Master 2 Recherche : Modélisation en Pharmacologie Clinique et Epidémiologie

Responsable : Pr. Eric BELLISSANT

Université de Rennes 1

Nom et prénom de l'étudiant	CLAUDE PELLEN
Parcours choisi (BS, EPI, PEPI, PCET)	PCET
Nom et n° de l'unité de recherche d'accueil Nom du directeur Adresse Tel Fax	Inserm CIC 1414 Bruno Laviolle CHU de Rennes, 2 rue Henri Le Guilloux, Rennes 02 99 28 43 21
Nom du Directeur de Recherche du stage Tel Fax	Florian Naudet 02 99 28 43 21

Sujet de recherche (1,5 pages maximum)

Background

In medicine, randomized controlled trials (RCTs) are considered high quality evidence and therefore availability of their data is expected to be crucial in the evaluation of health interventions (e.g. for individual patient data meta-analyses).

In June 2017, the International Committee of Medical Journal Editors (ICMJE) required a data sharing plan to be included in each paper (and pre-specified in study registration). As these new requirements for publishing experimental findings come into effect, it will be necessary to assess whether they have the intended effects.

Among the leading general medical journals, New England Journal of Medicine, Lancet, JAMA and JAMA Internal Medicine have had no specific policy for data sharing in RCTs until recently. BMJ and PLOS Medicine had already adopted stronger policies, beyond the ICMJE policy, that mandate data sharing for RCTs. We already know that their policy is still imperfect concerning data sharing rates (Naudet et al., in revisions).

More interestingly, Annals of Internal Medicine has encouraged (but not demanded) data-sharing since 2007. This is largely what is required by the new ICMJE policy. Therefore, a retrospective analysis of the RCTs published in Annals of Internal Medicine between 2007 and 2017 will provide the first (and already available) evidence on the expected impact of the ICMJE policy.

Objective

- To describe the impact of data-sharing practices at Annals of Internal Medicine over ten years under their policy;
- To assess the impact of data sharing on re-use of data;
- To assess the impact of data sharing on citation rates;

Data-bases searches

RCTs submitted and published at the Annals of Internal Medicine between April 2007 and December 2017 will be retrieved on PubMed following a search algorithm developed with an information specialist. A crude estimate of 257 potential studies was found in a preliminary search of PubMed using the following strategy: *(annals of internal medicine) AND ("2007/04/01"[Date - Publication]: "2017/12/31"[Date - Publication]) / limitation randomized controlled trial*.

We will include publications of RCTs, including parallel two group and multiple groups, cluster trials and cross-over studies, non-inferiority designs and superiority designs. All publications will be inspected to exclude secondary analyses (e.g. primary paper in NEJM or JAMA, secondary in Annals) and re-analyses of a RCT. The publications without reproducible research statement (i.e. those that were not eligible to the policy) will be excluded.

Then, we will use the Clarivate Web of Science database to identify secondary publications for these primary trials. It would be extremely unlikely for a secondary publication not to cite the primary trial. We will identify and record the total number of citations of each primary trial publication. We will record the number of citations by articles that use individual-level data from the original trial. For articles considered to potentially reflect secondary publications, the full text will be examined to confirm eligibility. Moreover, whenever a design article will be cited in the paper (an article describing the methods and protocol of a trial), we will enter those articles in Web of Science searches to identify additional citations as well.

Design

Survey of published RCTs. For each included study, information will be extracted on:

1) Characteristics of the study (year, country (USA/ Europe/Asia/other), intervention type (drug/device/complex intervention), control group (active/inactive), medical specialty (medicine/surgery/psychiatry), sample size, result on the primary outcome (positive/negative), funding source (academic/industry/charity/mixt));

2) Data sharing plan:

- A statement indicating that data can be available (i.e. intent to share data, 1/ yes, 2/ no) presented in the data-sharing plan;
- Type of data sharing plan (1/ directly available, 2/ available upon request) when investigators agree to share data;
- Material that is shared (1/ data-set, 2/ code, 3/ study protocol);
- Reasons for not sharing data when investigators don't want to share data;

3) PRIMARY OUTCOME:

Re-use of the data (yes/no) documented in all the citing papers, from Web of Science. This is a composite outcome defined by any secondary use 1/ in a re-analysis, 2/ in a secondary analysis and 3/ in an MIPD. This primary outcome will be limited to citing papers where the first, last and corresponding authors are not among the authors of the original paper.

4) SECONDARY OUTCOMES:

- Components (1, 2, and 3) of the primary outcome;
- Re-use of the data (yes/no) documented in all the citing papers, from Web of Science without the authorship restriction used for the primary outcome;
- Number and type of secondary use (re-analysis, secondary analysis, MIPD) in which lead authors (first, corresponding and last) of the paper were outside the team authoring the original Annals paper;
- Number and type of secondary use (re-analysis, secondary analysis, MIPD) by an independent team (no trial author in common);
- Number and type of secondary use (re-analysis, secondary analysis, MIPD) by a team where <50% of authors are from the original Annals paper;
- Number and type of secondary use (re-analysis, secondary analysis, MIPD) by a team where ≥50% of authors are from the original Annals paper;
- Number and type of secondary use (re-analysis, secondary analysis, MIPD) by a team where 100% of authors are from the original Annals paper;
- Mention in the secondary papers about the availability of the re-used data (yes/no) and how (qualitative data);
- Number of citations (from Web of Science);

First, we will describe 'intent to share data' rates over time with a model adjusted for country (USA/ Europe/Asia/other), intervention type (drug/device/complex intervention), control group (active/inactive), medical specialty (medicine/surgery/psychiatry), result on the primary outcome (positive/negative), funding source (academic/industry/charity/mixt).

