ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/323659054

Coûts unitaires standards ou coûts unitaires spécifiques : quels critères de choix pour l'évaluation économique de stratégies de santé dans les études multicentriques ?

Article *in* Revue d Épidémiologie et de Santé Publique · March 2018

DOI: 10.1016/j.respe.2018.02.002

CITATIONS
0

READS
31

20 authors, including:

Anne-Gaëlle Le Corroller Soriano
French Institute of Health and Medical Research
67 PUBLICATIONS **905** CITATIONS
SEE PROFILE

Benoit Dervaux
Université du Droit et de la Santé Lille 2
139 PUBLICATIONS **951** CITATIONS
SEE PROFILE

Camille Dutot
8 PUBLICATIONS **10** CITATIONS
SEE PROFILE

Pascale Guerre
Centre Hospitalier Universitaire de Lyon
12 PUBLICATIONS **8** CITATIONS
SEE PROFILE

Some of the authors of this publication are also working on these related projects:

ALD-Cancer View project

Counterfactual analysis of the added-value of Youth Employment Initiative (YEI) in France View project

All content following this page was uploaded by Lionel Perrier on 27 September 2018.

The user has requested enhancement of the downloaded file.

Disponible en ligne sur ScienceDirect www.sciencedirect.com

Elsevier Masson France EM|consulte www.em-consulte.com

Revue d'Épidémiologie et de Santé Publique 66 (2018) S101–S118

Revue d'Épidémiologie et de Santé Publique
Epidemiology and Public Health

Calcul et analyse des coûts hospitaliers : guide méthodologique

Méthodes d'analyse et de traitement des données de coût : approches par « micro-costing » et « gross-costing »

Methods for the analysis and treatment of cost data by micro- and gross-costing approaches

M. Morelle [a,*], M. Plantier [b], B. Dervaux [c], A. Pagès [d], F. Deniès [c], N. Havet [e], L. Perrier [a] on behalf of the French Costing Group [1]

[a] *GATE L-SE UMR 5824, centre Léon-Bérard, université de Lyon, université Claude-Bernard Lyon 1, 28, rue Laennec, 69008 Lyon, France*
[b] *ISFA, laboratoire SAF, université Lyon 1, 50, avenue Tony-Garnier, 69100 Lyon, France*
[c] *Cellule innovation, DRCI, CHRU de Lille, 2, avenue Oscar-Lambret, 59000 Lille, France*
[d] *Inserm UMR 1027, CHU de Toulouse, 37, allées Jules-Guesde, 31059 Toulouse, France*
[e] *ISFA, laboratoire SAF, université Lyon 1, 50, avenue Tony-Garnier, 69100 Lyon, France*

Disponible sur Internet le 9 mars 2018

Abstract

This work addresses the analysis of individual cost data in the setting of interventional or observational studies using statistical analysis software once the costs per patient have been estimated. It is in fact necessary to be able to present and describe data in an appropriate manner in each of the studied health strategies and to test whether the difference in costs observed between treatment groups is due to chance or not. Furthermore, cost analysis differs from conventional statistical analysis in that cost data have a certain number of specific properties, including their use by health decision-makers. This work also addresses the difficulties that generally arise in regard to the distribution of cost; it explains why the mathematical average constitutes the only relevant measure for economists; and it outlines which analyses are required for inter-strategy cost comparisons. It also covers the issue of missing or censored data, features that are inherent to information collected regarding costs and to sensitivity analyses.
© 2018 Elsevier Masson SAS. All rights reserved.

Keywords: Cost; Evaluation; Guidelines; Hospital; Production; Technology

Résumé

Ce travail traite de l'analyse des données de coût individuelles dans le cadre d'études interventionnelles ou observationnelles à l'aide de logiciels d'analyse de données et de traitement statistique dès lors que les coûts par patient ont été estimés. Il est, en effet, nécessaire de pouvoir les présenter et les décrire de façon appropriée dans chacune des stratégies de santé étudiées et de tester si la différence de coût observée entre les groupes de traitement est due au hasard ou non. De plus, l'analyse des données de coût se distingue des analyses statistiques « classiques » par un certain nombre de caractéristiques propres à ces données ainsi qu'à leur utilisation par les décideurs de santé. Ce travail présente également les difficultés que posent généralement les distributions de coût, explique pourquoi la moyenne arithmétique constitue la seule mesure pertinente pour les économistes, et décrit quelles analyses sont nécessaires pour la comparaison des coûts entre les stratégies. Il s'intéresse aussi à la question des données manquantes ou censurées, spécificités souvent inhérentes aux informations collectées sur les coûts et aux analyses de sensibilité.
© 2018 Elsevier Masson SAS. Tous droits réservés.

Mots clés : Coût ; Évaluation ; Hôpital ; Production ; Recommandation ; Technologie

* Auteur correspondant.
Adresse e-mail : M.Morelle[a*] magali.morelle@lyon.unicancer.fr">M.Morelle[a*] magali.morelle@lyon.unicancer.fr ().

[1] The French Costing Group: S. Baffert (Cemka), A-C. Bertaux (Dijon), J. Bonastre (Villejuif), N. Costa (Toulouse), F. Deniès (Lille), B. Dervaux (Lille), C. Dutot (Montpellier), P. Guerre (Lyon), N. Havet (Lyon), N. Hayes (Bordeaux), A-G. Le Corroller-Soriano (Marseille), C. Lejeune (Dijon), B. Lueza (Villejuif), J. Margier (Grenoble), G. Mercier (Montpellier), M. Morelle (Lyon), L. Perrier (Lyon), A. Pagès (Toulouse), M. Plantier (Lyon), V-P. Riche (Nantes).

https://doi.org/10.1016/j.respe.2018.02.003
0398-7620/© 2018 Elsevier Masson SAS. Tous droits réservés.

1. Introduction

Ce chapitre traite de l'analyse des données de coût individuelles, dans le cadre d'études interventionnelles ou observationnelles, à l'aide d'un logiciel d'analyse de données et de traitement statistique, une fois que les coûts par patient ont été estimés. Il est nécessaire de pouvoir :

- les présenter et les décrire de façon appropriée dans chacune des stratégies étudiées ;
- de tester si la différence de coût observée entre les groupes de traitement est due au hasard ou non.

L'analyse des données de coût se distingue des analyses statistiques « classiques » par un certain nombre de caractéristiques propres à ces données ainsi qu'à leur utilisation par les décideurs de santé. Dans ce qui suit, nous allons présenter les difficultés que posent généralement les distributions de coût, expliquer pourquoi la moyenne arithmétique constitue la seule mesure pertinente pour les économistes, et décrire quelles analyses sont nécessaires pour la comparaison des coûts entre les stratégies. Nous nous intéresserons ensuite à la question des données manquantes ou censurées, spécificités souvent inhérentes aux informations collectées sur les coûts. Nous finirons ce chapitre par les analyses de sensibilité.

Remarque : tout au long de ce chapitre, nous noterons *COÛT* la variable de coût, *X* les variables explicatives, *GROUP* la variable distinguant les différents groupes de patients à l'étude (en général, patients traités versus non traités).

2. Statistiques descriptives

Les statistiques descriptives des données continues comme les coûts consistent à présenter une mesure de positionnement, de la variabilité et les particularités de la forme de la distribution.

La distribution de coûts est généralement tronquée, avec une asymétrie (« skewness », en anglais) à droite. Ceci s'explique par plusieurs raisons : tout d'abord, un coût de traitement ne peut pas être négatif, tout comme les variables biologiques de poids ou de taille, qui suivent généralement une distribution normale. Mais à la différence de ces dernières, il peut y avoir un certain nombre de patients avec des coûts nuls. Ensuite, il existe généralement une proportion non négligeable de patients avec des coûts de traitements beaucoup plus élevés que les autres patients, du fait d'évènements indésirables, de comorbidités, de ré-interventions, qui produisent une longue queue sur la droite de la distribution. Enfin, il existe souvent un petit nombre de patients qui ont eu des évènements graves qui conduisent à des coûts qui se situent plusieurs déviations standards au-dessus de la moyenne et qui sont considérés comme des valeurs extrêmes.

Afin de mieux appréhender les distributions, il est souvent utile de représenter graphiquement les variables de coût, un histogramme étant particulièrement parlant. Chaque barre représente la fréquence ou la proportion de patients qui sont inclus dans un intervalle. La Fig. 1 représente la forme typique de la distribution d'une variable de coût.

En pratique, la construction d'un histogramme se réalise avec les commandes suivantes sous SAS ou STATA :

Réalisation d'histogramme sous SAS
PROC UNIVARIATE data = table;
HISTOGRAM COUT;
RUN;

Réalisation d'histogramme sous STATAtwo-way histogram COUT

Alors que la médiane (valeur de l'observation qui divise l'échantillon en deux parties d'effectifs équivalents) et la moyenne sont identiques dans une distribution symétrique (ex. : distribution normale), elles diffèrent dans les distributions asymétriques. Lorsque la distribution est asymétrique vers la droite comme sur la Fig. 1, la médiane est inférieure à la moyenne.

Alors qu'habituellement, la médiane est souvent utilisée pour décrire les données qui présentent une distribution asymétrique ou non normale, et bien qu'elle fournisse des informations descriptives intéressantes, elle ne permet pas aux décideurs de santé de déterminer le coût total du traitement pour un groupe de patients, car ce coût total est égal au coût moyen par patient multiplié par le nombre de patients dans le groupe.

Une revue de la littérature réalisée par Doshi et al. en 2006 [1] a montré que sur 115 évaluations médicoéconomiques, 110 (84 %) ont rapporté la moyenne arithmétiques des coûts, 14 (12 %) la moyenne et la médiane et 5 (4 %) le coût médian seulement.

Recommandation:

> Dans les évaluations économiques, la statistique qui doit être utilisée est le coût moyen par patient et ce, quelle que soit l'asymétrie de la distribution. Aucun autre paramètre de position, tels que la médiane, le mode ou la moyenne géométrique, ne permet une estimation du coût total qui constitue l'information pertinente pour les décideurs de santé.

La variabilité des coûts entre les patients est également importante à décrire. Plusieurs mesures de dispersion sont utilisées, comme l'étendue, l'intervalle interquartile, la variance et l'écart-type. L'étendue, correspond à la différence entre le coût le plus élevé (max) et le coût le plus faible (min). Cette statistique est facile à calculer, mais elle est totalement dépendante de la présence de valeurs extrêmes. L'écart-type (généralement noté SD, pour « Standard Deviation »), qui correspond à la racine carrée de la variance et qui mesure la dispersion autour de la moyenne, est une mesure souvent utilisée. Elle est exprimée en termes monétaires, tout comme la moyenne. Cependant, à cause de l'asymétrie, elle n'est pas idéale pour présenter la variabilité des coûts entre les individus. De ce fait, l'intervalle interquartile, qui correspond à la différence entre les 75^{e} (Q3) et 25^{e} percentiles (Q1) et qui écarte

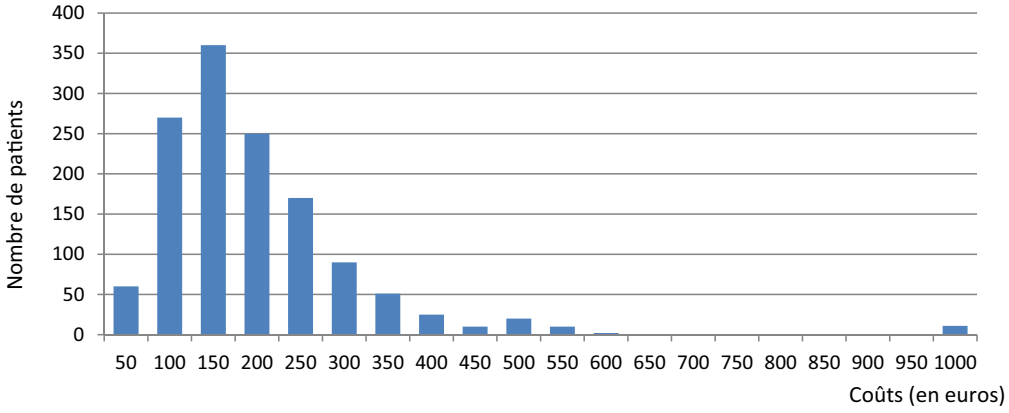

Fig. 1. Distribution typique des coûts par patient, en euros. Exemple fictif.

donc les valeurs de la distribution au-dessus et en dessous de 25 % est une mesure plus pertinente de la variabilité.

