisons de traitements (« Mixed Treatment Comparisons », MTC), les méta-analyses, les méta-régressions sont autant d'instruments mis à la disposition des spécialistes de santé publique.

2.1. Les mélanges de comparaisons de traitements

L'objectif des modèles de décision est de comparer l'efficacité et les coûts des différents programmes de santé envisageables. Or, il est quasiment impossible de disposer d'essais qui les compareraient tous en face à face. Des techniques de comparaison indirectes doivent donc être mises en œuvre. Autrement dit, les comparaisons indirectes s'avèrent utiles lorsque les stratégies auxquelles on s'intéresse, n'ont jamais été comparées de façon directe dans le cadre d'un essai randomisé, mais qu'il existe d'autres stratégies (ou traitements) au regard desquelles ils ont fait l'objet d'une confrontation en face à face. Le modèle de mélange de comparaisons de traitements [4, 5, 6] permet ainsi de mobiliser l'ensemble des essais disponibles, qu'ils permettent de comparer les traitements directement (en face à face) ou indirectement (par transitivité). En fait, le rôle des MTC est double. Le premier consiste à ren-

forcer l'inférence concernant l'efficacité relative de deux traitements en incluant des comparaisons directes et indirectes. Le second est de faciliter l'inférence simultanée de tous les traitements, afin de déterminer lequel est le meilleur. Les MTC peuvent être opérationnalisés grâce à la mise en œuvre de modèles hiérarchiques bayésiens.

Pour mieux fixer les idées, prenons un exemple. Supposons que les essais 1 et 2 comparent les traitements A et B et l'essai 3 compare les traitements B, C et D. Alors, par transitivité, nous pouvons obtenir une comparaison indirecte des traitements A et D, comme le montre la figure 1.

2.2. Méta-analyses

Dans un contexte un peu différent, une autre façon de synthétiser les données de la littérature est la méta-analyse. Elle permet d'estimer de manière fine un paramètre, à partir de plusieurs études, au lieu d'une seule. En outre, le caractère généralisable des résultats issus d'une méta-analyse est supérieur à celui des résultats issus d'une seule étude, puisque cette technique mixte des études réalisées sur des populations différentes et avec des protocoles différents. Le logiciel WinBUGS[®] est très utilisé pour réaliser les méta-analyses. Sa méthode de calcul est fondée sur la technique de Monte Carlo par chaînes de Markov.

Plusieurs paradigmes [7] (fréquentiste ou bayésien) et divers modèles (à effets fixes ou à effets aléatoires) peuvent être envisagés pour une méta-analyse. Les différentes méthodologies sont présentées ci-dessous.

2.2.1. Modèle à effets fixes versus modèle à effets aléatoires

Supposons que nous disposions de k études mesurant un paramètre d'intérêt, l'effet d'un traitement par exemple. Dans un modèle à effets fixes [8], nous supposons que chaque étude estime le même paramètre sous-jacent. Autrement dit, la variabilité observée entre les effets estimés sur chaque étude ne provient que de la précision avec laquelle les effets sont mesurés. Dans ce cas, l'effet global (ou « poolé ») du traitement correspond à la moyenne pondérée des effets estimés dans chaque étude, les poids associés étant égaux à la précision avec laquelle les effets sont mesurés. Si l'on note respectivement E_i et σ_i^2 l'estimateur de l'effet du traitement et sa variance dans l'étude i et θ l'estimateur « poolé » de l'effet du traitement, et si on fait l'hypothèse que l'effet du traitement suit une loi normale d'espérance θ et de variance σ_i^2 : nous pouvons représenter un modèle à effets fixes comme suit,

$$E_i \sim N(\theta, \sigma_i^2), i = 1, \dots, k$$

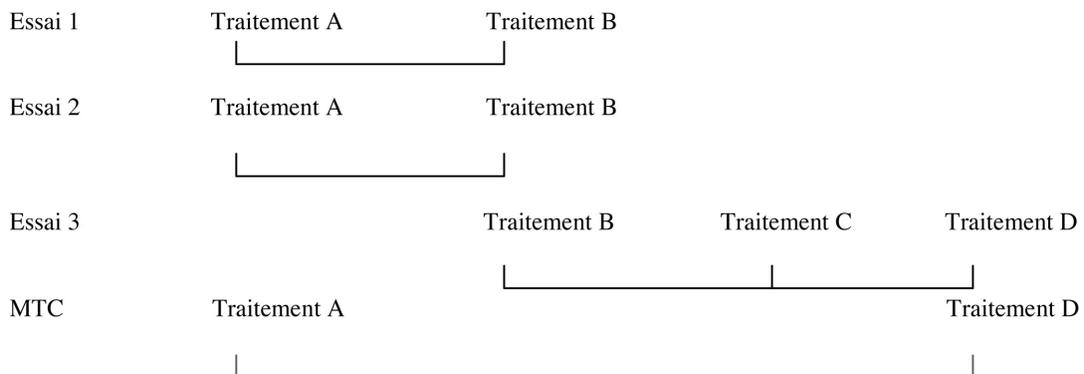


Figure 1 : Exemple de Mixed Treatment Comparisons.

De nombreux auteurs ont montré que l'hypothèse soutenant les modèles à effets fixes n'est pas réaliste. En effet, les études portant sur le même sujet ne sont pas des répliques les unes des autres, elles ne suivent pas forcément le même schéma d'étude et sont conduites différemment. Cela a donc des conséquences sur la mesure du critère d'efficacité du traitement.

Les modèles à effets aléatoires permettent de tenir compte de cette variabilité inter-études. Dans ces modèles, on abandonne l'hypothèse selon laquelle chaque étude estime exactement le même effet. En revanche, on suppose que chaque étude estime un effet différent, mais que tous ces effets sont issus d'une même distribution d'effets. Il est commode de représenter les incertitudes portant sur la nature incertaine des effets par une distribution de probabilité; il suffit alors d'estimer les valeurs d'un petit nombre de paramètres pour définir toute la distribution de probabilité. Si l'on fait l'hypothèse que les effets issus de chaque étude (θ_i) sont normalement distribués on, note $\theta_i \sim N(\mu, \tau^2)$, le vecteur paramètre à spécifier où μ désigne l'effet « poolé » du traitement, et τ^2 sa variance théorique. Ce qui revient à écrire le modèle de la manière suivante :

$$E_i \sim N(\theta_i, \sigma_i^2) \text{ et } \theta_i \sim N(\mu, \tau^2), i = 1, \dots, k$$

L'effet « poolé » issu d'un modèle à effets aléatoires est en général assez proche de celui issu d'un modèle à effets fixes. Le principal changement entre les deux modèles se situe au niveau de la précision de l'estimateur « poolé » : la prise en compte de la variabilité inter-études produit des intervalles de confiances plus larges dans les modèles à effets aléatoires. Ce dernier type de modèle est donc à privilégier.

