



Analyse des études d'observation : quelle place pour la méthode des variables instrumentales ?

Tom Poelman, Vakgroep Huisartsgeneeskunde en Eerstelijnsgezondheidszorg, UGent

Pour la plupart des lecteurs, c'est enfoncer une porte ouverte que de dire que les études randomisées contrôlées constituent la pierre angulaire de l'examen des avantages et des désavantages des interventions. Dans notre pratique quotidienne, nous faisons régulièrement l'expérience, que la réponse à certaines questions cliniques pertinentes ne puisse se trouver par le biais d'études randomisées contrôlées. En effet, il arrive que ces études existent mais qu'elles ne soient pas utilisables (trop peu de patients, méthodologie de qualité insuffisante), n'aient pas encore été menées, ou qu'elles soient irréalisables. Par exemple, pour évaluer les effets d'un traitement à (très) long terme ou pour en déceler les effets indésirables (très) rares, on devrait randomiser des milliers de participants pendant plusieurs années, ce qui est souvent impossible sur le plan pratique, financier ou éthique. Dans ces cas-là, l'étude d'observation est une alternative intéressante. Certains vont encore plus loin pour remettre les études d'observation à l'avant-plan en signalant les désavantages des études randomisées contrôlées, tels que leur caractère peu généralisable et leur coût élevé (1). Mais quelle est la fiabilité des données observationnelles concernant l'effet des traitements ?

Le principal problème des études d'observation est la présence de **facteurs de confusion**.

Contrairement à ce qui se passe dans les études d'intervention, dans les études d'observation portant sur l'effet d'un traitement, recevoir ou non le traitement dépend du choix du patient et/ou du médecin, sans intervention des investigateurs. Ainsi peut apparaître un déséquilibre quant aux facteurs pronostiques entre les personnes exposées et celles qui ne le sont pas, autrement dit entre le groupe traitement et le groupe contrôle, ce qui peut avoir une influence sur l'effet à étudier. Si l'on réussit à mesurer de manière précise les facteurs de confusion, comme l'âge, le sexe, la gravité de la maladie, la comorbidité, il est ensuite possible de neutraliser leur influence sur les critères de jugement à l'aide de différentes méthodes statistiques. Minerva a déjà traité certaines d'entre elles (2,3). Un article récent propose un aperçu des techniques les plus souvent utilisées (4). En voici un résumé.

La **stratification** est le processus consistant à répartir les personnes examinées en groupes homogènes (les strates), chaque fois selon un seul facteur pronostique (par exemple < 50 ans et ≥ 50 ans). Ensuite, on calcule la différence quant à l'effet entre le groupe traitement et le groupe contrôle (absence de traitement ou un autre traitement) dans toutes les strates, pour ensuite combiner toutes ces différences.

Supposons que, dans une étude de cohorte, 100 personnes soient exposées (groupe traitement) et 100 personnes ne soient pas exposées (groupe contrôle) à un traitement pendant une certaine période. Que 80% du groupe traitement ait < 50 ans, et 80% du groupe contrôle, ≥ 50 ans.

Que, dans le groupe traitement, 12 personnes décèdent (8 ayant < 50 ans et 4 ≥ 50 ans), et, dans le groupe contrôle, 18 personnes décèdent (2 ayant < 50 ans et 16 ≥ 50 ans). Cela revient à un risque relatif (RR) « non corrigé » de $12\%/18\% = 0,67$, soit une réduction relative de risque (RRR) « non corrigée » de 33% de décès dans le groupe traitement.

Mais, si l'on répartit les 200 personnes de l'étude de cohorte en une strate < 50 ans et une strate ≥ 50 ans, on voit que :

- dans la strate < 50 ans : la proportion de décès est de 8/80 dans le groupe traitement et de 2/20 dans le groupe contrôle, soit un RR_1 de $10\%/10\% = 1$
- dans la strate ≥ 50 ans : la proportion de décès est de 4/20 dans le groupe traitement et de 16/80 dans le groupe contrôle, soit un RR_2 de $20\%/20\% = 1$

La combinaison de RR_1 et de RR_2 donne un RR « corrigé » de 1. Si l'on tient compte de la différence d'âge entre le groupe traitement et le groupe contrôle, on voit que le traitement ne réduit pas la mortalité.

Afin de tenir compte des facteurs pronostiques, plus on voudra corriger plus il faudra prévoir de strates (nombre de strates = nombre de facteurs²). Il s'ensuit qu'à un certain moment, il devient plus simple d'utiliser une **analyse de régression multiple** (2). Cette comparaison maintenant constante la valeur de tous les facteurs de confusion, il est donc possible d'évaluer l'effet réel de l'intervention sur le critère de jugement. Lorsque le critère de jugement (ou une variable dépendante) est dichotomique, on utilise une **analyse par régression logistique**. Pour les critères de jugement continus, on utilise une analyse de régression linéaire, et lorsqu'on veut savoir dans quelle mesure l'intervention postpose un événement donné dans le temps, on applique un **modèle de Cox à risques proportionnels**. Le rapport est alors exprimé respectivement en rapport de cotes (RC, en anglais *Odds Ratio*, OR), en différence absolue de risque et en rapport de hasards (*Hazard Ratio*, HR).

