Commentary

Discrepancy between statistical analysis method and study design in medical research: Examples, implications, and potential solutions

John Mbotwa[1], Isaac Singini[2,3], Mavuto Mukaka[4,5]

1. Malawi University of Science and Technology, Thyolo, Malawi
2. Department of Statistical Sciences, University of Cape Town, Cape Town, South Africa
3. Malawi College of Medicine–Johns Hopkins University Research Project, College of Medicine, University of Malawi, Blantyre, Malawi
4. Mahidol–Oxford Research Unit (MORU), Bangkok, Thailand
5. Oxford Centre for Tropical Medicine and Global Health, University of Oxford, Oxford, United Kindgom

Correspondence: Dr Mavuto Mukaka (mmukaka@gmail.com)

Introduction

Medical research is the systematic, rigorous investigation of health-related problems in order to generate new knowledge or confirm existing knowledge, with the potential benefit of evidence-based medical practice and policy guidance.

The validity of the findings from medical research requires a thorough process from design, to data collection and data analysis.[1] However, the methods that researchers use during analyses are often unsuitable for the designs used during study conduct. Researchers are supposed to choose the study designs at the time of protocol development, before any investigation is carried out. Different study designs have different strengths and weaknesses.[2] The selected design should be the most appropriate design to answer the objectives of a study. This decision is crucial, as study design reflects directly on the hypothesis of interest. Sample size and other design aspects of the study are aimed at achieving valid conclusions of a trial or study.[1,3] It is therefore important to strive for compatibility between study design and analysis plan.

Many authors have reported on common discrepancies in medical research, specifically between analysis methods and study design.[4,5] For instance, after reviewing several published studies, Varnell et al. observed that many studies had applied inappropriate analysis methods.[5] Despite this review, such mismatched research reporting remained common about 7 years later when Suresh and colleagues also established that inappropriate analysis methods were commonplace in the scientific literature and highlighted the importance of ensuring that studies are adequately powered.[6] In their findings, common discrepancies included incorrect sample size estimation methods, inappropriate analysis methods, as well as ignoring clustering in cluster randomised trials (CRTs).

Calculating the sample size for a study depends on the design of the study. A study whose sample is too small may not be able to answer the research questions posed due to lack of power. In case of a randomised controlled trial, a small unpowered sample would be unethical, as it would be subjecting participants to an inferior treatment when the results will not be generalisable in the end anyway. On the other hand, it would be a waste of resources to use a sample that is too large. In addition, this may be unethical again in the context of a randomised controlled trial. More people than needed would be subjected to an inferior treatment or

to a treatment with many adverse events. As such, proper procedures must be carried out in order to come up with an appropriate sample size for a study.

Previous work by Mukaka and Moulton emphasised the need for consistency between design methods and analysis methods, and discussed how contrasting conclusions can be arrived at depending on the analysis method of choice.[7] Choice of an inappropriate analysis method will put the validity of research results into question. Such findings would be misleading and may not be accepted in the scientific community.

Chan et al. also described some of the most common discrepancies in medical research.[3] They emphasised the need to have a well-documented protocol long before the commencement of a trial and further advised researchers to avoid deviating from the protocol. The discrepancies that were cited are still generally common in contemporary medical research. These inconsistencies include failure to state, in a publication, whether interim analyses were performed or not; this has a direct impact on the interpretation of the final P-values, taking the probability of type I error into account. Methods of handling missing data is another problem that is commonly inadequately addressed in publications, and sometimes no attempt is made to look at the impact of missing data. These discrepancies sometimes lead to false findings.[4] The main reasons for false findings are failure to address issues of bias and how missing data are handled, as well as use of underpowered studies.[3,4,8]

Understanding the implications of the discrepancies between analysis method and study design, and how such discrepancies can potentially be avoided is therefore of great importance for principled research practice. The rationale of this short article is to discuss the implications of analyses that are inconsistent with the study design, giving some examples and then highlighting potential strategies to avoid design–analysis mismatch in research.

Examples of inconsistency between design and analysis, and the implications

The shortfall among researchers to match study design with appropriate analysis methods is caused by, among others things, inadequate training, negligence, and lack of coordination between coinvestigators.

