Journal of Clinical Epidemiology 144 (2022) 194–197

Journal of Clinical Epidemiology

LETTER TO THE EDITOR

A few things to consider when deciding whether or not to conduct underpowered research

Key findings:

- The statistical power of a meta-analysis of individually underpowered observational studies is determined. Additional factors to consider before undertaking an underpowered observational study are reviewed.

What this adds to what is known:

- A recent commentary (Hernán 2022) argues in favor of carrying out an underpowered observational study, if the study data are extracted from an existing database. Further nuance and caveats are presented around this recommendation.

What is the implication, what should change now:

- There are multiple factors to consider before embarking on an underpowered observational study, even if the study is based on an existing database.

Hernán [1], using a hypothetical example, argues that policies that prevent researchers from conducting underpowered observational studies using existing databases are misguided explaining that "[w]hen a causal question is important, it is preferable to have multiple studies with imprecise estimates than having no study at all." While we do not disagree with the sentiment expressed, caution is warranted. Small observational studies are a major cause of distrust in science, mainly because their results are often selectively reported. The hypothetical example used to justify Hernán's [1] position is too simplistic and overly optimistic. In this short response, we reconsider Hernán's [1] hypothetical example and offer a list of other factors–beyond simply the importance of the question–that are relevant when deciding whether or not to pursue underpowered research.

In Hernán's [1] hypothetical example, researchers wish to use a database of 6 million people to estimate the causal effect of a new COVID-19 vaccine on severe thrombosis. With these hypothetical data, an estimated risk ratio of 5.0 was obtained with a 95% confidence interval of [0.58, 43]. This result is compatible with observing 5 cases of thrombosis out of 3 million vaccinated individuals and 1 case amongst 3 million unvaccinated individuals.

Assuming a true baseline risk amongst the unvaccinated of 1 in 3 million and a true risk ratio of 5, a study with 3 million vaccinated individuals and 3 million unvaccinated individuals would have statistical power of about 37%. This is less than ideal. But one might still be able to "convince your colleagues" to conduct such a study by pointing out that, if one assumes a true risk ratio of 10 (a large but perhaps not implausible number to consider given the context of the question), the study will have 80% power.

What of an even smaller sample size? Suppose that the hypothetical study had only 1 million vaccinated and 1 million unvaccinated individuals. Then, assuming once again a true baseline risk amongst the unvaccinated of 1 in 3 million and a true risk ratio of 5, the hypothetical study would have power of about 8%. With a true risk ratio of 10, the study has power of about 32%. In order for the study to have a power of 80%, one must assume a true risk ratio of more than 20. Thus, under a wide range of plausible assumptions, all can agree that the study will be severely underpowered. Should one proceed with such a study? According to Hernán [1], it seems the answer is yes. After all, the importance of the question has not changed and if other similar studies become available, a meta-analysis (MA) could be done and may potentially provide a precise pooled effect estimate. This is not an unreasonable argument. However, it is worth considering just how many other similar studies would be needed for an adequately powered MA, and how similar these studies would need to be. We conducted a simple simulation study to answer this question.

Assuming that the true risk ratio in each study population is 5, we find that about 11 studies of equal sample size (each study having 2 million participants) would be required for a standard fixed-effect MA to have the often sought-after 80% power. This ignores the possibility of any across-study heterogeneity (e.g., due to difference in study design). Instead, if we assume that across all study populations, the true risk ratios range from 4 to 6 (evenly spaced), we find that about 14 studies of equal sample size would be required for a standard random-effects MA to have 80% power. However, given the fact that the literature is com-

meta analysis, statistical power; observational studies, evidence synthesis; healthcare databases, publication bias.

Conflict of interest: None.

