

# Observational studies: practical tips for avoiding common statistical pitfalls

Anna Freni Sterrantino

The Alan Turing Institute, 96 Euston Road, London NW1 2DB, UK



The Lancet Regional Health - Southeast Asia 2024;25: 100415

Published Online xxx  
<https://doi.org/10.1016/j.lansea.2024.100415>

## Summary

This Personal View is intended for early-career researchers who are not yet experts in statistics. The Personal View focuses on common but usually avoidable flaws in the context of observational studies. I point out how study design, data collection, and statistical methods impact statistical results and research conclusions. With particular attention to study planning, sample selection, biases, lack of transparency and results misinterpretations.

Copyright Crown Copyright © 2024 Published by Elsevier Ltd. This is an open access article under the CC BY-NC license (<http://creativecommons.org/licenses/by-nc/4.0/>).

**Keywords:** Observational study design; Inference; Statistical methodology; Good practices; Statistical checklist; Statistical pitfalls; Bias

## Introduction

The aim of statistical science is to improve understanding, draw inferences from data, and help make decisions on issues that involve uncertainty. We turn data into information, and we collect, analyse, interpret, present, and organise data to fulfil this aim. Statistics analysis is a set of methods and techniques for making inferences about the characteristics of a population based on a sample of data drawn from that population (see [Box 1](#)). The sample is then used to estimate a parameter (or parameters) of interest, like location, scale, and proportions. Further, data samples are modelled to assess association measures (i.e. correlation, coefficients, odds ratio, risk ratio, etc.) to quantify the relationship between dependent and independent variables, exposure or risk factors and outcomes. Samples are collected by study design that can be experimental or observational.

Experimental studies like randomised controlled trials (RCT)<sup>1–3</sup> are often reported as the gold standard for drawing inferences, as they minimise bias and ensure that potential confounding variables are evenly distributed between groups. The researcher can experimentally manipulate the independent variable (see [Box 2](#)) and randomly assign the subject to the experimental or control group. The researcher can minimise the effect of third variables by eliminating or keeping the confounding variables constant. The randomisation process enhances the study's internal validity and strengthens the causal inferences that can be drawn from the results.<sup>4</sup>

In observational studies, the researcher does not control the independent variable because of ethical concerns or logistical constraints. Hence, observational

studies are common in fields such as epidemiology, social sciences, public health, and health policy. A typical observational study investigates the possible effect of a treatment on participants where the assignment of participants into a treated group versus a control group is outside the investigator's control. Due to the lack of an assignment mechanism, these studies are inherently difficult to analyse inferentially. However, an appropriate study design can help minimise this issue and provide results comparable to a simple randomised study design, which is considered the gold standard.<sup>5</sup> Minimising biases in observational study design plays a pivotal role in the validity and strength of statistical inference. Researchers must be aware of the limitations of such designs, and claims based on these studies should be reviewed carefully.

In this Personal View, I discuss avoidable statistical flaws for studies based on 'classical' observational studies: case-control, cross-sectional, and cohort studies. First, I introduce the main observational studies, their applications, discuss study steps, biases, and use of checklist to keep track of the study aspects as well as transparent reporting. Then, I discuss good practice and caveats in the choice and execution of statistical methods. I conclude by drawing attention to some flaws seen in reporting and in discussing statistical results.

## The foundation: study design

The main observational studies are case-controls, cross-sectional, population-based cohorts, and variants of these three main designs.<sup>6–8</sup> In [Fig. 1](#), I reported the basic schematic representation of the main observational studies, more schemes and variants can be found in Goodman et al.<sup>9</sup> Briefly, (i) a case-control study compares two existing groups differing in outcome to

E-mail address: [afrenisterrantino@turing.ac.uk](mailto:afrenisterrantino@turing.ac.uk).

## Box 1.

### The main objectives of statistics can be summarised as follows:

- Describing data. Statistical analysis helps summarise and describe a dataset's main features, such as central tendency (mean, median, mode), dispersion (range, variance, standard deviation), and distribution shape.
- Inference and generalisation. It allows researchers to make inferences about a population based on a sample of data. It provides tools for estimating parameters and testing hypotheses about population characteristics.
- Decision making. Statistical methods are used to support decision-making processes by providing a basis for drawing conclusions from data. This is particularly important in medicine, public health and social sciences.
- Prediction. Statistical models can be used to predict future events or trends based on historical data.
- Comparisons and relationships. Statistical analysis helps compare groups, study relationships between variables, and identify patterns in data.
- Research design. Statistical analysis plays a vital role in designing experiments and surveys, helping researchers choose the appropriate data collection and analysis methods.
- Policy formulation. In social sciences and government, statistical findings are often used to formulate and evaluate policies. It provides a basis for understanding social, economic, and demographic trends.
- Risk assessment. Statistics assesses and manages public health risks, creates surveillance systems, and assesses potential threats.
- Driving further research. Statistical findings help to generate new hypotheses and drive further research.