When choosing an analysis method, investigators should ensure that they understand the main hypothesis that has been specified in the protocol. This will allow the

investigators to reflect on the most appropriate method of analysis to address the research objectives. It is also important to check the nature of data that will be collected, that is, whether the data will be collected at 1 time point or at several time points from the same subjects. Knowledge of the data to be collected informs the researcher of the most relevant approaches to data analysis. For example, suppose a researcher designs a randomised controlled trial with the aim of comparing the efficacy of 2 interventions. Study participants are randomly allocated into different study groups. If these participants are followed up for a particular period of time and multiple measurements are collected from each participant at predefined time points, then this leads to some form of statistical dependence or clustering. If all of the data points are to be used in data analysis, statistical methods that take clustering into account will often yield substantially different results from the methods that do not take the dependence of observations into account. In this case, taking clustering into account is the most appropriate approach, which means that publications that report the results from independent samples data analyses will be providing incorrect findings. Unfortunately, journal audiences often do not read the statistical methods and go straight to the conclusions section. This leads to incorrect messages being disseminated to the general public, leading to incorrect new beliefs. We emphasise the importance of choosing an appropriate statistical analysis method in order to arrive at valid conclusions.

As a second example, suppose a researcher designs a randomised controlled trial with the aim of comparing the efficacy of trial drugs. In this case, subjects are randomly allocated to 1 of the groups in order to receive 1 of the drugs under study. In such studies, participants are usually subject to high dropout rates and noncompliance issues, especially when the follow-up period is long. In malaria therapeutic efficacy studies, the length of follow-up may be 28 days or 42 days or 63 days. In other RCTs, the length of follow-up may well be a number of years. During the analyses of such data, a researcher might decide to exclude all participants with noncompliance and dropout issues, thereby reducing the original sample size and hence affecting the statistical power. Although it is still common practice for researchers to perform complete-case analysis (removing records with missing observations), the findings may no longer be representative of the population from which the study sample came from. In fact, such practice, though common, is against the Consolidated Standards of Reporting Trials (CONSORT) guidelines.[8] Findings from such studies may not answer the intended research questions. This kind of sloppiness in analysis can be easily avoided by strategising the analysis approach during protocol development. In line with this study design, it is recommended that the main strategy for analysis of RCTs is to use an intention to treat (ITT) approach, which emphasises that all randomised subjects should be included in the analysis, irrespective of whatever happens to them after randomisation.[9] The advantage of ITT is that not only will the statistical power of the study be maintained, but the principle of randomisation will also remain intact and consequently the treatment groups would be comparable with respect to baseline characteristics.[10] On the other hand, if the study was observational, ITT may not have a place and interpretation of the findings is not necessarily the same as those of a randomised trial. This again demonstrates the importance of paying attention to the study design when considering the analysis method.

Another common example is where a study is initally designed as a superiority trial only to be subsequently analysed as a noninferiority trial. Such practice is discouraged because the original design may not have enough power to answer a noninferiority research question.[11] Most likely, the choice of the post hoc noninferiority margin would be subjective, which can affect the validity of the research findings. In other cases crossover designs are analysed using methods applicable to parallel designs. The interpretation of such findings would be misleading because the design and the analysis approach do not match. Again, the findings may not answer the objectives of the study. In a related scenario, cross-sectional studies are sometimes analysed using risk ratios. The interpretation of the risk ratios in such scenarios are questionable, as it is often difficult to establish whether the exposure of interest preceded the outcome of interest.

Potential solutions

In order to curb the discrepancies between study designs and methods of analysis, a study protocol needs to be developed and effort should be made to have it published to enable peer review in good time before data analysis. The study protocol should clearly describe the study hypotheses, rationale for the study, methods for sample size calculation and data analysis, anticipated multiple testing, and planned interim analyses. Once a study protocol has been developed and approved by ethical review committees (and published if possible), a statistical analysis plan (SAP) should be developed. SAPs are documents that describe the planned analysis of a study, stating the methods of data analysis that are consistent with the design, as well as the software that will be used for specific analyses. It is also encouraged that SAPs contain dummy tables for results summaries. Peace et al. previously discussed some important aspects of SAPs.[12] Unfortunately, many researchers do not develop SAPs. This is likely because of a lack of awareness of SAPs or a lack of knowledge on how to prepare one. In some cases, SAPs are prepared without reference to the study protocol, either because the protocol cannot be traced or because the protocol does not exist (especially in the case of secondary data). SAPs that are not based on the study protocol may not be efficient. International Conference on Harmonisation (ICH), US Food and Drug Administration (FDA), and World Health Organization (WHO) guidelines emphasise the use of SAPs.[13] Even in cases where investigators understand the importance of developing an SAP, the timing is not paid attention to. Many researchers who make an effort to develop an SAP often develop it as the time to conduct analysis is approaching. In such cases, the utility of the SAP may be compromised. In such delayed SAPs, researchers often focus their discussion on the assumed correct methods in the analysis plan without emphasising the need for the design to match the method of analysis. This frequently leads to mismatched designs and analyses. We recommend that all properly designed studies have a protocol. We would also recommend that, as much as possible, researchers should make a statistical analysis plan and that the analysis plan should be developed as soon as the final version of the protocol is completed. Methods of analysis should be consistent with the design as reflected in the protocol. We urge researchers to follow conventional reporting guidelines, such as the CONSORT guidelines and STROBE (STrengthening the Reporting of OBservational studies in Epidemiology)[14] guidelines, as well

