Universiteit
Leiden
The Netherlands

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

Luijken, K.

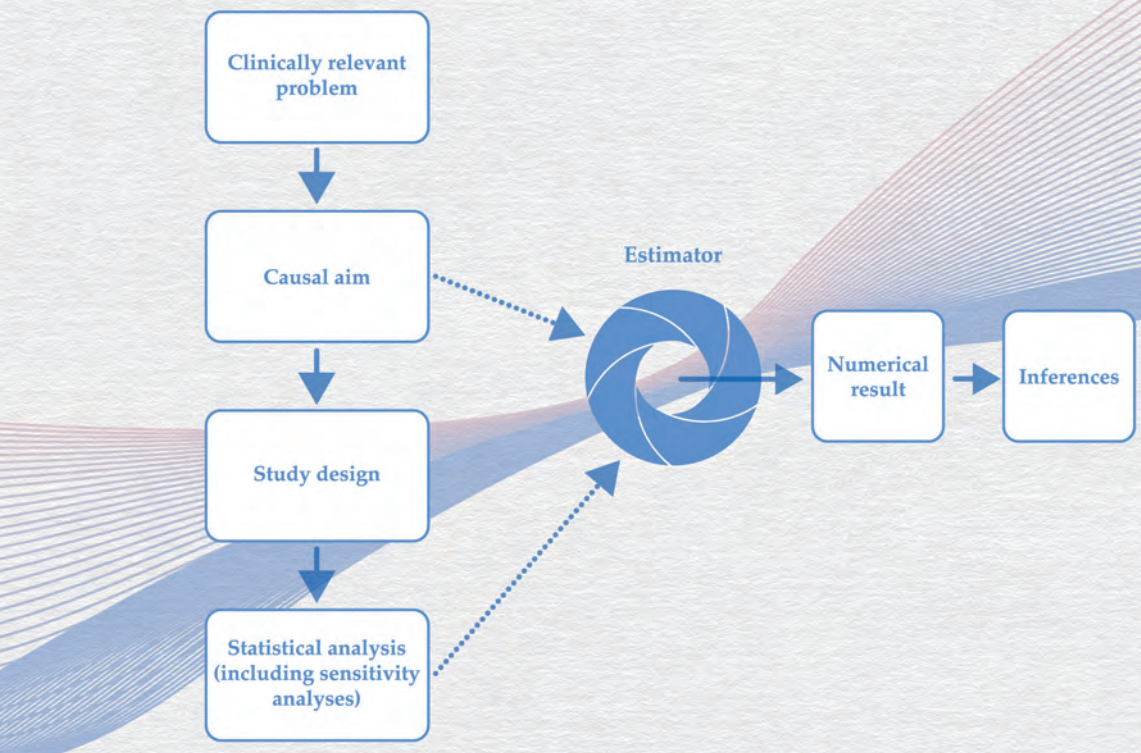
Citation

Luijken, K. (2022, May 19). *The impact of epidemiologic methods on findings in studies of causal effects and prediction modelling*. Retrieved from https://hdl.handle.net/1887/3304345

Version:	Publisher's Version
License:	Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden
Downloaded from:	https://hdl.handle.net/1887/3304345

Note: To cite this publication please use the final published version (if applicable).

3


Clinically relevant problem
Causal aim
Study design
Statistical analysis (including sensitivity analyses)
Estimator
Numerical result
Inferences

What harm is there in exploration? How to distinguish pernicious ad hoc analyses from valuable scientific contributions

Exploratory analyses run the risk of being sub-optimally conducted. Since exploratory analyses are typically done aiming to generate new hypotheses, it is tempting to quickly perform a statistical test (or multiple tests) to get a first answer to the problem. However, when such 'quick-test' results are presented in a publication, their interpretation may be ad hoc and unintentionally overconfident. We provide practical pointers for good practice in exploratory etiological research, such as the use of rigorous methodologic and statistical approaches and taking responsibility for exploratory findings by reporting a clear agenda for future research.

This chapter was based on: Luijken K, Dekkers OM, Rosendaal FR, Groenwold RHH. Exploratory analyses in etiologic research: considerations for assessment of credibility. *BMJ* (in press).