## Box 2.

### Roles of variables in statistical modelling.

- **Independent variable** is a variable that is manipulated or controlled by the researcher. It is the variable that is changed or varied in order to observe its effect on the dependent variable. It can often be coffee consumption or air pollution exposure, which is the presumed cause of a cause-and-effect relationship.
- **Dependent variable** is the variable being studied and observed for changes in response to the independent variable. It is the outcome or result that is measured or observed in an experiment or study. For example, is the health outcome like a cardiovascular event?
- **Confounding variable** is an extraneous variable in a statistical model that correlates with both the independent and dependent variables. It can obscure or distort the true relationship between the variables of interest. Controlling for confounding variables is important in statistical analysis to ensure accurate and reliable results. It is like smoking when we assess the relationship between coffee and cardiovascular events.

identify factors that may contribute to a condition by comparing subjects who have the condition with patients who do not have the condition but are otherwise similar.<sup>10,11</sup> (ii) A cross-sectional study is like 'photographing' a population at a specific time. It is employed to measure the prevalence of health outcomes, understand determinants of health, and describe features of a population.<sup>12–14</sup> (iii) A cohort study is a type of longitudinal study. It follows participants who share a specific characteristic(s) over a period (often many years).<sup>15,16</sup> During the study, participants get exposed to specific risk factors (for e.g. smoking or air pollution) and the associations between multiple outcomes are assessed.<sup>17,18</sup> In the prospective cohort, exposure is assessed at baseline and the researcher follows the subjects in time to study the development of disease or mortality. In a retrospective cohort, the researcher starts

the study at the end of the follow-up and retrospectively identifies the subject's eligibility and composition, and exposures are assessed at baseline. Each of these studies allows the estimation of specific parameters and associations between outcomes and exposure(s). In the hierarchy of evidence, after RCTs, we list prospective cohort, retrospective cohort, and case-control studies. Cross-sectional studies are rather descriptive and do not allow mimicking an RCT. To ensure robustness of results and the observational studies provide reliable inference as close to the RCTs, two main elements are crucial: (i) the definition of the target population and (ii) the assessment of the internal and external validity to allow for generalisation of the findings. These two elements are the pillars of observational studies and must be addressed in the planning process.

### Planning observational studies

Observational studies help to investigate research questions, find evidence for a hypothesis, and provide evidence for further investigations. To draw valid conclusions, statistical analysis requires careful planning from the beginning of the research process. Planning a research process offers a framework that will prevent major flaws: (i) specify your hypotheses and make decisions about your research question and design; (ii) define sample size, sampling procedure, and characteristics to collect.

The first step is to state the research question and associated hypothesis. The PICO format is an effective way to visualise and describe a research question, as it helps in directing data collection, analysis, interpretation, and application. PICO stands for Population or Problem (P), Intervention or Treatment of Interest (I), Comparison or Control (C) and Outcome (O). Although it is usually employed in clinical research questions, it offers a framework that can be adapted to observational studies or non-interventional studies. In a similar way, we identify the population and how large the sample is (P), what interventions or exposure to investigate (I), the definition of case/control groups (C), and what are the outcomes to measures (O) with the addition of the time component, hence when to measure (if needed). At the outset, a clear question and hypothesis will lead to an understandable and reproducible objective.

The second step is to define the sampling strategy and establish the criteria of inclusion or exclusion (often reported in flowcharts) to define the ability to generalise findings, hence ensuring external validity, i.e. extending findings to similar populations. Primary outcomes should be chosen carefully, based on the sample size and relevance to the research questions, as well as secondary outcomes that could provide proxies and aid in the interpretations. Exposures should be clearly defined, as incorrect or inadequate exposure can lead to measurement and misclassification biases. Because observational studies are uncontrolled, they are susceptible to

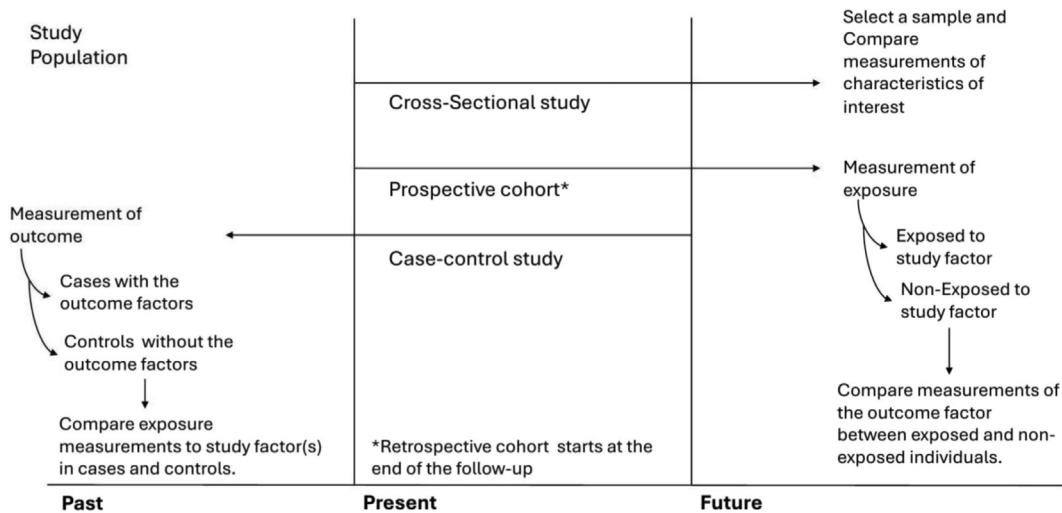


Fig. 1: Schematic representation of the processes of selection and measurement in cohort, case-control, and cross-sectional studies.

external factors affecting the relationships between outcome and exposure. Hence, variables introduced in the modelling should be closely examined for potential roles as confounders and effect modifiers. Variables that behave as confounders produce a spurious relationship between outcome and exposure. Effect modifiers are the variables associated with the outcome but not the exposure, for example, a drug works for females and not males.

These steps are potentially affected by bias, a systematic error in the design and method. Researchers should be aware of four main biases in their studies.

### Biases in observational studies

The main biases in observational studies are selection bias, information bias and measurement errors, confounding and Simpson's paradox.<sup>19,20</sup> All the biases compromise validity, i.e. selection bias compromises external validity while confounding compromises both internal and external validity.

