**Reporting methods of observational cohort studies in CMI**

Mical Paul [1], Angela Huttner [1], Julia A. Bielicki [1], Jesús Rodríguez-Baño [1], Andre C. Kalil [1], Mariska M. G. Leeflang, [1] Luigia Scudeller [1], Leonard Leibovici [1]

[1] Clinical Microbiology and Infection editorial office

Corresponding author:

Mical Paul, MD
Infectious Diseases Institute
Rambam Health Care Campus
The Ruth and Bruce Rappaport Faculty of Medicine
Technion – Israel Institute of Technology
Email: paulm@technion.ac.il
Phone: 972-50-2062140

In a previous editorial note, we presented the “do’s” and “don’ts” of observational studies, stressing the importance of transparent reporting of study methodology and correct terminology to avoid inference of effects and impact. [1] We still receive many observational studies that are poorly presented. As for any study design, good quality reporting significantly increases chances of publication.

The STROBE checklist and statement address the reporting requirements of observational studies. [2] We expand on the STROBE statement’s methods section to explain its items with reference to observational cohort studies in infectious diseases and to provide recommendations on the preferred order of item presentation in the methods section in CMI. The required methods components are shown in the table and explained in text.

**Study design**: The STROBE statement advises that authors "refrain from simply calling a study “prospective” or “retrospective"… and recommend that, whenever authors use these words, they define what they mean". [2] A retrospective study could have been prospectively defined, before start of data collection and before observation of the outcome, significantly strengthening it. A "prospective" study can collect only data documented in patients' charts (not very different from a well-planned retrospective study) or could include a measurement performed specifically for the purpose of study, adding to what would have been available from patients' charts alone. We sometime receive "prospective studies" defined as such because they are based on prospective surveillance systems designed for purposes other than the study question. These should be defined as retrospective cohort studies assessing prospectively collected data, if using these terms. Alternatively, declare transparently that the study question was formulated only after the data had been collected. Thus, a study assessing the association between pre-operative oral antibiotics and prosthetic joint infections appropriately reported that "This retrospective study was performed in… and that Patients, who had undergone an elective primary hip or knee replacement between September 2002 and December 2013 were identified from the local prospective joint replacement database." [3] There is a prospective registry of prosthetic joint replacement, but the analysis of oral antibiotics and PJI was retrospective.

We expect to be told whether the protocol and the research question were framed before or after the creation of data; whether the identification of patients was done in real time or in retrospect; whether data were collected prospectively, at pre-set times defined in the protocol; or in real time from files, patients and health care personnel; or in retrospect from files. Registering the study before outcomes have been determined and before start of data collection would lend support to the study and strengthen the confidence that outcome definitions or other data were not manipulated to show significance. [4] A study showed an association between an antibiotic stewardship intervention and lower antibiotic in leukemia unit. [5] Registering it before start of the intervention, would have ensured us that the study was not conducted because a reduction in antibiotic usage was observed.

**Setting**: Describe where the study was conducted and between what years. Further details on the setting should be provided as relevant in the context of the study. For example, a study examining the association between imipenem therapeutic drug monitoring (TDM) and clinical outcomes described that "Currently beta-lactam TDM is employed mainly by intensive care, infectious disease, and haematology-oncology specialists. It is performed on working days and is unavailable at night and on weekends; turnaround time for samples

drawn on workdays is roughly 7 hours…". [6] A study examining whether contact isolation in single-patient rooms is associated with less transmission of fluroquinolone- and cephalosporin-resistant *Escherichia coli* described the local hand-hygiene practice in the study setting. [7] In many studies, local epidemiology and resistance prevalence are important to understand the study and its external implications. Thus, in this same study on single-patient rooms, the baseline prevalence of fluroquinolone and cephalosporin-resistant *E. coli* is important, since the study results might not be applicable to settings with very different extended spectrum beta-lactamase (ESBL) endemicity levels.

**Ethics**: An ethics statement is mandatory for all observational cohort studies. Even if the study reports the results of standard surveillance, infection-control interventions, or changes in clinical practice, reporting should be approved by an ethics committee. If local regulations allow an observational study to be conducted without patient consent, or without the need for Ethics Committee approval, please say so explicitly.

