The impact of epidemiologic methods on findings in studies of causal effects and prediction modelling

Luijken, K.

1

General introduction

Developments in epidemiological practice

Quantitative scientific evidence is a key pillar of current evidence-based medicine to inform medical decision making[1]. To aid understanding of clinical questions through numerical results, quantitative approaches are continuously being developed in the field of clinical epidemiology and medical statistics, hereafter referred to as the field of 'epidemiology'. During most of the 20th century, study designs mainly revolved around operationalizations of the exposure and outcome variables, while statistical approaches focused on models that required no or minimal modelling assumptions[2]. Epidemiological approaches were used to address relatively broad, yet fundamental clinical questions, such as: 'does smoking cause lung cancer?'. As an example, this question was addressed in one of the first observational clinical studies, performed by the Sirs Doll and Hill[3]. It can be computed from their results that the odds of developing lung cancer when smoking was around 14 times the odds of developing lung cancer when not smoking: a substantial effect size.

Since then, the apparatus of epidemiological study designs and data analytical approaches has expanded considerably. The latter half of the twentieth century saw a strong rise of epidemiological evidence as a basis of medical decision making, where the hierarchy of evidence from clinical experience to epidemiological evidence essentially reversed[4,5]. More data sources became available for research purposes, including routinely collected clinical data[6]. Since the increase in data volume was not always paralleled by data quality[7], study designs and methods were increasingly refined to address variation in data collection procedures. Statistical techniques were developed to accommodate the higher complexity of the data, from multivariable models to machine learning, aided by an extraordinary increase in computing power. These refined epidemiological approaches are applied in a wide variety of public health and medical research fields to address specific questions, such as: 'is cotinine level (a biomarker for passive tobacco smoke exposure) associated with an increased 20-year risk of coronary heart disease in men who self-report to be non-smokers?'. A study by Whincup and colleagues found that the hazard of developing coronary heart disease over 20 years among non-smoking men with a cotinine concentration of 2.8 – 14.0 ng/mL was 1.57 times the hazard of developing coronary heart disease among men with a cotinine concentration of ≤ 0.7 ng/mL, conditional on established risk factors for coronary heart disease (with a 95% confidence interval from 1.08 to 2.28)[8]. Compared to the marginal odds ratio of 14 found by Doll and Hill, this association appears to be weak, meaning that methodological and statistical artefacts increasingly run the risk of bearing consequences for interpretation of findings.

The seemingly straightforward methodological principles of early epidemiological research and the current abundancy in methods are contrasted here to outline that the increasingly central position of formalized quantitative methods requires careful consideration of the analytical choices to be made during study conduct.

Analytical tools at the disposal of clinical questions

It is a challenge to keep abreast of all developments in epidemiological methods. To facilitate application of statistical approaches, a large number of tutorial series and coding vignettes break down statistical material into manageable endeavors (for instance, the 'Statistical Notes' series in the *British Medical Journal*[9] or vignettes of the R package epiR[10]). Some modelling procedures are even fully automated, such as for instance a high-dimensional propensity score algorithm for analysis of causal effects of pharmacological treatments in routinely collected health care databases[11,12], or software that automates prognostic modelling in observational databases using the Observational Medical Outcomes Partnership (OMOP) Common Data Model format[13], to the degree that manuscript output can be partly rendered by the statistical package[14]. These tutorials and packages ease the use of statistical tools, but the (default) method that is applied may not be optimal for a particular study[15]. Moreover, the choice of the analytical method directly influences the clinical interpretation of the results of the analysis and thus research findings, yet it is undesirable that technical decisions define the subject of the investigation.

The challenge of identifying a study design and statistical method for data analysis that appropriately helps answering the research question of clinical interest has long been recognized in epidemiological research. At the basis of epidemiological research is a careful definition of the research question and a plan how to conduct the study accordingly (Figure 1). Many epidemiology textbooks start from the premise that study conduct should be informed by a carefully defined research question, e.g., [16,17]. Statistical literature increasingly emphasizes that the study aim – identifying it as descriptive, predictive, or causal – is the basis for choosing an appropriate statistical model and interpreting the findings correctly[18-20]. An addendum to the ICH E9 guideline Statistical Principles for Clinical trials further explicated the link between clinical aims and analytical theory by clarifying what the specification of a study estimand entails[21].

Defining the research question and corresponding estimand provides a link between the study aims and its analysis and endows the numerical result with a clinically meaningful interpretation.