3. Comment les coûts doivent-ils être comparés ?

3.1. Analyses univariées

Les tests paramétriques standards (test *t* de Student, et analyse de la variance Anova, lorsque l'on compare plus de deux groupes) pour tester l'égalité de deux moyennes sont basés sur l'hypothèse que les échantillons sont tirés d'une distribution normale avec la même variance. Si le test *t* est connu pour être robuste dans une certaine mesure aux hypothèses de normalité/égalité des variances, il n'est pas clairement établi à partir de quel écart aux hypothèses le test *t* devient inapproprié. La robustesse à la non-normalité des tests paramétriques dépend de la taille de l'échantillon et de la sévérité de l'asymétrie. Pour des échantillons de plus de 150 patients, le test t est généralement robuste [2].

En pratique, le test *t* de Student se réalise avec les commandes suivantes sous SAS ou STATA :

Test de Student sous SAS			***Test de Student sous STATA***
PROC	**TTEST**	DATA = TABLE;	*t*-test COÛT, by (GROUP)
CLASS		GROUP;	
VAR		COÛT;	
RUN;			

Les tests non paramétriques (comme le test de Mann-Whitney ou le test de Kruskall-Wallis lorsque l'on compare plus de deux groupes), généralement recommandés par les statistiques conventionnelles [3] pour les données qui ne suivent pas une distribution normale, sont inappropriés pour les données de coût. En effet, si ces tests statistiques permettent de détecter des différences dans les distributions (forme, dispersion, médiane), ils ne permettent pas de tester si la moyenne arithmétique diffère entre les groupes. Or, comme nous l'avons dit plus haut, pour les économistes, ce sont les coûts moyens du traitement qui sont considérés comme les plus pertinents pour les décisions publiques et l'utilisation de statistiques non-paramétriques, peut conduire à des conclusions erronées [2].

Recommandation:

> L'utilisation de statistiques non paramétriques n'est pas appropriée pour les données de coût.

Pour pallier le problème de la non-normalité des coûts, certains auteurs ont transformé les coûts de façon à ce que la distribution de la variable transformée soit plus proche d'une distribution normale. Un exemple de ces transformations est la transformation logarithmique. Cependant, la transformation en logarithme soulève plusieurs difficultés pour les économistes. Premièrement, lorsque des observations sont nulles, le logarithme du coût est indéfini. La plupart des chercheurs ont résolu ce problème en rajoutant une constante (+1 par exemple) à toutes les observations avant de prendre le logarithme. Cependant, le choix de cette constante affecte certaines conclusions et la présence d'un grand nombre de zéros peut rendre cette approche problématique. De plus, de telles transformations ne permettent pas une comparaison des moyennes arithmétiques (une transformation en logarithme conduit à une comparaison des moyennes géométriques) et sont donc inappropriées pour les données de coûts [2].

Une alternative aux méthodes précédentes pour comparer les coûts entre deux groupes de traitement est le « bootstrap » non paramétrique [2].

La méthode du « bootstrap » non paramétrique également désignée sous le nom de « technique de ré-échantillonnage », permet une comparaison des moyennes arithmétiques des coûts des stratégies sans faire aucune hypothèse sur la distribution des données de coût. Le principe de cette technique est de tirer avec remplacement à partir de l'échantillon observé dans l'étude, un grand nombre d'échantillons aléatoires, de la même taille que l'échantillon initial. À chaque tirage, chaque sujet possède la même probabilité d'être tiré au sort : $1/n$. Une même observation peut apparaître plusieurs fois au sein d'un même échantillon (Fig. 2). Un échantillon « bootstrap » peut donc inclure le coût de certains patients plusieurs fois, alors qu'il exclut le coût d'autres patients. Le processus est répété un grand nombre de fois.

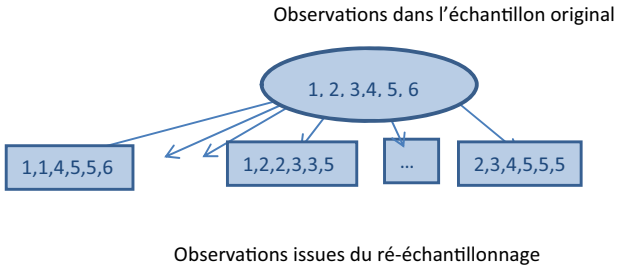

Fig. 2. Principe du « bootstrap ».

Généralement, le nombre d'itérations B (c'est-à-dire le nombre d'échantillons « bootstrap » construits) est de 1000 [4].

Pour chaque échantillon « bootstrap », la différence des coûts moyens est calculée, on obtient ainsi une distribution de B valeurs pour cette différence. L'approche la plus simple pour estimer l'intervalle de confiance de la différence des coûts moyens « bootstrap » est l'approche par percentile. Elle revient à classer les valeurs obtenues par ordre croissant et à déterminer les bornes basses et hautes de l'intervalle de confiance qui sont calculées en utilisant les $B.(\alpha/2)+1$ et $B.(1-\alpha/2)$ percentiles de la distribution (avec B = 1000 en général). Pour un intervalle de confiance à 95 %, les valeurs choisies sont celles qui correspondent à la 26^{e} et à la 975^{e} observation de la différence des moyennes des échantillons « bootstrapés » classées par ordre de valeur croissant. Ces points sont choisis car ils excluent les 25 valeurs (25/1000 = 2,5 %) de chaque côté de la distribution [5].

L'ouvrage très didactique de Gray et al. [6] détaille étape par étape sous Excel la réalisation du « bootstrap » sur le coût de deux stratégies et le calcul de l'intervalle de confiance des différences de coûts « bootstrapés » avec un exemple disponible en ligne[2].

Les commandes sous SAS, tirées de Cassel et al. [7], sont les suivantes :

Calcul de l'intervalle de confiance de la différence des coûts « bootstrapés » sous SAS

```
PROC SURVEYSELECT DATA = BOOTSTRAP OUT = OUTBOOTX
METHOD = URS                    /* TIRAGE AVEC REMISE */
SAMPRATE = 1                    /* N PARMI LES N */
OUTHITS
REP = 1000;                     /* NOMBRE DE REPLICATION */
STRATA BRAS;                    /* PAR STRATÉGIE */
RUN;
PROC MEANS
DATA = OUTBOOTX;
VAR COST;
OUTPUT
OUT = OUTALL1 MEAN = COÛTMOYEN1;

BY REPLICATE;

WHERE BRAS = 1;

RUN;

PROC MEANS DATA = OUTBOOTX MEAN;

VAR COST;

OUTPUT OUT = OUTALL2 MEAN = COÛTMOYEN2;

BY REPLICATE;

WHERE BRAS = 2;

RUN;

DATA FUSION;

MERGE OUTALL1 OUTALL2;

BY REPLICATE;

RUN;

DATA FUSION;

SET FUSION;

DIFF = COÛTMOYEN2-COÛTMOYEN1;

RUN;

PROC SORT DATA = FUSION OUT = TRIE;

BY DIFF;

RUN;

DATA CI_PERC;

SET TRIE END = EOF;

RETAIN CONF_LO CONF_HI;

IF _N_ = 26 THEN CONF_LO = DIFF;

IF _N_ = 975 THEN CONF_HI = DIFF;

IF EOF THEN OUTPUT;

KEEP CONF_LO CONF_HI;

RUN;

PROC PRINT DATA = CI_PERC;

RUN;
```

*http://www.herc.ox.ac.uk/downloads/supporting-material-for-applied-methods-of-cost-effectivenessanalysis-in-healthcare.

On conclut à un différentiel de coût statistiquement significatif lorsque l'intervalle de confiance ne contient pas la valeur 0.

Recommandation:

> Dans les échantillons de petite taille lorsque des écarts à la loi normale sont constatés, nous recommandons la mise en œuvre du « bootstrap » non paramétrique.

[2] http://www.herc.ox.ac.uk/downloads/supporting-material-for-applied-methods-of-cost-effectiveness-analysis-in-healthcare.

La revue de la littérature réalisée par Doshi et al. en 2006 [1] a montré que sur 115 évaluations médicoéconomiques recensées, seulement 92 (80 %) avaient réalisé une comparaison statistique des coûts entre les bras de traitement. Sur les 92 comparaisons statistiques, 62 (68 %) ont utilisé un test statistique approprié à la moyenne arithmétique des coûts (« bootstrap » non paramétrique [23 %] ou test paramétrique sur les données non transformées [45 %]). Des tests non paramétriques ont été effectués dans 18 % des cas.

3.2. Analyses multivariées

Les analyses multivariées complètent les analyses univariées décrites précédemment en recherchant les facteurs explicatifs des coûts autres que le facteur d'intérêt qui est le groupe de traitement. Si elles sont indispensables dans les études non-randomisées (cf. partie 6. Biais de sélection), l'apport de ce type d'analyses est peut-être moins évident pour le cas d'une étude randomisée mais n'est cependant pas inexistant. En effet, ces analyses, parce qu'elles sont plus complètes, permettent d'expliquer des variations de coûts qui seraient dues à d'autres causes (non équilibrées par la randomisation) et augmentent ainsi la puissance statistique liée au test de différence entre les groupes de traitement [8].

Les apports de l'analyse multivariée sont donc non négligeables, mais ne sont cependant valides qu'à condition d'une mise en œuvre adaptée. Pour cette raison, cette partie reviendra sur les méthodes d'analyses multivariées les plus utilisées dans le cadre des évaluations de coûts et mettra en avant les points importants qu'il est nécessaire de prendre en compte pour valider les résultats.

Dans leur revue de littérature de 2006, Doshi et al. [1] montrent que sur les 115 études médicoéconomiques recensées, 80 % ont utilisé un modèle de régression par les moindres carrés ordinaires (dont 70 % sur le coût et 10 % sur le logarithme du coût), 0 % ont utilisé les modèles linéaires généralisés, et 20 % d'autres modèles (dont le Tobit).

Nous partirons de la méthode qui a été la plus souvent utilisée dans la littérature [1] du fait de sa simplicité et de sa facilité de mise œuvre et nous présenterons ensuite des approches alternatives plus sophistiquées et plus adaptées aux données de coût.

Recommandation:

> Dans le cadre d'un essai non randomisé, les analyses multivariées permettent de prendre en compte le biais de sélection (cf. partie 4) et donc plus simplement de s'assurer du caractère non hasardeux d'un potentiel différentiel de coût observé entre les groupes de traitement.

3.2.1. La méthode des Moindres carrés ordinaires (MCO)

La technique des moindres carrés ordinaires (MCO) est la technique la plus simple pour estimer un modèle linéaire. Revenons rapidement sur la construction d'un tel modèle pour définir les concepts clés nécessaires à la compréhension de ce type de modélisations. Prenons tout d'abord le modèle linéaire le plus simple n'intégrant que la variable de traitement comme variable explicative du coût :

$$COUT_i = \alpha + \Delta_c GROUP_i + \varepsilon_i, \quad i = 1, \ldots, n$$

Avec :

- $COUT_i$ coût total observé pour le patient i sur la période d'étude ;
- α la constante ;
- $GROUP_i$ la variable de traitement prenant la valeur 1 si le patient est traité 0 sinon ;
- ε le terme d'erreur.

Le coefficient Δ_c associé à la variable de traitement mesure le coût incrémental du traitement. L'estimation d'un tel modèle revient à estimer la différence des coûts des deux bras de l'étude. À ce stade, l'intérêt par rapport à une comparaison issue de la différence des moyennes de coûts des différents bras est inexistant. L'avantage de la régression linéaire vient du fait qu'il est possible d'ajouter des variables autres que celle de traitement (on parle d'« ajustement »). On obtiendra alors le coût incrémental du traitement ajusté de l'effet d'autres variables explicatives (variables n'ayant pas pu être contre-balancées par la randomisation par exemple). Le modèle devient :avec x_{ik} les k variables explicatives pour le patient i.