1 | Background

Findings from medical research can sometimes find their way to practice very rapidly. This became clear during the outbreak of severe acute respiratory syndrome coronavirus 2, when clinical decisions sometimes had to be made on preliminary evidence combined with considerations regarding the pathophysiology of the disease. However, preliminary or exploratory findings may turn out to be incorrect and may even harm patients when implemented too early. The Hippocrates' oath "Primum non nocere" ("first, do not harm") applies to medical research just as well as it applies to clinical practice. Researchers bear responsibility for the impact their findings may have beyond the scientific debate, irrespective of the type of analyses, even if these are named 'exploratory'.

It is not uncommon to present multiple exploratory analyses in etiologic studies, generally with the aim to generate hypotheses for future research. Such hypotheses may often be considered scientifically harmless. However, even when researchers consider their study to be exploratory, a hypothesis is easily promoted to a fact. For instance, findings in journal articles can be exaggerated to more certain statements in press releases and news articles[1].

In the present paper, we discuss issues that complicate the interpretation of exploratory analyses in etiologic studies and argue that exploratory results may harm both clinical research and clinical practice. At the same time, we are aware that without exploration, there is certainly less progress in science. We provide practical pointers for researchers on how to conduct exploratory analyses and how to clarify what the exploratory results imply for future research and implementation in practice. We end with some thoughts on the delicate balance between what is pernicious and what is valuable when it comes to exploratory analyses.

2 | Exploratory analyses in etiologic research

The origin of exploratory data analysis can be traced back at least to Tukey in the 60's and 70's[2,3], who encouraged statisticians to develop visualization techniques for representing and capturing structures in data sets to establish new research questions. Tukey pioneered in motivating the value of data-driven questions and the development of methods for improving non-specific knowledge about these questions to more exact answers. In this, he seemed to be predominantly concerned with science more than

decision-making. His writing paid little attention to the role of complex models and increasing computing power – two aspects that allowed for more extensive exploratory research over the course of time.

In what follows, we use the term *exploratory analyses* to indicate analyses that provide preliminary information that will help defining new research questions, and which are not always specified prior to data analysis. Exploratory analyses are often conducted additional to planned primary analyses of a study, which we denote *confirmatory analyses*. Sensitivity analyses, in which the main hypothesis is evaluated under different assumptions, are not considered to be exploratory in this paper. Outcomes that are evaluated as a secondary objective but are correlated to the primary outcome are not considered exploratory either, because these analyses contribute to the investigation of the primary research question. Genome-wide association studies, in which the exploratory nature of is commonly accounted for by addressing multiple testing[4], are beyond the scope of this paper.

For randomized trials, preregistration of the study protocol is considered the norm[5] and guidance on cautious interpretation of subgroup analyses is increasingly available[6] as is guidance for reporting of exploratory work preceding the randomized trial in a feasibility or pilot study[7]. However, in observational etiologic research, exploratory research questions may arise during data-exploration or statistical analysis, and little guidance exists on how these questions should be studied and interpreted.

Published etiologic studies often contain numerous results. As an illustration, in the first issue of 2021 from four major epidemiology journals (25 original etiological articles from the American Journal of Epidemiology, Epidemiology, European Journal of Epidemiology and International Journal of Epidemiology), we found that these articles presented on average 33 (range 1 – 120) associations for the primary analysis, on average 30 (range 0 – 336) associations for sensitivity analyses, and on average 163 (range 0 – 1467) associations in additional analyses, mainly concerning subgroup or interaction analyses (details in Online Supplement). Categorizing the additional analyses as either 'confirmatory' or 'exploratory' was not straightforward. Most articles did not explicitly report which analyses were prespecified and only one study referred to a publicly available protocol[8]. Some subgroup analyses seemed to have been carried out thoughtfully, with the intention to evaluate exposure effect heterogeneity among well-established subgroups, while other subgroup analyses seemed to have been performed exhaustively across a large number of possible risk factors, and results were selectively reported in the main text based on statistical significance.