2.2.2. *Modèle fréquentiste versus modèle bayésien*

La description ci-dessus des modèles à effets fixes et à effets aléatoires se place dans un contexte classique, fréquentiste, c'est-à-dire que les paramètres sont supposés fixes et inconnus (θ et μ par exemple). Dans un paradigme

bayésien, au contraire, les paramètres sont supposés aléatoires, ce qui permet d'associer une distribution de probabilité sur chacun d'entre eux. Alors que l'école classique fonde le jugement sur une succession d'échantillon virtuels dont la loi statistique est conditionnelle a des paramètres fixes mais inconnus, l'école bayésienne considère que les paramètres sont des variables aléatoires. Il y a donc un sens à considérer leur distribution probabiliste et à les revoir à la lumière des nouvelles informations. Cette propriété ouvre des possibilités supplémentaires en terme de modélisation. Il devient possible de spécifier des distributions a priori pour chaque paramètre, fondées sur des connaissances antérieures. En effet, avant de conduire une étude, un chercheur peut avoir formé une intuition a priori concernant l'option thérapeutique étudiée. Cette intuition peut être fondée sur sa propre expérience ou sur des études antérieures menées sur le même thème. Les modèles bayésiens permettent de combiner ces intuitions ou connaissances a priori avec les données issues de l'étude menée (grâce au théorème de Bayes). Les résultats reflètent alors l'actualisation de l'état des connaissances sur l'effet de la stratégie considérée.

Si l'on reprend les notations précédentes du modèles à effets aléatoires, le modèle bayésien [9] est représenté ainsi :

$$E_i \sim N(\theta_i, \sigma_i^2) \text{ et } \theta_i \sim N(\mu, \tau^2), i = 1, \dots, k$$

$$\text{avec } \mu \sim [-, -] \text{ et } \tau^2 \sim [-, -]$$

Nous retrouvons exactement le même modèle que précédemment, à cette différence près que l'analyste transforme ses connaissances *a priori* sous la forme de deux paramètres μ et τ^2 indépendants des données observables. Ce sont des opinions qui sont formulées avant toute considération des résultats des plans d'expérience mis en place

Plusieurs types de lois a priori peuvent être utilisées. Lorsque aucune intuition a priori sur les paramètres n'est possible, des lois a priori dites non informatives sont employées.

Lorsque le critère de jugement est un pourcentage, la modélisation est simple. Par exemple, notons respectivement r_{1i} et r_{2i} le nombre de patients ayant une toxicité dans les bras avec le traitement 1 et le traitement 2 de l'étude i et n_{1i} et n_{2i} le nombre total de patients dans chaque bras pour cette même étude i . Alors, nous pouvons supposer que r_{1i} et r_{2i} suivent respectivement une loi binomiale de paramètres (p_{1i}, n_{1i}) et (p_{2i}, n_{2i}) où p_{1i} et p_{2i} représentent les pourcentages recherchés. A partir de la fonction logistique de p_{1i} et p_{2i} , il est ensuite possible de calculer un rapport de cotes (odds ratio). C'est ce nouveau critère qui sera « poolé », car plus facilement modélisable.

2.3. Prise en compte de facteurs explicatifs ou méta-régression

Le modèle à effets aléatoires que nous avons décrit précédemment permet de tenir compte de la variabilité inter-études sans expliquer l'origine de cette variabilité. Cela devient possible grâce à la méta-régression [10]. Cette technique possède les mêmes propriétés générales qu'une régression linéaire simple : elle vise à identifier des associations entre des caractéristiques liées à l'étude ou aux individus et le paramètre estimé, ce qui n'aurait pas été possible au sein d'une seule étude.

Pour rendre cette recherche d'association possible, il faut au préalable extraire les caractéristiques potentiellement liées au paramètre des publications retenues pour la méta-analyse. Dans le cas de l'étude d'un traitement par exemple, la durée du suivi et la dose initiale de traitement sont des paramètres qui peuvent influencer les résultats. En effet, plus la durée de suivi est longue, plus le pourcentage de patients répondeurs au traitement est susceptible d'être important. De la même manière, selon la dose initiale du traitement, la réponse est potentiellement modifiée.

Notons par exemple la durée de suivi de l'étude i par x_i . En reprenant les notations pré-

cédentes, le modèle de méta-régression s'écrit comme suit :

$$E_i \sim N(\delta_i, \sigma_i^2) \text{ avec } \delta_i = \theta_i + \beta x_i \\ \text{et } \theta_i \sim N(\mu, \tau^2), i = 1, \dots, k \\ \mu \sim [-, -], \tau^2 \sim [-, -] \text{ et } \beta \sim [-, -]$$

β est le coefficient de régression permettant d'évaluer l'impact de la durée de suivi sur l'effet du traitement. Une distribution *a priori* doit être spécifiée pour ce coefficient, en plus des distributions sur μ et τ^2 déjà nécessaires pour un modèle à effets aléatoires bayésien.

3. INCERTITUDE ET AIDE À LA DECISION

Les modèles de prise de décision permettent de rassembler à la fois les données issues des essais cliniques randomisés et celles relevées en situation réelle d'usage pour l'ensemble des actions de santé publique qui peuvent être envisagées dans un domaine d'intervention spécifique. L'efficacité relative des diverses actions de santé est étudiée dans un cadre analytique structuré et leurs répercussions analysées tant sur le plan clinique, que sur le plan humain (la qualité de vie) et budgétaire (les coûts).

