COMMENT

Let the question determine the methods: descriptive epidemiology done right

Sara Conroy [1] and Eleanor J. Murray[2]

What does it mean to control for confounding, and when do we actually need to do it? To answer this, we need a well-defined research question, driven by the goal of the study. For descriptive goals, we explain that confounding adjustment is often not just unnecessary but can be harmful.

MAIN

In 2017, Pedersen et al. published a study that explored the potential statistical association of frozen-shoulder and cancer diagnoses.[1] Recently, discussion of this paper on Twitter criticised the authors for not controlling for confounding.[2–4] But, what does it mean to control for confounding, and when do we actually need to do it? Confounding is typically defined as a variable that is related to both the main variable of interest and the outcome, but is not on the causal pathway. A directed acyclic graph is a useful tool to help determine if a variable is a confounder.[5] The decision to control for a confounder depends on the specific scientific question, and typically occurs when the focus of the research question is to investigate a causal relationship between the main variable of interest and the outcome. However, not all research requires a causal question.[6] Research studies that focus on describing a population of interest are essential building blocks for both causal and predictive frameworks, and do not typically require control for additional variables. To understand the purpose of a study (i.e., descriptive, causal or predictive), it is vital that the goals of the research be clearly explained.[7] The Pedersen article is a great example of a study with a research question that is not causal and therefore does *not* need to control for confounding. Here, we discuss why the authors were correct not to control for confounding, and how the research question should guide the methods, especially when conducting descriptive epidemiology studies.

What was the research question?

Pedersen et al.[1] were motivated by the observation that people who have cancer often develop frozen shoulder, but also that certain types of cancer may be misdiagnosed as frozen shoulder. With this ambiguity, the authors designed a study to address the question: "Is frozen shoulder a warning sign that can identify a group of people who might be at high risk for cancer?" This is not a causal question as the authors are not suggesting that frozen shoulder causes cancer, or that cancer diagnosis causes frozen shoulder.

Two main goals helped answer the research question. First, the authors wanted to describe the incidence of cancer among people with frozen shoulder. Second, they wanted to compare and contrast the incidence of frozen shoulder in the data with the general population (i.e., people who may one day get a cancer diagnosis).

What is the ultimate goal?

Ultimately, the researchers hope to improve early diagnosis of cancer. Cancer screening is the action/decision that the authors were trying to inform. After describing the population of interest and comparing with other relevant populations, if a difference is observed, there are multiple directions the next study could go. One could build a predictive model of cancer using frozen shoulder information or test a cancer-screening programme that focuses on shoulder patients compared with usual screening patients. These studies could help determine if a larger proportion of cancer patients could be found earlier. The main insight of those next steps is that any hypothetical trial building on the current study would not intervene on shoulder problems but instead on cancer screening.

But, what about confounding?

Discussion of this paper on Twitter has faulted the authors for not controlling for confounding.[3,4] However, we argue here that the authors were in fact correct to not control for confounding because confounding is precisely what they hoped to identify. The authors are attempting to find a statistical association between frozen shoulder and cancer. Any statistical relationship is expected to be influenced by other variables (e.g., preclinical cancer). Adjusting for confounding variables in this analysis runs the risk of getting the wrong answer as it might accidentally open a collider path and create an association that is not normally there.[8] An open collider path that could make it seem like frozen shoulder is a good early warning tool for cancer diagnosis when it actually is not.

The data set covered the entire Danish population, and there was no evidence to screen for cancer based on frozen shoulder. Generalisability would depend on the underlying cancer and frozen-shoulder incidence in other populations (outside of Denmark). For example, if the incidence of cancer was the same, but frozen-shoulder cases were higher in Canada, the association of

---

[1]Center for Biostatistics, Department of Biomedical Informatics, The Ohio State University, Columbus, OH, USA and [2]Department of Epidemiology, Boston University School of Public Health, Boston, MA, USA
Correspondence: Sara Conroy (Sara.Conroy@osumc.edu)

frozen shoulder and cancer may be smaller. If the incidence of cancer was the same, but frozen-shoulder cases were lower in Canada, the association of frozen shoulder and cancer might be higher compared with this population in this study. In a similar fashion, if both the common cause of frozen shoulder and cancer were higher in Canada, the association between frozen shoulder and cancer would probably be larger, and screening may be warranted.