**Participants**: Study eligibility criteria should be clearly presented. To check whether your definitions are clear, examine (or preferably let a colleague tell you) whether they are replicable, whether applying the definitions to the next patient will determine eligibility. Describe how potentially eligible patients were detected and how inclusion/exclusion criteria were applied. Participant numbers do not belong in the methods section, but should be described in the results section (e.g. in a flowchart), following the methods of patient detection and identification.

**Exposure**: An exposure variable is the equivalent of an intervention in a randomized controlled trial: e.g. treatment (of *C. difficile*) with fidaxomixin or vancomycin, [8] combination therapy with ciprofloxacin (for *Escherichia coli* meningitis in infants), [9] treatment with colistin vs. colistin-tigecycline (of *A. baumannii* bacteremia), an antibiotic stewardship program [5], appropriate empirical antibiotic treatment. [10] It might also relate to a patient characteristic that is observed: e.g. *P. aeruginosa* colonization (in patients with chronic obstructive pulmonary disease), [11] carbapenemase-producing Enterobacteriaceae colonization (among patients undergoing liver transplantation) [12]. It is the key risk factor on which the study focuses. Descriptive studies that do not perform a risk factor analysis do not have an exposure variable. Studies assessing risk factors may not have a single, pre-defined, exposure variable: e.g. a study that examined predictors for mortality among patients with drug-susceptible tuberculosis in the Netherlands, did not define key risk factor/s as an exposure. [13]

If examining an exposure, the exposure must be well-defined, well enough to be replicable and applicable. For example, observational studies comparing monotherapy to combination therapy (exposure) deal inherently with the problem that treatments were not standardized. [14] A good description of the exposure should convince us that the patients in the combination group did indeed receive the specified combination therapy, while the others did not. The definition should be clear enough to be applied to further patients. The study assessing combination therapy with ciprofloxacin for neonates with *E. coli* meningitis described that "Concerning ciprofloxacin treatment, we recorded the doses used, duration of treatment and delay between diagnosis and adjunct ciprofloxacin therapy." [9] This does not allow us to replicate the intervention. For full clarity, the definition should have addressed the start time relative to meningitis diagnosis, dosing and minimal duration of the exposure, as it would have had to in an interventional study where the treatment protocol

has to be clearly defined. The study assessing the management of *A. baumannii* bacteraemia reported that "The exposure variable was targeted treatment with monotherapy (colistin) versus combined therapy (colistin plus tigecycline). Only patients who began targeted therapy with colistin in the first 3 days following blood cultures and did not receive any other drug with potential activity against *A. baumannii* were included. The inclusion criterion for patients in the combination therapy group was the administration of tigecycline for >50% of the total treatment time." [15] The doses of the drugs were defined. This is a clear description of how patients were assigned to combination therapy.

**Outcome(s)**: The primary and secondary outcomes should be explicitly defined, addressing the time point of assessment. If the analysis is based on regression analysis, the primary outcome should concord with the dependent variable of the regression.

**Other study variables**: In traditional epidemiological teaching, the confounder is associated both with the exposure and the outcome, without being on the causal pathway. For example, being Asian or black was identified as a strong risk factor for ESBL colonization on admission screening to a hospital in London. [16] Clearly, this is not the cause for ESBL colonization. Confounders such as recent antibiotic use, travel, living conditions or other factors actually explain this association. An "effect" modifier (or association modifier for observational studies) is a variable that explains variability in the magnitude of the association between the exposure and the outcome. For example, in an observational study assessing treatment duration for *Staphylococcus aureus* bacteremia and 90-day mortality, complicated or uncomplicated bacteremia affected the magnitude of the association. [17] Sometimes it might be difficult to classify a variable as potential confounder or effect modifiers; expert knowledge and careful judgement is necessary, to allow the correct analysis to be applied. Predictors are other variables that predict the outcome, and in many studies other variables are collected to describe the cohort. It is critical to anticipate potential confounders and effect modifiers and plan to collect these, because we or the reviewers will probably ask for them. When reporting, while it is not mandatory to distinguish between different types of variables, it might be advantageous to highlight the variables that were collected specifically considering confounding and effect modification.

**Microbiology**: Microbiological methods for pathogen identification and resistance testing are special to studies in infectious diseases. Address these if relevant.