as other relevant reporting guidelines. If analyses will be done by someone who did not take part in the design of the study, the study protocol and the SAP should accompany the data. We advise that even if researchers know how to perform statistical analyses on their own, it is still good practice to consult a statistician to verify that analyses have been done correctly. We also recommend that a statistician be consulted to look at the statistical aspects of a manuscript before submission to a journal. In that way, any discrepancy between the study design and method of analysis may be avoided, and additional improvements can be made. Of course, consulting a statistician is easier said than done if researchers are not acquainted with medical statisticians who are willing and able to provide such expertise. On that note, we would highly encourage researchers to regularly include budget lines for (medical) statistical support in their grant applications. It should be noted that many statisticians are more than satisfied with the non-financial reward of being included in the author lists of work for which they contribute statistical guidance; this may be even more welcome than remuneration, as it allows statisticians to work on trials in their regular work time on the grounds that the work will contribute to their academic curriculum vitae and hence their career development.

We also note that peer reviewing of SAPs is often not considered. In this regard, we would encourage that SAPs be peer reviewed. This may be achieved by sending an SAP to statistician colleagues or by including an SAP with a protocol when submitting a protocol for peer review and publication. Among other journals, *Trials* publishes protocols and may be a useful channel for having your SAP reviewed alongside a protocol.

Conclusions

We emphasise that methods of analysis should correspond with study design. There is need for a well-planned and an early-timed statistical analysis plan, which can help address the issue of discrepancy between study design and the analysis method. We also encourage that researchers include a statistician as a collaborator or a coinvestigator in research studies. In that way, they become part and parcel of the study from the design stage, to data analysis and manuscript development.

Competing interests

The authors declare that they have no conflicts of interest.

References

1. Peace, K. E., Parrillo, A. V. & Hardy, C. J. Assessing the Validity of Statistical Inferences in Public Health Research: An Evidence-Based, 'Best Practices' Approach i. J. Ga. Public Health Assoc. 1, (2008).

2. Thadhani, R. & Tonelli, M. Cohort Studies : Marching Forward. 1117–1123 (2006). doi:10.2215/CJN.00080106

3. Chan, A.-W., Hróbjartsson, A., Jørgensen, K. J., Gøtzsche, P. C. & Altman, D. G. Discrepancies in sample size calculations and data analyses reported in randomised trials: comparison of publications with protocols. British Med. J. (2008). doi:10.1136/bmj.a2299

4. Ioannidis, J. P. A. Why Most Published Research Findings Are False. 2, (2005).

5. Varnell S, Murray D, Janega J, and B. J. Design and Analysis of Group-Randomized Trials: A Review of Recent Practices. Am. J. Public Health 94, 393–399 (2004).

6. Suresh K, Thomas SV, S. G. Design, data analysis and sampling techniques for clinical research. Ann. Indian Acad. Neurol. 287–290 (2011).

7. Mukaka, M. & Moulton, L. H. Comparison of empirical study power in sample size calculation approaches for cluster randomized trials with varying cluster sizes – a continuous outcome endpoint. J. Open Access Med. Stat. 1–7 (2016).

8. Schulz, K. F., Altman, D. G., Moher, D. & Group, C. CONSORT 2010 Statement : updated guidelines for reporting parallel group randomised trials. BioMed Cent. (2010).

9. SK, G. Intention-to-treat concept: A review. Perspect. Clin. Res. 2, (2011).

10. Armijo-olivo, S., Armijo-olivo, S., Warren, S. & Magee, D. Intention to treat analysis, compliance, dropouts and how to deal with missing data in clinical research: A review. (2009). doi:10.1179/174328809X405928

11. Piaggio, G., Elbourne, D. R., Pocock, S. J., Evans, S. J. W. & Altman, D. G. Reporting of Noninferiority and Equivalence Randomized Trials Extension of the CONSORT 2010 Statement. (2012).

12. Peace, K. E., Parrillo, A. V & Hardy, C. J. Assessing the Validity of Statistical Inferences in Public Health Research: An Evidence-Based, 'Best Practices' Approach i. J. Ga. Public Health Assoc. 1, (2008).

13. ICH Harmonised Tripartite Guideline. Ich harmonised tripatite guideline: Statistical principles for clinical trials E9. Int. Conf. Harmon. Techinical Requir. Regist. Pharm. Hum. Use (1998).

14. Elm, E. V. et al. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) Statement : Guidelines for reporting observational studies *. Int. J. Surg. 12, 1495–1499 (2014).