COMMENTARY

Causal analyses of existing databases: no power calculations required

Miguel A. Hernán [a,b,*]

[a] *CAUSALab, Harvard T.H. Chan School of Public health, Boston, MA 02115, USA*
[b] *Departments of Epidemiology and Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA 02115, USA*

Abstract

Observational databases are often used to study causal questions. Before being granted access to data or funding, researchers may need to prove that "the statistical power of their analysis will be high." Analyses expected to have low power, and hence result in imprecise estimates, will not be approved. This restrictive attitude towards observational analyses is misguided.

A key misunderstanding is the belief that the goal of a causal analysis is to "detect" an effect. Causal effects are not binary signals that are either detected or undetected; causal effects are numerical quantities that need to be estimated. Because the goal is to quantify the effect as unbiasedly and precisely as possible, the solution to observational analyses with imprecise effect estimates is not avoiding observational analyses with imprecise estimates, but rather encouraging the conduct of many observational analyses. It is preferable to have multiple studies with imprecise estimates than having no study at all. After several studies become available, we will meta-analyze them and provide a more precise pooled effect estimate. Therefore, the justification to withhold an observational analysis of preexisting data cannot be that our estimates will be imprecise. Ethical arguments for power calculations before conducting a randomized trial which place individuals at risk are not transferable to observational analyses of existing databases.

If a causal question is important, analyze your data, publish your estimates, encourage others to do the same, and then meta-analyze. The alternative is an unanswered question. © 2021 Elsevier Inc. All rights reserved.

Keywords: Causal analysis; Observational analysis; Meta-analysis; Causal inference; Observational studies; Statistical power; Sample size; Statistical significance

1. Introduction

Observational databases are often used to answer causal questions. For example, electronic health records have long been a data source for investigations about the effects of medical interventions such as treatments [1] and vaccines [2]. Before being granted access to data or funding, researchers may need to prove that the analysis will be sufficiently informative. Typically, this requirement is articulated as "show that the statistical power of your analysis will be high." Analyses expected to have low power, and hence generate imprecise estimates, will not receive the green light. This restrictive attitude towards observational analyses is misguided.

2. Hypothetical example

To see why, suppose a new COVID-19 vaccine is suspected of causing severe thrombosis in about 1 per million individuals. The suspicion, which is causing great social alarm, arose from reports of unusual thrombotic events among young vaccinees. Preapproval trials of the vaccine were too small to detect rare adverse events, so large well-designed observational studies are needed.

Your colleagues have access to the health records of all 6 million people in a health plan. You offer to help them with an analysis to estimate the effect of the vaccine on thrombosis risk. Your colleagues inform you that such study will not be conducted. "Why?," you ask. "Because you cannot detect such a small increase in risk in a population of 6 million. The study would have very low power." *Detect*? *Power*? Not the right words in this setting. The goal of this analysis is not to "detect" a causal effect, but to quantify it as unbiasedly and precisely as possible. Causal effects are not binary signals that are either detected or undetected; causal effects are numerical quantities that need to be estimated [3,4].

Conflict of interest statement: The author declares that there is no conflict of interest.

* Corresponding author. Tel.: +1 617 432 0101
  E-mail address: miguel_hernan@post.harvard.edu.

Suppose you finally convince your colleagues to conduct the observational analysis. The estimated effect of vaccine vs. no vaccine on thrombosis is, on the risk ratio scale, 5.0 with a 95% confidence or compatibility interval from 0.58 to 43. (For simplicity, we assume that the analysis is unbiased, e.g., there is no confounding, no selection bias, and no measurement error.) Thus, our effect estimate is extremely imprecise. The interpretation of these results would be: "Anything between a 42% lower risk and a 43-fold higher risk of thrombosis after vaccination is very compatible with our data."