The right methods for the right question

The authors could have also asked: "Is frozen shoulder predictive of a high risk of cancer above and beyond other known cancer risk factors?" This question would imply the use of frozen shoulder as a proxy for unknown causes of cancer, and in that case, control for known causes of cancer, or known predictors of cancer could be appropriate. Those other known predictors need not also be causes of frozen shoulder, and are thus not necessarily confounders.[8]

Finally, in keeping with current recommendations from the American Statistical Association and others,[9] the authors of this paper do not rely solely on *P* values for interpreting their findings. Instead, they consider the magnitude of the association, and conclude that, although statistically significant, the detected association is not big enough to warrant stratified screening in Denmark.[10]

In summary, the paper by Pedersen et al.[1] is an excellent example of descriptive epidemiology done right. We commend the authors for their clarity in explaining that they are not estimating causal effects, in the clear lack of causal language, and in the inclusion of specific discussion of whether the results suggest that stratified screening programmes could be a useful next step. We recommend this paper to those who teach descriptive epidemiology for use with their students.

ACKNOWLEDGEMENTS

AUTHOR CONTRIBUTIONS

E.M. had the initial idea for the commentary. S.C. wrote the first draft with notes from E.M. S.C. and E.M. revised and approved the final version.

ADDITIONAL INFORMATION

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

REFERENCES

1. Pedersen, A. B., Horváth-Puhó, E., Ehrenstein, V., Rørth, M. & Sørensen, H. T. Frozen shoulder and risk of cancer: a population-based cohort study. *Br. J. Cancer* **117**, 144–147 (2017).
2. function2fitnes. Massive population-based study. First nationwide cohort study to examine cancer risk in frozen shoulder patients from British Journal of Cancer. What is the risk of a cancer diagnosis after an incident diagnosis of frozen shoulder? Open Access' https://t.co/xEh2yy2irN; https://t.co/YZ3soly2GE. Available from https://twitter.com/function2fitnes/status/1214321809178464256 (2020).
3. giovanni_ef. There is no way in the world this study can even start answering this question. There is absolutely no information on how confounding was handled—which wasn't even mentioned as a study limitation. This is pure epidemiological rubbish. Available from https://twitter.com/giovanni_ef/status/1214332942417158144 (2020).
4. LiniearProbe. The risk of this study is clinicians & patients worrying that FS might be a forerunner of cancer when: A) this cannot be supported by this study B) may lead to unnecessary further Ix & referrals Treat FS as FS and use the same clinical skills to be alert to red flags as always. Available from https://twitter.com/LinearProbe/status/1214527944439283712 (2020).
5. Suzuki, E., Shinozaki, T. & Yamamoto, E. Causal diagrams: pitfalls and tips. *J. Epidemiol.* **30**, 153–162 (2020).
6. Lesko, C. R., Keil, A. P. & Edwards, J. K. The epidemiologic toolbox: identifying, honing, and using the right tools for the job. *Am. J. Epidemiol.* https://doi.org/10.1093/aje/kwaa030 (2020).
7. Hernán, M. A., Hsu, J. & Healy, B. A second chance to get causal inference right: a classification of data science tasks. *Chance* **32**, 42–49 (2019).
8. Hernán, M. A. & Robins, J. M. *Causal Inference: What If* (Chapman & Hall/CRC, Boca Raton, 2020).
9. Wasserstein, R. L. & Lazar, N. A. The ASA's statement on p-values: context, process, and purpose. *Am. Stat.* **70**, 129–133 (2016).
10. Greenland, S., Senn, S. J., Rothman, K. J., Carlin, J. B., Poole, C., Goodman, S. N. et al. Statistical tests, P values, confidence intervals, and power: a guide to misinterpretations. *Eur. J. Epidemiol.* **31**, 337–350 (2016).