*Selection bias* occurs when individuals, groups, or data are chosen for analysis in a not random way. This can lead to inaccurate results because the sample may not represent the targeted population. For example, if the study is at a participatory level, certain barriers, like socio-economic or access to resources may prevent the right population from participating as well as create the opposite effect, the self-selection bias. Self-selection, also known as volunteer-bias, arises in any research study in which participants choose if they want to be part of the sample. It is a common type of research bias and leads to a sample that is not representative of the population as a whole.<sup>21</sup> Online surveys are often flawed by voluntary selection bias. A similar remit is classified as coverage bias, which occurs when the target population does not coincide with the population sampled.

Coverage error can result from under-cover when intended members of the target population are excluded, or vice-versa with over-coverage. Both under-cover and over-coverage are biases that may distort inferences based on descriptive or analytical statistics. Weaknesses in the sampling frame or in survey implementation create coverage error by compromising the random selection and, thus, the representativeness of the populations. In this category, nonresponse biases can occur when subjects refuse to participate in a study. These biases should be identified early on and considered whenever possible. For example, nonresponse biases can be mitigated by a mix of responsive and adaptive design and non-response weighting.<sup>22</sup> These adjustments allow the researchers to compensate for the nonresponsive bias by mixing, for example, statistical adjustment. Statistical methods can help to reduce it by applying poststratification or weighting class adjustments, raking or weighting, generalised regression modelling and propensity score adjustment.<sup>23</sup> Other preventive steps include re-screening potential participants, implementing procedures to handle missing data from participants lost to follow-up, and checking intervention or exposure groups compared to the baseline.

*Information bias* refers to systematic errors in measurement (or misclassification) of the exposure or outcome. For example, if the subject is asked to report on past experiences, often recall-biases can happen when the subject is imprecise on past events. Answers can also be biased due to social desirability. A classic example in maternal and child health is the under-reporting of smoking habit by mothers during pregnancy.<sup>24</sup> All these lead to measurement errors, which tend to underestimate potential associations by making the groups more similar and tweaking the strength of the association.

*Confounding* occurs when a variable influences both the independent and dependent variables. Failing to account for confounding variables can cause you to estimate the relationship between your independent and dependent variables wrongly. For example, in a study where the aim is to find if drinking coffee is associated with coronary heart disease, smoking, which is also associated with coffee drinking, was not accounted for, and played a confounding role. This estimated association, i.e. ignoring that smoking is associated with both incidence and outcome, gives a false impression that coffee drinking and coronary heart disease are associated. The preventive measures to reduce the risk of confounding are randomisation, assigning subjects randomly to the treatment; stratification of the analysis (multiple regression by submitting the population: smokers versus non-smokers), and by ‘controlling’ for confounding variables in multivariate analysis, by including them in the regression models.<sup>25</sup> Residual confounding occurs when the distortion on the relationship between outcome and exposure remains, even though the research has included and controlled for all the other confounding variables.

Potential confounding also creates the statistical phenomenon known as *Simpson’s paradox*. The paradox states that an association between two variables in a population emerges, disappears, or reverses when the population is divided into subpopulations (as per gender, age, etc.). Indeed, not all biases can be mitigated or eliminated. Hence, results should be read considering the biases in the study, and conclusions should be expressed accordingly.<sup>19</sup> One way to keep track of all the potential biases and limits in observation studies is to rely on a checklist that can identify studies’ flaws early on and help mitigate these during the analysis.

### Did you forget anything? Use a checklist

Study-specific guidelines serve as a tool to help authors craft well-structured manuscripts that facilitate reader comprehension and critical assessment. Since the introduction of these guidelines, more journals require authors to adhere to the guidelines before submitting their manuscripts. It has been observed that authors who adhere to these guidelines enhance their chances of successfully publishing their findings in a journal.<sup>26,27</sup> To minimise incomplete and inadequate reporting, the EQUATOR (Enhancing the QUALity and Transparency Of health Research) Network, defined reporting guidelines as: ‘A checklist, flow diagram, or structured text to guide authors in reporting a specific type of research, developed using explicit methodology’.<sup>28,29</sup> Current checklists cover various studies, such as randomised trials, case reports, study protocols, quality improvement studies, qualitative research, and systematic and meta-analyses. Of interest here is the checklist STROBE ‘Strengthening the Reporting of Observational studies in Epidemiology’,<sup>30</sup> specifically designed to cover cohort,

case-control, cross-sectional studies, and conference abstracts. A good practice is to fill these checklists carefully with comments and details.

### The core: statistical methods

In this section, I draw attention to some aspects that need to be considered in conducting the statistical analysis that are sometimes overlooked.

#### Start simple

Statistical analysis is at the core of research articles. Before jumping into data modelling, summary data descriptions should be provided in tables and figures. A strong description of statistics drives the modelling. Often, this seems trivial, and some researchers jump directly to modelling. Descriptive statistics highlight covariate distribution, which can point to categorisation (if a variable is highly skewed), to assess non-linearity, missing data (with imputation methods necessary), univariate association and testing if two (or more) groups are statistically different. It helps to provide subsections describing all the elements: the data set, outcomes, covariates and their characteristics, statistical modelling, and additional sensitivity analysis. Formulas and equations should also be provided, particularly when variants of known methods are used, and it is harder for the reader to grasp what was done.

#### Modelling approaches

Research questions, sampling, and data collected will drive the modelling approaches. The board class of regression modelling is the cornerstone of many statistical analyses for estimating the association between exposures and outcomes. Researchers often fit simple models, but they do not provide any checks to ensure that model assumptions have been evaluated. Fitting regression models do not end at the coefficients table.

