



Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.

EDITORIAL

Methodological challenges in studying the COVID-19 pandemic crisis

The world is not the same as it was when we wrote last month's editorial. There is a great sense of urgency and a wish to contribute to the evaluation of the evidence base and to generating new knowledge for assessing and managing the COVID-19 pandemic. The focus of this journal is in the methods used to build this knowledge base. Below we list some ideas that we hope our community of readers and authors will tackle. We would welcome good papers on these important methodological challenges, to make a difference in effectively fighting the current pandemic and anticipating possible similar future crises. We consider a Covid -19 section based on these submissions, for which we will set up an expedited process for reviewing and publishing.

During the crisis, a general challenge is the necessity of making healthcare and policy decisions in situations of substantial uncertainty, when there is little knowledge and no direct evidence base to manage a yet largely unknown agent-related threat. In this context, a concrete challenge is providing methodological approaches/support for:

- high *speed scoping reviews* of potentially useful prior knowledge and experience from similar previous events (epidemiological, public health, biomedical, clinical);
- *surveillance to monitor the current pandemic threat and how to model the spread*: how to do this, what are useful (big) data sources, and how to assess risk of bias;
- *real-time studies of the occurrence* of COVID-19 (incidence, complications rate, lethality, with uncertainty ranges) and its biomedical and social determinants, in order to optimally predict (modeling) numbers and speed of spread, course and duration, relapse risk, and health care needs;
- *evaluation of diagnostic tests* for
 - active corona infection: accuracy and prediction of better health outcomes (addressing, e.g., reference standards and appropriate endpoints);
 - past corona infection: accuracy and implications for protection also of other people (knowledge of immunity is crucial for patient care, society and economy);
- *evaluation of (candidate) antiviral agents and vaccines*: how to ethically fast-track trials of benefit and assess adverse effects; and how to assess the

trade-off of the risk of adverse effects vs. the risk of more casualties or not getting a grip on the problem at all if the intervention is too late; what are appropriate (intermediate) outcome measures; how to recognize risk groups that need to be prioritized; and, how to implement appropriate ethical review;

- *evaluation of pandemic-specific interventions*, such as social distancing, personal protective equipment, isolation, apps for tracing COVID-19 contacts, and group-level interventions (e.g., closing schools);
- *standards for research and reporting* adapted to the particular situation, both to support researchers working under pressure and to prevent malpractice and abuse;
- *evaluating the strength of recommendations* for public health and health care practice;
- *decision making under uncertainty*: how to evaluate the effectiveness of, for example, medical decision making, shared care, and multilevel decision making approaches (taking individual-, risk group-, and population-specific interests into account).

After the current crisis when the viral pandemic is over or has turned into an intermittently occurring viral disease (that may still threaten older and frail patient with multiple chronic disorders), other methods challenges arise, including:

- *methodology to support establishing, monitoring and evaluating post-pandemic COVID-19 strategies*;
- *methods to retrospectively evaluate* COVID-19 occurrence over time and its determinants, long term prognosis, and the population health impact of the implemented risk group testing, and intervention (public health, drugs, vaccine) strategies.

Finally, learning points for the prevention/anticipation of new pandemics in the future are important, for example

- *improving the methodology of studying (effectiveness of) prevention and early detection* of emerging zoonoses and potential pandemic development;
- *anticipation and monitoring strategies* (integrating public health, clinical, and biomedical data, and big data sources);
- *improving the modeling and prediction* of a pandemic when the evidence is scarce;

- *improving the methodology of both timely and appropriate developing, evaluating and approving tests and vaccines;*
- *studying pandemic-related disruption of non-pandemic-related regular health care.*

Turning to the contents of this issue: Karl Popper made many contributions to medical science and one of them was his advocacy of counterfactuals [1]. In a Commentary designed as a tutorial, Bours uses this concept of counterfactuals to lay out the key concepts behind confounding and demonstrates this with causal diagrams such as Directed Acyclic Graphs; and also provides design and analysis strategies for combatting confounding along with examples such as confounding by indication.

