In the end, what matters most is not the metaresearch classification as primary or secondary research but the characteristics of this research that should be reliable, reproducible, and trustworthy.

Clovis Mariano Faggion Jr*
Faculty of Dentistry
Department of Periodontology and Operative Dentistry
University of Münster
Münster, Germany

Karla Tatiana Diaz
Master in Implant Dentistry
Faculty of Dentistry
Universidad Peruana Cayetano Heredia
Lima, Peru
Graduate School
Faculty of Dentistry
Universidad Privada San Juan Bautista
Lima, Peru

*Corresponding author. Faculty of Dentistry, University of Münster, Waldeyerstrasse 30, 48149 Münster, Germany. Tel.: +49-251-83-47061; fax: +49-251-83-47134.
E-mail address: clovisfaggion@yahoo.com (C.M. Faggion)

References

[1] Puljak L. Research-on-research studies, or methodological studies, are primary research. J Clin Epidemiol 2019. https://doi.org/10.1016/j.jclinepi.2019.04.013.

[2] Terry RF, van der Rijt T. Overview of research activities associated with the World Health Organization: results of a survey covering 2006/07. Health Res Policy Syst 2010;8:25.

[3] Cochrane. Available at: https://community.cochrane.org/glossary#letter-S. Accessed March 28, 2019.

[4] Glass GV. Primary, secondary, and meta-analysis of research. Educ Res 1976;5(10):3−8.

[5] Faggion CM Jr, Diaz KT. Overview authors rarely defined systematic reviews that are included in their overviews. J Clin Epidemiol 2019;109:70−4.

https://doi.org/10.1016/j.jclinepi.2019.04.005

Measuring clinical uncertainty as a preliminary step to randomized controlled trials

Reproducibility is a fundamental scientific property required of any instrument, diagnostic test, or prognostic score [1]. Treatment decisions or management recommendations made by clinicians are authoritative judgments with real-life impact on patients. Shouldn't we verify that they are reliable? How can this be done? Here we propose to study the reliability of clinical management decisions to measure uncertainty before the conduct of a randomized trial.

It seems natural to think that recommendations made by doctors are inevitably variable and ungeneralizable, for working on a case-by-case basis and with a complex and ineffable reasoning process, they must take into account the unique histories, characteristics, and circumstances of the particular patient [2]. However, this in itself does not make clinicians' recommendations fundamentally different from other clinical judgments such as diagnosis which equally concerns unique individuals: the clinician's verdict (the output of the process) often comes down to allocating the patient to one of a few categories, whether the judgment concerns diagnosis (disease present/absent) or management options (i.e., do not treat/treat medically/treat surgically). No matter the process behind the clinical decision, if it results in contradictory judgments or courses of action when the same patient is presented to different or to the same clinicians more than once, then the process is unreliable and the verdict uncertain.

The similarities in allocating a particular patient a diagnostic or management category are such that we naturally saw that the methodology typically used to study the reliability of imaging diagnoses could be concretely applied to clinical decision-making (Table 1).

In practice, a portfolio of diverse individual patients, all who share a similar clinical problem, and which cover a wide spectrum of clinical presentations, are independently submitted to a variety of clinicians who manage that problem. Clinicians are asked to choose one of the predefined management options to generate interobserver kappa statistics, whereas intrarater agreement can be assessed by a second independent evaluation at a later time. Variability and inconsistency in clinical decision-making need not be resolved through consensus sessions. The studies we propose are designed to transparently identify and measure the clinical uncertainty involved in the management of specific clinical problems, not to provide the a "truth" based on expert opinion. Such studies can be informative: the uncertainty can reveal gaps in medical knowledge or identify suboptimal practices that could be improved. The identification and estimation of such uncertainties can serve many purposes: i) clinicians and patients should be made aware that diverse options are actually proposed for the management of similar patients, if only to make alternative options available; ii) clinicians may be reassured when they realize they are not alone in being uncertain. This step may encourage members of the community to proceed with the clinical research that will addresses that uncertainty; iii) when a reliability study is designed with a randomized trial in view [7−10], it can provide empirical evidence of Freidman's notion of "clinical equipoise", that "an honest professional disagreement among expert clinicians about the preferred treatment" exists [11].

The methodology we propose here is inspired from diagnostic agreement studies, and its application to management decisions permits the demonstration of

---

Conflict of interest: All authors declare that they have no significant financial, professional, or personal interests that might have influenced the performance or presentation of the work described in this article.