Every model works based on specific assumptions; for example, residual analysis in linear regression models assesses that the model assumption has not been violated, and we can trust estimates. Similarly, in survival analysis, the Cox model requires the proportionality of the hazard to be true. Otherwise, a different model must be fitted.

Data collection drives the modelling choice; for example, data collected as clusters or by group, like individuals in different hospitals or regions, or nested data like children in class, schools and regions, will have to be analysed considering mixed effect (multilevel or hierarchical) models to account for this difference between schools or classes or regions, by including random effects. Furthermore, subgroup analysis can help identify Simpson’s paradox.

Data transformation changes the interpretation. For example, in a linear regression model, the log transformation of the outcome (dependent variable) requires

the exposure coefficients to be exponentiated and interpreted as a multiplicative factor. If both outcome and exposure are log-transformed, then the coefficients are interpreted as a per cent increase rather than an independent unit increase.

Data Missed at Random (MAR) occurs when there is no value for the observation collected, but often, missing data do not happen randomly. Hence, an imputation (interpolation) approach is indicated for a small portion of missing data, while data removal can occur for larger portions. Missing it at random requires the researcher to understand the potential mechanism, as these can impact the findings and unravel some mechanisms or errors in the data collection. In surveys, two known variables that are quite often missed, not at random, are income and smoking habit for obvious privacy and stigma reasons. Educational levels are often used then as a proxy of the socio-economic status and in ecological studies, tobacco sales per area level are usually accounted as proxy for smoking habit.

Other errors and flaws can interfere with regression analysis, such as categorical covariates passed as numerical into the model fit software rather than ensuring it is coded as a factor. Writing the statistical methods provides only the software function used in the statistical software rather than model equations. Forgetting to report sample size for subgroups or sensitivity analysis. Producing tables only reports the exposure estimates and not the full model. Not addressing the role of potential effect modifiers or confounding variables. If data has been imputed, the model estimates for the imputed and not-imputed sample have not been reported, or missing data has been ignored.

Finally, uncertainty cannot be ignored in statistical analysis. In an observation study, there are two main layers of uncertainty: direct about specific facts, numbers, and science (both absolute and relative) and indirect about the quality of our underlying knowledge. The uncertainty associated with statistical estimates should be clearly provided and communicated: scale measures, confidence and credible intervals, and residual analysis or residual maps in disease mapping. However, indirect uncertainty should also be reflected in the description of the findings. The wording used to evaluate and interpret results should be conveyed using appropriate qualifying statements in the discussion section.<sup>31</sup>

### Modelling approaches for correlated data

Many applications in epidemiology, infectious diseases, and exposure modelling rely on data with spatial support, like spatial sampling, geo-localised curves census data and in cases of surveillance on time series to capture changes or rate of changes in adverse outcomes. The main feature of spatiotemporal observations is their correlation. Hence, the observation is

correlated in time or space (or both). Spatio-temporal data aims to capture the distribution and relationships of physical or geographical features within a specific area or identify particular interest clusters. This type of study is useful for understanding spatial patterns, trends, and variations in a location. Spatio-temporal statistics methods tend to employ Bayesian approaches<sup>32,33</sup> rather than frequent ones. Bayesian methods offer a powerful framework for analysing spatio-temporal data by incorporating prior information, flexibly modelling spatial processes, quantifying uncertainty, and making reliable predictions.<sup>34</sup> The core difference is that frequentist statistics relies on long-run frequencies and treats parameters as fixed, unknown values, while Bayesian statistics incorporates subjective beliefs, treats parameters as random variables, and updates these beliefs based on observed data using Bayes' theorem. While there is no hard rule on what type of approach should be used in tackling specific statistical analysis, Bayesian approaches are a popular choice for spatio (and temporal) analysis. The main feature of spatial data is their intrinsic spatial correlation, where things that are closer to the first geographic law tend to be more similar. In other words, spatial data is characterised by losing independence between units as with temporal time series, where past observations are highly correlated closer to the present date. While there is no damage in using geographical approaches or frequentist approaches to model time series, for the combination of spatio-temporal data, the advice here is to use Bayesian approaches to modelling.

### Show me, don't tell me

Figures and tables present results succinctly, so they should be able to stand alone. This means that a figure that is well done has labelling, a colour legend that is easy to grasp, is not too cluttered, and the caption helps to explain the additional details. Tables should follow the same rules. A good reminder is to present a summary statistics table early on, to provide an overview of the measures' ranges and distributions. Do not attempt to rewrite what is in the table but report the main data features.

It is good practice to polish the software output with appropriate and ordered labelling, avoid scientific number representations (for example, reporting scientific numerical representation), and remember to report how multivariate models were adjusted for, define the reference classes, or treat categorical variables as numerical.

A flowchart representing how the sample has been identified is more practical than describing all the steps in the text, but it is quite often missed. Providing access to codes and data is always a good sign. The statistical reviewers appreciate it, and hosting websites like GitHub are well suited for this type of material. A



pseudo dataset could also be used if data is not publicly available.

When working with spatial data and geo-localised data, maps are a must. But as any figure, maps should be made assuming that not all the readers are familiar with the specific area or regions; insets and country area maps should be included (check Fig. 1 in the study by Indu and colleagues<sup>35</sup>). Guidance on the geographical area classifications should be provided as not all the readers will be familiar with the country's geography.