As the distinction confirmatory / exploratory is not always straightforward in etiologic research, we propose to describe analyses in terms of a continuum of scrutiny (Figure 1), relating to exploratory and confirmatory studies alike. Where an analysis is situated on this spectrum from 'ad hoc' to 'targeted' depends on the nature of the research question and methodological rigor of the analysis. An analysis qualifies as being 'targeted' when the research question is well-advised by theory and the methodology and statistical analysis are designed accordingly, leading to fair credibility of the resulting evidence. While exploratory analyses are generally situated more on the 'ad hoc' end of this scale, they can be moved towards the 'targeted' side by conducting the study rigorously, as is described in more detail later.

3 | Exploratory analyses require directions for the reader

Exploratory analyses seem to be perceived harmless if their "hypothesis-generating" nature is made clear. No doubt, an open-minded approach in exploratory research leads to new insights and scientific progress that cannot be achieved by a rigid system of confirmatory research alone[9]. Even though confirmation can be considered a safeguard against the incorrect promotion of a hypothesis to a fact, there is a practical downside to such an approach as it is not feasible to conduct a confirmatory study for each of the large number of hypotheses that are currently generated in exploratory analyses. It can be easily understood that if a confirmatory study is needed for each of the aforementioned 163 results from additional analyses that we found on average in recently published epidemiological studies, this would overcharge available research capacity, if only in terms of independent data sets.

Exploratory results are tentative findings that should not form a basis for implementation in clinical practice. Some associations discovered during exploration of the data are expected to be false positive findings. Rapid implementation of exploratory findings into policy can have unwarranted consequences. Delay in dissemination of knowledge from science to clinical practice may detain patients from health benefits, but informing medical decisions based on preliminary evidence should be the exception rather than the rule to avoid harm by scientifically unjustified policies given the risk of false positive results.

Presentation of multiple analyses requires pointers for readers. For a judgement on the credibility of exploratory findings, readers of a study depend on unambiguous reporting of the nature and conduct of exploratory analyses, from which they can assess

whether the scrutiny of the analysis was 'ad hoc' or 'targeted' (Figure 1). Particularly for those that are more of the 'ad hoc' type, it is the responsibility of the author to prioritize which of the results are worth conducting a confirmatory study for and to report what that prioritization is based on (e.g., pathophysiological mechanism). Taking this responsibility clarifies how research resources should be allocated under minimally expected research waste[10-12].

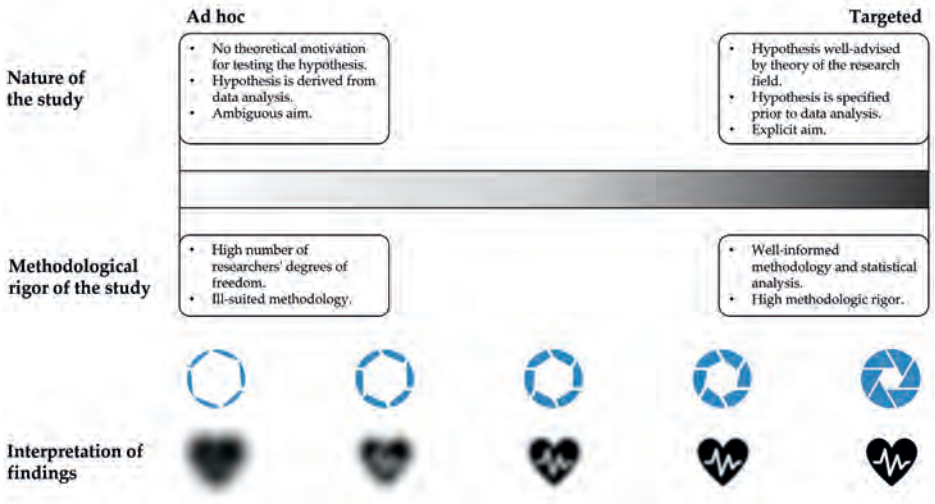


Figure 1. The continuum of scrutiny in conduct of etiologic studies. The nature of a research question and the degree to which it is aligned with the study design and statistical analysis determine the degree to which findings can be interpreted in a meaningful way. Although exploratory questions are generally situated more on the 'ad hoc' side of this scale, the interpretation of results could be clarified by improving methodological quality of the data analysis.