Prospective registration provides an important safeguard against selective reporting of positive studies, changing primary research questions and cherry picking of analyses. It is now increasingly required for clinical trials and systematic reviews of intervention and researcher designed observational studies, but what about the database studies using administrative datasets not designed for research that are being widely encouraged to provide Real World Evidence [RWE]? Zarin et al. review the situation and argue that four key needs urgently need addressing: (1) an unambiguous list of studies; (2) well-defined and broadly enforced policies to achieve a comprehensive listing of studies; (3) international collaboration on a registration system for the prevention of informational “chaos”; (4) pre-specified planned analyses with a defined level of detail to distinguish pre-specified from post hoc analyses. The authors call for constructive engagement of the various relevant stakeholders to agree on what is ‘good enough’ that identifies minimal requirements that balances caution against new regulations being too burdensome.

Predatory publishing is one of the evils on the 21st century that we have commented on before [2]. Hayden reports on an under-recognised new threat namely the challenge to the trustworthiness of systematic reviews due to the inclusion of primary studies published in predatory journals. As she points out this is a serious problem since it dilutes the credible literature through citation and further dissemination of untrustworthy evidence; increases the amount of poor quality data, or studies with unusable data; and includes duplicated or fraudulent data in systematic reviews. The magnitude of this is unknown but in one Cochrane review by the author of this commentary, 5 of 65 references when checked were found to be papers from suspected predatory journals. Hayden recommends that action be taken urgently. This could include using search strategies that limit or exclude articles from journals with inadequate peer review processes, using one of the accredited checklists [3] to identifying and excluding reports of trials that exhibit characteristics of predatory publishing, and using sensitivity analyses to test the impact of questionable studies on interpretation of evidence.

It is well accepted that prognostic criteria derived from one cohort need to be confirmed in an independent replication cohort. Iwakami et al. report that current approaches are suboptimal and show that more attention to careful matching of the sampling frame and recruitment details in the replication cohort leads to better agreement. They demonstrated how this was achieved on the 30 day mortality rate from cohort databases of consecutive patients after admission for acute heart failure using two instruments to do this: the critical appraisal tool Quality In Prognosis Studies (QUIPS) for risk-of-bias assessment and the data extraction tool CHcharacteristics in prediction Appraisal and data extraction for systematic Reviews of prediction Modeling Studies [CHARMS]).

Reproducibility of research is a basic tenet of the scientific method. There are various approaches to this using either the same database or collecting new data. Goodman et al. [4] propose the term of ‘inferential reproducibility’ for making of knowledge claims of similar strength from either a study replication or reanalysis of original data. Schreijenberg et al. report on an example of the latter, a re-analysis of original data from a trial with over 1,500 patients to study the efficacy of paracetamol in low back pain. This was the first and only trial showing that paracetamol had no effect on the outcomes of acute low back pain which contradicted most clinical guidelines so was important to check this for robustness. A second new study was attempted but failed to recruit enough patients so the data from the first study were given to an independent set of investigators who carried out a re-analysis using a different analytic approach and selected different primary outcomes [pain and quality of life] felt to be of greater relevance to clinicians than the primary outcome selected for the first analysis [time to recovery]. Similar results were obtained bolstering the conclusions of the first analysis.

This journal encourages systematic reviews of methods but although there is a considerable literature on optimal searching strategies for clinical questions on treatments diagnosis and prognosis, there is little guidance on literature searching strategies for systematic reviews of methods. Bas et al. report that the usual approach of starting with title, abstract and key word only detects less than 50% of relevant articles when they used this approach to search for studies on methods such as propensity score methods, inverse probability weighting, marginal structural modeling, multiple imputation, Kaplan-Meier estimation, number needed to treat, measurement error, randomized controlled trials, and latent class analysis. These authors then showed how text mining can substantially improve such searches.