La construction de ce modèle permet ainsi d'estimer le coût incrémental du traitement contrôlé des effets des k variables explicatives introduites dans le modèle.

En termes techniques, pour fournir des résultats robustes, cet estimateur doit être précis – *à variance minimale* — et sans biais — *l'espérance de l'estimateur coïncide avec la vraie valeur du paramètre* [9]. Ces deux caractéristiques sont obtenues sous certaines hypothèses sous- jacentes à l'utilisation des MCO (cf. Encadré 1). L'utilisation de cette technique d'estimation recommande la mise en œuvre de certains tests pour vérifier ces hypothèses.

En pratique, la régression linéaire se réalise avec les commandes suivantes sous SAS et STATA :

Régression linéaire sous SAS
PROC REG DATA = TABLE;
MODEL **COÛT = X_1 X_2 X_3… X_k;**
RUN;
Régression linéaire sous STATAreg COÛT X_1 X_2 X_3… X_k

Avant même d'envisager la mise en œuvre des tests décrits dans l'Encadré 2, les caractéristiques de la distribution des données de coûts décrites précédemment suffisent à remettre en cause certaines de ces hypothèses et donc plus généralement l'utilisation des MCO comme technique d'estimation. Plus précisément, deux hypothèses principales sont en général violées par la distribution de coûts observée : la normalité des résidus et l'hypothèse d'homoscédasticité [10]. Le premier réflexe de l'économiste pour pallier ces deux problèmes — hétéroscédasticité et non-normalité des données — est généralement d'opérer une transformation logarithmique :

Encadré 1. Les hypothèses sous-jacentes à l'utilisation des MCO

Les variables explicatives X_k sont non stochastiques, c'est-à-dire qu'il ne s'agit pas de variables aléatoires.
L'espérance mathématique des termes d'erreurs u est nulle : $E(\beta_j) = 0$, cela signifie que les termes d'erreur positifs et négatifs se compensent en moyenne et donc que le modèle est bien spécifié.
La variance des termes d'erreurs u β est constante et ne dépend pas des observations : hypothèse d'homoscédasticité.
Les termes d'erreurs sont indépendants et donc non corrélés entre eux : hypothèse d'absence d'autocorrélation.
L'erreur est indépendante des variables explicatives, c'est-à-dire Cov $(X_j, \beta_j) = 0$: hypothèse d'exogénéité.
Les erreurs sont indépendamment et identiquement distribuées selon une loi normale, il s'agit de l'hypothèse de normalité des erreurs.
Le respect de ces hypothèses fait de l'estimateur MCO un estimateur sans biais et à variance minimale.
Pour plus d'explications concernant ces hypothèses, se référer au manuel de Green [9].

L'estimation se réalise en ayant recours à la méthode des MCO, présentée ci-dessus en utilisant simplement le logarithme népérien de la variable de coût comme variable à expliquer au lieu de la variable en niveau.

Comme cette méthode soulève cependant un certain nombre de problèmes déjà évoqués, nous proposons le recours à des méthodes permettant une plus grande flexibilité en termes de spécification – *les modèles linéaires généralisés*. Ils permettent de répondre aux problèmes de la non-normalité de la distribution des coûts, du nombre important de valeurs nulles, ainsi qu'à la question de la comparaison des moyennes arithmétiques (et non géométriques comme le permet la transformation logarithmique).

Recommandations:

Dans le cadre d'une analyse multivariée sur les données de coût, l'application de la méthode des MCO sur les données de coût ne semble pas appropriée et donc le recours à des modèles plus sophistiqués est préférable.

Le choix du modèle de coût n'est pas toujours évident en pratique : l'analyste doit estimer et reporter plusieurs modèles, fournissant au lecteur une mesure du degré de variation dans les estimations [8].

3.2.2. Les modèles linéaires généralisés – « Generalized linear models »

Bien que leur utilisation soit rare dans le cadre des évaluations de coûts [1], les modèles linéaires généralisés peuvent apporter une réponse satisfaisante à la question de la modélisation des données de coûts. Ils consistent à utiliser une transformation mathématique des données de coûts en tenant compte de la véritable distribution des erreurs et non en imposant l'hypothèse de normalité fortement contestable. Leur premier avantage est d'offrir un degré de flexibilité quant à la modélisation de la moyenne et de la variance, ce qui résout une partie des problèmes générés par les spécificités des données de coût. Leur deuxième avantage est de modéliser directement la moyenne arithmétique des coûts, qui, rappelons-le est la seule mesure pertinente.

Sur le plan pratique, la mise en œuvre de ces modèles implique une estimation par maximum de vraisemblance et le choix d'une fonction de lien — « link function » — ainsi que d'une famille de distribution — « family » — en fonction des données étudiées (cf. Encadré 2).

Pour la modélisation des dépenses de santé, la fonction de lien « log » associée à une variance distribuée selon une loi gamma est le cadre le plus usuellement utilisé [10–12]. La loi gamma est en effet une loi de probabilité des variables aléatoires réelles positives. Elle permet également de traiter les variables dont la distribution est asymétrique. Ces deux conditions semblent parfaitement correspondre aux caractéristiques de la distribution des données de coûts. Il existe cependant certaines techniques pour s'assurer de la validité de ce choix par rapport à d'autres possibilités (cf. Encadré 2). Pour plus de détails, on peut se reporter par exemple à l'article de Manning [11] qui propose un algorithme permettant de choisir le meilleur estimateur.

Encadré 2. Mise en œuvre des modèles linéaires généralisés

La **famille** renvoie à la distribution caractérisant le lien entre la moyenne et la variance des données. On peut, par exemple, avoir recours à une loi gaussienne indiquant une variance constante, à une loi de Poisson indiquant une variance proportionnelle à la moyenne — adaptée aux variables de comptage —, à une loi gamma indiquant une variance proportionnelle au carré de la moyenne, à une loi gaussienne inverse indiquant une variance proportionnelle au cube de la moyenne, etc.
Le « Modified Parks Test » est un test pour sélectionner la famille appropriée aux données dont on dispose [11]. Le logiciel STATA permet par exemple de choisir entre sept fonctions de lien.
La **fonction de lien** exprime la relation fonctionnelle entre les coûts et les variables retenues en tant que prédicteur (variables explicatives).
Le choix de cette fonction de lien est l'une des principales difficultés relatives à la mise en œuvre de ces modèles. La démarche suivante peut être appliquée :

- mise en œuvre successive des mêmes modèles en changeant seulement la fonction de lien ;
- recours à des tests tels que « Pregibon Link Test » [13], « Modified Hosmer-Lemeshow Test » [14], ou encore « Copas Test » [15].

En pratique, la mise en œuvre de modèles linaires généralisés se réalise avec les commandes suivants sous SAS et STATA :

Modèle linéaire généralisé sous SASproc genmod data = -table;model COÛT = X_1 X_2 X_3... X_k/link = ''linkname'' dist = ''familyname'';run;

Modèle linéaire généralisé sous STATAglm COÛT X_1 X_2 X_3... X_k, family(familyname) link(linkname)

3.2.3. Les modèles en deux étapes – « Two-part Models »

Les données de coûts peuvent se caractériser par un nombre relativement important de valeurs nulles. Ce problème peut être traité par le recours à une modélisation en deux étapes (« Two-part Models » en anglais). Comme son nom l'indique, cette méthode implique une estimation en deux étapes : la première consiste à prédire la probabilité que le patient soit associé à un coût nul ou non et la deuxième consiste à estimer le montant — comme dans une modélisation simple — conditionnellement au fait que ce coût soit non nul (cf. Encadré 3). L'intérêt de ce type de modèle est d'exploiter la dimension mixte de la distribution des données de coûts (part non négligeable de coûts nuls et autre partie de la distribution plus continue avec des montants positifs). Cette méthode permet ainsi de savoir si la baisse de coûts associée à un traitement est attribuable à une modification (à la hausse) de la probabilité d'avoir un coût nul, à une diminution directe du niveau des coûts lui-même ou bien encore à un mixte des deux.

En pratique, l'estimation d'un Tobit se réalise avec les commandes suivantes sous SAS ou STATA :

Estimation d'un Tobit sous SAS

PROC QLIM DATA = TABLE;

MODEL *COÛT* = X_1 X_2 X_3... X_k;

ENDOGENOUS Y ∼ CENSORED (LB = 0);

RUN;

Ou encore

PROC LIFEREG DATA = TABLE;

MODEL (LOW, COÛT) = X1 X2 X3... XK/DIST = NORMAL;

RUN;

Estimation d'un Tobit sous STATA

TOBIT COÛT X_1 X_2 X_3... X_k, LL(*) UL(0*)

Il est nécessaire de spécifier au moins un des deux éléments suivants : ll(*), ul(*) où ll(*) correspond à la spécification de la valeur limite (*) de la censure à gauche et ul(*) à la spécification de la valeur limite (*) de la censure à droite. Dans notre cadre d'analyse, il s'agit de spécifier ll(0) pour indiquer la censure à gauche relative à la valeur nulle.

3.3. Les analyses multiniveaux

Les évaluations de coûts s'inscrivent de plus en plus souvent dans le cadre d'essais multicentriques et internationaux. Dans ce cadre-là, même en présence d'un essai randomisé, se pose la question des effets centres et des effets pays évoqués précédemment. Ces effets renvoient au fait que d'un centre à l'autre ou d'un pays à l'autre, les pratiques médicales peuvent différer de même que le « case-mix » des patients ou encore la tarification associée aux soins. Sans précaution particulière, une évaluation de coût suppose que les données relatives aux patients sont indépendantes du lieu de collecte des données. Or

Encadré 3. Mise en œuvre des modèles en deux étapes

Sur le plan pratique, la première étape consiste à estimer :

Prob ($COÛT_i > 0$) = ϕ ($x'\beta$)

Où ϕ représente la fonction de distribution cumulée de la loi normale – *dans le cas de l'utilisation d'un modèle Probit.*

La deuxième partie du modèle, qui se focalise sur les coûts conditionnellement au fait que ceux-ci soient non nuls — information obtenue suite à la première étape — est estimée à l'aide d'un modèle linéaire standard (modèles présentés précédemment), qui laisse à nouveau une certaine flexibilité en termes de spécification comme évoqué dans le cadre des modèles linéaires généralisés.

La méthode d'Heckman en deux étapes permet de réaliser ce type d'analyses.

Les logiciels de statistiques permettent d'estimer simultanément ces deux équations, donc plus rapidement et plus précisément. Il s'agit de l'estimation d'un modèle Tobit.

Pour aller plus loin...

Il existe en réalité deux types de modèles Tobit : le Tobit type 1 et le Tobit type 2 ou encore le Tobit simple et le Tobit généralisé type 2.

Le Tobit 1 permet de tenir compte spécifiquement des coûts nuls pour estimer de manière plus adaptée l'impact des variables explicatives. L'impact alors mesuré ne dissocie pas l'impact des variables explicatives sur la probabilité que le coût soit nul — ou non — (étape 1) et sur la variabilité des montants des coûts positifs (étape 2). Il correspond à une sorte de moyenne de ces deux effets combinés.

Le Tobit 2 permet un raffinement par rapport à cette analyse : il permet une décomposition des impacts liés aux deux étapes d'estimation. Ce modèle permet ainsi d'expliquer d'une part la probabilité que le coût soit nul, et d'autre part la variabilité de ces coûts dès lors qu'ils sont supérieurs à zéro. Autrement dit, lorsqu'un Tobit 1 estime l'effet moyen d'une variable sur la probabilité que le coût soit nul et son impact sur la variabilité des coûts non nuls, le Tobit 2 décompose cet effet en deux et est ainsi plus explicatif. Cette décomposition offre également la possibilité de différencier les variables explicatives entre l'étape 1 et l'étape 2 si bien que si l'on suppose que certaines variables jouent uniquement sur la probabilité que le coût soit nul et pas sur la variabilité des coûts non nuls, le modèle Tobit 2 permet de prendre en compte cette hypothèse en incluant ces variables qu'en étape 1. Pour l'estimation d'un Tobit type 2, il est d'ailleurs recommandé de trouver une variable qui serait introduite dans l'étape 1 mais pas dans l'étape 2 (c'est-à-dire explicative de la probabilité de supporter ou non un coût, mais pas du montant en lui-même).

