

EDITORIAL

Open Access

Methodological challenges of analysing COVID-19 data during the pandemic



Martin Wolkewitz¹ and Livia Puljak^{2*}

Editorial

On March 11, 2020, the World Health Organization (WHO) declared that COVID-19 can be characterized as a pandemic [1]. The disease is caused by the novel coronavirus SARS-CoV-2, which rapidly overwhelmed the entire world. The virus was first described in China in December 2019, in early January it was already characterized, and already on January 30, 2020, the outbreak was declared a Public Health Emergency of International Concern, which later evolved into a pandemic [1].

Devastating and unpredictable spread of COVID-19 throughout the world has caused unprecedented global lockdowns and immense burden for healthcare systems. The WHO called for immediate research actions including “immediately assess available data to learn what standard of care approaches are the most effective” and “evaluate as fast as possible the effect of adjunctive and supportive therapies” [1].

This pandemic is now an enormous challenge for researchers, clinicians, health-care workers, epidemiologists and decision-makers. BMC Medical Research Methodology would like to contribute to this global endeavour by setting up a collection of articles called “Methodologies for COVID-19 research and data analysis”. As Guest Editors of the Collection, we would like to offer our views regarding methodological challenges where researchers can help.

Statistical challenges of analysing COVID-19 data

Statistical models will play a major role in “fighting panic with information” [2] to avoid or at least minimize the risk of bias which is a common threat in clinical and epidemiological studies. In this article, we describe the

most striking challenges for statisticians and data analysts who want to provide support in this pandemic with their expertise.

Getting proper clinical data of active and closed COVID-19 cases

After the outbreak in Wuhan, China (available as open access epidemiological data [3]), clinical data can be prospectively collected in a cohort study design. Merging and cleaning of data from large multi-centre hospitals is crucial and requires sophisticated data management. Artificial intelligence and deep learning algorithm might be suitable to tackle this challenge. Data security, patients consent, ethics statements are essential in non-pandemic situation but they are bureaucratic barriers to get rapid access to clinical data. Pandemic situations require specific handling of these issues and should be discussed on national level.

We have to distinguish between active (still hospitalized) and closed (discharged or dead) COVID-19 cases. Case report forms (CRF) for patients with suspected or confirmed COVID-19 are needed to collect and store their data in a standardised way. There are two main initiatives which created protocols for the investigators, the ‘International Severe Acute Respiratory and emerging Infection Consortium (ISARIC)’ (isaric.tghn.org) and the ‘Lean European Open Survey on SARS-CoV-2 Infected Patients (LEOSS)’ (leoss.net). In these two initiatives, it is planned that only closed COVID-19 cases are stored.

Understanding the complexity of clinical endpoints

Endpoints in patients with severe pneumonia are challenging [4]. For COVID-19 patients, the most relevant clinical endpoints are the admission to intensive care, invasive ventilation and survival. Less relevant endpoints include the need of supportive oxygen. The analysis of

* Correspondence: livia.puljak@unicath.hr

²Center for Evidence-Based Medicine and Health Care, Catholic University of Croatia, Ilica 242, 10000 Zagreb, Croatia

Full list of author information is available at the end of the article



© The Author(s). 2020 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

these endpoints requires complex models which handles the time-dependent dynamic of the data.

Understanding common statistical pitfalls in clinical epidemiology

Clinical data are highly time-dependent and require advanced statistical methods to avoid common pitfalls such as selection, length, immortal-time and competing risk bias [5–8].

Developing appropriate analysis strategies

In the same way as data should be collected in a standardised way, data should also be analysed in a standardised way. Statisticians are encouraged to develop suitable analytical strategies to analyse data which were collected from standardised protocols (such as ISARIC and LEOSS).

Communicating statistical effects and distinguishing them from artefacts

Communicating statistics, especially in hectic times during a pandemic, is very challenging. Statisticians are encouraged to support this with clear and transparent statements.

Learning from similar studies about SARS, MERS and influenza A(H1N1pdm09)

As in other outbreaks such as SARS in 2002–2003, clinicians are confronted with new diseases for which there is limited knowledge of effective treatment options [9]. Since there is no targeted agent for COVID-19 in such an early outbreak phase, repurposing of available antiviral drugs and corticosteroids is discussed [9–16], based on case series [17–23]. Until promising targeted randomized controlled trials exist, it is expected that large observational clinical studies will be performed to evaluate potential treatment effects as it was done, for instance, for SARS, MERS and influenza A(H1N1pdm09) on hospital mortality [24–27]. Observational studies cannot replace randomized controlled trials due to their limited ability to draw causal conclusions. However, they can be used to stimulate further research on the effectiveness of potential treatment options.