Although good systematic reviews are recommended for all important health and health care practice and policy questions, the length of time these take [typically a year or more] has been a major drawback for time-sensitive decisions. It is thus very encouraging to see the paper by Clark et al. demonstrating that with the use of ten automated processes a medium-sized systematic review

(classified as 2,000–3,000 search results and 10–20 included studies), on increased fluid intake on urinary tract infection (UTI) recurrence was completed by a dedicated team in 61 person hours over 10 working days. This proof-of-concept case study needs to be more extensively tested and assessed for quality. If satisfactory this could be the model for fast-track services such as the one that Cochrane [formerly the Cochrane Collaboration] provides.

In response to specific challenges that complex interventions were not adequately addressed in the main CONSORT reporting of trials guideline, this was supplemented in 2004 by the publication of the CONSORT extension for Non-Pharmacological Trials (CONSORTnpt). This covers psychotherapy, surgical, and rehabilitation trials. Alvarez et al. review the impact of this extension on the reporting of trials of manual therapy before and after this CONSORTnpt extension was published. There was improvement in the frequency of a flowchart diagram, the estimated effect size, precision descriptions, and the description of intervention procedures. However, sample sizes remain small with few providing a sample size calculation, and there was persistent poor reporting of randomisation, protection against allocation bias, handling of missing data and assessment of harms. The authors call for journals to make article submission conditional on the inclusion of all the information required by the guidelines.

Another important advance in speeding up systematic reviews is the establishment of Cochrane Crowd, Cochrane's citizen science platform, which hosts a global community of more than 14,000 citizens and researchers who undertake identification. Gartlehner et al. report on studying the success of this 'crowd' approach using a sample of ten reviewers for each of 100 abstracts each of pharmacologic and public health papers. After completing a training exercise, participants screened abstracts online based on predefined inclusion and exclusion criteria. Although many systemic reviewer organisations including U.S. Agency for Healthcare Research and Quality (AHRQ), the Campbell Collaboration, Cochrane, and the National Institute for Health and Care Excellence allow a single person to screen articles for inclusion, previous studies have shown that with one screener that over 10% of studies get missed. This paper shows that with at least two reviewers/screeners

with the level of expertise in the Cochrane Crowd, the missed article rate drops to less than 3%.

Faroutani et al. in the 28th JCE article on GRADE describe how the same methodologic issues as for assessing certainty of estimates of treatment benefit and risk (risk of bias, imprecision, inconsistency, indirectness, and publication bias, as well as the domains for rating up) also should be used for assessing the certainty of evidence about prognostic factors. Examples used to demonstrate this approach include the association between delirium and mortality, cervical cancer in indigenous and non-indigenous women, and smoking and venous thromboembolism.

Takwoing et al. report on the 2018 PRISMA–DTA [5] extension of the PRISMA reporting guideline for transparent reporting systematic reviews of diagnostic test accuracy studies. In addition to demonstrating that most such studies in the past have been flawed, they provide guidance on how to apply the new criteria as well as identifying five exemplar reviews that will help authors of this type of study.

J. André Knottnerus

*Department of General Practice
Netherlands School of Primary Care Research
Maastricht University
P.O. Box 616, 6200 MD Maastricht, The Netherlands*

Peter Tugwell

*Department of Medicine
School of Epidemiology and Public Health
University of Ottawa
Ottawa Hospital Research Institute
Bruyere Research Institute
Ottawa, ON, Canada*

References

- [1] Popper K. *The Logic of Scientific Discovery*. *Logik der Forschung*; 1935.
- [2] Knottnerus JA, Tugwell P. Evidence-based medicine: achievements and prospects. *J Clin Epidemiol* 2017;84:1–2. <https://doi.org/10.1016/j.jclinepi.2017.02.006>.
- [3] Submit TC. Choose the right journal for your research. Available at <https://thinkchecksubmit.org/>. Accessed October 4, 2019.
- [4] Goodman SN, Fanelli D, Ioannidis JP. What does research reproducibility mean? *Sci Transl Med* 2016;8(341):341ps12.
- [5] McInnes MDF, Moher D, Thoms BD, McGrath TA, Bossuyt PM, PRISMA-DTA Group. Preferred reporting items for a systematic review and meta-analysis of diagnostic test accuracy studies: the PRISMA-DTA statement. *JAMA* 2018;319:388–96.