3.1. Une analyse décisionnelle renouvelée

Plusieurs types de modélisation sont envisageables. Tout d'abord, la modélisation peut, soit intégrer les résultats des méta-analyses dans un modèle de type Monte Carlo réalisé sous TreeAge®, soit être réalisée à la suite de la méta-analyse, dans un contexte entièrement bayésien sous WinBUGS®. La première approche, dite à deux niveaux, est la plus courante. Elle consiste à intégrer les résultats de la méta-analyse dans un modèle de décision : il y a donc deux étapes, réaliser la méta-analyse et construire un modèle de décision. La seconde repose sur l'emploi de modèles de décision complets intégrant à la fois le traitement statistique des données, l'analyse des

actions possibles et la quantification de leurs conséquences réalisés au sein d'un seul et même modèle bayésien [11, 12, 13, 14].

Quatre étapes forment ce modèle [15]:

- ❑ La réalisation d'une méta-analyse des données pertinentes: cette partie est identique à celle déjà décrite précédemment.
- ❑ L'estimation de toutes les données d'entrée pour le modèle (efficacité, probabilités de transition, coûts): cette étape permet de mettre en relation les variables issues de la méta-analyse pour obtenir l'espérance de coût, l'espérance d'efficacité, les probabilités de transitions...
- ❑ La réalisation d'une analyse de sensibilité par rapport aux données et aux spécifications du modèle: l'analyse de sensibilité est double, au niveau de l'incertitude sur l'estimation des variables (comme dans un modèle à deux niveaux) et au niveau des hypothèses faites dans le modèle. Il est par exemple possible d'évaluer l'impact d'un changement de distribution *a priori* sur les résultats pour évaluer la robustesse du modèle à ce type d'hypothèse.
- ❑ L'évaluation du modèle développé: pour un modèle de décision, l'espérance de coût et l'espérance d'efficacité sont calculées pour chaque branche en multipliant le coût et l'efficacité de la branche par la probabilité correspondante. Pour un modèle de Markov, à chaque cycle, les valeurs associées à chaque état sont multipliées par la portion de cohorte présente dans cet état, puis sommées sur l'ensemble des états.

Comparée à l'approche à deux niveaux, ce type de modèle possède multiples avantages. Le fait de n'avoir qu'un seul modèle facilite la prise en compte de toutes les sources de preuves disponibles (essais cliniques randomisés, études observationnelles, opinions d'ex-

perts...), fournit un cadre systématique pour mettre en relation l'incertitude liée aux paramètres du modèle (exprimée au travers de distribution de probabilité) avec l'incertitude liée à aux actions possibles et à leur conséquences potentielles; intègre les inter-relation entre les paramètres du modèle; et facilite les analyses de sensibilité, puisque tout changement, n'importe où dans le modèle, se propage automatiquement aux résultats.

En pratique, les estimations obtenues par les deux approches sont similaires. Toutefois, la différence se situe essentiellement sur la précision de ces dernières: l'approche à deux niveaux a tendance à produire des intervalles de confiance plus serrés, car toute l'incertitude n'a pas pu être prise en compte.

3.2. Analyse de sensibilité paramétrique

Une fois le modèle construit, quelle que soit la méthodologie adoptée, il convient ensuite d'identifier les paramètres qui ont le plus fort impact sur les estimations, afin de tester la solidité des conclusions obtenues sous les hypothèses centrales. Une analyse de sensibilité peut être conduite de plusieurs manières [16].

La méthode la plus simple était dans le passé l'analyse de sensibilité univariée, où l'on modifiait une par une les valeurs de chaque variable en gardant la valeur des autres fixes. Cette méthode permettait d'identifier les variables clés dont le choix des valeurs était susceptible d'inverser les conclusions de la modélisation. L'analyse de sensibilité de type seuil consistait à faire varier la valeur d'une variable jusqu'à atteindre le point où l'alternative choisie n'apparaît plus comme préférable aux autres alternatives évaluées. Elle était particulièrement utile lorsque l'on cherchait à définir des groupes cibles d'un traitement à partir des valeurs d'une certaine variable. Les analyses de sensibilité multidimensionnels consistaient à faire varier simultanément les valeurs de plusieurs variables. Les valeurs

attendues étaient alors calculées pour chaque combinaison de valeurs des variables sujettes à l'analyse. Aujourd'hui, ces analyses de sensibilité, dites ponctuelles, sont périmées [17].

Seules les analyses probabilistes sont acceptées par les comités de lecture des revues internationales. On peut entre autres utiliser des techniques de Monte Carlo pour construire n échantillons de bootstrap et déterminer expérimentalement la distribution recherchée. Il existe trois types de simulation de Monte Carlo [18].

- La simulation de Monte Carlo de premier ordre consiste à tirer au hasard le devenir d'un patient. Le patient va se déplacer dans l'arbre, en fonction de la probabilité de survenue de chaque événement, jusqu'à atteindre un nœud terminal. En tirant au hasard le devenir d'un grand nombre de patients, les valeurs attendues des coûts et de l'efficacité auront la même moyenne que celles obtenues en reliant l'arbre. Dans une simulation de premier ordre, il n'existe aucune incertitude sur les valeurs choisies pour caractériser les paramètres de l'arbre. Les résultats permettent juste de déterminer la probabilité de survenue de chacun des événements décrits dans l'arbre et d'obtenir un écart-type pour les conséquences étudiées.
- La simulation de Monte Carlo de second ordre consiste à attribuer à chacune des variables de l'arbre une distribution bien définie. Modèle paramétrique veut dire ici que la distribution de la variable aléatoire obéit à une forme régie par des paramètres. Plusieurs types de lois existent pour rendre compte des variables [19, 20].
 - Les distributions de Poisson : elles permettent de décrire des événements discrets, en particulier les événements rares (probabilité de survenue proche

de zéro). La spécification de la distribution ne dépend que de la moyenne du nombre d'événements observés.

- Les distributions normales : leur choix est plus judicieux dans le cas où les événements sont fréquents.
- Les distributions Bêta : ce sont des distributions bornées et non symétriques, utiles dans la caractérisation des proportions.

Une fois que chaque variable est caractérisée par une distribution, on simule n réalisations de chacune des lois par Monte Carlo et on calcule les espérances d'efficacité et de coût de chaque stratégie à partir de ces réalisations. On obtient ainsi un échantillon de n valeurs d'efficacité et de coût qui est accessible à tous les traitements statistiques connus. Dans une simulation de second ordre, l'incertitude sur les variables est donc prise en compte.