Updating reporting guidelines for observational studies during a pandemic

In pandemic situation, rapid and valid information flow and reporting is crucial. Long-lasting reporting guidelines might do more harm than good. Specific reporting guidelines are needed for pandemic settings.

Statistical support for randomized trial

The first randomized trial about Lopinavir–Ritonavir for Covid-19 patients has already been published and

showed no promising effect [28]. Statistical expertise is needed to understand potential effects on the complexity of clinical endpoints.

Other methodological challenges in research on COVID-19

Beyond challenges related to data analysis, there are many other methodological challenges related to research on SARS-CoV-2 and COVID-19.

Searching for relevant information sources

We are witnessing tremendous growth of articles published on this topic, already counting in thousands. For methodologists and researchers in the field of evidence synthesis, the challenge will be searching for the relevant information sources. Creating specialized, publicly accessible collection of studies with original studies about COVID-19 can surely help in this. For example, WHO has set up a collection of articles about COVID-19, compiled in a publicly available database. On March 30, 2020 this database had already included 3294 articles.

Source of those articles is described by WHO as [quote]: “We update the database daily from searches of bibliographic databases, hand searches of the table of contents of relevant journals, and the addition of other relevant scientific articles that come to our attention” [29]. However, by 6 April 2020 it was not publicly reported which databases and journals are searched for this purpose. The WHO web site offers several crude search filters available, for searching these articles. The WHO also offers filtering for “Newest updates”, but it is not clear how new are the newest updates, i.e. there is no search by date. The articles in the database can be downloaded, but cursory look at those articles indicates that the majority of them do not have original data; instead it appears that the majority are news, commentaries and opinions. Thus, it would be useful to separate articles in this database that actually report original data. At the time when this article went to publication, multiple other collections of evidence on COVID-19 were being announced and set up, indicating that multiple teams globally are creating the same or similar evidence collections, leading to needless waste of human resources.

Synthesizing evidence rapidly

In a world where each day brings hundreds of new articles on a hot topic, conducting evidence synthesis will be particularly challenging. Systematic reviews are considered by many as the highest-level of evidence in the hierarchy of evidence in medicine, but their production often takes years [30, 31]. However, multiple systematic reviews about COVID-19 have already been published. It remains to be seen what is the quality of those rapidly produced systematic reviews.

Producing evidence syntheses on a short time scale usually requires cutting corners with methodology, and for this reason, rapid reviews have evolved. Rapid reviews are conducted with a condensed timeline, sacrificing certain aspects of systematic review methodology for speed [32]. Pilot study has shown, for example, that rapid research needs appraisal can be conducted within 5 days in the case of an infectious disease outbreak [33]. However, it has also been shown that transparency and inadequate reporting are the major limitations of rapid reviews [34].

Ensuring adequate quality of published research

Journal editors are currently under pressure to publish relevant articles on COVID-19 quickly, which has been described as “rather maddening”. It has been argued that this could also be advantageous in a long run, as it can help journals to become more efficient in future.

However, haste is likely to be detrimental to the quality of publications. Speed is not necessarily a friend of good science. Articles may be assembled too quickly, publishing processes may be hastened, and quality of peer-review may not be adequate. Anecdotal reports indicate that highly specialized experts in the field may be swamped with requests for peer-review that they are unable to accommodate, which may lead to inviting less specialized peer-reviewers, to the detriment of manuscript quality check. We will need to wait to find out how many corrections and retractions there will be for journals published hastily on the topic of COVID-19, and whether methodological and reporting quality of those articles will be lower compared to the articles on other topics. In the times of emergency, researchers should still pay attention to transparency and adequate reporting of their research, to ensure its reproducibility.

Data sharing

To enable analysis of data gathered during COVID-19 pandemic, principles of open science and raw data sharing will be of utmost importance. Global norms have been proposed [35] for data sharing during global health emergencies, and it remains to be seen whether researchers will be more likely to share their raw data publicly in articles covering COVID-19.

In conclusion, there are many methodological challenges related to producing, gathering, analysing, reporting and publishing data in condensed timelines required during a pandemic. We certainly did not mention all of them, but we hope that researchers willing to contribute to research methodology related to COVID-19 will help us address those other issues as well. It is customarily said that each crisis is also an opportunity, and therefore we hope that the BMC Medical Research Methodology will have an

opportunity to publish research articles that will help the humanity win the battle against SARS-CoV-2.

Acknowledgements

None.

Authors' contributions

MW and LP: manuscript design, writing the first draft, revising the first draft, approval of the version to be published. The authors read and approved the final manuscript.

Funding

No extramural funding.

Availability of data and materials

Not applicable.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

Martin Wolkewitz and Livia Puljak are guest editors of the BMC Medical Research Methodology collection “Methodologies for COVID-19 research and data analysis”. Livia Puljak is a Section Editor of the BMC Medical Research Methodology.