4 | 'Exploratory' does not imply 'less rigorous'

Prioritization of future research cannot be done based on numeric results of exploratory analyses only, because seemingly convincing results easily fool us into taking an observed association as something real and finding a clinical explanation that does not follow from the statistical evidence[13,14]. The statistical properties of exploratory tests are less well known than those of confirmatory tests[15]. For instance, the expected number of false positives (i.e., type I error rate) is likely inflated when statistical tests are not specified prior to data analysis or when the choice for a particular test depends on patterns in the data. While procedures have been developed for correction of multiple

testing in confirmatory settings, consensus on how to prevent false positive findings in exploratory settings has not been established.

What is more, the design and methods applied in an exploratory analysis may be less optimal than for the primary analysis of the study, which further complicates interpretation. For instance, when various exposure-outcome associations are explored, this likely requires more consideration than combining another set of covariates in a model. Analytic decisions should be reconsidered for each exposure-outcome combination that is studied, for instance the selection of confounders, specification of functional forms, model specification, and evaluation whether assumptions can be invoked[16,17].

The fact that an analysis is exploratory does not mean it can be taken as a loophole to avoid setting up a rigorous statistical analysis plan. Designing a study to test a hypothesis usually requires time and effort. When additional analyses are performed, either more resources should be spent to execute and report them rigorously, or it should be determined upfront how to restrict the number of analyses to match the available research capacity. Probably, to impose a strict methodological standard for exploratory analyses will induce a reduction in these analyses.

5 | Some good practices for exploratory analyses

Table 1 describes recommendations for good practice in conducting exploratory analyses at different stages of study conduct.

5.1 | Protocol

Protocols may contain a section describing exploratory analyses. It is tempting to presume that the plan to analyze a hypothesis will clarify itself once the data can be accessed. Of course, not every detail can be thought of and specified in advance, but interpretation of results provided by data can be challenging when no question was clearly articulated prior to seeing the answer. A way to prevent this confrontation is to pre-specify the analysis as thoroughly as possible[18,19]. The continuum of scrutiny suggests that pre-specification of a hypothesis (or an exposure – outcome relation) alone is insufficient for an analysis to qualify as 'targeted' and to render credible results. Furthermore, at the design stage of the study, it might be worthwhile to consider the degree of information that can be gained from exploratory analyses: if they will

provide little extra knowledge, why not refrain from performing these analyses? If truly interesting, why not work out a detailed protocol focused on that research question?[20,21]

Table 1 Practical recommendations for exploratory analyses

Research stage	Recommendation
Study protocol	- Limit the number of exploratory analyses. - Prespecify the aims and conduct of data analysis as thoroughly as possible in a research protocol. - Preregister the protocol (see [8] for an example of good practice*).
Statistical analysis	- When the idea to perform an analysis comes up after running the primary analysis, take a time-out to establish a targeted analysis plan for each exploratory research question. - Avoid conducting analyses ad hoc by formulating and analyzing exploratory questions as rigorous as possible. - Make sure that inferences are supported by the applied methodology and in line with their nature (i.e., confirmatory or more exploratory).
Reporting	- Present the results of every exploratory analysis including the methodological rigor with which the question was examined (ad hoc or targeted), possibly in supplementary files (see [23] for an example of good practice*). - Set a research agenda: which of the generated hypotheses are worth studying in confirmatory follow-up research, give arguments for that and direction on how should they be studied? (see [24] for an example of good practice*).
Peer-review and journals	- Credit methodological rigor rather than the number of results reported.

* The examples of good practice were identified from etiologic studies published in the first issues of 2021 of four major epidemiological journals (see Online Supplement).

5.2 | Statistical analysis

An obvious yet relevant good practice in performing statistical analyses is to stick to the prespecified analysis plan as closely as the data allow for[22]. In the situation where the idea to perform an analysis comes up after running the primary analysis, i.e., in case of a *post-hoc* analysis, it is of importance to design and analyze this exploratory question as rigorously as possible, similar to the way a confirmatory analysis would have been planned. We recommend taking a time-out to establish a targeted analysis plan for each exploratory research question, to avoid performing post-hoc analyses using ad hoc methodology. Importantly, the results need to be interpreted in line with their exploratory nature and communicated (e.g., in a research paper) as such.