### The aftermath: reporting and discussing results

The results and interpretations are presented once all the methods and analyses have been carried out. Presenting and discussing results are important parts of the process and help drive interventions and future studies. This is also the part where biases and a critical approach to the studies should be discussed with the readers and conclusion drafted.

### Everything has its place

Reviewers—also known as the first readers—value order, clarity, logical steps, consistency, simplicity, transparency, and reproducibility.

The main step is to write in good scientific English: with short sentences, active form rather than passive and strong verbs. Crafting a well-written article improves readability, and there are plenty of resources<sup>36,37</sup> and available free courses on the main learning platforms.

Presenting methods and results should follow the same rules as a cooking recipe and should not be mixed up. The method should contain a subsection of the data definition. For example, Tandon and colleagues<sup>38</sup> used data from India's Fifth National Family Health Survey (NFHS-5) and extracted a subset of women who have experienced an adverse pregnancy outcome and reported it in a table. When the targeted population is identified from hoping cohorts or registries, flowcharts with eligibility criteria are the way to convey this type of information; as an example, check the supplementary information in the cohort study by Karuniawati and colleagues<sup>39</sup> and Figures 1 and 2 in the cohort study by Thornton and colleagues.<sup>40</sup> The following subsection may be used to describe the outcomes and covariates confounders and exposure that are relevant to the study (see for example<sup>35,40–42</sup>). Providing the readers with a detailed list of the steps done and choices made in carrying out the statistical analysis increases transparency and supports reproducibility. As mentioned before, sharing of code fulfils these requirements.

Once all the elements are listed, then the statistical analysis paragraph can be written. It should start with simple descriptive statistics (in epidemiology, sometimes these are known as univariate), moving to the modelling or multivariate analyses. Secondary outcomes

and sensitivity, interaction and subgroup analysis should also be described jointly with the rationale. Listing the modelling equation will prevent reviewer questions. Sometimes, the analysis description can be confusing, and it is hard to understand what was actually done. For consistency, the reviewer expects to see the results section mimicking the statistical methods: starting with simple statistics, i.e. descriptive, finishing with sensitivity (sometimes, as these are to corroborate the results obtained in the primary analysis, the tables and figures can be reported in the supplementary material [SM]). A good practice is to report all the useful tables and figures in the SM. Statistical reviewers have been accustomed to journals with word limits and authors reporting just the main steps in a paragraph. But the statistical reviewers are pleased to see extra details in SM.<sup>43,44</sup>

Hence, to ensure clarity, explain the rationale and specific analysis choices. A good practice is to address all the loose end in the statistical analysis. What was done about missing observations? Was an imputation made, or were they discarded? Have the covariates been categorised, yes/no, why, and if so, what are the categorical variable reference levels, and so on? How many levels have been selected, where collapsed, why? etc.

### No favouritism, please

Statistical examples in books are often crafted in a way to minimise loose ends: lack of data, low variability, equally distributed categorical variables, etc. However, once data is collected, adjustments must be made to the statistical analysis, and the study conclusion must consider the potential issues that have arisen in due course. Quite often, however, some researchers commit a logical fallacy known as 'cherry picking'.<sup>45</sup>

The phenomenon happens when researchers focus only on evidence supporting their stance while ignoring evidence contradicting it. Typical behaviour is seen when researchers choose data supporting a conclusion while ignoring contradicting data. This leads to statistical fallacies and undermines the objectivity of analysis with faulty conclusions and misguided actions. It also damages the trust in statistical studies, affecting their credibility. Hence, transparency in the data collection and analysis process reduces cherry-picking and increases reader trust.

### Watch your language!

In the context of observational studies, we are interested in finding associations between variables rather than concluding a potential causality statement. Causality is not part of the statistical analysis<sup>46</sup>; rather, it is a narrative that we add to the statistical analysis and the scientific and statistical hypothesis that we are interested in refuting. However, for the sake of fairness, in recent years, researchers have been proposing methods to bend observational studies to mimic experimental

studies.<sup>47–49</sup> Nonetheless, these approaches require statistical expertise, a deep understanding of the potential confounding variables matching approaches, and control over caveats that could hinder the study's conclusions. Hence, causality should not be addressed in the research article where the aim is to assess associations, and it does not seem to be useful to address the lack of causality due to the type of study conducted.

Researchers should familiarise themselves with statistical lexicography. In particular, the use of 'statistical significance' and the difference between 'correlation' and 'association'. The word 'significant' means sufficiently great or important to be worthy of attention, noteworthy, consequential, or influential. We use the term 'statistically significant' to comment on the result of a statistical test. Specifically, we say that we had observed 'statistically significant' results when we observed a low probability that the null hypothesis is true (traditionally, p-value less than 0.05), we reject the null hypothesis. Hence, comments on statistical results that are only described as 'significant' are improper and inaccurate. Results can be significant in the sense of being 'important' for clinical or social reasons, for example. But if the intention is to comment or discuss the null hypothesis result, then the wording 'statistically significant' should be used.

Some authors confuse 'association' with 'correlation' or use them interchangeably and misinterpret them as 'causation'. Association means that one variable provides information about another variable. A relationship between two variables that we estimated via linear coefficients, relative risk, odds ratio, and hazard ratio. On the contrary, correlation specifically measures the trend decreasing or increasing between two variables. A Pearson correlation coefficient (often denoted by "r") quantifies the relationship between linearly related variables. While non-parametric correlation measures, Kendall and Spearman correlation coefficient assess monotonic relationships. Pearl et al.<sup>46</sup> define causation as follows: we define that variable X is a cause of variable Y, if Y in any way relies on X for its value. Hence, correlation quantifies a relationship between two variables, not causation, while association is the same as dependence and may be due to direct or indirect causation. 'Correlation implies association, but not causation. Conversely, causation implies association, but not correlation'.<sup>50</sup> Hence, when reporting or commenting on model estimates, use association; when commenting on one of the three correlation measures, use correlation (or association) and use causality if you have conducted an experimental study design.