- La *simulation de Monte Carlo de troisième ordre* consiste en un mélange de la première et de la seconde stratégie. Le cheminement d'un patient sera déterminé comme lors d'une simulation de premier ordre, mais à partir de probabilités tirées au hasard comme lors d'une simulation de second ordre

3.3. Courbe d'admissibilité au remboursement

Le rapport coût-efficacité différentiel est le critère d'intérêt dans toute analyse coût-efficacité confrontant deux traitements [21, 22, 23]. Le remplacement d'une stratégie par une autre entraîne à la fois une différence de coût et une différence d'efficacité. C'est le quotient de ces différences qui définit le rapport coût-efficacité différentiel.

Dans le cadre d'une analyse déterministe le classement des stratégies étudiées les unes

par rapport aux autres était obtenu par application des algorithmes de dominance forte et faible. Une stratégie était dite fortement dominée par une autre lorsqu'elle est moins efficace et plus chère, ou plus onéreuse et d'une efficacité identique. Une stratégie était faiblement dominée si, et seulement si, le rapport coût-efficacité différentiel décroissait lorsque le coût total du traitement augmentait ; auquel cas, il convenait de l'éliminer et de recalculer les différences de coûts et d'efficacités qu'entraînait un nouveau choix thérapeutique par rapport à la dernière stratégie non dominée. Une stratégie était finalement dite efficiente dès lors qu'il n'existait aucune autre stratégie permettant d'obtenir un résultat supérieur pour un coût moindre.

L'incertitude autour du rapport coût-efficacité calculé n'était pas estimable avec les méthodes standards. Il est en effet difficile d'appréhender la distribution d'un rapport [24, 25, 26, 27, 28]. Le bootstrap paramétrique (ou non paramétrique si l'on dispose de données individuelles) permet de contourner le problème. En réalisant B échantillons de bootstrap comprenant les couples (coût ; années de vie sauvées) à partir des distributions de probabilités qui sont associées à toutes les variables, il est possible de construire B rapports et de les représenter dans le plan (coût différentiel ; ΔC – efficacité différentielle ΔE). Ce graphique permet d'apprécier la dispersion du rapport. A partir de ces mêmes échantillons de bootstrap, il est également possible de construire une courbe d'acceptabilité du nouveau traitement [29]. Cette courbe donne, en fonction de la disposition à payer du décideur, la probabilité que le traitement soit efficient. Pour une disposition à payer λ , cette probabilité est égale à la proportion d'échantillons de bootstrap tels que le rapport calculé sur cet échantillon est inférieur à λ . Cette courbe fournit une autre mesure de l'incertitude liée à l'estimation ponctuelle du rapport coût-efficacité. Cette courbe, dont l'interprétation n'a réellement de sens que dans un

contexte bayésien, représente la probabilité que le nouveau traitement soit coût-efficace par rapport à l'ancien, en fonction de la disposition à payer de la collectivité. L'avantage de ce type de représentation est qu'il est plus facilement mobilisable pour un décideur, qui peut choisir son niveau de disposition à payer.

L'analyse coût-efficacité est une analyse de rendement ; elle se propose d'étudier si les résultats clinique, humain et financier qui sont obtenus en contre partie de l'investissement réalisé sont à la hauteur des efforts déployés (rapport performance investissement) ; ou lorsque l'on prend l'inverse de la formule précédente si la valeur des moyens additionnels mis en œuvre n'est pas disproportionnée par rapport au surcroît d'efficacité constaté lorsqu'on administre un nouveau traitement à la place des anciens (rapport coût-efficacité). Cette tension lancinante entre les 2 paramètres sur la base desquels chacun fait ses choix dans la vie courante, suppose qu'il existe un seuil à ne pas dépasser ou un intervalle de valeur à respecter en fonction des indications reconnues comme étant plus ou moins prioritaire en termes de santé publique.

Le chiffre le plus couramment avancé est celui de 50 000 \$ par année de vie gagnée ajusté sur la qualité de vie (QALY). Sa valeur reflétait, en 1997, le coût d'une des interventions médicales les plus coûteuses : la dialyse rénale. Une estimation plus réaliste du coût actuel de celle-ci fixerait le seuil des dépenses acceptables entre 74 000 et 95 000 \$ par QALY [30].

L'OMS a proposé un autre seuil en fonction d'un critère différent, le coût par année de vie sans incapacité (DALY) [31]. Sa valeur est exprimée en proportion du produit intérieur brut par habitant (3 PIB par habitant). Sa valeur varie donc en fonction de la richesse des pays où les analyses coût-efficacité sont conduites : 108 000 \$ pour les États-Unis, 79 500 \$ pour la France, 53 000 \$ pour la Nouvelle Zélande [32].

4. INTEGRATION DES INFORMATIONS DE CONTEXTE

La méthode de Cochrane, en répartissant au hasard les individus entre les deux bras d'une étude, neutralise les caractéristiques personnelles des sujets qui seraient susceptibles d'introduire des biais dans l'interprétation des résultats. En revanche, très souvent les critères d'inclusion sont trop restrictifs, les critères de jugement trop étriqués, la période de suivi trop brève, les protocoles de surveillance trop rigoristes et les populations trop homogènes. De tels plans d'expériences ne sont pas représentatifs des malades tout-venant. Leur validité externe en est réduite d'autant.

Si l'on veut évaluer l'efficacité réelle d'une action de santé en situation courante d'usage, il faut mettre en place des études observationnelles rigoureuses, mais non randomisées, qui devraient toujours intégrer un groupe témoin. Les études analytiques quasi expérimentales [33, 34, 35] sont susceptibles d'offrir la réponse à la question posée. Elles reposent toutes sur le même principe l'exposition ou la non exposition à une innovation thérapeutique ou organisationnelle.