https://doi.org/10.1016/j.jclinepi.2021.11.038

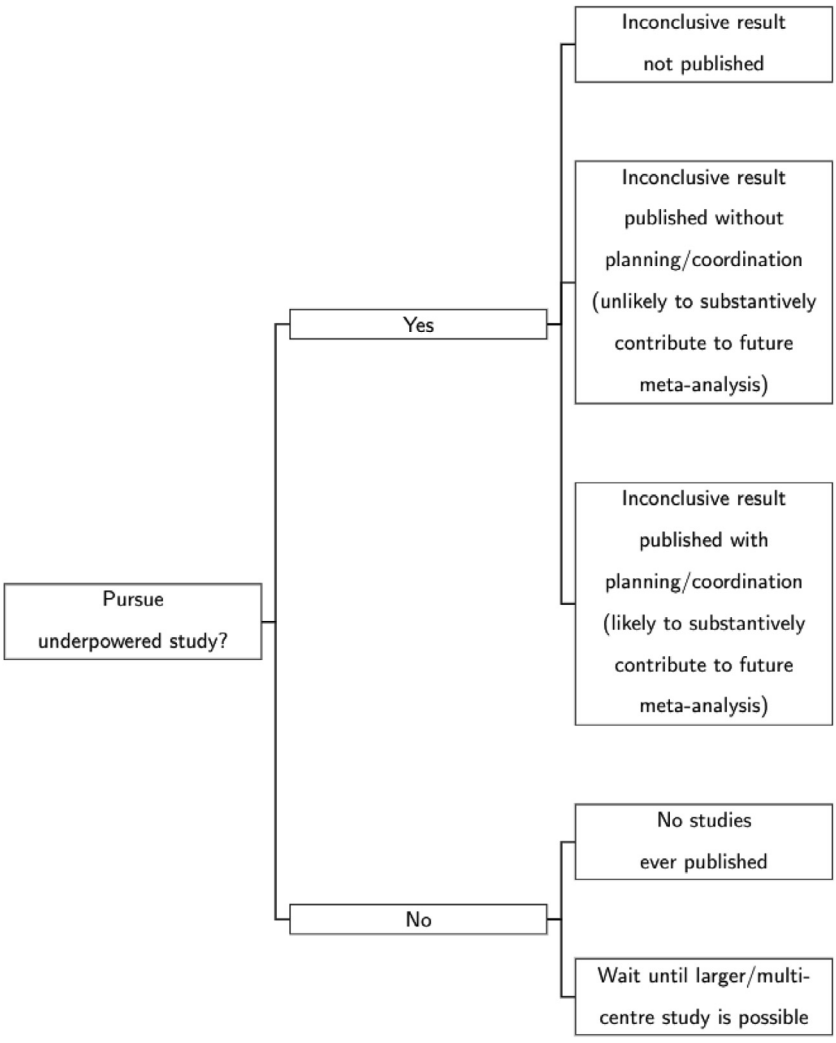

Fig. 1. Flowchart of likely outcomes from deciding whether or not to conduct underpowered research.

promised by publication bias, a standard random-effects MA may not be adequate.

The trim and fill robust random-effect MA (the most popular approach to correct for potential publication bias [2]) will have less power: our hypothetical analysis with 14 studies will have only about 65% power. This presumes all 14 studies are actually published, that is, the robust analysis is protecting against publication bias which actually isn't manifested.

As must be, power would be lower if publication bias actually is manifested [3]. Starting with the 14 studies as above, say that those yielding P-value >0.05 have a 50% chance of being published, while the remainder have a 100% chance of being published. The trim-and-fill random-effects MA of the published studies will have about 50% power. Indeed, under these conditions one would need about 30 studies (a total of 60 million participants) for the MA to have 80% power. And publication bias is of course, only one reason why a study might not be included in a MA [4,5]. With all this in mind, the advantages of pursuing the underpowered study seem much less obvious.

The alternative, according to Hernán [1], is simply an "unanswered question." But this is not necessarily the case. One must think about the potential outcomes of pursuing (or failing to pursue) underpowered research; see Fig. 1. If one is not able or willing to conduct a small underpowered study, one might instead choose to invest more time and resources to design a prospective study with increased sample size [6]. Or one might attempt to coordinate with other interested researchers to conduct a multicenter study [7].

We would argue that the primary consideration when deciding whether or not to conduct an underpowered observational study should be the likelihood that the study will eventually contribute in a substantive way to an adequately powered MA [8]. Therefore, we suggest that when deciding whether or not to pursue an underpowered observational study (be it based on an existing database on not),

one should consider a few things in addition to simply the importance of the question, see Box 1. These mostly align with the general recommendation made for randomised trials by Schulz and Grimes [9] which calls for "[s]ome shift of emphasis from a fixation on sample size to a focus on methodological quality." Finally, we note that while our simple simulation study was able to shed some light on Hernán [1] hypothetical example, a more elaborate simulation study, similar to IntHout et al. [10] consideration of clinical trials, would no doubt be worthwhile.

Box 1 A few things to consider when deciding whether or not to conduct underpowered observational research:

-A commitment to pre-specify all of the intended analyses and publish a study protocol [11,12].

-A commitment to publish, or at the very least make public, the results of the (pre-specified) analyses regardless of the outcome obtained. (We note that achieving/enforcing such commitments can be difficult in practice, even for preregistered clinical trials; see DeVito et al. [13].)

-The availability of resources (i.e., some budget/personnel) to facilitate future reviews (e.g., by offering help with a re-analysis of the data adjusting for additional/alternative covariates so as to make published estimates more comparable).

-The ability to provide the individual participant data (IPD) for future research so as to facilitate a future IPD-MA. IPD-MAs for allow more powerful and consistent analyses relative to MAs based on aggregate data; see Smith et al. [14].