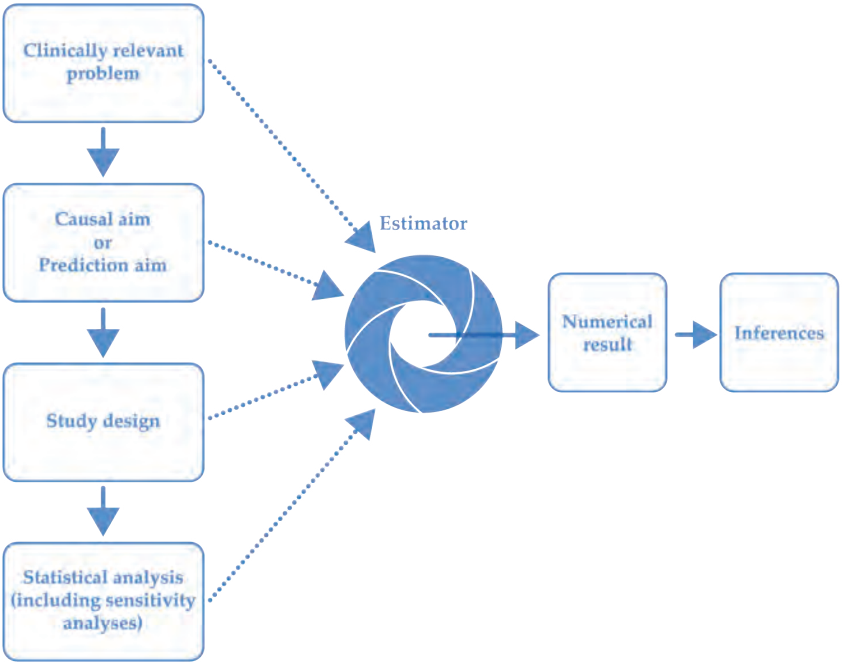


Figure 1. Schematic depiction of stages of research conduct of a quantitative study. The estimator is depicted as a diaphragm, resembling the aperture around the lens through which a research problem is studied in a quantitative analysis. The way the diaphragm is tuned by the clinical context, study aim, study design, and statistical analysis plan determines what can be observed through the lens and thus directly affects inferences that can be made from the estimate. A scheme like this should be accompanied by the remark that "Trying to cash out a full-blown picture of inquiry that purports to represent all contexts of inquiry is a fool's errand. [...] If one is not to land in a Rube Goldberg mess of arrows and boxes, only to discover it's not pertinent to every inquiry, it's best to settle for pigeonholes roomy enough to organize the interconnected pieces of a given inquiry" [22, p. 86-7].

Challenges in generating clinically meaningful estimates

Defining an estimand and choosing an adequate study design and statistical analysis that align with the clinical aim of a study is clear in principle yet can be complicated in practice. The clinical interpretation of estimates has been clearly described for some estimators[23-28], but the link between technical procedures and substantive interpretation sometimes remains implicit in methodological guidance. As an illustration, most risk of bias tools that are widely implemented to assess the validity of a study do not explicitly address how to assess the potential for bias with respect to the aim of a particular study[29-32].

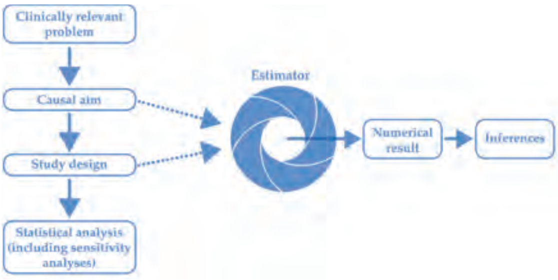
Having a deeper understanding of the impact that data analytical decisions can have on the interpretation of numerical results of a study would help to use methodology as a means to a clinical end. Comprehension of the influence that design choices might have on interpretation of estimates would be insightful for etiologic research and prediction research in particular, since most discussions on estimands and alignment with study conduct to date focused on therapeutic research.

Thesis aim and outline

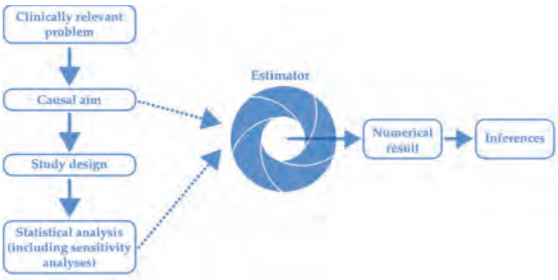
As described above, epidemiological methods have been increasingly refined, but it may not always be clear what the impact is of (habitual) choices regarding the design and statistical analysis of a study on the meaning of its numerical results. The aim of this thesis is to investigate this impact, where we focus separately on research into causal effects (Chapters 2-5) and prediction research (Chapters 6-8).

Part I: impact of applied methods on the meaning of numeric estimates in studies of causal effects

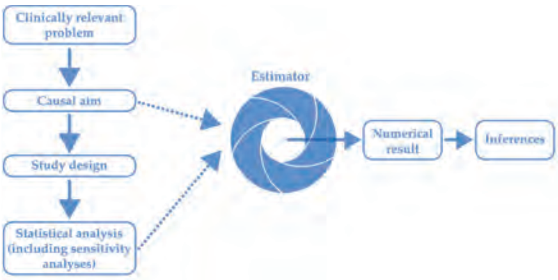
The first set of case studies (Chapter 2 – 5) focuses on causal research.