**Data sources/measurements:** We prefer separation between the study variable definitions and their data sources. Data sources of all study variables described above should be presented. In observational studies, study participants are often drawn from a larger database such as a microbiology database or admission/discharge databases. In this case, describe the original database and how patients were selected from it. The data sources for the exposure variable should reflect the confidence that patients were exposed to the intended intervention or had the exposure characteristic. The exposure variable in a case-control study assessing risk factors for *Clostridioides difficile* infection (CDI) was remote cholecystectomy. The authors report that "manual chart review was necessary to accurately ascertain distant cholecystectomy". [18] Indeed, obtaining the data merely from admission/ discharge diagnoses would probably not have been precise. Outcome data sources might be different from other variables' data sources; methods to collect post-discharge outcome data should be described.

While we suggest reporting of all data sources in one section, refrain from repetition if patient selection methods were previously presented under participants, and if exposure or outcome sources were described under their respective items.

**Sample size/power**: A sample size justification is required. The study comparing fidaxomicin to vancomycin for CDI reported: "Assuming a 35% combined clinical failure/recurrence rate for vancomycin, at least 134 participants were necessary in each arm to find a 20% combined clinical failure/recurrence rate for fidaxomicin…". [8] This is a calculation appropriate for a randomized trial, not for an observational study that requires more than a crude comparison between fidaxomicin vs. vancomycin. Nonetheless, this estimate strengthens the authors' conclusions on the lack of difference between the 213 patients treated with fidaxomixin compared to 639 patients treated with vancomycin after propensity-score-matching. Since sample size calculation for prediction models and other observational studies is not standardized, we will accept observational studies without a formal sample size calculation, but we nonetheless require reporting of how investigators arrived at the final sample size (e.g. 10,411 confirmed Crimean- Congo haemorrhagic fever cases were reported between 2004 and 2017 in Turkey [19]). A justification of why an available, known, sample of patients is sufficient to answer the study question will strengthen the study, especially when the available sample is small. In any case, we will appraise whether the sample is reasonable to answer the study question based on precision of the results (confidence interval spread) and require interpretation of negative results considering the sample size/ power. Note that the number of included patients does not belong in the methods section, unless justifying the number of patients in a known available sample (as in the Crimean- Congo haemorrhagic fever example).

**Statistical methods**: Make sure to define all the analyses presented in the manuscript in the methods section. We expect separation between skewed continuous data (e.g. hospitalization duration) and normally-distributed data. Statistical assumptions and variables planned for the construct of regression and propensity score models should be explicitly described. See also our guidance on reporting of multivariable analyses in CMI [16].

**Subgroup and sensitivity analyses**: Important association modifiers can be addressed through subgroup analyses and known methodological limitations can be addressed through sensitivity analyses. Defining these in the methods is necessary if performing subgroup or sensitivity analyses. Importantly, we would like to know whether these were planned in advance or driven by the results.

While, all the above relates to the final stage of reporting a completed study, it cannot be stressed enough how writing the protocol methods according to reporting guidelines will improve the study methodology and allow high-quality reporting at the end. Addressing carefully each item of the checklist will raise questions best addressed before start of the study. A study planning to examine risk factors can be strengthened by defining prospectively one (or few) risk factor(s) of interest that led to the risk-factor analysis Defining an exposure variable allows hypothesis generation and provides a focus that lacks in studies exploring a dearth of potential risk factors. For example, a study assessing risk factors for carbapenemase-producing enterobacteriaceae (CPE) infections following liver transplantation, hypothesized that colonization by CPE before or after transplantation is a significant risk factor, defining it as the exposure. [12] The definition of an "intervention" type of exposure will determine the number of exposed patients; at the protocol stage the researchers can consider strict criteria for an informative exposure or broader criteria with more exposed patients. When defining the outcomes, authors might want to consult

consensus statements on the relevant outcomes, such as the COMET (http://www.comet-initiative.org/) and others. Sample size calculations can only be performed prospectively, and guidance for sample size/power calculations for prediction models have been proposed. [20, 21] At the protocol stage an item should be added to the methods section, that in the full report is addressed in the discussion section: potential biases. An explicit consideration of any biases specific to the study and any measures taken to prevent/counteract bias would enhance robustness of results and causal inference. A pre-defined statistical analysis plan, analogous to that required for clinical trials, would increase quality of research and facilitate the final reporting. [22]

A methods section organized according to this scheme is easier to follow and ensures complete reporting. An unorganized methods section leads to missing information, repetitions and inconsistencies. Although seemingly a long list of items, writing in line with the STROBE scheme actually allows significant shortening of the text to the concise format necessary for scientific reporting. The items can be presented under subheadings or the text can flow more smoothly, as long as all the items are addressed and preferably in the above order. The STROBE scheme should be applied to the methods section of the abstract, to the methods section of a brief report, and to full papers.