Additional exploratory analyses introduce issues regarding multiplicity of analyses that have no straightforward solution but should be taken into consideration.

Which analyses should be statistically corrected for multiple testing and in what way? Should the planned primary analyses be corrected for multiple testing once additional exploratory analyses are performed? Performing multiple tests without a statistical correction inflates the risk of drawing false-positive conclusions, but too strict correction for multiple-testing can increase the probability of false-negative findings too (i.e., the type II error rate)[25]. This could occur, for instance, when an analysis of various positively correlated hypotheses is corrected for multiple testing as if all hypotheses were independent (for example by applying a *Bonferroni correction*). The decision to statistically correct for multiple testing depends on i.a. the total number of tests performed in the same dataset, the correlation between the hypotheses being tested, and the sample size.

5.3 | Reporting

Apart from clearly indicating which analyses were exploratory in nature, the report of a study should provide further guidance on the level of evidence for each of the hypotheses tested. A researcher is eminently aware to what degree the quality of the data and statistical methods were suitable to answer the research question and therefore can judge firsthand the credibility of exploratory finding[26]. Since the prior likelihood of an effect is generally low in exploratory observational studies, the posterior evidence is likely not astonishingly credible. Such knowledge can be communicated by stating a research agenda containing prioritization of future research and how this should be set up, thus allowing researchers to take responsibility for the presented exploratory findings and future research that should be performed, avoiding the empty statement that 'more research is needed'.

5.4 | Peer review and publication

Results are still often decisive for publication of a study[27] and presentation of multiple exploratory findings seems to match this reward system; articles that contain myriad results might contain interesting findings that make the study attractive for publication. A study that carefully specifies a sound research question and implements apt methodology may present fewer results but can make a valuable contribution and is potentially less harmful. Peer reviewers and journal editors should take this into account during the evaluation of a manuscript submitted for publication.

By no means should a manuscript reporting on a multitude of analyses be rejected for publication. Sensitivity analyses evaluating the primary analysis under different

assumptions can still lead to numerous results, but these analyses add to the credibility of the conclusions. Moreover, although allowing studies to report only a handful of additional analyses may improve the methodological quality (as there is more attention for each analysis separately), this may do more harm than good if coherent bodies of results become fragmented into multiple publications ('salami-slicing'[28]).

6 | Balancing opportunities and perils of exploratory analyses

Exploration is indispensable to the progress of science. Strict confirmatory studies are a powerful mechanism for final evaluations of existing evidence before implementation in clinical practice, yet will likely not spark new ideas[29]. In other words, research that is situated at the extreme targeted side of the continuum presented in Figure 1 provides most straightforward interpretations, but likely adds little to scientific understanding and progress. Open-minded exploratory analyses can lead to serendipitous discoveries and resourceful innovations of epidemiological science. Yet, this requires that exploratory questions are being answered using rigorous methodologic and statistical approaches. Even then, implications for clinical practice remain uncertain (and must so). Especially in medical science, where study results are sometimes quickly implemented in clinical practice, it is essential that researchers take responsibility for the results they report by minimizing the number of ad hoc analyses, designing all analyses thoroughly, and clearly explaining which exploratory findings should be investigated in future research and how. Only when exploratory analyses are conducted and interpreted *lege artis* will they unfold their full value.

Online Supplementary Files

The supplementary files referred to in this Chapter are available online at https://github.com/KLuijken/Dissertation_Online_Supplements/tree/main/Chapter_3