### Do not sugar coat

Given the study design and the data collection, there will most likely be biases that must be addressed regarding limitations in the discussion section and in drawing conclusions from the data collection and analysis.

While identifying strengths is less complicated, writing the limitations paragraph requires wearing the auto-critical hat. Limitations should focus on reviewing the generalisability of the results by assessing and evaluating the external and internal validity of the study.

For example, if the study is conducted in a specific district area or using data from a single-centre (like a health centre or a hospital),<sup>51,52</sup> it adds to the evidence of the research literature. However, because of the low national representativeness, extrapolating conclusions to suggest national-level intervention or extending the findings to the whole country population—based on such a small sample—is inappropriate. Questions that help to evaluate it critically should focus on what could have been done better. To evaluate the study critically, look for areas where confounding and biases may have hampered the results. A way to critically evaluate the study is to identify and recognise which areas of study prevent mocking an RCT. Limitations should also address potential error measurements and the choice made by the researchers in using certain variables as proxies, for example. In recent years, I have noticed a tendency to include as strength the phrases: 'To our knowledge, this is the first ...' and in study limitations, the clause that 'the study (classic observational) does not permit causal conclusion'. Both convey irrelevant information and add to the impression of overselling the work, rather than with the actual science and a result's validity.<sup>53</sup> As a reviewer, I recommend avoiding them at all costs.

### Conclusion

In this paper, I briefly addressed the common statistical pitfalls in observational studies (see [Box 3](#)), similar to the article by Mansournia and Nazemipour<sup>54</sup> that provides advice on accurate reporting in medical research statistics and the article by Sydes and Langley<sup>4</sup> on how to avoid pitfalls in the design and reporting of clinical trials. I highly recommend reading both. Carelessness in conducting robust statistical analysis can invalidate the efforts and benefits that would have been derived. Hence, a well-conducted statistical analysis provides crucial evidence for driving interventions and improving

#### Box 3.

##### Summary on how to avoid statistical pitfalls.

- Identify biases early on in the study.
- Always adhere to reporting guidelines as well as complete and submit study checklist: STROBE, PRISMA, etc.
- Use appropriate statistical methods.
- Provide as many details on statistical analysis as possible (more details in supplementary material).
- Be consistent in presenting methods and results.
- Do not ignore contradicting evidence and mention limitations of the study.
- Use an appropriate statistical lexicon.

public health, thus enhancing population well-being while identifying key areas of concern.

## Contributors

AFS conceptualised and wrote the paper.

## Declaration of interests

This work was supported by the Ecosystem Leadership Award under the EPSRC Grant EP/X03870X/1 & The Alan Turing Institute, particularly the Turing Research Fellowship scheme under that grant. The author declares no conflicts of interest.

## Acknowledgements

These personal views are based on my experience as a statistical reviewer and on the frustrations of reading great papers with important research questions, but poor statistics. I am grateful to the Editor and the two anonymous reviewers for their comments and feedback.