4.1. Schémas D'études quasi expérimentaux

Leur modalité pratique de mise en œuvre est fonction de la réponse positive ou négative qui est apportée au cinq questions suivantes : 1) la comparaison exposé/non exposé est elle faite au même moment et sur le même groupe ? (coupe transversale) ; 2) est-elle faite à des moments différents sur le même groupe en une fois ou en plusieurs fois ? (étude avant-après ou séries temporelles interrompues) ; 3) est-elle faite à des moments différents sur deux groupes différents avec affectation contrôlée ? ; en cas de réponse négative à cette question, 4) comment sont définis les groupes ? à partir de la survenue des événements cliniques ? (étude cas — témoins) ou à partir des facteurs de

risques ? ; 5) s'il y a définition à partir des facteurs de risque, ceux-ci sont-ils critériés de façon très générale ? (étude avant-après avec groupe contrôle, séries intertemporelles interrompues avec séries appariées) ou rigoureusement définis pour obtenir dans le même temps des populations très homogènes ? (études de cohorte prospective ou rétrospective).

La mise en place d'études avant-après ou de séries temporelles interrompues avec groupe contrôle est susceptible de fournir une réponse adaptée aux besoins des promoteurs d'innovations thérapeutiques ou organisationnelles qui souhaitent disposer d'outils opérationnels et peu coûteux pour apporter la preuve de leur efficacité en pratique médicale quotidienne. Ce type de protocole permet d'établir un relation causale entre la mise en œuvre de la nouvelle organisation et les résultats médicaux et-économiques obtenus.

4.2. Recueil prospectif à l'aide de cahier informatique

Le caractère multicentrique des études nécessite souvent la mise en place d'un système informatique qui permet à la fois :

- une collecte décentralisée dans chaque centre d'investigation ;
- un regroupement des données au niveau national ;
- une alerte immédiate sur les données manquantes ou les incohérences dont la présence est signalée à l'écran par l'apparition de « pense bêtes ».

Les techniques de programmation dynamique autorisent une interaction permanente entre les investigateurs et le centre chargé du traitement de l'information. Traditionnellement, le recueil des données s'accompagne de nombreux déplacements dans les centres investigateurs et de procédures de contrôles successifs. La réalisation de telles études sur le Net optimise la circulation et le partage de l'in-

formation entre les différents intervenants. Par ailleurs, l'utilisation d'interfaces de saisie formatées facilite le recueil d'informations par un accès rapide aux champs de saisie, et permet un contrôle à chaque saisie par l'investigateur lui-même.

Pour garantir la confidentialité des données collectées, les dossiers-patient sont identifiés à l'aide d'un numéro unique appelé « identifiant patient » généré par le médecin investigateur qui seul connaît la correspondance entre cet identifiant et l'identité du patient. De la même façon, chaque investigateur a un identifiant et un mot de passe. Ceci ne permet un accès restreint à la base de données qu'aux utilisateurs référencés, avec des droits d'accès définis.

4.3. Exploitation rétrospective des bases de données médico-administratives

L'exploitation des bases de données rétrospectives doit pouvoir contribuer à renforcer cette démonstration. Les bases de données médico administratives permettent d'étudier l'impact des décisions thérapeutiques sur des populations importantes, dont les caractéristiques sont plus proches de celles observées en population générale que ne le sont celles retenues au titre des critères d'inclusion des essais cliniques ; la validité externe des résultats obtenus en est améliorée d'autant. Ces banques de données peuvent être d'origine médicale (RESALIS, Cub-Réa, OMIT), administrative (le PMSI) ou être constituées à partir des demandes de remboursement adressées par les assurés sociaux aux organismes payeurs (SNIIR-AM). Ces bases fournissent des données sur les consommations médicales allant de la consommation des médicaments à celle des autres soins médicaux (consultations, examens biologiques, hospitalisations, etc.). Les données sont recueillies de façon rétrospective une fois que les décisions thérapeutiques ont été prises et que le traitement a été administré de sorte que ce recueil n'a aucune incidence

sur le déroulement de la prise en charge clinique. Elles offrent donc la possibilité de mesurer en pratique réelle les conséquences des décisions individuelles.

4.4. Un impératif dans tous les cas : Travailler sur des populations comparables

Tout comme dans un essai clinique, on peut comparer de façon longitudinale sur une même période d'observation, les résultats obtenus sous différentes options thérapeutiques. Contrairement à un essai clinique randomisé, rien ne garantit dans de telles études observationnelles que la population des patients ou des ressortissants inclus dans une phase « Avant » ou inclus « Ici » ou exposés à un « mode de prise en charge expérimental » soient comparables à celle des patients inclus dans la phase « Après », ou « Ailleurs » ou ayant bénéficié d'une prise en charge conventionnelle puisque la décision d'inclure les patients dans l'un ou l'autre de ces bras de l'étude ne résulte pas d'un processus aléatoire. On parle alors de biais de recrutement. Le biais de recrutement est ainsi un type de biais de sélection, toute analyse de coûts ne prenant pas en compte ces différences aboutira à des estimateurs erronés.

La notion d'équilibre d'une variable entre les deux phases permet d'appréhender l'existence de ce type de biais. Elle se mesure au travers des différences standardisées : pour une variable donnée, une différence standardisée supérieure à 10 % traduit un déséquilibre de cette variable entre les deux phases et donc un potentiel biais de recrutement. Pour une variable quantitative, la différence standardisée se mesure comme suit :

$$d = \frac{\overline{X}_{\text{Expérimentation}} - \overline{X}_{\text{Contrôle}}}{\sqrt{\frac{S_{\text{Expérimentation}}^2 + S_{\text{Contrôle}}^2}{2}}}$$

Il s'agit donc de la différence entre les moyennes, pondérée par l'écart-type commun. La généralisation au cas d'une variable binaire

est immédiate. Pour les variables qualitatives à plus de deux modalités, la valeur de la différence standardisée dépend de la modalité prise comme référence. L'équilibre d'une telle variable se traduira par la valeur absolue de la différence standardisée maximale parmi celles calculées en prenant chaque modalité en référence.

