

Les études portant sur les bases de données médico-administratives sont-elles reproductibles ?

L'hypothèse d'une inconduite éthique en santé publique

Vincent Looten

► La reproductibilité méthodologique fait référence à la capacité à obtenir exactement les mêmes résultats, en reproduisant le même protocole d'étude sur les mêmes données. Nous avons voulu évaluer la reproductibilité méthodologique des études publiées entre 2008 et 2017 dans la *Revue d'épidémiologie et de santé publique*, qui font appel à des données issues du système national des données de santé. Nos résultats suggèrent que seules 49 % des études portant sur ce système national pourraient être reproduites sans le recours aux auteurs initiaux. L'absence de partage systématique des programmes peut révéler un manque de préoccupation quant à la finalité de la recherche en santé publique. Il est difficile d'attribuer la responsabilité de ce manque de reproductibilité aux seuls chercheurs, et nous faisons donc l'hypothèse d'une inconduite éthique instituée. ◀



Inserm, Centre de Recherche des Cordeliers, UMRS 1138, équipe « Sciences de l'information au service de la médecine personnalisée », Université Paris Descartes, Sorbonne Paris Cité, 15, rue de l'École de Médecine, 75006 Paris, France.
looten@hotmail.fr

réfuter une théorie est au cœur de la démarche scientifique. En ce sens, la reproductibilité est un critère opposable de scientificité et d'intégrité scientifique [3,4]. La polysémie¹ du terme « reproductibilité » a amené Goodman *et al.* à en décrire trois dimensions [5] : la reproductibilité méthodologique, la reproductibilité des résultats et la reproductibilité des conclusions.

La reproductibilité méthodologique fait référence à la capacité à reproduire exactement les mêmes résultats, en utilisant le même protocole d'étude sur les mêmes données. La reproductibilité des résultats est la capacité à reproduire des résultats semblables au travers d'expériences de réplication indépendantes. La reproductibilité des conclusions est la capacité des chercheurs à aboutir aux mêmes conclusions à partir des mêmes résultats.

Les bases de données administratives sont incontournables pour réaliser des études en vie réelle. Issues de la population générale (*population-based*), elles sont de dimension

Les bases de données administratives sont le support des politiques de santé publique en France, en particulier des politiques de remboursement et d'organisation des soins [1]. Qu'advierait-il si des résultats d'études produites à partir de ces bases étaient faux ? Quelles en seraient les conséquences humaines, et qui serait responsable ? Un véritable séisme eut lieu en 2002 quand John Ioannidis publia son article « *Pourquoi la plupart des résultats scientifiques sont faux* » [2]. La communauté scientifique s'est dès lors réappropriée la question de la reproductibilité de ces expériences. La possibilité d'invalider ou de

Vignette (Photo © Inserm - Koulikoff, Frédérique/Pinci, Alexandra).

¹ Caractéristique d'un mot ou d'une expression qui a plusieurs sens ou significations.



Critère	Mesure	Description
C1	M1	Les dates d'inclusion sont-elles suffisamment précises ? (jour, mois, année)
C1	M2	Les dates d'extraction sont-elles précisées ? (liés à l'historicisation)
C1	M3	La population de l'étude est-elle précisée ? (âge, caisses de remboursement, région)
C1	M4	Les listes des codes issues des terminologies internationales et nationales sont-elles à disposition ?
C2	M5	Les programmes d'extraction des données sont-ils disponibles ?
C2	M6	Les programmes d'analyse des données extraites sont-ils disponibles ?

Tableau I. Mesures de reproductibilité méthodologique.

suffisante pour estimer correctement les tailles d'effet (*effect-size*), sont historicisées (capacité de retrouver l'état exact des données à une date antérieure) et elles sont partagées au travers du Système national des données de santé (SNDS) [1], qui les rend accessibles à toute la communauté scientifique. Les études portant sur ces bases ne présentent *a priori* aucun obstacle à leur reproductibilité méthodologique. L'objectif de notre travail était d'étudier la reproductibilité méthodologique des études réalisées à partir de ce Système national des données de santé.