Author details

¹Institute of Medical Biometry and Statistics, Faculty of Medicine and Medical Center, University of Freiburg, 79104 Freiburg, Germany. ²Center for Evidence-Based Medicine and Health Care, Catholic University of Croatia, Ilica 242, 10000 Zagreb, Croatia.

Accepted: 7 April 2020

Published online: 14 April 2020

References

- World Health Organization (WHO). <https://www.who.int/>. Accessed 12 Apr 2020.
- The Lancet. COVID-19: fighting panic with information. *Lancet*. 2020; 395(10224):537. [https://doi.org/10.1016/S0140-6736\(20\)30379-2](https://doi.org/10.1016/S0140-6736(20)30379-2).
- Xu B, Kraemer MUG, Xu B, et al. Open access epidemiological data from the COVID-19 outbreak. *Lancet Infect Dis*. 2020. [https://doi.org/10.1016/S1473-3099\(20\)30119-5](https://doi.org/10.1016/S1473-3099(20)30119-5).
- Timsit J-F, de Kraker MEA, Sommer H, et al. Appropriate endpoints for evaluation of new antibiotic therapies for severe infections: a perspective from COMBACTE's STAT-net. *Intensive Care Med*. 2017;43(7):1002–12. <https://doi.org/10.1007/s00134-017-4802-4>.
- Wolkewitz M. Avoidable statistical pitfalls in analyzing length of stay in intensive care units or hospitals. *Crit Care*. 2014;18(1):408.
- Wolkewitz M, Schumacher M. Survival biases lead to flawed conclusions in observational treatment studies of influenza patients. *J Clin Epidemiol*. 2017; 84:121–9. <https://doi.org/10.1016/j.jclinepi.2017.01.008>.
- Wolkewitz M, Allignol A, Harbarth S, de Angelis G, Schumacher M, Beyersmann J. Time-dependent study entries and exposures in cohort studies can easily be sources of different and avoidable types of bias. *J Clin Epidemiol*. 2012;65(11):1171–80.
- Wolkewitz M, Cooper BS, Bonten MJ, Barnett AG, Schumacher M. Interpreting and comparing risks in the presence of competing events. *BMJ*. 2014;349:g5060.
- Stockman LJ, Bellamy R, Garner P. SARS: systematic review of treatment effects. *Low D, ed. PLoS Med*. 2006;3(9):e343. <https://doi.org/10.1371/journal.pmed.0030343>.
- Vetter P, Eckerle I, Kaiser L. Covid-19: a puzzle with many missing pieces. *BMJ*. 2020;1:m627. <https://doi.org/10.1136/bmj.m627>.
- Zumla A, Hui DS, Azhar EI, Memish ZA, Maeurer M. Reducing mortality from 2019-nCoV: host-directed therapies should be an option. *Lancet*. 2020; 395(10224):e35–6. [https://doi.org/10.1016/S0140-6736\(20\)30305-6](https://doi.org/10.1016/S0140-6736(20)30305-6).