Pour plus d'informations, voir l'ouvrage de Verbeek, Wiley et Sons [16] ainsi que celui de Crépon et Jacquemet [17].

ces différences de pratiques peuvent être la cause d'une corrélation entre patients appartenant au même sous-groupe que constitue le centre, le pays, etc. Dans cette configuration l'hypothèse d'indépendance des observations est violée. Une réponse possible à cette problématique peut être apportée par les analyses multiniveaux.

Recommandation:

> Dans le cadre d'essais multicentriques et internationaux, il est important de prendre en compte de potentiels effets centre ou effets pays grâce à l'utilisation d'analyses multiniveaux.

En termes techniques, les modèles multiniveaux permettent de structurer les données sur différents niveaux pour permettre une décomposition en effets fixes et effets aléatoires. On peut ainsi percevoir plus finement l'hétérogénéité non observée entre ces sous-groupes. Ces modèles vont donc permettre d'intégrer la dimension hétérogène des coûts d'un centre à l'autre ou d'un pays à l'autre par exemple. On peut ainsi prendre en compte des variables explicatives différentes selon le niveau du modèle considéré. Par exemple au niveau du patient, on intègre des variables relatives aux caractéristiques du patient (sexe, âge, etc.) et au second niveau, au niveau du centre par exemple, on intègre des variables caractéristiques de l'établissement (statut, type d'activité, etc.). L'intérêt de ces modèles est d'évaluer l'hétérogénéité des coûts à différents niveaux (celui du patient, du centre ou encore du pays par exemple) et ainsi d'éliminer l'impact de chaque caractéristique du patient des éventuelles différences de traitement des patients appartenant à un même établissement ou un même pays. Ne pas prendre en compte un effet multiniveau revient notamment à sous-estimer la variance des effets centre, pays, etc., et donc à surestimer leur significativité.

En pratique, les analyses multiniveaux se réalisent avec les commandes suivantes sous SAS ou STATA :

Analyses multiniveaux sous SAS

Via une procédure mixed dans le cadre linéaire

Via les procédures glimmix et nlmixed dans le cadre non linéaire

Analyses multiniveaux sous STATA

Via la commande xtmixed dans le cadre linéaire

Via les commandes (par exemple) xtmelogit et xtpoisson dans le cadre non linéaire

4. Traitements des données manquantes

Comme pour beaucoup d'études en sciences sociales, les données manquantes sont fréquemment rencontrées dans les évaluations économiques. On parle de valeurs manquantes monotones si la variable Y_j manquante pour un individu, implique que toutes les variables suivantes Y_k pour k > j sont manquantes pour cet individu. Les données manquantes sont dites arbitraires ou non-monotones lorsqu'elles suivent une structure non monotone et sont réparties uniformément dans la base de données. Ignorer les données manquantes peut entraîner outre une perte de précision, de forts biais dans les analyses. En économie, bien que le principe des données manquantes ne soit pas différent des autres types de données manquantes, la forme de la distribution des données de coût — généralement fortement asymétrique — constitue un défi supplémentaire pour la personne chargée d'analyser les données. Afin de minimiser les non-réponses, il est recommandé de soigner le questionnaire économique avec un souci de clarté et de concision, en vérifiant la pertinence et la compréhensibilité des items. Il est également conseillé de limiter les informations à recueillir aux données essentielles à l'analyse et d'éliminer les données superflues, un trop grand nombre de données à recueillir pouvant conduire à un abandon par lassitude en cours de recueil. Le protocole de l'étude doit prévoir également un contrôle qualité adéquat des données économiques, avec un envoi de requêtes en cas de donnée manquante ou d'incohérence. Des vérifications sur site et/ou des entretiens téléphoniques doivent également être prévus afin de vérifier les informations recueillies et de mener des actions préventives pour limiter d'autres données manquantes.

Recommandation:

> L'économiste doit mettre tout en place pour ne pas avoir de valeurs manquantes [18], puisqu'il n'existe pas de méthode totalement efficace pour traiter de ce problème.

4.1. Rappel général du problème de données manquantes

Trois mécanismes de données manquantes ont été décrits par Little et Rubin [19] :

- les données manquantes sont dites « Missing Completely At Random » (MCAR) si la probabilité d'avoir une donnée manquante sur Y est une constante indépendante des valeurs de la variable d'intérêt Y et des autres variables observées. Cela signifie que l'échantillon des données observées est représentatif de l'ensemble des données, et donc que les individus avec des valeurs manquantes sont un échantillon aléatoire de l'échantillon initial. Les données MCAR entraînent une perte de puissance statistique mais pas de biais puisqu'il n'y a pas de différence systématique entre ceux ayant des valeurs observées et ceux ayant des données manquantes. *Exemple de données MCAR :* soit un questionnaire distribué aux patients pour recueillir leurs consommations de soins à la suite d'une intervention. Les non-réponses suivent un mécanisme MCAR si les raisons de ne pas compléter le questionnaire ne sont reliées à aucune des variables de l'étude ;
- les données manquantes sont dites « Missing At Random » (MAR) si la probabilité d'avoir des observations manquantes sur la variable d'intérêt Y dépend des autres variables observées X. Les données MAR entraînent une perte de

puissance, mais pas de biais si des méthodes statistiques appropriées sont appliquées. *Exemples de données MAR* : lorsque les personnes âgées refusent de donner leurs revenus ;
- les données manquantes sont « Missing Not At Random » (MNAR) si la probabilité d'avoir des observations manquantes dépend des valeurs non observées de la variable d'intérêt, *i.e.* des vraies valeurs. Cette situation est plus problématique car elle entraîne une perte de puissance et des estimations biaisées. *Exemples de données MNAR :* lorsque les personnes avec un revenu important refusent de dévoiler leur revenu.

Recommandations:

Il est conseillé de discuter avec toutes les personnes en charge du projet des mécanismes possibles des données manquantes. Il est généralement impossible de savoir à partir des données observées le vrai mécanisme (MCAR, MAR ou MNAR). La méthode décrite dans la littérature pour tester si les données manquantes sont de type MCAR est de comparer les individus manquants et non manquants sur toutes les covariables en utilisant des tests de Student. Si des différences significatives de moyennes entre les deux groupes sont trouvées, le mécanisme MCAR peut être rejeté [19,20]. Si le mécanisme MCAR a été réfuté, il reste à vérifier si le mécanisme est MAR ou MNAR. Ces deux mécanismes sont très difficiles à distinguer étant donné que les données non observées sont, par définition, inconnues [19,21–23].

4.2. Les différentes méthodes de traitement des données manquantes

Cette partie expose les principales méthodes utilisées pour traiter les données manquantes, en en détaillant les hypothèses sur lesquelles elles reposent. En dessous de 5 % de données manquantes, quel que soit le mécanisme, les données peuvent être gérées comme si elles obéissaient à un mécanisme MCAR, donc par les méthodes les plus simples (analyse des cas complets et imputation simple) sans risque de biais important. Au-delà, la méthode à privilégier est l'imputation multiple [24].

4.2.1. Analyse des cas complets (« Complete Case Analysis »)

Cette méthode consiste à utiliser dans les analyses uniquement les cas qui ne comportent aucune donnée manquante. Cette technique, du fait de sa simplicité est fréquemment l'option implicite pour l'analyse dans la plupart des logiciels statistiques. Deux limites peuvent être faites sur cette approche. Premièrement, cette méthode sacrifie un grand nombre de données (beaucoup d'observations peuvent disparaître), entraînant une diminution de la puissance statistique. Deuxièmement, à moins que le mécanisme de génération des données ne soit complètement aléatoire (MCAR), les statistiques, telles que la moyenne ou la variance, seront fortement biaisées [25].

Cependant, cette technique reste raisonnable si les données manquantes sont en faible proportion (un seuil de 5 % est souvent cité dans la littérature).

4.2.2. Procédures de remplacement

Les procédures de remplacement visent à se ramener à une base complète, en trouvant un moyen adéquat de remplacer les valeurs manquantes. On nomme ce procédé « imputation », « complétion » ou « substitution ». Trois méthodes d'imputation seront présentées ci-après : l'imputation simple, les méthodes du maximum de vraisemblance et l'imputation multiple.

4.2.2.1. Imputation simple. L'imputation simple consiste à remplacer les données manquantes par une seule valeur. La donnée imputée est considérée comme une donnée observée. Ces méthodes entraînent généralement une diminution de la variance et peuvent altérer la force des associations identifiées. Il existe à ce jour une multitude de techniques développées à cet effet (prédiction par la moyenne ou par la médiane, prévision par un modèle de régression, « Last observation carries forward » (LOCF), « hot-deck », etc.).

4.2.2.2. Les méthodes du maximum de vraisemblance (« Maximum Likelihood Approaches »). Cette méthode repose sur l'hypothèse que les données sont MAR. Sous sa forme la plus simple, l'approche par maximum de vraisemblance pour analyser des données manquantes, suppose que les données observées sont tirées d'une distribution normale multivariée. Le principe de l'algorithme de maximisation d'espérance (EM) est d'estimer les paramètres d'intérêt suivant la méthode du maximum de vraisemblance à partir des données complètement observées ; on estime ensuite les données manquantes à partir de cette première estimation des paramètres. Puis une seconde estimation des paramètres d'intérêt est reconduite sur l'ensemble des individus (complets et manquants) et une nouvelle estimation des données manquantes est réalisée à partir des nouvelles estimations des paramètres, et on répète le processus jusqu'à ce que la convergence soit atteinte. À la fin de l'exécution de l'algorithme, on dispose non seulement des paramètres de notre modèle, mais également d'une matrice de données complétée. Cette technique est très coûteuse en temps de calculs. De plus, elle demande la spécification d'un modèle de génération des données. Cette tâche implique de faire un certain nombre d'hypothèses, ce qui est toujours délicat. Pour ces raisons, l'application d'EM pour remplacer les données manquantes n'est pas toujours envisageable.

4.2.2.3. Imputation multiple. Cette méthode repose sur l'hypothèse que les données sont MAR. L'imputation multiple consiste, comme son nom l'indique, à imputer plusieurs fois les valeurs manquantes afin de combiner les résultats pour diminuer l'erreur due à l'imputation. Généralement, tous les processus d'imputation multiple ont en commun trois phases :

- phase d'imputation, dans laquelle on attribue des valeurs aux données manquantes en utilisant un modèle aléatoire adapté ;
- phase d'analyse séparée, en utilisant les méthodes statistiques standards pour l'analyse des données complètes pour estimer les quantités d'intérêt (par exemple, coût attendu dans chaque groupe de traitement sur l'horizon temporel). À la fin de cette seconde étape, m paramètres sont obtenus à partir de m analyses séparées ;
- phase d'analyse combinée : elle consiste à combiner les m estimations en une seule.

La Fig. 3 illustre les trois étapes de l'imputation multiple.

Plus le nombre d'imputations est élevé, plus les estimateurs seront précis. En pratique, le manuel de référence du logiciel Stata (StatCorp, Release 13) recommande d'utiliser au moins 20 imputations pour réduire l'erreur d'échantillonnage mais un petit nombre d'imputations (5 à 20) peut être suffisant lorsque le pourcentage de données manquantes est faible. De hautes proportions de données manquantes nécessitent au moins 100 imputations. Lorsque c'est possible, il est recommandé de varier le nombre d'imputations pour voir si cela affecte les résultats.

L'imputation peut se faire soit au niveau des consommations de ressource, soit au niveau de la valorisation des ressources, mais étant donné qu'il est généralement recommandé de présenter les consommations de ressources, et que différents types de ressources peuvent avoir différents schémas de données manquantes, l'imputation au niveau des consommations de ressources peut être une meilleure approche [26].