Plusieurs méthodes concurrentes permettent de réduire les biais de recrutement : l'appariement, la stratification et l'ajustement. Les deux premières méthodes doivent être mises en place au moment de l'inclusion des patients. La dernière est la méthode la plus utilisée. Elle consiste à introduire les facteurs de confusion potentiels dans un modèle multivarié. La méthode de la probabilité prédite ou score de propension (SP) constitue une alternative aux méthodes précédentes [36, 37, 38]. On peut interpréter le SP comme la probabilité, pour une personne de caractéristiques données, d'être exposée à un traitement. Le SP réduit ainsi un nombre potentiellement élevé de covariables en une seule variable composite qui résume correctement l'ensemble des caractéristiques mesurées. L'objectif premier du SP est de simuler un essai clinique randomisé, en s'assurant que les patients traités et non traités ne diffèrent pas quant à leurs caractéristiques de base. Cependant, alors qu'un essai randomisé assure en principe une totale comparabilité des patients (à l'exception du traitement reçu) grâce à la randomisation, le SP ne permet que d'assurer la comparabilité des patients sur les variables mesurées. Tout facteur de confusion non relevé dans le protocole peut donc biaiser les résultats.

Le SP peut s'employer de trois principales manières pour contrôler les biais de sélection : soit pour *stratifier* les patients en sous-populations aux caractéristiques comparables ; soit pour *appairer* a posteriori les patients exposés et non exposés (au traitement) ; soit pour *ajuster* les modèles multivariés. Certaines analyses peuvent être plus complexes et combiner plusieurs de ces usages (appairer une popu-

lation puis construire des modèles ajustés sur le SP, par exemple). Par ailleurs, on peut également construire plusieurs SP si l'on a plusieurs expositions majeures à explorer.

En cas d'*appariement*, la difficulté technique réside dans le choix de la méthode pour faire correspondre un patient non exposé à un patient exposé : le SP étant continu et compris entre 0 et 1, la probabilité de trouver deux patients dont le SP soit identique est nulle. Il faudra par ailleurs trouver un compromis entre un appariement plutôt strict, garantissant une bonne comparabilité des patients mais réduisant de beaucoup la taille d'échantillon (on parle alors d'appariement incomplet, puisque certains patients seront exclus de l'analyse faute d'avoir trouvé leur paire), et un appariement moins strict, permettant d'obtenir un échantillon plus important mais moins comparable (on parle alors d'appariement inexact, puisque certaines paires s'avèreront dissemblables). L'appariement inexact est considéré comme préférable lorsque l'on s'intéresse à inférer des résultats sur la population. Il faudra alors avoir recours à des analyses supplémentaires (en pratique, l'ajustement) pour gérer la moindre comparabilité des sujets.

4.5. Délimitation du périmètre de définition des coûts

Une erreur fréquente chez les professionnels de santé consiste à penser que – dans une perspective sociétale – les indemnités journalières, l'assurance chômage et les autres prestations sociales doivent figurer dans les coûts (ou les bénéfices) indirectement associés au traitement. Les économistes n'incluent pas ces prestations dans les estimations de coûts qui sont faits dans ce contexte. Traditionnellement, on admet que de telles sommes ne bénéficient pas directement au secteur sanitaire. Elles sont prélevées sur certains assujettis pour être redistribuées à d'autres et servent à financer des dépenses qui alimentent l'activité économique générale. Elles ne correspondent en aucune

façon à une affectation spécifique de ressources au bénéfice de la santé. Une seconde erreur peut être commise lorsque, dans un souci honorable d'exhaustivité, on introduit dans les coûts indirects du traitement les pertes de productivité liées à l'absentéisme, au renouvellement du personnel ainsi que la perte de revenu imputable à un décès prématuré. La plupart des échelles et notamment le QWB qui explorent ces domaines prennent en compte les répercussions de la maladie sur les rôles sociaux et les occupations habituelles. Valoriser ces absences sur les lieux de travail en les portant sous forme de coût au numérateur, reviendrait à les décompter deux fois. Le risque serait grand d'aboutir à des doubles emplois entre le numérateur et le dénominateur d'un ratio coût-efficacité exprimé en QALYs.

Pour ne pas alourdir le cahier d'observation, les évaluateurs renoncent parfois à intégrer dans l'analyse les coûts non médicaux directs, c'est-à-dire les coûts restants à la charge des familles. Le périmètre de définition des coûts retenu, se borne à étudier l'impact de l'innovation sur les dépenses de consommation de biens et services médicaux du système de soins [39, 40, 41] – Ce périmètre est plus large que celui défini par les dépenses reconues et remboursées par l'assurance maladie puisqu'il intègre les restes à charge; mais en même temps il est plus étroit qu'il aurait pu l'être si l'ensemble des prestations en nature ou en espèce de l'assurance maladie avait été pris en compte. Sa spécificité par rapport à la perspective sociétale est tout aussi réelle. Les transferts en sont exclus, comme ils l'étaient quand on se plaçait du point de vue de la collectivité, mais à la différence des calculs qui sont habituellement faits dans cette optique, ni les coûts non médicaux directs, ni les coûts indirects ne figurent dans les décomptes soit pour des raisons de commodité, soit pour éviter les doubles emplois. Ce point de vue qui s'ajoute aux trois perspectives classiques (Hôpital, Assurance maladie, Société) sans les recou-

vrir, est désigné sous l'appellation « perspective système de soins ».

4.6. Comparaison des coûts

La méthode la plus classique pour comparer les coûts moyens entre deux groupes de traitement est le test de Student. Ce test est toutefois fondé sur l'hypothèse d'indépendance des données et de normalité des distributions de coûts dans chacun des groupes comparés, ce qui va à l'encontre de la nature de la distribution des coûts, le plus souvent dispersée vers la droite. L'hypothèse de normalité a donc toutes les chances de ne pas être respectée.

Le test de Mann-Whitney pourrait éventuellement être utilisé puisqu'il ne fait aucune hypothèse sur la distribution des données. Toutefois, il ne compare pas les coûts moyens mais les coûts médians. De plus, les résultats sont sensibles à l'hétérogénéité des variances des deux groupes.

Il serait enfin possible d'obtenir un autre estimateur de la moyenne en supposant que les données sont log-normales. Toutefois, la comparaison des coûts associée à cette méthode n'est justifiée que si la loi log-normale est proche de la vraie distribution des coûts.

Une alternative à ces méthodes paramétriques et non-paramétriques [42, 43, 44, 45, 46] de comparaison des coûts entre deux groupes de traitement est le bootstrap non paramétrique; c'est celle qui doit être mise en œuvre si des écarts à la loi normale sont constatés (test de Shapiro-Wilk).