**Checklist: Expanded methods of observational cohort studies**

| | |
|---|---|
| Study design | In addition to specifying the study as an observational cohort study, clarify whether the submitted study was planned before data collection, which parts of the data collection were prospective and which were retrospective. |
| Setting | Describe the study location(s) and setting, including start and end dates. No need to repeat these data in results. Describe the relevant epidemiology, infection control or other management features of the setting, as relevant in the study context. |
| Ethics and registration | Provide a statement addressing the ethical approval of the study, whether informed consent was necessary and registration details if the study was registered. |
| Participants | Define the inclusion and exclusion criteria of the study. |
| Exposure | When examining an "intervention", treatment or specific patient characteristic, define this as the exposure variable and provide a thorough description of this variable such that it could be reproduced. |
| Outcome/ endpoint | Define all outcomes reported in the study, specifying which is the primary outcome, including the time point at which the outcomes were measured. In observational studies performing regression analysis, this is also the dependent variable and can be presented as such. |
| Other variables | Other study variables may include confounders, association modifiers, other predictors and cohort descriptors. Justify dichotomization of continuous variables and describe whether they were planned ahead of analysis. |
| Microbiology | Provide the microbiological methods of the study, if relevant. |
| Data sources/ measurements | Describe the sources of data, including participant identification, outcome data, exposure and study variables. Address the methods to collect missing data or end of follow-up data (such as post-discharge outcomes). Describe how measurements were performed. |
| Sample size/ power | Provide a sample size calculation. If none was performed, describe how you arrived at the final sample size. Unless known before the study, in the context of power justification, do not provide data on the number of included patients in the methods section. |
| Statistical methods | Describe all statistical methods, including those used to control for confounding and time-dependent exposures. Explain how missing data were addressed. |
| Subgroup and sensitivity analyses | Define subgroup and sensitivity analyses, if performed. Describe whether planned per protocol or added *post hoc*. |

