Editorial

Reporting Standards for Clinical and Translational Research

Fengyu Zhang and Claude Hughes[1]

Transparency in reporting the results of clinical and preclinical research is critical for unbiased publications. Funding agencies, publishers, and regulators have the responsibility to advocate and implement reporting standards for rigorous design. While individual study protocols may have included these standards, the items reported in the respective publications have often been inconsistent or lack transparency. This editorial intends to provide some specific guidelines for reporting results of clinical research with standards required for a rigorous study design. We recommend that reporting clinical research should include sufficient information on study design and analysis plan that contains data processing, quality assurance, and appropriate methods used for rigorous statistical analysis or modeling. Any discrepancy between publications and original study design should be disclosed and discussed. Additionally, recent advances in the analysis of outcome with repeated measurements and statistical modeling should be employed to obtain an unbiased estimates. Finally, we briefly discuss some issues reporting real-world evidence in clinical research.

Keywords

Reporting standards, clinical research, reporting transparency

Research publications on clinical research are the primary essential vehicle to disseminate knowledge that drives further research and have real-world impacts on patients, clinicians, regulators, and policymakers. Selective reporting of the results of clinical research has been noted (1, 2) within academia, funding and regulatory agencies, and biopharmaceutical institutions and organizations. Examples include that outcomes have been measured and collected but not reported and that investigators have analyzed the data but only reported positive results (3, 4). Discrepancies between publications and other study documents such as study protocol and statistical analysis plan are not trivial (1) and may distort the results and lead to bias in subsequent meta-analyses. In addition, inconsistent findings in both clinical and pre-clinical research have been noted. While the problems of selective reporting might be reduced by broad encouragement to publish negative results or by data sharing, as a general principle, reporting of consistent results may depend upon the rigor of study design.

A group of leading scientists from the US National Institutes of Health (NIH) and stakeholders have made a call for transparent reporting of preclinical studies and proposed a set of standards for reporting of rigorous design (5). The NIH held a joint workshop with the then Nature Publishing Group and *Science* in Bethesda, MD, and discussed the issues of reproducibility and rigor in research, with editors from more than 30 journals in basic or translational sciences (6, 7). They came to a consensus on a set of principles and guidelines for reporting preclinical research and a considerable number of journal editors have endorsed the principles and guidelines. Soon afterward, these principles and guidelines were adopted and extended by other societies such as the Biophysical Society and the Center for Open Science. Some journals have started to implement the guidelines and require authors to report some specific items when a manuscript is considered for a potential publication. This helps editors assess the study design and transparency in reporting.

In this editorial, we advocate for transparency in reporting clinical research and suggest some specific recommendations for publications. Our primary recommendations focus on rigorous statistical analysis that should reflect the recent advances in statistical methods, then the reporting standards for a rigorous study design in clinical research, and finally, additional measures to assure transparent reporting for publications in clinical research. Some of these recommendations can be adapted for observational studies including population-based health research.

Rigorous statistical analysis

Study design determines what outcomes to collect and what statistical analysis to perform. In most clinical research, biostatisticians are involved and may have a protocol and analysis plan in place before beginning a study. Once data is collected, the data processing and statistical

[1] Correspondence to: F Zhang, zhangfy@gcatresearch.org or C Hughes, cluade.hughes@iqvia.com

analysis may have to be carried out by or under the supervision of biostatisticians. In recent years, there has been discussions regarding inconsistent replication of research findings in terms of p-value (8). Statistical analysis involves a series of steps, in which mishandling occurring at any step will lead to an incorrect p-value (Box 1**)**. Any discrepancy between report and protocol should be noted and discussed as to how the discrepancy affect the study results, and what measures taken to ensure unbiased results including controls for potential confounding.

Box 1. The process of design, measurement, and analysis

1. Design: Study population and sample, clinical diagnosis, inclusion and exclusion criteria, sample size calculation, randomization, blinding;
2. Measurement: Clinical evaluation of primary and secondary outcomes, measurement, quality control, biospecimen collection and testing, and covariates that may require adjustments;
3. Data process: Data entry and management software, variable types (numeric, string, etc.), coding book, data cleaning, detection of outliers or data errors, data recording, summary statistics, cross-tabulation;
4. Analysis: Single variable analysis and statistical modeling: data science to evidence-based analysis;
5. Replication: independent replications are required for a cross-sectional study

Statistical methods selected for data analysis should fit the type of outcome measure. An outcome could be measured in a form of continuous, binary, categorical (or nominal), time or duration to an event. The general linear model including analysis of variance (ANOVA) and analysis of covariance (ANCOVA) or multivariate linear model is the first choice of methods for analyzing continuous variables. Duration data such as time to failure or death, or time to stop using medication is the most common primary outcome in clinical trials of cancer and is often analyzed using the Cox proportional hazard regression. In addition, a binary outcome (e.g., remission or not, hypertension or not) can be analyzed using logistic or Logit regression, or Probit regression if a latent variable (i.e., intermediate endpoint) underlying a dichotomous outcome follows a normal distribution (9). Categorical variables, meaning an outcome measured in more than two categories such as a choice of methods, can be analyzed using multinomial logistic regression (Table 1).