## References

- Kim K, Bretz F, Cheung YKK, Hampson LV. *Handbook of statistical methods for randomized controlled trials*. CRC Press; 2021. <https://doi.org/10.1201/9781315119694>.
- Hariton E, Locascio JJ. Randomised controlled trials—the gold standard for effectiveness research. *BJOG An Int J Obstet Gynaecol*. 2018;125(13):1716. <https://doi.org/10.1111/1471-0528.15199>.
- Matthews JN. *Introduction to randomized controlled clinical trials*. CRC Press; 2006. <https://doi.org/10.1201/9781420011302>.
- Sydes MR, Langley RE. Potential pitfalls in the design and reporting of clinical trials. *Lancet Oncol*. 2010;11(7):694700. [https://doi.org/10.1016/s1470-2045\(10\)70041-3](https://doi.org/10.1016/s1470-2045(10)70041-3).
- Rosenbaum PR. *Design of observational studies*. Springer series in statistics. Springer International Publishing; 2021. Available from: [https://books.google.pl/books?id=srp\\_zgEACAAJ](https://books.google.pl/books?id=srp_zgEACAAJ).
- Rothman KJ. *Epidemiology: an introduction*. Oxford University Press; 2012.
- Rothman KJ, Greenland S, Lash TL. *Modern epidemiology*. Wolters Kluwer Health/Lippincott Williams & Wilkins; 2015. Available from: <https://books.google.co.uk/books?id=MSTgnQAACAAJ>.
- Krickeberg K, Van Trong P, Hanh PTM. *Epidemiology: key to public health*. Springer; 2019. <https://doi.org/10.1007/978-3-03016368-6>.
- Goodman CS. *HTA 101 Introduction to health technology assessment*; 2014. Available from: <https://www.nlm.nih.gov/nichsr/hta101/ta10103.html>.
- Nilsson L, Farahmand B, Persson P, Thiblin I, Tomson T. Risk factors for sudden unexpected death in epilepsy: a case control study. *Lancet*. 1999;353(9156):888–893. [https://doi.org/10.1016/s0140-6736\(98\)05114-9](https://doi.org/10.1016/s0140-6736(98)05114-9).
- Intawong K, Chariyalertsak S, Chalom K, et al. Effectiveness of heterologous third and fourth dose COVID-19 vaccine schedules for SARS-CoV-2 infection during delta and omicron predominance in Thailand: a test-negative, case-control study. *Lancet Reg Health Southeast Asia*. 2023;10:100121. <https://doi.org/10.1016/j.lansea.2022.100121>.
- Sharif H, Sheikh SS, Seemi T, Naeem H, Khan U, Jan SS. Metabolic syndrome and obesity among marginalised schoolgoing adolescents in Karachi, Pakistan: a cross-sectional study. *Lancet Reg Health Southeast Asia*. 2024;21:100354. <https://doi.org/10.1016/j.lansea.2024.100354>.
- Sakboonyarat B, Rangsin R. Characteristics and clinical outcomes of people with hypertension receiving continuous care in Thailand: a cross-sectional study. *Lancet Reg Health Southeast Asia*. 2024;21:100319. <https://doi.org/10.1016/j.lansea.2023.100319>.
- Wang X, Cheng Z. Cross-sectional studies: strengths, weaknesses, and recommendations. *Chest*. 2020;158(1):S65–S71.
- Barrett D, Noble H. What are cohort studies? *Evid Based Nurs*. 2019;22(4):95–96. <https://doi.org/10.1136/ebnurs-2019-103183>.
- Setia MS. Methodology series module 1: cohort studies. *Indian J Dermatol*. 2016;61(1):21. <https://doi.org/10.4103/00195154.174011>.
- Sarfraz A, Jamil Z, Ahmed S, et al. Impact of diarrhoea and acute respiratory infection on environmental enteric dysfunction and growth of malnourished children in Pakistan: a longitudinal cohort study. *Lancet Reg Health Southeast Asia*. 2023;15:100212. <https://doi.org/10.1016/j.lansea.2023.100212>.
- Gupta V, Singh A, Ganju S, et al. Severity and mortality associated with COVID-19 among children hospitalised in tertiary care centres in India: a cohort study. *Lancet Reg Health Southeast Asia*. 2023;13:100203. <https://doi.org/10.1016/j.lansea.2023.100203>.
- Hammer GP, du Prel JB, Blettner M. Avoiding bias in observational studies: part 8 in a series of articles on evaluation of scientific publications. *Deutsches Arzteblatt Int*. 2009;106(41):664. <https://doi.org/10.3238/arztebl.2009.0664>.
- Pandis N. Bias in observational studies. *Am J Orthod Dentofacial Orthop*. 2014;145(4):542–543. <https://doi.org/10.1016/j.ajodo.2014.01.008>.
- Cheung KL, Ten Klooster PM, Smit C, de Vries H, Pieterse ME. The impact of non-response bias due to sampling in public health studies: a comparison of voluntary versus mandatory recruitment in a Dutch national survey on adolescent health. *BMC Public Health*. 2017;17(1):1–10. <https://doi.org/10.1186/s12889-017-4189-8>.
- Brick JM, Tourangeau R. Responsive survey designs for reducing nonresponse bias. *J Off Stat*. 2017;33(3):735–752. <https://doi.org/10.1515/jos-2017-0034>.
- Greenacre ZA. The importance of selection bias in internet surveys. *Open J Stat*. 2016;6(3):397. <https://doi.org/10.4236/ojs.2016.63035>.
- Wong M, Koren G. Bias in maternal reports of smoking during pregnancy associated with fetal distress. *Can J Public Health*. 2001;92:109–112. <https://doi.org/10.1007/bf03404942>.
- Skelly AC, Dettori JR, Brodt ED. Assessing bias: the importance of considering confounding. *Evid Base Spine Care J*. 2012;3(1):9–12. <https://doi.org/10.1055/s-0031-1298595>.
- Sharp MK, Glonti K, Hren D. Online survey about the STROBE statement highlighted diverging views about its content, purpose, and value. *J Clin Epidemiol*. 2020;123:100–106. <https://doi.org/10.1016/j.jclinepi.2020.03.025>.
- Cuschieri S. The STROBE guidelines. *Saudi J Anaesth*. 2019;13(Suppl 1):S31. <https://doi.org/10.4103/sja.sja54318>.
- Hope DL, King MA. *The 'so what' of reporting guidelines*. UK: Oxford University Press; 2022. <https://doi.org/10.1093/jipp/riac035>.
- Simera I, Moher D, Hirst A, Hoey J, Schulz KF, Altman DG. Transparent and accurate reporting increases reliability, utility, and impact of your research: reporting guidelines and the EQUATOR network. *BMC Med*. 2010;8(1):24. <https://doi.org/10.1186/17417015-8-24>.
- Von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP. The Strengthening of Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *Lancet*. 2007;370(9596):1453–1457. <https://doi.org/10.1097/ede.0b013e3181577654>.
- Van Der Bles AM, Van Der Linden S, Freeman AL, et al. Communicating uncertainty about facts, numbers and science. *R Soc Open Sci*. 2019;6(5):181870. <https://doi.org/10.1098/rsos.181870>.
- Stern HS. Bayesian statistics. In: Wright JD, ed. *International encyclopedia of the social behavioral sciences*. 2nd ed. Oxford: Elsevier; 2015:373–377. <https://doi.org/10.1016/B978-0-08-097086-8.42003-9>. Available from: <https://www.sciencedirect.com/science/article/pii/B9780080970868420039>.
- Kokolakis G. Bayesian statistical analysis. In: Peterson P, Baker E, McGaw B, eds. *International encyclopedia of education*. 3rd ed. Oxford: Elsevier; 2010:37–45. <https://doi.org/10.1016/B978-0-08-044894-7.01308-7>. Available from: <https://www.sciencedirect.com/science/article/pii/B9780080448947013087>.
- Louzada F, Nascimento DC, Egbon OA. Spatial statistical models: an overview under the Bayesian approach. *Axioms*. 2021;10(4):307. <https://doi.org/10.3390/axioms10040307>.
- Indu PS, Anish TS, Chintha S, et al. The burden of dengue and force of infection among children in Kerala, India; seroprevalence estimates from Government of Kerala-WHO Dengue study. *Lancet Reg Health Southeast Asia*. 2023;22:100337. <https://doi.org/10.1016/j.lansea.2023.100337>.
- Schimmel J. *Writing science: how to write papers that get cited and proposals that get funded*. USA: OUP; 2012.
- White EB, Strunk W. *The elements of style*. Open Road Media; 2023.
- Tandon A, Roder-DeWan S, Chopra M, et al. Adverse birth outcomes among women with 'low-risk' pregnancies in India: findings from the Fifth National Family Health Survey, 2019–21. *Lancet Reg Health Southeast Asia*. 2023;15:100253. <https://doi.org/10.1016/j.lansea.2023.100253>.
- Karuniawati A, Pasaribu AP, Lazarus G, et al. Characteristics and clinical outcomes of patients with pre-delta, delta and omicron SARS-CoV-2 infection in Indonesia (2020–2023): a multicentre prospective cohort study. *Lancet Reg Health Southeast Asia*. 2024;22:100348. <https://doi.org/10.1016/j.lansea.2023.100348>.