Chapter 2 evaluates how the time origin of a study design affects the interpretation of numerical results through a systematic review of reporting practices of the estimand of interest and the definition of study time origin in observational comparative effectiveness and safety cohort studies into effects of pharmacological treatments.



Chapter 3 discusses the implications of being ambiguous about the study aim in exploratory etiologic studies. We define a continuum of scrutiny with which exploratory statistical analyses are conducted and provide practical pointers for good practice in exploratory etiological research.

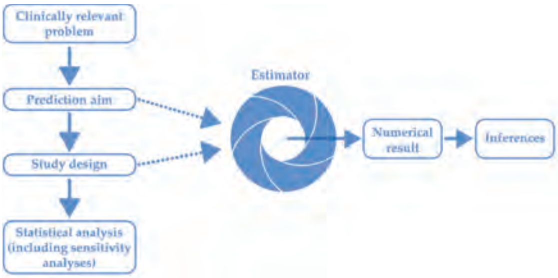


Chapter 4 evaluates the impact of the choice of a statistical approach on interpretation of results, by investigating under which conditions causal effect estimation in observational studies improves by using backward elimination on a prespecified set of potential confounders.

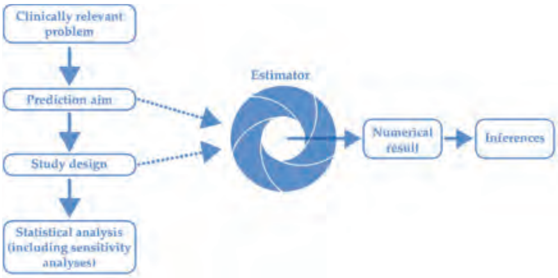
Chapter 5 describes how the link between the research question and study conduct can be assessed in studies of operative interventions, for instance as part of a systematic review. We propose an easy-to-use concise set of items derived from existing tools that is intended to perform an initial assessment of the applicability and methodological quality of research into the effects of surgical interventions.

Part II: impact of applied methods on the meaning of numeric estimates in prediction modelling studies

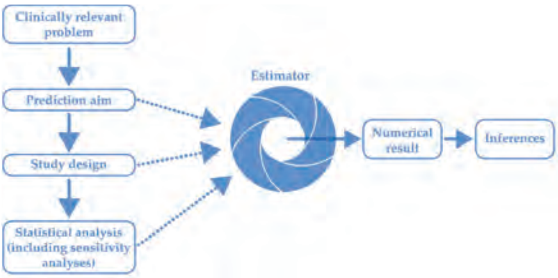
Chapters 6 – 8 form an in-depth case study in prediction research evaluating the impact of predictor measurement procedures on the performance of prediction models.



Chapter 6 uses an established taxonomy of measurement error models to define and clarify *predictor measurement heterogeneity*: differences in predictor measurement strategies across settings of derivation and validation. It is investigated whether predictor measurement heterogeneity affects the predictive performance of binary logistic prediction models, using analytical expressions and simulations.



Chapter 7 provides an illustration of the impact of predictor measurement heterogeneity on predictive performance in clinical examples using previously developed prediction models for diagnosis of ovarian cancer, mutation carriers for Lynch syndrome, and intrauterine pregnancy.



Chapter 8 presents a quantitative prediction analysis that can be used in validation studies to anticipate the impact of predictor measurement heterogeneity on predictive performance of time-to-event prediction models at implementation in practice if one of the predictor measurements is expected to deviate from the prediction target.

Chapter 9 summarizes the impact of applied methodology on interpretation of results described in these examples and ends with a general discussion on strengthening research question formulation in clinical studies.