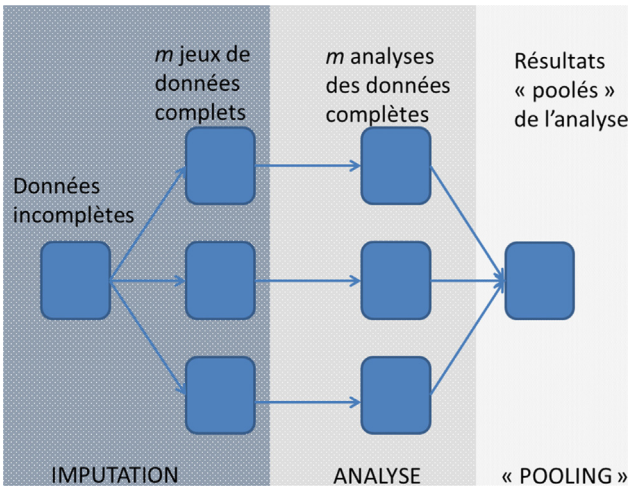

Fig. 3. Schéma d'une imputation multiple.

Recommandations:

Il est recommandé de débuter par une analyse descriptive des données manquantes. Une description appropriée des données manquantes doit contenir :

- le nombre et la proportion de données manquantes par variable et par stratégie de traitement ;
- le type de variables (quantitative, continue) et la forme de la distribution avec et sans données manquantes ;
- le schéma des données manquantes (monotone versus non monotone) ;
- le mécanisme des données manquantes (MCAR, MAR,...).

Le choix de la méthode d'imputation repose en effet sur les schémas de non-réponse et le type de variables imputées. Par exemple, pour les données qui ont un motif de non-réponse monotone, l'imputation d'une variable continue peut se faire avec des méthodes paramétriques sous l'hypothèse de normalité multivariée, comme la méthode de régression [27] et la méthode « Predictive mean matching » [28,29]), ou bien avec des méthodes non paramétriques, comme le score de propension [30]. Lorsque les données sont catégorielles (binaires ou ordonnées), une méthode de régression logistique peut être appliquée.

Quand les données sont manquantes de façon arbitraire, on peut utiliser la méthode « Markov Chain Monte Carlo » (MCMC) [31] sous l'hypothèse de normalité multivariée ou bien par équations chaînées (« Multiple Imputation by Chained Equations », MICE) dénommée aussi « Fully conditionnal specification » si la donnée manquante est catégorielle ou non normale [32–35].

Le Tableau 1 synthétise les principales méthodes d'imputation multiple.

Un schéma de type monotone simplifie grandement la tâche d'imputation. Une imputation multivariée peut être formulée comme une séquence d'imputations univariées. Si le motif des données manquantes est monotone, la complétion sera séquentielle. Si le motif des données manquantes est non monotone, il existe deux possibilités : soit de rendre les données monotones avec la méthode MCMC puis utiliser une méthode pour données manquantes monotones, soit compléter entièrement les données manquantes uniquement par MCMC.

Bien que la régression ou la méthode MCMC reposent sur l'hypothèse de normalité, les inférences basées sur l'imputation multiple peuvent être robustes à la non normalité si le taux de données manquantes n'est pas élevé, car le modèle d'imputation est appliqué non pas à l'ensemble des données mais seulement à sa partie manquante [31]. Il est possible aussi de transformer les données pour satisfaire l'hypothèse de

Tableau 1
Méthodes d'imputation multiples.

Schéma de données manquantes	Type de la variable à imputer	Méthode d'imputation possible
Monotone	Continue	Méthodes paramétriques qui reposent sur l'hypothèse de normalité multivariée : régression monotone ; « Predictive mean matching » Méthode non paramétrique : score de propension
	Ordinal	Régression logistique
Arbitraire (non monotone)	Continue	« Markov Chain Monte Carlo » (MCMC) : imputation totale (« full-data imputation ») ou partielle[a] FCS (« Fully Conditional Specification ») regression FCS « Predicted mean matching »
	Ordinal	FCS « logistic regression »

Tableau modifiée tirée de « The MI Procedure Imputation Methods, SAS 9.3, The SAS Institute http://support.sas.com/documentation/cdl/en/statug/63962/HTML/default/viewer.htm#statug_mi_sect019.htm

[a] Imputation des données non monotones avec MCMC afin d'obtenir une disposition des données manquantes monotones. Puis imputer le reste à l'aide de méthodes monotones.

normalité. Les variables sont transformées avant le processus d'imputation et sont retransformées pour créer les données imputées. Une autre option est d'utiliser la méthode du « Predictive mean matching », qui peut être plus appropriée que la méthode de régression quand l'hypothèse de normalité n'est pas respectée [36]. Dans cette méthode, on impute par la valeur observée la plus proche de la valeur prédite issue du modèle de simulation. Cette méthode permet d'éviter les valeurs aberrantes (comme les coûts qui sont toujours positifs). Les modèles en deux étapes (« Two-part Models ») peuvent être utilisés pour les variables avec une large proportion de zéro (voir partie 3.2 Analyses multivariées).

La stratégie de sélection des co-variables pour le modèle d'imputation consiste à prendre le plus de variables possibles liées à la variable à imputer. Par exemple, Briggs et al. [37], sur une étude randomisée comparant deux traitements du diabète de type 2, ont régressé le logarithme de la durée de séjour sur l'année, l'âge et l'indice de masse corporelle des patients, la spécialité (cardiologie, chirurgie, oncologie, etc.). Faria et al. [26], sur une étude randomisée comparant deux stratégies sur les reflux gastro-œsophagiens avec une durée de suivi de cinq ans, ont utilisé pour imputer les coûts à cinq ans comme co-variables les caractéristiques de base des patients, les coûts à 1–4 ans et QALYs à 1–5 ans.

L'imputation multiple avec SAS s'effectue avec la procédure « MI ». Par défaut, la procédure utilise la méthode MCMC pour créer cinq imputations. Les données peuvent être ensuite analysées avec des procédures standard de statistique par imputation (by _imputation). La procédure « MIANALYZE » combine ensuite les résultats des analyses des imputations et génère des inférences statistiques valides. Des exemples sous SAS peuvent être trouvés *SAS/STAT(R) 9.2 User's Guide, Second Edition*. Chapitre 11. The MI procedure et dans l'ouvrage « Multiple Imputation of Missing Data Using SAS » de Berglund et al. [38]. Des exemples STATA peuvent être trouvés dans le manuel « Multiple-Imputation référence Manual-Release 13 »[3].

5. Coûts censurés

Les données censurées peuvent être vues comme un type particulier de données manquantes. Lors d'un essai clinique, un recueil des coûts est prévu sur un laps de temps défini ou jusqu'au décès du sujet. Durant la période de suivi, il arrive que certains sujets quittent l'étude précocement (les perdus de vue). Dès lors, on parle de données censurées. La probabilité de censure croît évidemment avec la durée de suivi. En présence de données censurées, le coût moyen observé constitue une sous-estimation du coût moyen réel puisque les coûts survenant après la sortie d'étude ne sont pas pris en compte dans le calcul. Restreindre l'analyse aux seuls sujets suivis sur l'ensemble de la période d'intérêt ne permet pas de résoudre correctement le problème car un poids trop important est alors accordé aux sujets décédés au cours de l'étude. En effet, les sujets survivants ont une probabilité plus importante d'être censurés que les sujets décédés puisqu'ils sont observés plus longtemps. Plusieurs méthodes sont proposées ci-dessous pour traiter ce type de problèmes.

5.1. Modèle de Lin, 1997

Lin et al. [39] ont proposé une méthode par intervalle pour analyser les données de coût censurées lorsque les dates d'apparition des coûts sont connues (historique des coûts). Ce cas de figure est le plus fréquent en pratique. En effet, le protocole prévoit généralement la collecte des dates auxquelles les soins sont dispensés dans le cahier de recueil clinique ou a minima le recueil des données de coût lors des visites intermédiaires de suivi. Premièrement, la période de suivi est divisée en intervalles de temps les plus réduits possible, par exemple, en mois. Deuxièmement, on calcule la probabilité que les participants survivent au début de chaque intervalle, par exemple au début du mois, en utilisant la méthode de Kaplan-Meier. Troisièmement, on calcule le coût moyen par intervalle. Ces coûts moyens par intervalle sont ensuite multipliés par la

[3] http://www.stata.com/manuals13/mi.pdf.

probabilité de survie de Kaplan-Meier de leur intervalle. Enfin, il faut additionner les coûts pondérés par intervalle pour obtenir le coût moyen total par patient.

Dans le même article, Lin et al. [39] ont proposé une deuxième approche si l'on ne dispose pas de l'historique des coûts pour chaque patient (mais uniquement du coût cumulé par patient). Dans cette méthode, l'estimateur sans historique des coûts, un coût moyen pour chaque intervalle est calculé en faisant la moyenne des coûts cumulés des patients décédés dans l'intervalle. Ces coûts moyens par intervalle sont multipliés par la probabilité de décéder durant l'intervalle correspondant (différence des probabilités de survie de Kaplan-Meier au début et à la fin de l'intervalle). Il convient enfin d'additionner les coûts pondérés par intervalle pour obtenir le coût moyen total par patient.

Ces méthodes sont considérées comme appropriées lorsque le mécanisme de censure est de type MCAR.

5.2. Modèle de Bang et Tsiatis, 2000

Cette méthode est une variante de la méthode précédente [40]. Deux estimateurs sont proposés, selon que les dates d'apparition des coûts sont connues (« Partitioned estimator ») ou non (« Simple weighted estimator »). Cette méthode est considérée comme appropriée lorsque le mécanisme de censure est de type MCAR.

- pour l'estimateur avec historique de coût (« Partitioned estimator ») il convient premièrement, de diviser le temps de suivi en intervalles, deuxièmement, de calculer pour chaque intervalle, la probabilité d'être observé (donc de ne pas être censuré) grâce à la méthode de Kaplan-Meier inversé et troisièmement de multiplier pour chaque patient les coûts observés par l'inverse de la probabilité d'être observé correspondante à l'intervalle dans lequel le coût est apparu. Les coûts ainsi pondérés sont additionnés puis divisés par le nombre total de patients pour obtenir un coût moyen total ;
- pour l'estimateur sans l'historique des coûts (« Simple weighted estimator »), seuls les coûts cumulés des patients avec un suivi complet (décédés ou suivis durant toute la durée de l'étude) sont pondérés par l'inverse de la probabilité d'être observé. Les coûts ainsi pondérés sont additionnés puis divisés par le nombre de patients (censurés et non censurés) pour obtenir un coût moyen total.

5.3. Modèle de Carides, 2000

De manière générale, on peut supposer que les coûts totaux des patients sont liés au temps de survie. La méthode de régression développée par Carides et ses collègues se fonde sur cette hypothèse [41]. Le calcul de l'estimateur est réalisé en deux étapes, uniquement sur les patients avec un suivi total. Dans un premier temps, il convient de réaliser un modèle de régression dans lequel le coût total par patient est expliqué par le délai jusqu'au décès. Les auteurs ne préconisent pas de méthode particulière, des méthodes tant paramétriques que non paramétriques (avec ou sans transformation des variables) sont utilisables pour correspondre au mieux aux données. Ensuite, la durée de suivi est divisée en intervalles, les coûts prédits pour chaque intervalle sont déterminés par le modèle de régression à partir des coûts cumulés par patient. Les coûts par intervalle ainsi obtenus sont multipliés par la probabilité de survie de Kaplan-Meier de l'intervalle correspondant. Enfin, le coût total est obtenu par addition des coûts pondérés de chaque intervalle.

La méthode de Carides peut être considérée comme appropriée lorsque les censures sont de type MAR, sous l'hypothèse qu'aucune autre variable que le délai jusqu'au décès n'est lié à la censure.

5.4. Modèle de Lin, 2000

Lorsque des variables influençant la présence de censures sont connues, il est possible d'en tenir compte dans le modèle de prédiction des coûts. Un nouveau modèle a été proposé par Lin pour tenir compte de ce phénomène [42]. Cette méthode parait appropriée pour des cas où la censure est de type MAR. Comme pour les méthodes non paramétriques déjà décrites [39,40] deux estimateurs sont proposés selon que les délais d'apparitions des coûts sont connus ou pas.