REFERENCE LIST


1. Paul M, Leibovici L. Observational studies examining patient management in infectious diseases. Clin Microbiol Infect. 2017;23(3):127-8.
2. Vandenbroucke JP, von Elm E, Altman DG, Gotzsche PC, Mulrow CD, Pocock SJ, et al. Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): explanation and elaboration. Ann Intern Med. 2007;147(8):W163-94.
3. Honkanen M, Jamsen E, Karppelin M, Huttunen R, Syrjanen J. The effect of preoperative oral antibiotic use on the risk of periprosthetic joint infection after primary knee or hip replacement: a retrospective study with a 1-year follow-up. Clin Microbiol Infect. 2019;25(8):1021-5.
4. Loder E, Groves T, Macauley D. Registration of observational studies. BMJ. 2010;340:c950.
5. So M, Mamdani MM, Morris AM, Lau TTY, Broady R, Deotare U, et al. Effect of an antimicrobial stewardship programme on antimicrobial utilisation and costs in patients with leukaemia: a retrospective controlled study. Clin Microbiol Infect. 2018;24(8):882-8.
6. Bricheux A, Lenggenhager L, Hughes S, Karmime A, Lescuyer P, Huttner A. Therapeutic drug monitoring of imipenem and the incidence of toxicity and failure in hospitalized patients: a retrospective cohort study. Clin Microbiol Infect. 2019;25(3):383 e1-e4.
7. Biehl LM, Higgins P, Wille T, Peter K, Hamprecht A, Peter S, et al. Impact of single-room contact precautions on hospital-acquisition and transmission of multidrug-resistant Escherichia coli: a prospective multicentre cohort study in haematological and oncological wards. Clin Microbiol Infect. 2019;25(8):1013-20.
8. Gentry CA, Nguyen PK, Thind S, Kurdgelashvili G, Skrepnek GH, Williams RJ, 2nd. Fidaxomicin versus oral vancomycin for severe Clostridium difficile infection: a retrospective cohort study. Clin Microbiol Infect. 2019;25(8):987-93.
9. Tauzin M, Ouldali N, Levy C, Bechet S, Cohen R, Caeymaex L. Combination therapy with ciprofloxacin and third-generation cephalosporin versus third-generation cephalosporin monotherapy in Escherichia coli meningitis in infants: a multicentre propensity score-matched observational study. Clin Microbiol Infect. 2019;25(8):1006-12.
10. Wiggers JB, Sehgal P, Pinto R, MacFadden D, Daneman N. The association of adequate empirical treatment and time to recovery from bacteraemic urinary tract infections: a retrospective cohort study. Clin Microbiol Infect. 2019;25(10):1253-8.
11. Eklof J, Sorensen R, Ingebrigtsen TS, Sivapalan P, Achir I, Boel JB, et al. Pseudomonas aeruginosa and risk of death and exacerbations in patients with chronic obstructive pulmonary disease: an observational cohort study of 22 053 patients. Clin Microbiol Infect. 2019.
12. Giannella M, Bartoletti M, Campoli C, Rinaldi M, Coladonato S, Pascale R, et al. The impact of carbapenemase-producing Enterobacteriaceae colonization on infection risk after liver transplantation: a prospective observational cohort study. Clin Microbiol Infect. 2019;25(12):1525-31.
13. Pradipta IS, Van't Boveneind-Vrubleuskaya N, Akkerman OW, Alffenaar JWC, Hak E. Predictors for treatment outcomes among patients with drug-susceptible tuberculosis in the Netherlands: a retrospective cohort study. Clin Microbiol Infect. 2019;25(6):761 e1- e7.
14. Paul M, Carmeli Y, Durante-Mangoni E, Mouton JW, Tacconelli E, Theuretzbacher U, et al. Combination therapy for carbapenem-resistant Gram-negative bacteria. J Antimicrob Chemother. 2014;69(9):2305-9.
15. Amat T, Gutierrez-Pizarraya A, Machuca I, Gracia-Ahufinger I, Perez-Nadales E, Torre-Gimenez A, et al. The combined use of tigecycline with high-dose colistin might not be

associated with higher survival in critically ill patients with bacteraemia due to carbapenem-resistant Acinetobacter baumannii. Clin Microbiol Infect. 2018;24(6):630-4.
16. Otter JA, Natale A, Batra R, Tosas Auguet O, Dyakova E, Goldenberg SD, et al. Individual- and community-level risk factors for ESBL Enterobacteriaceae colonization identified by universal admission screening in London. Clin Microbiol Infect. 2019;25(10):1259-65.
17. Abbas M, Rossel A, de Kraker MEA, von Dach E, Marti C, Emonet S, et al. Association between treatment duration and mortality or relapse in adult patients with Staphylococcus aureus bacteraemia: a retrospective cohort study. Clin Microbiol Infect. 2019.
18. Wang Y, Li J, Zachariah P, Abrams J, Freedberg DE. Relationship between remote cholecystectomy and incident Clostridioides difficile infection. Clin Microbiol Infect. 2019;25(8):994-9.
19. Ak C, Ergonul O, Gonen M. A prospective prediction tool for understanding Crimean-Congo haemorrhagic fever dynamics in Turkey. Clin Microbiol Infect. 2020;26(1):123 e1- e7.
20. Riley RD, Snell KI, Ensor J, Burke DL, Harrell FE, Jr., Moons KG, et al. Minimum sample size for developing a multivariable prediction model: PART II - binary and time-to-event outcomes. Stat Med. 2019;38(7):1276-96.
21. van Smeden M, Moons KG, de Groot JA, Collins GS, Altman DG, Eijkemans MJ, et al. Sample size for binary logistic prediction models: Beyond events per variable criteria. Stat Methods Med Res. 2019;28(8):2455-74.
22. Thomas L, Peterson ED. The value of statistical analysis plans in observational research: defining high-quality research from the start. JAMA. 2012;308(8):773-4.