L'étude, les résultats et les hypothèses

Le Système national des données de santé comprend principalement les bases de données de l'Assurance maladie (Système national d'information inter-régimes de l'Assurance maladie ou SNIIRAM) et la base de données des hospitalisations issues du Programme de médicalisation des systèmes d'informations (PMSI). Les études reposant sur le SNDS peuvent inclure uniquement les données du PMSI, la totalité des données du SNIIRAM chaînées avec les données du PMSI (*datamart consommation inter-régimes*, DCIR), ou un échantillon de ce dernier appelé échantillon généraliste des bénéficiaires (EGB).

Les études qui reposent sur le SNDS sont de trois types : (1) des rapports destinés aux autorités de santé (littérature grise) ; (2) des publications scientifiques qui ne sont pas référencées, en langue française ; et (3) des publications scientifiques référencées et publiées dans des revues à comité de lecture. Dans le cadre de notre étude, nous avons limité notre analyse à la série d'articles portant sur tout ou partie du SNDS et publiés dans la *Revue d'épidémiologie et de santé publique (RESP)* de l'éditeur Elsevier, entre 2008 et 2017. Cette revue a été choisie car elle constitue l'un des principaux canaux de communication de la communauté d'information médicale et en santé publique en France.

Nous nous proposons dans cette étude de répondre à deux questions : (1) les données publiées sont-elles suffisantes pour reproduire l'étude considérée (critère C1) ? (2) les programmes d'extraction et d'analyses de données sont-ils mis à disposition dans la publication (critère C2) ? Si les programmes d'extraction et d'analyses ont été mis à disposition (critère C2), alors les données publiées sont suffisantes pour reproduire l'étude (critère C1). À noter que la réciproque n'est pas forcément vraie. Le *Tableau I* présente 8 mesures correspondant, pour ces deux critères de sélection, à des questions dont les réponses peuvent être soit oui, soit non.

Parmi les 363 numéros de la *Revue d'épidémiologie et de santé publique* publiés entre 2008 et 2017 (hors numéro spécial et hors résumé), 42 répondent aux critères d'inclusion (C1 et C2). Cinq études, qui ont été menées en région Provence-Alpes-Côte d'Azur (PACA), concernent des études sur le SNIIRAM. Dix-neuf études sur 39 (soit 49 %), qui concernent toutes des études portant sur le PMSI, combinent toutes les critères de reproductibilité méthodologique (M1-4). Aucune étude ne propose un partage des codes (M5-6) (*Tableau II*).

Ces résultats suggèrent que seules 49 % des études publiées dans *RESP* et portant sur le SNDS pourraient être reproduites sans le recours aux auteurs initiaux. Ils sont particulièrement marquants pour les études portant sur l'échantillon généraliste des bénéficiaires (EGB) et les données du SNIIRAM chaînées avec celles du PMSI (DCIR). La reproductibilité méthodologique ne semble pas s'améliorer avec le temps. Nous avons identifié cinq études portant sur la région PACA, qui auraient pu être généralisées à l'ensemble du territoire français. Néanmoins, si elles avaient satisfait les critères M5 (les programmes d'extraction des données sont-ils disponibles ?) et M6 (les programmes d'analyse des données extraites sont-ils disponibles ?), cette généralisation aurait pu être immédiate. La plupart des études ne reproduisant pas régulièrement leurs analyses, cela diminue leur intérêt en santé publique (indicateurs obsolètes).

La principale limite de notre étude concerne le mode de sélection des articles. Nous n'en avons, en effet, pas réalisé une revue systématique. Cette étude n'a donc que la valeur que lui donne l'analyse d'une série de cas. Une autre limite concerne le nombre de relecteurs, puisqu'elle n'a pas bénéficié d'une double relecture. Enfin, bien que la base de données ayant servi à cette étude soit disponible en accès ouvert², les mesures de reproductibilité méthodologique que nous proposons n'ont pas fait l'objet d'une validation scientifique.

² github.com/vlooten/reproductibilite



Études publiées dans RESP (N=42)	
Données :	
PMSI	24 (57 %)
EGB	2 (5 %)
DCIR-PMSI	16 (38 %)
Années de publication :	
2008-2012	19 (45 %)
2014-2017	23 (55 %)
M1	38 (97 %) (NA = 3)
M2	2 (12 %) (NA = 25)
M3	38 (97 %) (NA = 3)
M4	29 (74 %) (NA = 3)
M5	0 (NA = 4)
M6	0 (NA = 3)

Tableau II. Description des publications et des mesures de reproductibilité.