References

1. Dickersin K, Straus SE, Bero LA. Evidence based medicine: increasing, not dictating, choice. *British Medical Journal*. 2007;334:s10.
2. Hill AB. *Principles of Medical Statistics.* London: Lancet; 1937.
3. Doll R, Hill AB. Smoking and carcinoma of the lung. *British Medical Journal.* 1950;2(4682):739.
4. Guyatt G, Cairns J, Churchill D, et al. Evidence-based medicine: a new approach to teaching the practice of medicine. *Journal of the American Medical Association.* 1992;268(17):2420-2425.
5. Porter TM. *Trust in numbers: The pursuit of objectivity in science and public life.* Princeton University Press; 2020.
6. Mooney SJ, Westreich DJ, El-Sayed AM. Epidemiology in the era of big data. *Epidemiology (Cambridge, Mass).* 2015;26(3):390.
7. Hand DJ. Statistical challenges of administrative and transaction data. *Journal of the Royal Statistical Society: Series A (Statistics in Society).* 2018;181(3):555-605.
8. Whincup PH, Gilg JA, Emberson JR, et al. Passive smoking and risk of coronary heart disease and stroke: prospective study with cotinine measurement. *British Medical Journal.* 2004;329(7459):200-205.
9. Katrak P, Bialocerkowski AE, Massy-Westropp N, Kumar VS, Grimmer KA. A systematic review of the content of critical appraisal tools. *BMC Medical Research Methodology.* 2004;4(1):1-11.
10. Stevenson M. Sample Size Calculations Using epiR. https://cran.r-project.org/web/packages/epiR/vignettes/epiR_sample_size.html. Published 2021. Accessed 28-05-2021.
11. Schneeweiss S, Rassen JA, Glynn RJ, Avorn J, Mogun H, Brookhart MA. High-dimensional propensity score adjustment in studies of treatment effects using health care claims data. *Epidemiology (Cambridge, Mass).* 2009;20(4):512.
12. Schneeweiss S. Automated data-adaptive analytics for electronic healthcare data to study causal treatment effects. *Clinical Epidemiology.* 2018;10:771.
13. Stang PE, Ryan PB, Racoosin JA, et al. Advancing the science for active surveillance: rationale and design for the Observational Medical Outcomes Partnership. *Annals of Internal Medicine.* 2010;153(9):600-606.
14. Reps JM, Schuemie MJ, Suchard MA, Ryan PB, Rijnbeek PR. Design and implementation of a standardized framework to generate and evaluate patient-level prediction models using observational healthcare data. *Journal of the American Medical Informatics Association.* 2018;25(8):969-975.
15. Feinstein AR. “Clinical judgment” revisited: the distraction of quantitative models. *Annals of Internal Medicine.* 1994;120(9):799-805.
16. Westreich D. *Epidemiology by design: a causal approach to the health sciences.* Oxford University Press; 2019.
17. Lash TL, VanderWeele TJ, Haneause S, Rothman K. *Modern epidemiology*. Lippincott Williams & Wilkins; 2020.
18. Shmueli G. To explain or to predict? *Statistical Science.* 2010;25(3):289-310.
19. Hernán MA, Hsu J, Healy B. A second chance to get causal inference right: a classification of data science tasks. *Chance.* 2019;32(1):42-49.
20. Van der Laan MJ, Rose S. *Targeted learning: causal inference for observational and experimental data.* Springer Science & Business Media; 2011.
21. ICH E9 working group. ICH E9 (R1): addendum on estimands and sensitivity analysis in clinical trials to the guideline on statistical principles for clinical trials. Published 2020. Accessed 28-05-2021.
22. Mayo DG. *Statistical inference as severe testing.* Cambridge University Press; 2018.
23. Ratitch B, Bell J, Mallinckrodt C, et al. Choosing estimands in clinical trials: putting the ICH E9 (R1) into practice. *Therapeutic Innovation & Regulatory Science.* 2020;54(2):324-341.

24. Mallinckrodt C, Bell J, Liu G, et al. Aligning estimators with estimands in clinical trials: putting the ICH E9 (R1) guidelines into practice. *Therapeutic Innovation & Regulatory Science.* 2020;54(2):353-364.

25. Ratitch B, Goel N, Mallinckrodt C, et al. Defining efficacy estimands in clinical trials: examples illustrating ICH E9 (R1) guidelines. *Therapeutic Innovation & Regulatory Science.* 2020;54(2):370-384.

26. Petersen ML, van der Laan MJ. Causal models and learning from data: integrating causal modeling and statistical estimation. *Epidemiology (Cambridge, Mass).* 2014;25(3):418.

27. Greifer N, Stuart EA. Choosing the Estimand When Matching or Weighting in Observational Studies. *arXiv preprint arXiv:210610577.* 2021.

28. Goetghebeur E, le Cessie S, De Stavola B, Moodie EE, Waernbaum I, initiative ttgCIotS. Formulating causal questions and principled statistical answers. *Statistics in Medicine.* 2020;39(30):4922-4948.

29. Whiting PF, Rutjes AW, Westwood ME, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Annals of Internal Medicine.* 2011;155(8):529-536.

30. Hayden JA, van der Windt DA, Cartwright JL, Côté P, Bombardier C. Assessing bias in studies of prognostic factors. *Annals of Internal Medicine.* 2013;158(4):280-286.

31. Higgins JP, Altman DG, Gøtzsche PC, et al. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. *British Medical Journal.* 2011;343.

32. Wells GA, Shea B, O'Connell D, et al. The Newcastle-Ottawa Scale (NOS) for assessing the quality of nonrandomised studies in meta-analyses. http://www.ohri.ca/programs/clinical_epidemiology/oxford.asp. Published 2000. Accessed 28-05-2021.