12. del Rio C, Malani PN. 2019 novel coronavirus—important information for clinicians. *JAMA*. 2020. <https://doi.org/10.1001/jama.2020.1490>.
13. Sheahan TP, Sims AC, Leist SR, et al. Comparative therapeutic efficacy of remdesivir and combination lopinavir, ritonavir, and interferon beta against MERS-CoV. *Nat Commun*. 2020;11. <https://doi.org/10.1038/s41467-019-13940-6>.
14. Russell CD, Millar JE, Baillie JK. Clinical evidence does not support corticosteroid treatment for 2019-nCoV lung injury. *Lancet*. 2020;395(10223):473–5. [https://doi.org/10.1016/S0140-6736\(20\)30317-2](https://doi.org/10.1016/S0140-6736(20)30317-2).
15. Shang L, Zhao J, Hu Y, Du R, Cao B. On the use of corticosteroids for 2019-nCoV pneumonia. *Lancet*. 2020;395(10225):683–4. [https://doi.org/10.1016/S0140-6736\(20\)30361-5](https://doi.org/10.1016/S0140-6736(20)30361-5).
16. Bouadma L, Lescure F-X, Lucet J-C, Yazdanpanah Y, Timsit J-F. Severe SARS-CoV-2 infections: practical considerations and management strategy for intensivists. *Intensive Care Med*. February 2020. <https://doi.org/10.1007/s00134-020-05967-x>.
17. Yang X, Yu Y, Xu J, et al. Clinical course and outcomes of critically ill patients with SARS-CoV-2 pneumonia in Wuhan, China: a single-centered, retrospective, observational study. *Lancet Respir Med*. 2020. [https://doi.org/10.1016/S2213-2600\(20\)30079-5](https://doi.org/10.1016/S2213-2600(20)30079-5).
18. Xu X-W, Wu X-X, Jiang X-G, et al. Clinical findings in a group of patients infected with the 2019 novel coronavirus (SARS-Cov-2) outside of Wuhan, China: retrospective case series. *BMJ*. 2020;1:m606. <https://doi.org/10.1136/bmj.m606>.
19. Chen N, Zhou M, Dong X, et al. Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study. *Lancet*. 2020;395(10223):507–13. [https://doi.org/10.1016/S0140-6736\(20\)30211-7](https://doi.org/10.1016/S0140-6736(20)30211-7).
20. Young BE, Ong SWX, Kalimuddin S, et al. Epidemiologic features and clinical course of patients infected with SARS-CoV-2 in Singapore. *JAMA*. 2020. <https://doi.org/10.1001/jama.2020.3204>.
21. Huang C, Wang Y, Li X, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet*. 2020. [https://doi.org/10.1016/S0140-6736\(20\)30183-5](https://doi.org/10.1016/S0140-6736(20)30183-5).
22. Wang D, Hu B, Hu C, et al. Clinical characteristics of 138 hospitalized patients with 2019 novel coronavirus-infected pneumonia in Wuhan, China. *JAMA*. 2020. <https://doi.org/10.1001/jama.2020.1585>.
23. Guan W, Ni Z, Hu Y, et al. Clinical characteristics of coronavirus disease 2019 in China. *New Engl J Med*. 2020. <https://doi.org/10.1056/NEJMoa2002032>.
24. Arabi YM, Shalhoub S, Mandourah Y, et al. Ribavirin and interferon therapy for critically ill patients with Middle East respiratory syndrome: a multicenter observational study. *Clin Infect Dis*. <https://doi.org/10.1093/cid/ciz544>.
25. Arabi YM, Mandourah Y, Al-Hameed F, et al. Corticosteroid therapy for critically ill patients with Middle East respiratory syndrome. *Am J Respir Crit Care Med*. 2018;197(6):757–67. <https://doi.org/10.1164/rccm.201706-1172OC>.
26. Delaney JW, Pinto R, Long J, et al. The influence of corticosteroid treatment on the outcome of influenza a(H1N1pdm09)-related critical illness. *Crit Care*. 2016;20. <https://doi.org/10.1186/s13054-016-1230-8>.
27. Muthuri SG, Venkatesan S, Myles PR, et al. Effectiveness of neuraminidase inhibitors in reducing mortality in patients admitted to hospital with influenza a H1N1pdm09 virus infection: a meta-analysis of individual participant data. *Lancet Respir Med*. 2014;2(5):395–404.
28. Cao B, Wang Y, Wen D, et al. A trial of Lopinavir–ritonavir in adults hospitalized with severe Covid-19. *New Engl J Med*. 2020. <https://doi.org/10.1056/NEJMoa2001282>.
29. Global research on coronavirus disease (COVID-19). <https://www.who.int/emergencies/diseases/novel-coronavirus-2019/global-research-on-novel-coronavirus-2019-ncov>. Accessed 4 Mar 2020.
30. Runjic E, Behmen D, Pieper D, et al. Following Cochrane review protocols to completion 10 years later: a retrospective cohort study and author survey. *J Clin Epidemiol*. 2019;111:41–8. <https://doi.org/10.1016/j.jclinepi.2019.03.006>.
31. Runjic E, Rombey T, Pieper D, Puljak L. Half of systematic reviews about pain registered in PROSPERO were not published and the majority had inaccurate status. *J Clin Epidemiol*. 2019;116:114–21. <https://doi.org/10.1016/j.jclinepi.2019.08.010>.
32. Garritty C, Stevens A, Hamel C, Gofam M, Hutton B, Wolfe D. Knowledge synthesis in evidence-based medicine. *Semin Nucl Med*. 2019;49(2):136–44. <https://doi.org/10.1053/j.semnuclmed.2018.11.006>.
33. Sigfrid L, Moore C, Salam AP, et al. A rapid research needs appraisal methodology to identify evidence gaps to inform clinical research priorities in response to outbreaks—results from the Lassa fever pilot. *BMC Med*. 2019;17(1):107. <https://doi.org/10.1186/s12916-019-1338-1>.
34. Kelly SE, Moher D, Clifford TJ. Quality of conduct and reporting in rapid reviews: an exploration of compliance with PRISMA and AMSTAR guidelines. *Syst Rev*. 2016;5:79. <https://doi.org/10.1186/s13643-016-0258-9>.
35. Modjarrad K, Moorthy VS, Millett P, Gsell P-S, Roth C, Kiely M-P. Developing global norms for sharing data and results during public health emergencies. *PLoS Med*. 2016;13(1):e1001935. <https://doi.org/10.1371/journal.pmed.1001935>.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