References

1. Sumner P, Vivian-Griffiths S, Boivin J, et al. The association between exaggeration in health related science news and academic press releases: retrospective observational study. *British Medical Journal*. 2014;349.
2. Tukey JW. The future of data analysis. *The Annals of Mathematical Statistics*. 1962;33(1):1-67.
3. Tukey JW. *Exploratory Data Analysis*. Vol 2: Reading, Mass.; 1977.
4. Sham PC, Purcell SM. Statistical power and significance testing in large-scale genetic studies. *Nature Reviews Genetics*. 2014;15(5):335-346.
5. Zarin DA, Tse T, Williams RJ, Rajakannan T. Update on trial registration 11 years after the ICMJE policy was established. *New England Journal of Medicine*. 2017;376(4):383-391.
6. Kent DM, Paulus JK, Van Klaveren D, et al. The predictive approaches to treatment effect heterogeneity (PATH) statement. *Annals of Internal Medicine*. 2020;172(1):35-45.
7. Eldridge SM, Chan CL, Campbell MJ, et al. CONSORT 2010 statement: extension to randomised pilot and feasibility trials. *British Medical Journal*. 2016;355.
8. Keys MT, Serra-Burriel M, Martínez-Lizaga N, et al. Population-based organized screening by faecal immunochemical testing and colorectal cancer mortality: a natural experiment. *International Journal of Epidemiology*. 2021;50(1):143-155.
9. Tukey JW. We need both exploratory and confirmatory. *The American Statistician*. 1980;34(1):23-25.
10. Schuit E, Roes KC, Mol BW, Kwee A, Moons KG, Groenwold RH. Meta-analyses triggered by previous (false-) significant findings: problems and solutions. *Systematic Reviews*. 2015;4(1):1-8.
11. Lakens D. The reproducibility project: a model of large-scale collaboration for empirical research on reproducibility. In Stodden V, Leisch F, Peng RD, editors. *Implementing reproducible research*. Osa Roca: Taylor and Francis Ltd. 2014:299-324. (Chapman & Hall/CRC The R Series).
12. Ioannidis JP. Why most published research findings are false. *PLoS medicine*. 2005;2(8):e124.
13. Goldacre B, Drysdale H, Marston C, et al. COMPare: Qualitative analysis of researchers' responses to critical correspondence on a cohort of 58 misreported trials. *Trials*. 2019;20(1):1-13.
14. Greenland S, Senn SJ, Rothman KJ, et al. Statistical tests, P values, confidence intervals, and power: a guide to misinterpretations. *European Journal of Epidemiology*. 2016;31(4):337-350.
15. Goeman JJ, Solari A. Multiple testing for exploratory research. *Statistical Science*. 2011;26(4):584-597.
16. Westreich D, Greenland S. The table 2 fallacy: presenting and interpreting confounder and modifier coefficients. *American Journal of Epidemiology*. 2013;177(4):292-298.
17. 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.
18. Shmueli G. To explain or to predict? *Statistical Science*. 2010;25(3):289-310.
19. Wagenmakers E-J, Wetzels R, Borsboom D, van der Maas HL, Kievit RA. An agenda for purely confirmatory research. *Perspectives on Psychological Science*. 2012;7(6):632-638.
20. Loder E, Groves T, MacAuley D. Registration of observational studies. British Medical Journal. 2010;340:c950.
21. Wang SV, Pinheiro S, Hua W, et al. STaRT-RWE: structured template for planning and reporting on the implementation of real world evidence studies. *British Medical Journal*. 2021;372.
22. Huebner M, le Cessie S, Schmidt CO, Vach W. A contemporary conceptual framework for initial data analysis. *Obs Stud*. 2018;4:171-192.
23. Rojas-Saunero LP, Hilal S, Murray EJ, Logan RW, Ikram MA, Swanson SA. Hypothetical blood-pressure-lowering interventions and risk of stroke and dementia. *European Journal of Epidemiology*. 2021;36(1):69-79.

24. Morrison CN, Kaufman EJ, Humphreys DK, Wiebe DJ. Firearm homicide incidence, within-state firearm laws, and interstate firearm laws in US counties. *Epidemiology.* 2020;32(1):36-45.

25. Groenwold RH, Goeman JJ, Le Cessie S, Dekkers OM. Multiple testing: when is many too much? *European Journal of Endocrinology.* 2021;184(3):E11-E14.

26. Mayo DG, Spanos A. Severe testing as a basic concept in a Neyman–Pearson philosophy of induction. *The British Journal for the Philosophy of Science.* 2006;57(2):323-357.

27. Smaldino PE, McElreath R. The natural selection of bad science. *Royal Society open science.* 2016;3(9):160384.

28. Editorial. The cost of salami slicing. *Nature Materials.* 2005;4(1):1.

29. Vandenbroucke JP. Observational research, randomised trials, and two views of medical science. *PLoS med*. 2008;5(3):e67.