Table 1
Examples of agreement studies for management decisions

Management decision under study	Patients	No. of clinicians	Judgments	Kappa	Corresponding RCT	Ref
Clip or coil for unruptured aneurysms	41	28	1,148	<0.4	CURES	[3]
Flow diversion of aneurysms	35	40	1,400	<0.3	FIAT	[4]
Clip or coil for ruptured aneurysms	42	85	3,570	<0.3	ISAT-2	[5]
Management of arteriovenous malformations	35	30	1,050	<0.3	TOBAS	[6]
Management of acute stroke	41	86	3,526	<0.5	EASI	[7]

We have performed several agreement studies (references in last column) for various clinical dilemmas for which there was a lack of evidence (displayed in first column). We built a portfolio of a sufficient number of patients and sent it to several clinicians (numbers displayed in second and third columns, respectively) to gather enough judgments (4th column) to measure agreement/uncertainty with kappa statistics. The lack of agreement for each of these dilemmas (demonstrated by the low kappa values, as shown in 5th column) demonstrates the need for each of the corresponding RCTs that have subsequently been launched (6th column).

the lack of agreement which underlines the need for randomized controlled trials. We encourage trialists to consider using this method as a preliminary study to show the medical community/ethical boards/funders the need for RCTs appropriate to confront the clinical dilemma.

Robert Fahed
Department of Interventional Neuroradiology
Fondation Rothschild Hospital
Paris, France

Tim E. Darsaut
Division of Neurosurgery, Department of Surgery
University of Alberta Hospital
Mackenzie Health Sciences Centre
Edmonton, Alberta, Canada

Behzad Farzin
Department of Radiology
Centre Hospitalier Universitaire de Montréal – CHUM
Service of Interventional Neuroradiology
Montreal, Canada

Miguel Chagnon
Department of Mathematics and Statistic
Université de Montréal
Montreal, Quebec, Canada

Jean Raymond*
Department of Radiology
Centre Hospitalier Universitaire de Montréal – CHUM
Service of Interventional Neuroradiology
Montreal, Canada

*Corresponding author. Department of Radiology, Centre Hospitalier Universitaire de Montréal – CHUM, Service of Neuroradiology, 1000 Saint-Denis Street, Room D03-5462B, Montreal, Quebec H2X 0C1, Canada. Tel.: +1-514-890-8000x27235; fax: +1-514-412-7489.
E-mail address: jean.raymond@umontreal.ca (J. Raymond)

References

[1] Kottner J, Audige L, Brorson S, Donner A, Gajewski BJ, Hrobjartsson A, et al. Guidelines for reporting reliability and agreement studies (GRRAS) were proposed. J Clin Epidemiol 2011;64:96–106.

[2] Pellegrino ED. The Anatomy of Clinical Judgments. Clinical Judgment: A Critical Appraisal. Dordrecht: Springer; 1977:169–94.

[3] Darsaut TE, Estrade L, Jamali S, Bojanowski MW, Chagnon M, Raymond J. Uncertainty and agreement in the management of unruptured intracranial aneurysms. J Neurosurg 2014;120(3):618–23.

[4] Darsaut TE, Gentric JC, McDougall CM, Gevry G, Roy D, Weill A, et al. Uncertainty and agreement regarding the role of flow diversion in the management of difficult aneurysms. AJNR Am J Neuroradiol 2015;36(5):930–6.

[5] Darsaut TE, Fahed R, Macdonald RL, Arthur AS, Kalani MYS, Arikan F, et al. Surgical or endovascular management of ruptured intracranial aneurysms: an agreement study. J Neurosurg 2018;1:1–7.

[6] Fahed R, Batista AL, Darsaut TE, Gentric JC, Ducroux C, Chaalala C, et al. The Treatment of Brain Arteriovenous Malformation Study (TOBAS): a preliminary inter- and intra-rater agreement study on patient management. J Neuroradiol 2017;44(4):247–53.

[7] Ducroux C, Fahed R, Khoury N, Gevry G, Kalsoum E, Labeyrie M, et al. Intravenous thrombolysis and thrombectomy decisions in acute ischemic stroke: an interrater and intrarater agreement study. Rev Neurol (Paris) 2019;175(6):380–9. https://doi.org/10.1016/j.neurol.2018.

[8] Darsaut TE, Jack AS, Kerr RS, Raymond J. International subarachnoid aneurysm trial - ISAT part II: study protocol for a randomized controlled trial. Trials 2013;14:156.

[9] Darsaut TE, Magro E, Gentric JC, Batista AL, Chaalala C, Roberge D, et al. Treatment of brain AVMs (TOBAS): study protocol for a pragmatic randomized controlled trial. Trials 2015;16:497.