RESP : *Rev Épidémiol Santé Publique* ; PMSI : Programme de médicalisation des systèmes d'informations ; DCIR : *datamart consommation inter-régimes* ; EGB : échantillon généraliste des bénéficiaires. Certains critères n'étaient pas adaptés aux objectifs des articles, nous les avons indiqués comme « non applicable » (NA) dans le tableau. M1 à M6 font référence aux mesures du *Tableau I*.

Les deux principales composantes de la reproductibilité méthodologique, énoncées par Peng *et al* [6], sont le partage des données (*data sharing*) et le partage des programmes (*code sharing*). Dans le cadre des études portant sur le SNDS, seule la question du partage des programmes demeure. Il n'existe pas d'argument technique expliquant le non-partage des programmes [7, 8]. Or, de nombreuses solutions permettant de partager le code existent. Nous sommes probablement en présence d'un « dilemme du prisonnier ». Cette situation a déjà été formalisée dans le cadre du partage des données [9]. Les chercheurs sont en compétition et ne perçoivent pas l'intérêt de partager leurs programmes. Pour autant, s'ils les rendaient accessibles, s'exposant ainsi à la critique, tous pourraient améliorer leurs pratiques et les enrichir des expériences des autres. Un probable gain découlerait de cette coopération. Les résultats seraient plus compréhensibles et plus transparents vis-à-vis des décideurs, notamment.

Une inconduite éthique instituée ?

Au-delà de la mise en place de mesures d'incitation au partage des programmes, c'est la place de la valeur « partage » en recherche qui est ici posée, à plus forte raison en santé publique où la réflexion éthique est constamment discutée [10-12]. L'absence de partage systématique des programmes peut révéler un manque de préoccupation quant à la finalité de la recherche en santé publique, celui de maximiser les résultats

permettant d'améliorer la santé du plus grand nombre. À l'objectif de bénéfice collectif est substitué un objectif de performances individuelles : maximiser le nombre de ses publications en maintenant des oligopoles d'expertise ou accroître son influence pour obtenir de plus grands financements pour son équipe. Il semble difficile d'attribuer la responsabilité de ce manque de reproductibilité aux seuls chercheurs [13].

Nous émettons ainsi l'hypothèse d'une « inconduite éthique instituée ». Paul E. Smaldino et Richard McElreath [14] proposent un modèle multi-agents, expliquant la pression de sélection en recherche, qui conduit à l'utilisation de méthodologies de moins bonne qualité et à des taux de plus en plus élevés de découvertes se révélant finalement fausses. Ils montrent que le changement de pratiques de la communauté des chercheurs ne peut être obtenu en l'absence de mesures incitatives institutionnelles. Ils montrent également que la réplication des études ralentit, mais n'arrête pas, le processus de détérioration méthodologique. Leurs résultats appuient donc notre hypothèse.

Conclusion

Toutes les études portant sur le Système national des données de santé (SNDS) ne satisfont pas à l'ensemble des critères de reproductibilité méthodologique. Pourtant, les données qui en sont issues sont partagées. L'hypothèse d'une inconduite éthique instituée pourrait expliquer l'absence de partage des programmes, ce qui nuit à la reproductibilité méthodologique. Au-delà de la nécessité de mesures incitatives institutionnelles, le partage des programmes des travaux portant sur le SNDS devient un impératif catégorique, se justifiant de lui-même indépendamment des objectifs particuliers d'une étude. ♦

SUMMARY

Are studies of claims databases reproducible? The hypothesis of an instituted ethical misconduct in public health

Methodological reproducibility refers to the ability to reproduce exactly the same results by reproducing the same study protocol on the same data. The aim of this study was to assess the methodological reproducibility of studies published in the *Revue d'Épidémiologie et de Santé Publique* between 2008 and 2017 and using data from the national health data system. Our results suggest that only 49% of the studies could be replicated without the help of the authors. The results may reveal a lack of concern about the purpose of public health research. It is difficult to attribute responsibility for this default of reproducibility solely to researchers, so we hypothesize an instituted ethical misconduct. ♦