Si l'historique des coûts n'est pas connu, il est proposé de réaliser une régression linéaire des moindres carrés pondérés (« Weighted ordinary least squares ») pour tenir compte des variables influençant les coûts. Ainsi, les coûts totaux des patients avec un suivi complet sont expliqués par des caractéristiques cliniques des patients. La pondération est effectuée au moyen de l'inverse de la probabilité d'être observé (méthode de Kaplan-Meier inversé) dans l'intervalle. Le modèle est appliqué à tous les patients afin d'obtenir un coût prédit par patient permettant de calculer un coût moyen par patient.

Dans le cas où les délais d'apparition des coûts sont décrits, il est nécessaire de faire la régression sur chaque intervalle, avant d'additionner les coefficients obtenus. Enfin, la régression avec les coefficients additionnés est appliquée à tous les patients pour obtenir in fine un coût moyen par patient.

Des améliorations à ce modèle ont été proposées dans les années suivantes notamment en intégrant des variables dépendantes du temps grâce à des modèles de données de panels [43] ou en utilisant une régression linéaire généralisée à la place de la régression des moindres carrés ordinaires [44].

5.5. Comparaisons des modèles

Une comparaison des estimateurs des quatre méthodes présentées ainsi que des coûts sans tenir compte de la censure a été réalisée en faisant varier la durée des intervalles temporels et le type de censure [45]. Les auteurs démontrent que les meilleures méthodes pour estimer les coûts moyens sont l'estimateur par intervalle de Bang et Tsiatis (« Partitioned estimator »), l'estimateur sans historique de coûts de Lin, 1997 et la méthode de Carides. Cependant, en cas de censure partielle (suivi inégal des événements cliniques et des coûts), l'estimateur par intervalle de Bang et Tsiatis se révèle le plus

efficace. L'auteur conclut finalement que l'estimateur par intervalle de Bang et Tsiatis est le plus performant tant pour l'estimation des coûts moyens que de l'erreur standard à la moyenne c'est-à-dire l'incertitude autour du coût moyen.

Recommandations:

> Les estimateurs prenant en compte l'historique des coûts sont plus efficaces que ceux utilisant les coûts cumulés par patient. L'estimateur avec historique de coûts (« Partitioned estimator ») de Bang et Tsiatis est à privilégier par rapport à celui de Lin, 1997 si le nombre d'intervalles temporels est faible [46]. L'estimateur de Lin, 1997 avec historique des coûts et l'estimateur par intervalle de Bang et Tsiatis, 2000 sont les plus fiables si le taux de censure est important [47].

6. Biais de sélection

Lors de l'évaluation médicoéconomique des interventions et des technologies de santé, il arrive fréquemment que les données mobilisées pour évaluer les coûts ne soient pas randomisées, soit parce qu'il s'agit de bases de données médico-administratives qui n'ont pas été conçues, au départ, à des fins de recherche (cf. Chapitre 2), soit que l'on se situe dans le cadre d'études dites observationnelles où la randomisation peut être difficile à envisager ou à mettre en place pour de multiples raisons (faisabilité, problème éthique ou financier).

Les études randomisées sont souvent considérées comme la référence pour l'évaluation des interventions et des technologies de santé [48]. Dans un essai randomisé, les individus sont répartis au hasard entre les deux groupes ou plus. Le groupe intervention reçoit le nouveau traitement, tandis que le groupe témoin reçoit le traitement alternatif. Le processus de randomisation permet d'équilibrer les variables entre les groupes comparés. La randomisation est le seul moyen de s'assurer que les groupes de sujets sont comparables sur les paramètres connus et inconnus. Ainsi, une différence observée entre les bras de traitement peut seulement être attribuée à l'effet du traitement.

Les études non randomisées ont souvent des critères d'inclusions plus larges, et peuvent être plus représentatives de la population, avoir une période de suivi plus longue, et sont généralement moins coûteuses et conduites sous des conditions plus réalistes. Cependant, dans les études observationnelles, les investigateurs ne contrôlent pas l'allocation du traitement, ce qui peut entraîner d'importantes différences entre les groupes de patients traités et non traités, à la fois en ce qui concerne leurs caractéristiques observées, mais aussi leurs caractéristiques inobservées. Par conséquent, les différences observées dans les résultats entre groupes peuvent être soit liées au traitement, soit dues à des différences entre les caractéristiques comparées. Si tel est le cas, l'étude est entachée de biais (biais de sélection).

Recommandations:

> Dans les études non randomisées, il est nécessaire de contrôler le biais de sélection. Le biais de sélection fait référence à une différence systématique des caractéristiques des individus entre les groupes comparés, ces différences affectant les résultats de coût et d'efficacité. Le biais de sélection dans les études n'est pas seulement dû aux caractéristiques observées (biais « mesurable », « overt bias »), mais aussi aux variables inobservées ou omises (biais caché ou « hidden bias »).

Dans ce qui suit, on distingue les méthodes proposées pour réduire le biais de sélection « observable », des méthodes utilisées pour contrôler le biais issu de variables « non observables ».

6.1. Méthodes utilisées pour réduire le biais de sélection « observable »

Dans la situation où la plupart des variables pertinentes ont été identifiées dans le protocole de l'étude, les biais sont « mesurables », et les analyses de régressions standards ainsi que des méthodes utilisant le score de propension permettent de les neutraliser.

6.1.1. Analyses de régression multivariée

L'approche traditionnelle pour contrôler le biais de sélection est d'utiliser des modèles de régression. Nous avons vu, en partie 3.2 (Analyse multivariée), qu'un modèle de régression multivariée estime de combien chaque variable indépendante est liée à la variable dépendante, toutes choses étant égales par ailleurs. Une des principales limites des analyses de régression est qu'elles ne peuvent ajuster que sur les variables observées, il n'y a pas d'ajustement possible sur les variables inobservées.

6.1.2. Analyse par les scores de propension

Rosenbaum et Rubin [49] ont défini le score de propension comme la probabilité d'un sujet de recevoir un traitement spécifique conditionnellement à ses caractéristiques observées. En d'autres termes, un modèle est estimé pour prédire la probabilité qu'un individu soit assigné au groupe traitement, comparativement au groupe témoin, basé sur les valeurs des variables observées.

En estimant le score de propension, on peut créer une expérimentation « quasi-randomisée ». C'est le concept de « pseudo-randomisation ». En comparant chaque patient traité avec un patient non traité qui a le même score de propension, alors la différence de coût ou d'efficacité entre les deux patients est un estimateur non biaisé de l'effet du traitement.

Le score de propension résume toutes les caractéristiques (incluses dans les calculs) des patients, en un indice unique. Quand il y existe un chevauchement suffisant des scores de propension entre les groupes, alors le groupe traitement et le groupe témoin ont la même distribution sur toutes les variables observées. Le biais de sélection est alors réduit lorsque les comparaisons sont réalisées entre des groupes avec les mêmes scores de propension. S'il y a un chevauchement insuffisant du

score de propension entre les groupes, alors les deux groupes ne sont pas vraiment comparables, et cette méthode ne peut pas résoudre le problème.

Le score de propension peut être estimé pour chaque patient à partir d'un modèle de régression Probit ou Logit. Dans ce type de modèle, la variable dépendante est l'attribution du traitement (0 = groupe témoin, 1 = groupe traitement) et les variables indépendantes sont les caractéristiques du patient, comme l'âge, le sexe, etc.

Contrairement à la randomisation, l'utilisation du score de propension ne permet de contrôler le biais de sélection que sur les différences observées [27]. De plus, le score de propension ne fournit pas de test direct du biais de sélection, ni d'estimation de l'importance du biais de sélection si celui-ci est présent. Les méthodes utilisant le score de propension fonctionnent mieux sur des grands échantillons, car le chevauchement des scores entre le groupe traitement et le groupe témoin en termes de caractéristiques observées augmente avec la taille de l'échantillon.

Une fois calculé, le score de propension est appliqué de plusieurs façons pour réduire le biais de sélection : appariement (« matching »), stratification, ajustement par régression, ou pondération par l'inverse du score de propension [50]. Ces méthodes sont brièvement décrites ci-après.

6.1.2.1. Appariement par le score de propension. La méthode d'appariement par le score de propension consiste à ordonner les patients des groupes traitement et contrôle par leur score de propension et ensuite à apparier chaque patient qui reçoit le traitement à un patient du groupe témoin avec un score de propension similaire. Cette analyse conduit cependant à exclure les patients qui n'ont pas pu être appariés. Plusieurs procédures d'appariement sont proposées dans la littérature. La procédure d'appariement du plus proche voisin (« Nearest-Neighbor Matching ») consiste à apparier un patient dans le groupe traitement à un patient dans le groupe témoin qui a le score de propension le plus proche. Ceci peut être réalisé sans remplacement (un patient du groupe témoin ne peut être apparié qu'une seule fois avec un patient du groupe traitement) ou avec remplacement (un patient du groupe témoin peut être sélectionné plus d'une fois). Dans l'appariement de la « cinquième à la première décimale » (« 5 to 1 digit matching »), le sujet témoin à apparier au sujet traité est d'abord recherché parmi les sujets ayant un score similaire à la cinquième décimale. S'il n'y en a pas, on recherche parmi ceux ayant un score de propension égal à la quatrième décimale, et ainsi de suite jusqu'à la première décimale. Dans la méthode d'appariement appelée « radius matching », les patients du groupe témoin sont sélectionnés dans un rayon de valeurs prédéfini.

Une fois que les patients traités sont appariés, la prochaine étape est de calculer les coûts moyens dans chaque groupe puis de faire la différence pour obtenir l'effet moyen du traitement.

Des exemples très détaillés de programmation sous SAS des méthodes d'appariement par le score de propension sont fournis dans l'article de Coca-Perraillon [51].

6.1.2.2. Stratification par le score de propension. Une fois que les scores de propension sont obtenus pour chaque patient, cette méthode consiste à grouper des patients par strate, avec approximativement le même nombre total de patients dans chaque strate. Une strate contient des patients de chaque groupe avec des scores de propension similaires. L'investigateur doit décider les valeurs limites pour les différentes strates (les strates sont généralement divisées en fourchettes égales de score de propension). Selon Cochrane [33], l'utilisation de cinq strates élimine plus de 90 % du biais. Une fois que les strates sont définies, les patients du groupe traitement et du groupe témoin qui sont dans la même strate sont comparés directement pour estimer l'effet du traitement dans chaque strate. Ces résultats sont ensuite combinés pour estimer l'effet global du traitement.

6.1.2.3. Ajustement du modèle de régression par le score de propension. Dans cette méthode, une fois que le score de propension a été obtenu pour chaque patient, il est introduit dans un modèle de régression comme variable explicative ainsi que la variable indicatrice du traitement [50]. Il est possible d'inclure des variables explicatives supplémentaires. Le choix du modèle de régression dépend de la nature de la variable d'intérêt. L'effet du traitement est déterminé à partir du coefficient estimé de la régression ajustée.

6.1.2.4. Pondération par l'inverse du score de propension. - L'idée de la méthode de pondération par l'inverse du score de propension (« Inverse Probability of Treatment Weighting », IPTW) est de pondérer les sujets traités et les témoins par leurs scores de propension respectifs dans le but d'obtenir un échantillon artificiel dans lequel la distribution des co-variables de base est indépendante de l'allocation du traitement. Le poids affecté aux sujets réellement traités est l'inverse de leur score de propension, et le poids affecté aux sujets réellement non traités est l'inverse de leur probabilité de ne pas être traité (c'est-à-dire 1 moins le score de propension). Par rapport à l'appariement, cette technique présente l'avantage de conserver l'ensemble de l'échantillon initial dans l'analyse et de maintenir la puissance statistique [52]. Certaines méthodes combinent l'estimation par la régression et par pondération par l'inverse du score de propension pour estimer l'effet causal du traitement [53]. Dans ce contexte, l'IPTW fait partie d'une famille plus large de méthodes, permettant d'inférer une relation causale et connues sous le nom de modèles structuraux marginaux [54,55].