4.7. Analyse de sensibilité non paramétrique

Principes théoriques [47]: Considérons le cas où nous tentons d'estimer la moyenne d'une caractéristique particulière X sur une population de taille N (par exemple, la différence de coûts moyenne entre deux bras de traitement). Une estimation classique est la moyenne arithmétique sur un échantillon aléatoire de taille

n. Cet estimateur possède une distribution F_m , qui dépend de la distribution F de la variable aléatoire X , dont les valeurs observées sont x_1, \dots, x_n . Comme F est inconnue, on doit l'estimer. Si nous ne pouvons attribuer un modèle paramétrique aux données, nous utilisons $\hat{F} = F_n$ comme approximation de F , où F_n est la fonction de distribution empirique. On génère alors B échantillons de taille n provenant de F_n , appelés échantillons bootstrapés, et on calcule les B moyennes $\bar{x}_1, \dots, \bar{x}_B$ sur chacun de ces échantillons. On montre alors que la distribution empirique des valeurs simulées $\bar{x}_1, \dots, \bar{x}_B$ fournit une approximation de la distribution de la vraie moyenne.

Mise en œuvre pratique : Pour obtenir les échantillons bootstrapés, il suffit de tirer aléatoirement avec remise n individus de l'échantillon initial et de répéter ce procédé B fois. Chaque individu possède une même probabilité de tirage, égale à $1/n$, et un même individu peut apparaître plusieurs fois dans un même échantillon bootstrapé. On calcule alors la moyenne de la caractéristique d'intérêt sur chacun des B échantillons ainsi obtenus ($\bar{x}_1, \dots, \bar{x}_B$). L'estimation de la moyenne sur la population totale est alors donnée par la moyenne arithmétique des moyennes bootstrapées :

$$\bar{x}_{boot} = \frac{1}{B} \sum_{i=1}^B \bar{x}_i^*$$

La technique est opérationnalisée en quatre étapes. On tire aléatoirement avec remise n_j individus appartenant au groupe contrôle et on calcule le coût moyen et l'efficacité moyenne pour cet échantillon. On tire aléatoirement avec remise n_k individus appartenant au groupe traitement innovant et on calcule le coût moyen et l'efficacité moyenne correspondantes. Puis on calcule par différence le ratio coût efficacité différentiel entre les deux échantillons bootstrapés. L'opération est enfin répétée B fois pour obtenir une estimation de la distribution du RCED dans le plan ACE.

5. CONCLUSION

Chaque type d'étude offre des avantages spécifiques et a sa place dans les différentes phases qui caractérisent le cycle de vie d'une innovation thérapeutique. Au moment de la mise sur le marché, on ne dispose que de données prévisionnelles quant aux bénéfices médicaux et économiques attendus d'un traitement. Les essais cliniques randomisés de phase III et des méta-analyses sont alors les seules sources d'informations disponibles.

Après que l'autorisation de mise sur le marché a été accordée, les informations d'utilisation en pratique courante s'accumulent et les études rétrospectives de base de données médico-administratives permettent d'évaluer les résultats obtenus par le traitement quand celui-ci est mis en œuvre dans un système de santé complexe. Elles offrent alors un moyen efficace de vérifier si les résultats promis sont bien au rendez-vous puisque rien ne dit que les conclusions d'un essai soient généralisables à toutes les situations cliniques et à tous les systèmes santé.

6. RÉFÉRENCES

- [1] Beck RJ, Pauker SG. The Markov process in medical prognosis. *Medical Decision Making* 1983;3:419-58.
- [2] Sonnenberg FA, Beck JR Markov models in medical decision making. A practical guide. *Medical Decision Making* 1993;13:322-38.
- [3] Launois R, Croutsche JJ, Mégnigbêto A, Le Lay K, Portafax C. L'apport indispensable de l'épidémiologie clinique aux modèles de Markov. *Journal d'Economie Médicale* 1999;17(5):343-361.
- [4] Ades AE, Welton N, Lu G. Introduction to mixed treatment comparisons. *MRC Health Services Research Collaboration*.
- [5] Glenny AM, Altman DG, Song F, Sakarovitch C, Deeks JJ, D'Amico R, Bradburn M, Eastwood AJ. Indirect comparisons of competing interventions. *Health Technology Assessment* 2005;9(26).