[10] Raymond J, Gentric JC, Darsaut TE, Iancu D, Chagnon M, Weill A, et al. Flow diversion in the treatment of aneurysms: a randomized care trial and registry. J Neurosurg 2017;127(3):454–62.

[11] Raymond J, Darsaut TE, Altman DG. Pragmatic trials can be designed as optimal medical care: principles and methods of care trials. J Clin Epidemiol 2014;67:1150−6.

https://doi.org/10.1016/j.jclinepi.2019.04.011

Reducing waste and increasing value through embedded replicability and reproducibility in systematic review process and automation

Clovis Mariano Faggion Jr. points out a serious problem starting a valuable discussion suggesting two approaches to facilitate the reproducibility of the systematic reviews [1]. Because performing systematic reviews are already time-consuming [2] and resource-rating [3], it is arguable how adding more steps such as reproducibility test that requires even more time and human resources could reduce waste and increase value compared to excluding some steps [4]. Here, I discuss how the replicability of methods and reproducibility of results (RMRR) have been embedded within the systematic reviewing and how "semi-automation" and "sharing" could solve RMRR issues [5].

The masterminds who developed the process of systematic reviewing considered involving at least between two or three people in screening and data extraction steps. Although the purpose of double-checking could be to reduce the errors [6,7], it also means the screening and data extraction are being repeated or replicated by at least one other member of the team to ensure the reproducibility of the same results in each step; when there are discrepancies, either these two members reconsider the decision for the third time or they ask the third member's opinion. These two steps enjoy RMRR as embedded within the methodology. But how do we know if what has been said in the systematic review has actually been done? We usually trust the researchers but using the existing online semi-automated platforms that document the steps of the systematic reviews [8−12] could help the transparency if the team share the processes and methods openly, and share the results in findable, accessible, interoperable, and reusable format as advised by FAIR principles [13]. This is not the best practice right now [14] but we have what it takes to do the systematic reviews once without being worried about RMRR that is also a requirement in update step.

Following the protocol and sharing the data, [15,16] on the other hand, the meta-analysis step—based on established math embedded within software programs—can be repeated conveniently. It only leaves vulnerable search step behind. I intentionally kept the search, the first step, to discuss at last.

Like meta-analysis the search is rooted in computerized systems with certain differences:

- Unlike the computer programs for meta-analysis, the databases are not freely accessible to develop the search strategies or to repeat them;
- Preferred Reporting Items for Systematic Reviews and Meta-Analyses (RISMA) mandates reporting of search strategy for at least one database [17], so RMRR is possible for only one database not all the databases;
- Last but not least, even if the authors decide to be generous in reporting the search strategies for all databases, they do not share the search results [14]. The main excuse for not sharing the search results is that the abstracts are copyrighted; however, it is and was always possible to share the search results excluding the copyrighted abstract after deduplication in RIS (RefMan/Reference Manager) format.

Apart from those review teams who have the privilege of using a study-based register to conduct a register-based study [4], the search step is the weakest point in terms of RMRR in evidence synthesis [18].

Although many follow PRISMA guideline in reporting the systematic reviews, it is not currently the primary purpose of PRISMA to ensure the RMRR in systematic reviews. My suggestion is for PRISMA 2019 to include items that enforces the scientific principles of RMRR through public data/methods sharing based on FAIR principles and using the online automated platforms where they support public accessibility to documented processes, methods, and data as recommended within seven available strategies for reproducibility of systematic reviews [18].

Farhad Shokraneh*
Research Fellow
Division of Psychiatry and Applied Psychology
Institute of Mental Health, School of Medicine
University of Nottingham
Nottingham, UK
Division of Psychiatry and Applied Psychology,
Institute of Mental Health, School of Medicine,
University of Nottingham, Triumph Road,
Nottingham NG7 2TU, UK.
Tel.: +44 115 74 86747.
E-mail address: *Farhad.Shokraneh@nottingham.ac.uk*

References

[1] Faggion CM. Should a systematic review be tested for reproducibility before its publication? J Clin Epidemiol 2019. [Epub ahead of print].
[2] Sampson M, Shojania KG, Garritty C, Horsley T, Ocampo M, Moher D. Systematic reviews can be produced and published faster. J Clin Epidemiol 2008;61:531−6.
[3] Borah R, Brown AW, Capers PL, Kaiser KA. Analysis of the time and workers needed to conduct systematic reviews of medical interventions using data from the PROSPERO registry. BMJ Open 2017;7(2):e012545.

Funding: None.
Conflict of interest: None.
Declarations of interest: None.