Then, we will compare the different outcomes between papers intending to share data and papers not intending to share data by adjusting on year, country (USA/ Europe/Asia/other), intervention type (drug/device/complex intervention), control group (active/inactive), medical specialty (medicine/surgery/psychiatry), result on the primary outcome (positive/negative), funding source (academic/industry/charity/mixt).

General strategy for modeling: a univariate model will be used to explore the relationship between the dependent and explanatory variables. All associations with a p-value <0.25 will be subsequently explored in a multivariate model. In all models, we will have to consider studies that might be published by the same team (by adding a random effect).
All analyses will be done in R version 3.4.1 (2017-06-30, The R Foundation for Statistical Computing).

Expected impact

There is a need for some monitoring of data sharing practices to ensure that the new ICMJE's requirements are effective and useful. To this end, one should keep in mind that data-availability is a surrogate of the expected benefit of having open data. Proving that data-sharing rates impacts re-use of the data is a further step. Proving that this reuse might translate into discovery that can change care without generating false positive findings (e.g. in series of unreliable a posteriori subgroup analyses) is even more challenging.
It is necessary to provide robust evidence that data-sharing has more benefits than the associated risks (e.g. problems with the privacy of clinical trial participants). This study will be a first step in exploring these issues.

Collaboration

David Moher, ORCID: 0000-0003-2434-4206
John P.A. Ioannidis, ORCID: 0000-0003-3118-6859

References

Devereaux, P. J., G. Guyatt, H. Gerstein, S. Connolly, and S. Yusuf. "Toward Fairness in Data Sharing." [In eng]. N Engl J Med 375, no. 5 (Aug 04 2016): 405-7.
Haug, C. J. "Whose Data Are They Anyway? Can a Patient Perspective Advance the Data-Sharing Debate?" [In eng]. N Engl J Med 376, no. 23 (Jun 08 2017): 2203-05.
Kalager, M., H. O. Adami, and M. Bretthauer. "Recognizing Data Generation." [In eng]. N Engl J Med 374, no. 19 (May 12 2016): 1898.
Laine, C., S. N. Goodman, M. E. Griswold, and H. C. Sox. "Reproducible Research: Moving toward Research the Public Can Really Trust." [In eng]. Ann Intern Med 146, no. 6 (Mar 20 2007): 450-3.
Platt, R., and J. Ramsberg. "Challenges for Sharing Data from Embedded Research." [In eng]. N Engl J Med 374, no. 19 (May 12 2016): 1897.
Rockhold, F., P. Nisen, and A. Freeman. "Data Sharing at a Crossroads." [In eng]. N Engl J Med 375, no. 12 (Sep 22 2016): 1115-7.
Rosenbaum, L. "Bridging the Data-Sharing Divide - Seeing the Devil in the Details, Not the Other Camp." [In eng]. N Engl J Med 376, no. 23 (Jun 08 2017): 2201-03.
Taichman, D. B., J. Backus, C. Baethge, H. Bauchner, P. W. de Leeuw, J. M. Drazen, J. Fletcher, et al. "Sharing Clinical Trial Data: A Proposal from the International Committee of Medical Journal Editors." [In eng]. PLoS Med 13, no. 1 (Jan 2016): e1001950.
Taichman, D. B., P. Sahni, A. Pinborg, L. Peiperl, C. Laine, A. James, S. T. Hong, et al. "Data Sharing Statements for Clinical Trials." [In eng]. BMJ 357 (Jun 05 2017): j2372.

Appendix 1: examples of Reproducible Research Statement

Reproducible Research Statement:
Study protocol: See Supplement 1 (available at Annals.org).
*Statistical code:*Available from the Supplement 2 (available at Annals.org) or from Dr. Olsen (vog.av@neslo.neram).
Data set: Available through a data use agreement from Dr. Olsen (vog.av@neslo.neram).

Reproducible Research Statement:
Study protocol: See the Supplement (available at www.annals.org).
Statistical code and data set: Available from Dr. Wang (e-mail, gro.retneclacidemstfut@ 2gnawc).

Reproducible Research Statement:
Study protocol and statistical code: Not available.
Data set: Will be made available for public access according to National Institutes of Health policy.

Reproducible Research Statement:
Study protocol, statistical code, and data set: Available through written agreement with the Division of AIDS.

Reproducible Research Statement:
Protocol: Available to interested readers by contacting Diana Buist, PhD at gro.chg@d.tsiub.
Statistical Code: Available to interested readers by contacting Melissa Anderson, MS at gro.chg@assileM.nosrednA.
Data: not available.

Reproducible Research Statement:
Protocol: Available to interested readers by contacting Dr. Kutner at ude.revnedcu@rentuk.naej. Message and simple touch protocols are available as electronic appendices at annals.org.
Statistical code: Available to interested readers by contacting Dr. Fairclough at ude.revnedcu@hguolcriaf.enaid.
Data: Available to interested readers by contacting Dr. Kutner at ude.revnedcu@rentuk.naej.

Reproducible Research Statement:
Study protocol: Available from Dr. Goldbach-Mansky (e-mail, vog.hin.liam@rcabdlog).
Statistical code: Available from Dr. Wesley (e-mail, vog.hin.cc@yelsewb).
Data set: Available to qualified investigators by contacting Dr. Goldbach-Mansky (e-mail, vog.hin.liam@rcabdlog).