Si on veut corriger pour tenir compte d'un nombre trop élevé de facteurs de confusion, par rapport au nombre d'observations, les résultats découlant de l'analyse de régression multiple seront peu valides. On peut aborder ce problème en commençant par attribuer à chaque patient un score de propension (0 à 1) (3). Ce score, qui exprime la probabilité de recevoir un traitement, dépendra des autres facteurs pronostiques (par exemple : les patients âgés gravement malades seront plus rapidement traités que des patients jeunes moins malades). En corrigeant pour tenir compte de ce score, on va pouvoir tenir compte d'un grand nombre de facteurs pronostiques sans perdre en validité. Mais, on a beau corriger, il restera toujours des facteurs pronostiques « inconnus » responsables d'un certain déséquilibre. Et ce n'est pas tout.

Dans une étude d'observation, il peut y avoir des facteurs qui influencent le fait de recevoir ou non le traitement, mais qui ne sont pas associés au pronostic. On les appelle des variables instrumentales. Elles ont pour conséquence que le traitement, dans une certaine mesure, a été attribué au hasard (pseudo-randomisation), ce qui fait qu'un équilibre a pu s'établir également pour les facteurs pronostiques inconnus dans les groupes étudiés (5).

Supposons que, dans une étude de cohorte, 80% des patients d'un centre X soient traités avec le médicament A, et 20%, avec le médicament B tandis que, dans le centre Y, 40% soient traités avec le médicament A, et 60%, avec le médicament B. En admettant que les patients arrivent au hasard dans le centre X et dans le centre Y et si les autres aspects du traitement ne diffèrent pas entre le centre X et le centre Y, on peut considérer la différence de 40% quant à l'utilisation du médicament A et du médicament B entre les patients des 2 centres comme une pseudo-randomisation.

Maintenant, supposons que 48% des patients du centre X ne présentent pas les symptômes étudiés, versus 42% des patients du centre Y. La différence de 6% dans l'effet peut alors être attribuée pour 40% à l'utilisation du médicament A à la place du médicament B. Supposons maintenant que tous les patients aient reçu le médicament A au lieu du médicament B (100%), selon les estimations, il y aurait eu 15% (6%/0,40) de patients en plus qui n'auraient pas présenté de symptômes avec le médicament A au lieu du médicament B.

Cet exemple fictif permet de déduire certaines conditions importantes pour l'utilisation d'une variable instrumentale fiable :

- les investigateurs doivent indiquer clairement combien de patients ont reçu l'intervention et combien ne l'ont pas reçue
- il doit y avoir une différence substantielle dans le statut de la variable instrumentale (par exemple probabilité nettement différente de recevoir une intervention dans l'hôpital A que dans l'hôpital B)

- la variable instrumentale elle-même ne peut pas avoir d'influence sur le pronostic futur (par exemple, il ne faut pas que, dans l'hôpital A, on administre, en plus du traitement à l'étude, d'autres traitements susceptibles d'influencer la guérison des patients)
- les facteurs pronostiques doivent être équitablement répartis en fonction du statut de la variable instrumentale (par exemple: il ne faut pas que l'hôpital A admette principalement des patients âgés).

On recense 5 sortes de variables instrumentales : des différences régionales en matière de soins, des habitudes hospitalières différentes, des différences dans les pratiques des médecins (généralistes) (dans les 3 cas, il y a une analogie avec la randomisation en grappes), des antécédents de patients n'influençant pas le critère de jugement étudié ainsi que des différences de délais (comme le délai d'approbation d'un médicament).

Comme l'analyse selon la méthode des variables instrumentales permet également de tenir compte des facteurs de confusion non mesurés et non mesurables, l'évaluation sera plus proche de la réalité qu'avec les autres techniques. Mais dans quelle mesure ? Seule une étude randomisée permet de le savoir.

Conclusion

Il existe un large éventail de méthodes statistiques garantissant la fiabilité de l'analyse des résultats des études d'observation. Comme la méthode des variables instrumentales permet également de corriger en tenant compte des facteurs de confusion non mesurés et non mesurables, les résultats seront plus proches de ceux d'une étude randomisée contrôlée.

Références

1. Goodman SN, Schneeweiss S, Baiocchi M. Using design thinking to differentiate useful from misleading evidence in observational research. *JAMA* 2017;317:705-7. DOI: 10.1001/jama.2016.19970
2. Poelman T. La différence entre régression et corrélation. *MinervaF* 2016;15(2):51-3.
3. Poelman T. Appariement sur les scores de propension. *MinervaF* 2013;12(8):103.
4. Agoritsas T, Merglen A, Shah ND, et al. Adjusted analyses in studies addressing therapy and harm. *Users' guides to the medical literature. JAMA* 2017;317:748-59. DOI: 10.1001/jama.2016.20029
5. Boef AG, le Cessie S, Dekkers OM. Instrumentele-variabele-analyse. *Ned Tijdschr Geneeskd* 2013;157:A5481.