Recommandation:

> Quand le score de propension est utilisé, les auteurs doivent expliquer comment ils ont constitué leur modèle de score de propension. En particulier, les variables incluses doivent être détaillées et leur choix doit être justifié.
> L'objectif étant d'obtenir des groupes de patients traités et non traités équilibrés sur leurs caractéristiques cliniques, il est primordial que cet équilibre soit rapporté et correctement évalué, en calculant par exemple la différence standardisée (d) de chaque variable entre le groupe témoin et le groupe traité, avant et après appariement.

Pour une variable continue, la différence standardisée est définie comme :

avec les moyennes et les variances respectivement dans le groupe des patients traités et dans le groupe témoin.

Pour une variable dichotomique, la différence standardisée est définie comme :

Où et désignent la prévalence de la variable dichotomique dans les sujets traités et non traités.

On estime que l'appariement est un succès si la différence standardisée résiduelle après appariement est limitée pour l'ensemble des variables. Une valeur de d inférieure ou égale à 10 % a été déterminée comme acceptable [52]. Une représentation graphique permet de faire une synthèse lisible pour le lecteur. Gayat et Porcher [52] soulignent qu'il est important que le lecteur soit capable d'estimer l'impact qu'a eu l'analyse par le score de propension sur l'estimation de l'effet du traitement ; il est donc intéressant de rapporter aussi les résultats de l'analyse brute du critère de jugement. Des analyses de sensibilité sont souhaitables pour explorer l'incertitude générée par la construction et l'utilisation du score de propension.

Shah et al. [56] ont réalisé une revue systématique pour déterminer si le score de propension donnait des résultats différents des modèles de régression. Ils ont trouvé que les deux méthodes produisaient des résultats similaires, même si le score de propension conduisait à des estimations de l'effet du traitement légèrement plus faibles. Par rapport à la régression, l'avantage de l'appariement par le score de propension est que ce dernier ne repose pas sur l'hypothèse de linéarité (relation constante entre les résultats de coût et les co-variables dans chacun des groupes de traitement), tandis que la régression fait cette hypothèse [57]. Le score de propension fonctionne mieux sur des grands échantillons, car un chevauchement entre les deux groupes en termes de caractéristiques observées augmente avec la taille de l'échantillon [58]. L'importance de la réduction du biais dépend crucialement de la disponibilité et de la qualité des variables de contrôle.

Gayat et Porcher [35] ont proposé plusieurs recommandations pour l'utilisation du score de propension. L'appariement sur le score de propension constitue la technique la plus populaire et la plus performante dans différentes situations [59,60]. Cependant, il peut conduire à exclure des patients qui n'ont pas pu être appariés.

6.2. Méthodes utilisées pour réduire le biais de sélection lié aux caractéristiques inobservées

Dans la situation d'une étude observationnelle où des variables importantes pouvant jouer sur l'allocation du traitement ont été omises au départ et ne peuvent pas être récupérées, il existe alors un risque de biais « caché » que les approches de régression classique ou de score de propension sont incapables de contrôler [61]. En effet, lorsque l'on régresse le résultat d'intérêt Y (tel que le coût ou la durée du séjour), en fonction des caractéristiques du patient et du traitement prédit, l'omission de variables importantes qui jouent sur le traitement (comme par exemple, la sévérité de la maladie) vont se retrouver dans le terme d'erreur et conduisent à une corrélation entre le terme d'erreur et la variable de traitement (T) ; on dit alors qu'il y a un problème d'endogénéité et les coefficients de la régression sont par conséquent biaisés.

Méthode des variables instrumentales, « Control function », modèle d'inclusion des résidus en deux étapes

Une solution consiste à utiliser la méthode des variables instrumentales (VI) qui nécessite de disposer de variables non corrélées avec les termes d'erreurs (variables dites instrumentales ou instruments). La méthode d'estimation classique des variables instrumentales appliquée généralement sur les modèles linéaires est la *méthode des doubles moindres carrés ordinaires* (MC2O). La première étape de la méthode VI consiste à trouver une (ou des) variable(s) instrumentale(s) « valide(s) » (Z) qui sont à la fois :

- très fortement corrélées avec le fait de recevoir le traitement (T) ;
- non corrélées avec le résultat d'intérêt (Y).

Conceptuellement, le mécanisme d'estimation peut être décomposé en deux parties.

Dans une première partie, « la forme réduite », la variable *T* d'affectation au traitement est expliquée par l'instrument Z et les caractéristiques observables (*X*).

À partir de cela, on obtient :

Dans un deuxième temps, l'équation structurelle est estimée en remplaçant T par:

Les modèles traditionnels MC2O ne sont pas valides pour les modèles non linéaires [62]. Par exemple, quand le traitement est binaire, l'utilisation des MC2O ne permet pas une bonne prédiction du traitement ou du critère de jugement en présence de plusieurs co-variables. Il y a un risque d'obtention de valeurs prédites inférieures ou supérieures à 1 [61]. Une variante de la méthode des VI adaptée aux modèles non linéaires a été proposée par Terza et al. [63], qui est l'inclusion des résidus η en deux étapes « Two Stage Residual Inclusion » (2SRI). Ce modèle 2SRI est un cas particulier de l'approche « Control Fonction » (CF) dans laquelle le biais est contrôlé en utilisant une fonction $f(\eta)$ [64,65]. Après avoir estimé l'équation [3] à l'aide d'un modèle Probit, on prend les valeurs des résidus η et on estime l'équation suivante :

Si on fait l'hypothèse d'une normalité bivariée, la fonction de contrôle est l'inverse du ratio de Mills.

Recommandations:

La taille de l'échantillon, la disponibilité des variables de contrôle, la forme de la distribution du traitement et de la variable d'intérêt comme le coût vont permettre de déterminer la méthodologie à utiliser pour neutraliser le biais de sélection [62]. Rappelons que dans la situation où la plupart des variables pertinentes ont été identifiées dans le protocole de l'étude, les biais sont « mesurables », et les analyses de régressions standards ainsi que celles utilisant le score de propension permettent de les contrôler.

Contrairement aux méthodes de régression classiques et au score de propension, les méthodes des VI permettent de réduire le biais à la fois pour des différences dans les caractéristiques observées et inobservées. La nécessité de recourir à une méthode de VI dépend donc de l'importance des variables inobservées dans l'étude. Elle doit tout d'abord être évaluée d'un point de vue scientifique avec les investigateurs cliniques [66]. De plus, en pratique, les données ne disposent pas toujours de variables qui satisfassent les conditions requises pour constituer une bonne variable instrumentale. Il est fortement conseillé d'abandonner cette méthode si l'intensité de l'association entre l'instrument et le traitement est faible. Son utilisation sur des échantillons de petite taille est également déconseillée dans la mesure où elle ne produit des estimations efficientes qu'asymptotiquement [61].

Il est impossible de tester l'absence d'association entre l'instrument et le terme d'erreur. Par contre si on dispose de plusieurs instruments potentiels et si le nombre d'instruments est strictement supérieur au nombre de variables endogènes, il est possible d'effectuer un test appelé test de « suridentification » (test de Sargan) permettant de se prononcer sur la qualité des instruments candidats.

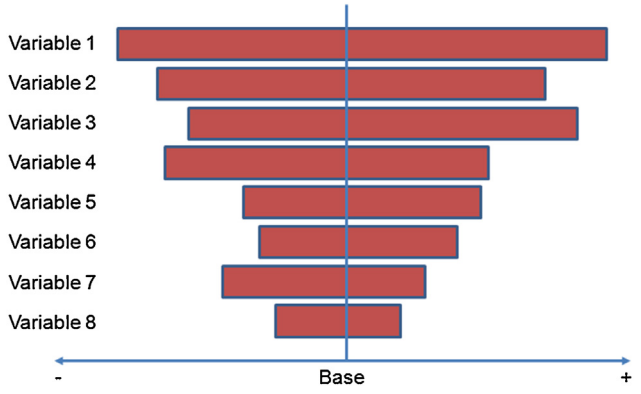

Fig. 4. Représentation graphique d'un diagramme de Tornado.

7. Analyse de sensibilité

Une analyse de sensibilité permet de comparer l'influence relative de plusieurs variables ou paramètres (les paramètres d'entrée) sur notre variable d'intérêt, la variable de coût (paramètre de sortie). Ce type d'analyse va permettre d'estimer la sensibilité des calculs à la variation d'un ou plusieurs paramètres, et ainsi d'alimenter la discussion autour de la robustesse des conclusions. On relève deux sources principales d'incertitude qui méritent ainsi de faire l'objet d'une analyse de sensibilité : l'incertitude sur les paramètres et les choix méthodologiques. L'analyse de sensibilité va donc permettre de caractériser l'incertitude sur ces deux éléments et donc de tester la robustesse des conclusions de l'évaluation vis-à-vis de ces deux choses.

Selon les recommandations de la Haute Autorité de Santé (HAS) [67], la valeur du taux d'actualisation doit, par exemple, faire l'objet d'une analyse de sensibilité de manière à apprécier la robustesse des conclusions de l'évaluation à l'égard de ce paramètre. L'analyse de sensibilité fait partie intégrante des conclusions d'une évaluation économique.

Le graphique de Tornado est la représentation graphique d'une analyse de sensibilité déterministe univariée. Un graphique de Tornado est une forme spécifique d'histogramme qui permet de comparer l'incertitude générée par chacun des

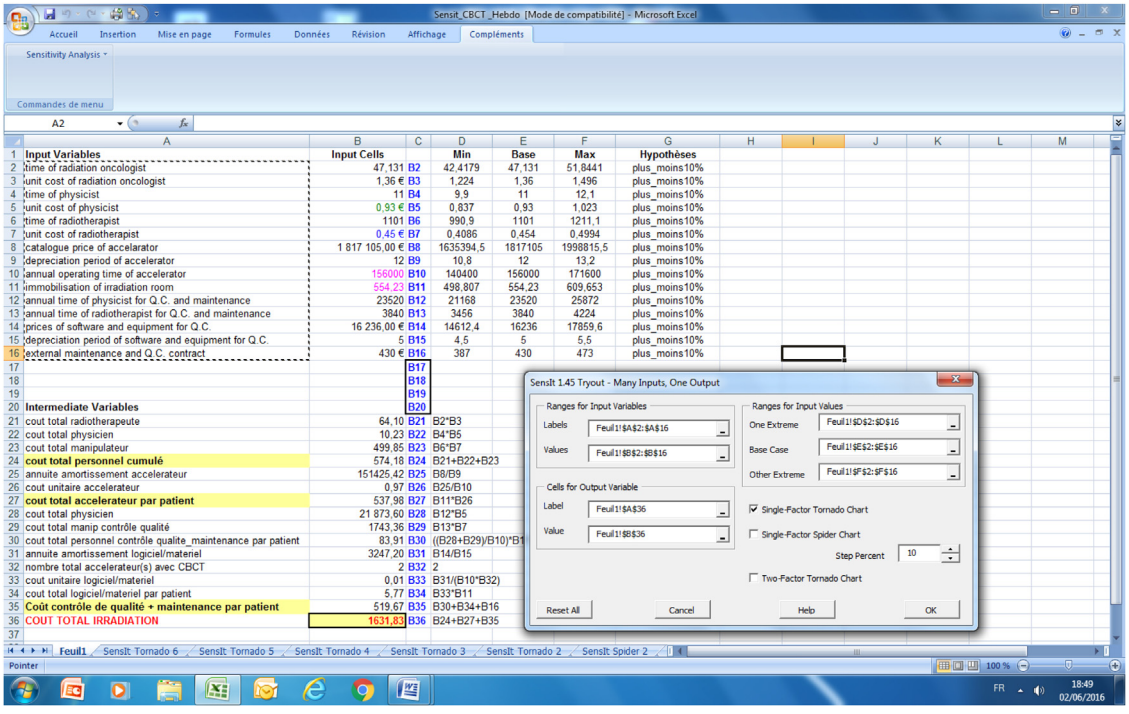

Fig. 5. Interface de l'extension d'Excel sensIT.

paramètres ; les autres paramètres étant fixés à leur valeur retenue dans l'analyse de référence. Il permettra d'illustrer l'impact d'un certain niveau de variation (10 % par exemple) de la valeur de chaque paramètre ou variable pris en compte dans l'analyse de sensibilité. L'interprétation de ce type de diagramme sera donc la suivante : la longueur de chaque barre représente l'importance de la sensibilité du coût moyen, représenté par la ligne verticale du diagramme de Tornado, axe vertical « Base » (Fig. 4) et servant de valeur de base sur laquelle l'influence des variations est testée, par rapport aux paramètres inclus dans l'analyse. La représentation du diagramme est telle que l'influence des paramètres est classée de manière décroissante : le paramètre le plus influent étant en haut (Variable 1 – Fig. 4) et le moins influent étant en bas (Variable 8 – Fig. 4) [68].