- 40 Thornton HV, Cornish RP, Lawlor DA. Non-linear associations of maternal pre-pregnancy body mass index with risk of stillbirth, infant, and neonatal mortality in over 28 million births in the USA: a retrospective cohort study. *EClinicalMedicine*. 2023;66:102351. <https://doi.org/10.1016/j.eclinm.2023.102351>.
- 41 George PE, Thakkar N, Yasobant S, Saxena D, Shah J. Impact of ambient air pollution and socio-environmental factors on the health of children younger than 5 years in India: a population-based analysis. *Lancet Reg Health Southeast Asia*. 2024;20:100328. <https://doi.org/10.1016/j.lansea.2023.100328>.
- 42 Chaudhary M, Sharma P. Abdominal obesity in India: analysis of the national family health survey-5 (2019–2021) data. *Lancet Reg Health Southeast Asia*. 2023;14:100208. <https://doi.org/10.1016/j.lansea.2023.100208>.
- 43 Ghosh RE, Freni-Sterrantino A, Douglas P, et al. Fetal growth, stillbirth, infant mortality and other birth outcomes near UK municipal waste incinerators; retrospective population based cohort and case-control study. *Environ Int*. 2019;122:151–158. <https://doi.org/10.1016/j.envint.2018.10.060>.
- 44 Freni-Sterrantino A, Ghosh R, Fecht D, et al. Bayesian spatial modelling for quasiexperimental designs: an interrupted time series study of the opening of Municipal Waste Incinerators in relation to infant mortality and sex ratio. *Environ Int*. 2019;128:109–115. <https://doi.org/10.1016/j.envint.2019.04.009>.
- 45 Elston DM. Cherry picking, HARKing, and P-hacking. *J Am Acad Dermatol*. 2021. <https://doi.org/10.1016/j.jaad.2021.06.844>.
- 46 Pearl J, Glymour M, Jewell NP. *Causal inference in statistics: a primer*. John Wiley & Sons; 2016.
- 47 Hernán MA, Alonso A, Logan R, et al. Observational studies analyzed like randomized experiments: an application to postmenopausal hormone therapy and coronary heart disease. *Epidemiology*. 2008;19(6):766–779. <https://doi.org/10.1097/ede.0b013e3181875e61>.
- 48 Dominici F, Zigler C. Best practices for gauging evidence of causality in air pollution epidemiology. *Am J Epidemiol*. 2017;186(12):1303–1309. <https://doi.org/10.1093/aje/kwx307>.
- 49 Vandembroucke JP, Broadbent A, Pearce N. Causality and causal inference in epidemiology: the need for a pluralistic approach. *Int J Epidemiol*. 2016;45(6):1776–1786. <https://doi.org/10.1093/ije/dyv341>.
- 50 Altman N, Krzywinski M. Points of significance: association, correlation and causation. *Nat Methods*. 2015;12(10):899–900. <https://doi.org/10.1038/nmeth.3587>.
- 51 Krishnan A, Asadullah M, Kumar R, Amarchand R, Bhatia R, Roy A. Prevalence and determinants of delays in care among premature deaths due to acute cardiac conditions and stroke in residents of a district in India. *Lancet Reg Health Southeast Asia*. 2023. <https://doi.org/10.1016/j.lansea.2023.100222>.
- 52 Ansari N, Kabir F, Khan W, et al. Environmental surveillance for COVID-19 using SARSCoV-2 RNA concentration in wastewater—a study in District East, Karachi, Pakistan. *Lancet Reg Health Southeast Asia*. 2024;20:100299. <https://doi.org/10.1016/j.lansea.2023.100299>.
- 53 Grant MD. This is the first study. *Lancet*. 2004;364(9440):1126. [https://doi.org/10.1016/s0140-6736\(04\)17096-7](https://doi.org/10.1016/s0140-6736(04)17096-7).
- 54 Mansournia MA, Nazemipour M. Recommendations for accurate reporting in medical research statistics. *Lancet*. 2024;403(10427):611–612. [https://doi.org/10.1016/s0140-6736\(24\)00139-9](https://doi.org/10.1016/s0140-6736(24)00139-9).