-Assurances that the cost/risk for participants has been minimized. In some settings (e.g., use of registry or administrative data that was not collected for research purposes), risks may indeed be very limited (but not zero, e.g., privacy breach risks). Other observational studies, however, may have participant burdens that more resemble those of RCTs; see Norris et al. [15].

-A reasonable likelihood that other comparable studies will occur. If the intended study is the first in its class, then one should consider the ability of the study to spur the research community to conduct similar studies in other populations, via cogent arguments for the importance of the question.

-The ability to reduce heterogeneity across studies (e.g., through coordination and design choices that align well with past studies and/or would be relatively simple for future study investigators to adopt).

-Whether the specific population is "hard to reach" (i.e., is at high risk of being excluded from other studies). If so, then the potential contribution to a future MA is arguably greater; see Crosby et al. [16].

Code

Note that the code used for the simulation study is available at https://github.com/harlanhappydog/underpowered/.

Harlan Campbell*
Department of Statistics, University of British Columbia, Vancouver, British Columbia, Canada

Valentijn M.T. de Jong
Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht University, Utrecht, the Netherlands

Thomas P.A. Debray
Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht University, Utrecht, the Netherlands
Cochrane Netherlands, Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht University, Utrecht, the Netherlands

Paul Gustafson
Department of Statistics, University of British Columbia, Vancouver, British Columbia, Canada

*Corresponding author. Harlan Campbell, phone: 1 604 822 0570 fax: 1 604 822 6960
E-mail address: campbell@stat.ubc.ca (H. Campbell)

References

[1] Hernán MA. Causal analyses of existing databases: no power calculations required. J Clin Epidemiol 2021.
[2] Peters JL, Sutton AJ, Jones DR, Abrams KR, Rushton L. Performance of the trim and fill method in the presence of publication bias and between-study heterogeneity. Stat Med 2007;26(25):4544–62.
[3] Sutton AJ, Duval SJ, Tweedie R, Abrams KR, Jones DR. Empirical assessment of effect of publication bias on meta-analyses. The BMJ 2000;320(7249):1574–7.
[4] Song F, Parekh S, Hooper L, Loke YK, Ryder J, Sutton AJ, et al. Dissemination and publication of research findings: an updated review of related biases. Health Technol Assess (Rockv) 2010;14(8):1–220.
[5] Sterne JA, Gavaghan D, Egger M. Publication and related bias in meta-analysis: power of statistical tests and prevalence in the literature. J Clin Epidemiol 2000;53(11):1119–29.
[6] Williamson EJ, Walker AJ, Bhaskaran K, Bacon S, Bates C, Morton CE, et al. Factors associated with COVID-19-related death using OpenSAFELY. Nature 2020;584(7821):430–6.
[7] Sprague S, Matta JM, Bhandari M. on Behalf of the Anterior Total Hip Arthroplasty Collaborative (ATHAC) Investigators. Multicenter collaboration in observational research: improving generalizability and efficiency. J Bone Jt Surg 2009;91(Supplement 3):80–6.

[8] Jackson D, Turner R. Power analysis for random-effects meta-analysis. Res Synth Methods 2017;8(3):290–302.
[9] Schulz KF, Grimes DA. Sample size calculations in randomised trials: mandatory and mystical. Lancet North Am Ed 2005;365(9467):1348–53.
[10] IntHout J, Ioannidis JP, Borm GF. Obtaining evidence by a single well-powered trial or several modestly powered trials. Stat Methods Med Res 2016;25(2):538–52.
[11] Dal-Ré R, Ioannidis JP, Bracken MB, Buffler PA, Chan A-W, Franco EL, et al. Making prospective registration of observational research a reality. Sci Transl Med 2014;6(224):224cm1.
[12] Loder E, Groves T, MacAuley D. Registration of observational studies. The BMJ 2010;340(c950).
[13] DeVito NJ, Bacon S, Goldacre B. Compliance with legal requirement to report clinical trial results on clinicaltrials.gov: a cohort study. Lancet North Am Ed 2020;395(10221):361–9.
[14] Smith CT, Marcucci M, Nolan SJ, Iorio A, Sudell M, Riley R, et al. Individual participant data meta-analyses compared with meta–analyses based on aggregate data. Cochrane Database Syst Rev 2016(9).
[15] Norris A, Jackson A, Khoshnood K. Exploring the ethics of observational research: the case of an HIV study in Tanzania. AJOB Prim Res 2012;3(4):30–9.
[16] Crosby RA, Salazar LF, DiClemente RJ, Lang DL. Balancing rigor against the inherent limitations of investigating hard-to-reach populations. Health Educ Res 2010;25(1):1–5.