Have we detected an effect? No. Have we *not* detected an effect? No [5]. What we have done is to identify a range of values of the effect that are very compatible with the data. Unfortunately, that range is very wide, which makes the estimate only weakly informative (all we can say is that large harmful effects are much more compatible with the data than preventive effects). One less direct way to state this conclusion is saying that our study is (grossly) underpowered. Given the rarity of the event under study, the 95% confidence interval of almost every observational study will be quite wide. The solution cannot be refusing to carry out each of those observational analyses with imprecise estimates, but rather to encourage the conduct of many such observational analyses. After several studies become available, we can meta-analyze them and provide a more precise pooled effect estimate [6]. In our example, if three additional research groups had conducted analyses in their respective databases of about 6 million people and had obtained results similar to ours, the combination of the estimates from the four studies would have resulted in a pooled estimate with a 95% confidence interval from 1.7 to 14.7. The interpretation would be: "Anything between a 70% increased risk and a 15-fold higher risk of thrombosis after vaccination is very compatible with our data." In other words, the pooled effect estimate would be precise enough to guide vaccination policy, especially if alternative vaccines are available.

3. Observational databases and randomized trials

When a causal question is important, it is preferable to have multiple studies with imprecise estimates than having no study at all. Refusing to conduct an observational study because "it'll be underpowered" prevents the accumulation of scientific evidence. The justification to withhold an observational analysis of preexisting data cannot be that our estimates will be imprecise, but rather that the question is not important enough, that we are devoting our resources to other important questions, or that no additional evidence is expected from other research groups. Since we have no control over whether other research groups will conduct their observational analyses, the best we can do is to conduct our own. If it is likely that several research groups will estimate effects, we would ideally coordinate our efforts so that the results can be easily pooled. Like for randomized

trials [7], we can agree on a master protocol that defines the event of interest (e.g., thrombosis), the eligibility criteria (e.g., age 20–65, no prior comorbidities), the choice of measures (e.g., 28-day risks), the pooling procedures…

In fact, we can sometimes view our observational studies as provisional evidence while waiting for the results of randomized trials with a coordinated design. Note that those trials, unlike the observational studies, are subject to ethical constraints on human experimentation. Individuals invited to participate in a randomized trial need the guarantee that their (possibly risky) participation will produce usable scientific evidence. Since a trial's investigators cannot ensure that other investigators will do similar trials, the trial must be designed to provide informative results on its own. This requirement is usually articulated as "the trial is adequately powered,", though it'd be better articulated as "the trial will produce estimates with a guaranteed level of precision." Thus, in the absence of coordination across trials, there is a strong ethical argument for power— or, better, precision [8]—calculations before conducting a randomized trial which may place patients at risk. This ethical argument is not transferable to observational analyses of existing databases.

Another pragmatic argument against preanalysis calculations of precision in observational analyses: truly exact calculations (as opposed to textbook exercises) are impossible when adjustments for measurement error, selection bias, and confounding are required, especially in complex settings with time-varying treatments and confounders. In those cases, we would need complex simulation studies to obtain approximate estimates of precision, or skip them when designing realistic simulations would take longer than doing the observational analyses.

4. Conclusion

In summary, for an important causal question, analyze your data no matter how imprecise you expect your estimate to be, publish your estimate, encourage others to do the same, and then meta-analyze. The alternative is an unanswered question.

Acknowledgments

References

[1] Jick H, Walker AM, Watkins RN, D'Ewart DC, Hunter JR, Danford A, et al. Oral contraceptives and breast cancer. Am J Epidemiol 1980;112(5):577–85.

[2] Walker AM, Jick H, Perera DR, Thompson RS, Knauss TA. Diphtheria-tetanus-pertussis immunization and sudden infant death syndrome. Am J Public Health 1987;77(8):945–51.

[3] Greenland S. Invited commentary: the need for cognitive science in methodology. Am J Epidemiol 2017;186(6):639–45.

[4] Hernán MA. Does water kill? A call for less casual causal inferences. Ann Epidemiol 2016;26:674–80.

[5] Lash TL. The harm done to reproducibility by the culture of null hypothesis significance testing. Am J Epidemiol 2017;186(6):627–35.

[6] Deeks JJ, Higgins JPT, Altman DG (editors). Chapter 10: analysing data and undertaking meta-analyses. In: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA (editors). Cochrane handbook for systematic reviews of interventions version 6.2 (updated February 2021). Cochrane, 2021. Available from www.training.cochrane.org/handbook. (accessed September 17, 2021).

[7] Woodcock J, LaVange L. Master protocols to study multiple therapies, multiple diseases, or both. N Engl J Med 2017;377:62–70.

[8] Rothman KJ, Greenland S. Planning study size based on precision rather than power. Epidemiology 2018;29(5):599–603 erratum in *Epidemiology* 2019; 30(1):e1.