- [6] Ades AE, Claxton K, Sculpher M. Evidence synthesis, parameter correlation and probabilistic sensitivity analysis. 2004. *MRC Health Services Collaboration*.
- [7] Thompson SG Controversies in meta-analysis: the case of the trials of serum cholesterol reduction. *Statistical methods in medical research* 1993;2:173-192.
- [8] Sutton AJ, Abrams KR, Jones DR. An illustrated guide to the methods of meta-analysis. *Journal of Evaluation in Clinical Practice* 2000;7;2:135-148.
- [9] Sutton AJ, Abrams KR Bayesian methods in meta-analysis and evidence synthesis. *Statistical Methods in Medical Research* 2001;10:277-303
- [10] Thompson SG, Higgins J How should meta-regression analyses be undertaken and interpreted? 2001. *MRC Biostatistics Unit*. Cambridge
- [11] Spiegelhalter DJ, Myles JP, Jones DR, Abrams KR. Bayesian methods in health technology assessment: a review. *Health Technology Assessment* 2000;4(38).
- [12] Spiegelhalter DJ, Best NG. Bayesian approaches to multiple sources of evidence and uncertainty in complex cost-effectiveness modelling. 2002.
- [13] Parmigiani G. Measuring uncertainty in complex decision analysis models. *Statistical Methods in Medical Research* 2002;11:513-537
- [14] Cooper NJ, Abrams KR, Sutton AJ, Turner D, Lambert PC A Bayesian approach to Markov modeling in cost-effectiveness analyses: application to taxane use in advanced breast cancer. *JR Statist Soc A* 2003;166, part 3:389-405.
- [15] Cooper NJ, Sutton AJ, Abrams KR, Turner D *et al.* Comprehensive decision analytical modeling in economic evaluation: a bayesian approach. *Health economics* 2004;13:203-226.
- [16] Weinstein MC, Siegel JE, Gold MR, Kamlet MS, Russel LB. Recommendations of the panel on cost-effectiveness in health and medicine. *JAMA* 1996;276 :1253-8.
- [17] Briggs AH. Handling uncertainty in cost-effectiveness models. *Pharmacoeconomics* 2000;17: 479-500.
- [18] Doubilet P, Begg CB, Weinstein MC, Braun P, McNeil BJ. Probabilistic sensitivity analysis using Monte Carlo simulation. *Med Decis Making* 1985;5:157-77
- [19] Claxton K, Sculper M, McCabe C *et al.* Probabilistic sensitivity analysis for NICE technology assessment: not an optimal extra. *Health Economics* 2005;14:339-347.
- [20] Chilcott J, Brennan A, Booth A, Karnon J, Tapenden P. The role of modelling in prioritising and planning clinical trials. *Health Technology Assessment* 2003;7(23).
- [21] Launois R La recherche de l'efficience : un impératif déontologique. *Médecine Sciences (NS)* août 1990:25-30.
- [22] Launois R Qu'est-ce que la régulation médicalisée. *Journal d'économie médicale* 1994;12:63-74.
- [23] Launois R, Lhoste F Quelle place pour l'évaluation économique en matière de politique de santé ? *Journal d'Economie Médicale* 2002;20(2): 128-135.
- [24] Briggs AH, O'Brien BJ The death of cost-minimisation analysis ? *Health Econ* 2001;10:179-184.
- [25] Glick HA, Briggs AH, Polsky D Quantifying stochastic uncertainty and presenting results of cost-effectiveness analyses. *Expert Rev. Pharmacoeconomics Outcomes Res.* 1(1), 2001.
- [26] O'Hagan A, Stevens JW, Montmartin J. Inference for the Cost-Effectiveness Acceptability Curve and Cost-Effectiveness Ratio. *Pharmacoeconomics* 2000 Apr.;17(4):339-349.
- [27] Briggs AH A bayesian approach to stochastic cost-effectiveness analysis. *Health Economics* 1999;8(3):257-261.
- [28] Heitjan DF, Moskowitz AJ, Whang W Bayesian estimation of cost-effectiveness ratios from clinical trials. *Health Economics* 1999;8:191-201.
- [29] M Löthgren, N Zethraeus. Definition, Interpretation and calculation of cost-effectiveness acceptability curves. *Health Economics* 2000;9:623-630.
- [30] Hirth RA, Chernew ME, Miller E, Fendrick AM, Weissert WG. Willingness to pay for a quality-adjusted life year: in search of a standard. *Med Decis Making* 2000 Jul-Sep;20(3):332-42.
- [31] World Health Organization. *Making choices in health: WHO guide to cost effectiveness analysis*. 2003. Edited by T Tan-Torres Edejer, R Baltussen, T Adam, R Hutubessy, A Acharya, DB Evans, CJL Murray.

- [32] Eichler HG, Kong SX, Gerth WC, Mavros P, Jönsson B. Use of cost-effectiveness analysis in health-care resource allocation decision making: How are cost-effectiveness thresholds expected to emerge? *Value In Health* 2004;7(5):518-528.
- [33] Launois R, Cartraud A, Perez V. Evaluation of integrated health delivery systems – example from France”. In Preuß KJ, Rübiger J, Sommer JH (Eds): *Managed Care – Evaluation und Performance-Measurement integrierter Versorgungsmodelle*. Schattauer. Stuttgart 2001;Chap. 12:171-190.
- [34] Launois R, Mègnigbèto AC, Pérez V, Roy M, Camus A, Gailhardou M, Lançon F, Quéniart M Expérimentation de la prise en charge de l'asthme en réseau de soins coordonnés® bilan de l'expérience RESALIS dans l'Eure, 18 mois avant et 12 mois après intervention. *Journal d'Economie Médicale* 2002;20(6):333-352.
- [35] Launois R L'évaluation socio-économique une discipline entre recherche clinique et études observationnelles. *Bull cancer* 2003;90(1):97-104.
- [36] Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. *Biometrika* 1983;70:41-55.
- [37] Normand ST, Landrum MB, Guadagnoli E, Ayanian JZ, Ryan TJ, Cleary PD, McNeil BJ. Validating recommendations for coronary angiography following acute myocardial infarction in the elderly: a matched analysis using propensity scores. *J Clin Epidemiol* 2001 Apr;54(4):387-98.
- [38] D'Agostino RB Jr. Propensity score methods for bias reduction in the comparison of a treatment to a non-randomized control group. *Stat Med* 1998 Oct 15;17(19):2265-81.
- [39] Sznajder M, Aegerter P, Launois R, Merliere Y, Guidet B, CubRea. A cost-effectiveness analysis of stays in intensive care units. *Intensive Care Med* 2001 Jan;27(1):146-53.
- [40] Launois R. Un coût, des coûts, quels coûts? *Journal d'Economie Médicale* 1999;17(1):77-82.
- [41] Launois R, Vergnenègre A, Garrigues B Notions et mesure des coûts en fonction de la perspective choisie. *Bull Cancer* 2003;90(11):946-54 repris dans Guide méthodologique pour l'évaluation des actions de santé, recommandations méthodologiques – CES Paris 2003. Annexe I.
- [42] Thomson SG, Barber JA. How should cost data in pragmatic randomised trials be analysed? *British Medical Journal* 320, 1197-1200 (2000).
- [43] O'Hagan A, Stevens JW. Assessing and Comparing Costs: How Robust are the bootstrap and the methods based on asymptotic normality? *Health Economics* 12, 33-49 (2003).
- [44] Zimmerman DW Statistical Significance Levels of Nonparametric Tests Biased by Heterogenous Variances of Treatment Groups. *Journal of General Psychology*, Oct 2000.
- [45] Zhou XH, Melfi CA, Hui SL. Methods for comparison of cost data. *Ann Intern Med* 1997;127(8 Part 2):752-756.
- [46] Manning WG, Mullahy J. Estimating log models: to transform or not to transform? *J Health Econ*. 2001 Jul;20(4):461-94.
- [47] Desgagné A, Castilloux AM, Angers JF, LeLorier J. The Use of the Bootstrap Statistical Method for the Pharmacoeconomic Cost Analysis of Skewed Data. *Pharmacoeconomics* 1998;13:487-497.