Ce type de représentation graphique (Fig. 4) peut être réalisé de manière simple et intuitive grâce à une extension d'Excel, sensIT[4] (voir interface – Fig. 5).

Recommandations:

L'analyse de sensibilité fait partie intégrante de l'évaluation économique. Une analyse de sensibilité déterministe univariée est systématiquement réalisée sur les paramètres susceptibles d'influencer les conclusions du modèle. La sélection des paramètres soumis à une analyse de sensibilité, ainsi que la plage de valeurs retenue pour tester ces paramètres, sont présentées et justifiées [67].

Déclaration de liens d'intérêts

Les auteurs déclarent ne pas avoir de liens d'intérêts.

Références

[1] Doshi JA, Glick HA, Polsky D. Analyses of cost data in economic evaluations conducted alongside randomized controlled trials. Value Health 2006;9:334–40.
[2] Barber JA, Thompson SG. Analysis of cost data in randomized trials: an application of the non-parametric bootstrap. Stat Med 2000;19:3219–36.
[3] Altman DG, Gore SM, Gardner MJ, Pocock SJ. Statistical guidelines for contributors to medical journals. Br Med J Clin Res Ed 1983;286:1489–93.
[4] Efron B, Tibshirani RJ. An Introduction to the Bootstrap. New York: Chapman and Hall/CRC; 1994.
[5] Briggs A, Gray A. The distribution of health care costs and their statistical analysis for economic evaluation. J Health Serv Res Policy 1998;3:233–45.
[6] Gray AM. Applied methods of cost-effectiveness analysis in healthcare. Oxford, New York: OUP Oxford; 2010.
[7] Cassell DL. Don't be loopy: re-sampling and simulation the SAS Way. SAS Global Forum; 2007 [Accédée le 2 novembre 2017 ; disponible sur : http://www.2.sas.com/proceedings/forum2007/183-2007.pdf].
[8] Glick HA, Doshi JA, Sonnad SS. Economic evaluation in clinical trials (2 edition). Oxford: Oxford University Press; 2014.
[9] Greene WH. Econometric analysis (7 edition). Boston: Pearson; 2011.
[10] Blough DK, Madden CW, Hornbrook MC. Modeling risk using generalized linear models. J Health Econ 1999;18:153–71.
[11] Manning WG, Mullahy J. Estimating log models: to transform or not to transform? J Health Econ 2001;20:461–94.
[12] Manning WG, Basu A, Mullahy J. Generalized modeling approaches to risk adjustment of skewed outcomes data. J Health Econ 2005;24:465–88.
[13] Pregibon D. Goodness of link tests for generalized linear models. J R Stat Soc Ser C Appl Stat 1980;29:14–5.
[14] Hosmer Jr DW, Lemeshow S, Sturdivant RX. Applied Logistic Regression (3rd edition). Hoboken, New Jersey: Wiley-Blackwell; 2013.
[15] Copas JB, Regression. Prediction and shrinkage. J R Stat Soc Ser B Methodol 1983;45:311–54.
[16] Verbeek M. A guide to modern econometrics (4th Edition). Hoboken, NJ: John Wiley & Sons; 2012.
[17] Jacquemet N, Crépon B. Econometrie : méthode et applications (première édition). Bruxelles: De Boeck Université; 2010.
[18] Fichman M, Cummings JN. Multiple imputation for missing data: making the most of what you know. Organ Res Methods 2003;6:282–308.
[19] Little RJA, Rubin DB. Statistical analysis with missing data (2nd edition). Hoboken, NJ: Wiley-Blackwell; 2002.
[20] Graham JW. Missing data analysis: making it work in the real world. Annu Rev Psychol 2009;60:549–76.
[21] Allison PD. Missing data (1 edition). Thousand Oaks, Calif: SAGE Publications, Inc; 2001.
[22] Little RJA. Modeling the drop-out mechanism in repeated-measures studies. J Am Stat Assoc 1995;90:1112–21.
[23] Little RJ, Wang Y. Pattern-mixture models for multivariate incomplete data with covariates. Biometrics 1996;52:98–111.
[24] Schafer JL. Multiple imputation: a primer. Stat Methods Med Res 1999;8:3–15.
[25] Magnani M. Techniques for dealing with missing data in knowledge discovery tasks. University of Bologna, department of Computer Science; 2004 [Accédée le 2 novembre 2017 ; disponible sur : http://www.magnanim.web.cs.unibo.it/data/pdf/missingdata.pdf].
[26] Faria R, Gomes M, Epstein D, White IR. A guide to handling missing data in cost- effectiveness analysis conducted within randomised controlled trials. Pharmacoeconomics 2014;32:1157–70.
[27] Rubin DB. Multiple imputation for non-response in surveys. John Wiley & Sons, Inc; 1987.
[28] Heitjan DF, Little RJA. Multiple imputation for the fatal accident reporting system. J R Stat Soc Ser C Appl Stat 1991;40:13–29.
[29] Schenker N, Taylor JMG. Partially parametric techniques for multiple imputation. Comput Stat Data Anal 1996;22:425–46.
[30] Lavori PW, Dawson R, Shera D. A multiple imputation strategy for clinical trials with truncation of patient data. Stat Med 1995;14:1913–25.
[31] Schafer JL. Analysis of incomplete multivariate data (1 edition). London; New York: Chapman and Hall/CRC; 1997.
[32] Raghunathan TE, Lepkowski JM, Hoewyk JV, Solenberger P. A multivariate technique for multiply imputing missing values using a sequence of regression models. Survey Methodology 2001;27:85–95.
[33] van Buuren S. Multiple imputation of discrete and continuous data by fully conditional specification. Stat Methods Med Res 2007;16:219–42.
[34] van Buuren S, Boshuizen HC, Knook DL. Multiple imputation of missing blood pressure covariates in survival analysis. Stat Med 1999;18:681–94.
[35] van Buuren S, Brand JP, Groothuis-oudshoorn CG, Rubin DB. Fully conditional specification in multivariate imputation. J Stat Comput Simul 2006;76:1049–64.
[36] Horton NJ, Lipsitz SR. Multiple imputation in practice: comparison of software packages for regression models with missing variables. Am Stat 2001;55:244–54.
[37] Briggs A, Clark T, Wolstenholme J, Clarke P. Missing... presumed at random: costanalysis of incomplete data. Health Econ 2003;12:377–92.
[38] Berglund P, Heeringa S. Multiple imputation of missing data using SAS (third edition). North Carolina, USA: SAS Institute Inc., Cary; 2014.
[39] Lin DY, Feuer EJ, Etzioni R, Wax Y. Estimating medical costs from incomplete follow-up data. Biometrics 1997;53:419–34.

[4] https://www.math.unl.edu/~tshores1/Public/Math428S08/SensIt/sensit.pdf.

[40] Bang H, Tsiatis AA. Estimating medical costs with censored data. Biometrika 2000;87:329–43.
[41] Carides GW, Heyse JF, Iglewicz B. A regression-based method for estimating mean treatment cost in the presence of right-censoring. Biostat Oxf Engl 2000;1:299–313.
[42] Lin DY. Linear regression analysis of censored medical costs. Biostatistics 2000;1:35–47.
[43] Başer O, Gardiner JC, Bradley CJ, Yüce H, Given C. Longitudinal analysis of censored medical cost data. Health Econ 2006;15:513–25.
[44] Lin DY. Regression analysis of incomplete medical cost data. Stat Med 2003;22:1181–200.
[45] Young TA. Estimating mean total costs in the presence of censoring: a comparative assessment of methods. PharmacoEconomics 2005;23: 1229–42.
[46] O'Hagan A, Stevens JW. On estimators of medical costs with censored data. J Health Econ 2004;23:615–25.
[47] Raikou M, Mc Guire A. Estimating medical care costs under conditions of censoring. J Health Econ 2004;23:443–70.
[48] Haute Autorité de Santé. Guide méthodologique. In: Choix méthodologiques pour le développement clinique des dispositifs médicaux; 2013.
[49] Rosenbaum PR, Rubin DB. Reducing bias in observational studies using subclassification on the propensity score. J Am Stat Assoc 1984;79: 516–24.
[50] Austin PC. An introduction to propensity score methods for reducing the effects of confounding in observational studies. Multivar Behav Res 2011;46:399–424.
[51] Coca-Perraillon M.In: Matching with propensity score to reduce biais in observational studies; 2007 [Disponible sur : http://www.2.sas.com/proceedings/forum2007/185-2007.pdf ; accédée le 2 novembre 2017].
[52] Gayat E, Porcher R. Comparaison de l'efficacité de deux thérapeutiques en l'absence de randomisation: intérêts et limites des méthodes utilisant les scores de propension. Reanimation 2011;21:109–16.
[53] Joffe MM, Have TRT, Feldman HI, Kimmel SE. Model selection, confounder control, and marginal structural models: review and new applications. Am Stat 2004;58:272–9.
[54] Hernán MA, Brumback B, Robins JM. Marginal structural models to estimate the causal effect of zidovudine on the survival of HIV-positive men. Epidemiol Camb Mass 2000;11:561–70.
[55] Hernán MA, Brumback BA, Robins JM. Estimating the causal effect of zidovudine on CD4 count with a marginal structural model for repeated measures. Stat Med 2002;21:1689–709.
[56] Shah BR, Laupacis A, Hux JE, Austin PC. Propensity score methods gave similar results to traditional regression modeling in observational studies: a systematic review. J Clin Epidemiol 2005;58:550–9.
[57] Mistry H. Exploring two cost-adjustment methods for selection bias in a small sample: using a fetal cardiology dataset. Int J Technol Assess Health Care 2014;30:325–32.
[58] Mistry H. Economic issues associated with the operation and evaluation of telemedicine (thesis); 2011 [Disponible sur : http://www.bura.brunel.ac.uk/handle/2438/583; accédée le 2 novembre 2017].
[59] Austin PC. The performance of different propensity score methods for estimating marginal odds ratios. Stat Med 2007;26:3078–94.
[60] Austin PC. The performance of different propensity-score methods for estimating relative risks. J Clin Epidemiol 2008;61:537–45.
[61] Ghabri S, Launois R. Évaluation quasi-expérimentale des interventions médicales : méthode des variables instrumentales. J Gest Econ Med 2015;32:371–88.
[62] Crown WH. Propensity-score matching in economic analyses: comparison with regression models, instrumental variables, residual inclusion, differences-in-differences, and decomposition methods. Appl Health Econ Health Policy 2014;12:7–18.
[63] Terza JV, Basu A, Rathouz PJ. Two-stage residual inclusion estimation: addressing endogeneity in Health Econometric Modeling. J Health Econ 2008;27:531–43.
[64] Heckman J, Navarro-Lozano S. Using matching, instrumental variables, and control functions to estimate Economic Choice Models. Rev Econ Stat 2000;86:30–57.
[65] Garrido MM, Deb P, Burgess JF, Penrod JD. Choosing models for health care cost analyses: issues of nonlinearity and endogeneity. Health Serv Res 2012;47:2377–97.
[66] Baiocchi M, Cheng J, Small DS. Instrumental variable methods for causal inference. Stat Med 2014;33:2297–340.
[67] Haute Autorité de santé. Guide méthodologique. In: Choix méthodologiques pour l'évaluation économique à la HAS; 2011.
[68] Eschenbach PE. Engineering economy: applying theory to practice (3 Har/Cdr). New York, USA: Oxford University Press; 2010.

View publication stats