The number of adverse events or episode of relapses throughout a trial may be considered as an outcome in clinical research. This type of data can be treated as ordinal outcomes and analyzed using ordinal Probit or Logit regression model, which usually requires an assumption of proportional hazard between individual categories. While they can be treated as continuous and analyzed using a general linear regression model, a negative predicted value likely occurs; and especially when the count is fewer (e.g., <10), least square regression may produce a bias in the results. Poisson regression and negative binomial regression may provide optimal analysis(10).

So far, most clinical trials have still been using classical statistical methods. However, recent advances in statistical methods have not been well reflected in data analysis. For example, multicenter clinical trials have often been used (11). Patients recruited from multiple centers may be more heterogeneous than those from a single-center study and thus a larger sample size may be required; but in turn, results from a multi-center study may be generalized to a broader "real world" population. Patients within the same center may have some dependence due to sharing diagnosis or treatment under the same physician or subject to a local standard-of-care within the same hospital; and as such, they may share some common unobserved or even unobservable heterogeneity at a level of physician or hospital (12). This dependence violates the underlying assumption of independence among observations for a parametric statistical model; and if not corrected through analysis, it will lead to a underestimating of the standard errors for parameter estimates. However, very few multicenter clinical trials have employed a particular approach to consider the unobserved heterogeneity at higher levels, which consequently, cause false positive findings.

Table 1. Type of outcome measure and methods for analysis

Type of measure	Example	Methods
Continuous	Blood pressure, cognitive score	Linear regression, ANOVA, ANCOVA
Binary	Case vs. control	Logistic or Logit, Probit regression
Categorical	Choice of methods (A, B, C)	Multinomial Logit regression
Ordinal	Number of adverse outcomes	Ordinal Logit or Probit regression
Count data	Number of adverse events, episode of relapse	Poisson or negative binomial regression
Duration data	Time to failure, death, or stop the medication	Cox proportion regression model

In recent years, people have expressed concerns about possible biased results in publications (13), which has provoked a series of debates across multiple disciplines on the reliability of the statistical error, p-value (14). In response to this, the American Statistical Association issued a statement on the interpretation of p-value in 2016. Later, a group of methodologists proposed to lower the routine p-value threshold from 0.05 to 0.005 for

designating statistical significance for a discovery(15). The occurrence of false positives likely arises from data quality and inappropriate data analysis rather than the p-value threshold, which has served as a gold standard for nearly a century. Shrinking the p-value threshold may not help reduce false positive findings. Instead, such a change will increase not only false negatives but also the number of patients required for clinical studies. Therefore, there is an urgent need to call for rigorous statistical analysis and analysts should have a better understanding of the nature and quality of the data, the assumptions for a specific statistical method, and model diagnosis.

Hierarchical statistical modeling (also known as multilevel modeling) has been an active field of statistical and methodological research (16) and population-based research over the past decades (17, 18). This concept can be applied to analyzing all types of outcomes that are collected in multi-center clinical trials. With the development of computing power and computational programming, hierarchical statistical modeling has been implemented in major standard statistical packages such as SAS and R package (12). Nevertheless, performing such model-based analyses and interpretations of results may require a sophisticated statistician.

Standards for reporting transparency

Publishers or funding agencies can play a critical role in reinforcing transparency in reporting of rigorous study designs. A core set of standards for reporting have been proposed for preclinical studies (5). We suggest implementing specific guidelines adapted for clinical research (see Table 2 and below).

Subjects and design. The investigators/authors need to define the study population clearly from which the subjects are sampled or recruited for a clinical study. This will determine the population that results of a sample study can be generalized *ex post facto***.** In general, the study population for a single-center study may be different from that for a multi-center study. Investigators/authors are required to state 1) the criteria for inclusion and exclusion of individual subjects, 2) reliability of criteria for the diagnosis of patients, and for clinical evaluations of the primary and secondary outcomes, and 3) the measurements of primary and secondary outcomes. Any significant discrepancy between the report and the original study protocol should be discussed in term of its influence on potential biases of results. For a secondary analysis or retrospective clinical studies based on medical records, investigators /authors are required to state how subjects are selected for analysis if the study sample is not all possible subjects within the institution(s) during a specified period. Purposive selection of subjects into the analysis will result in biased results.

Randomization and blinding. Randomization, blinding, and placebo-control are vital components to minimize bias in results. The investigators/authors are required to state how the recruited subjects are randomized and with what methods (e.g., simple randomization, block randomization, stratified randomization, covariate adaptive randomization), and if blinding is used and what type, such as single or double blinding, triple blinding if data analysts are also unaware of treatment the patient received.

Table 2. List of the items for reporting in publications

Category	Items
Subjects and design	Criteria for inclusion and exclusion;
	Diagnosis and clinical evaluation to collect the data
	Primary and secondary outcomes or measurement
	Ethics approval and trial registration
	Treatment group
Sample size and power	Methods for determining the sample size
	Post-hoc power analysis if not meeting the required sample size
Randomization and blinding	Randomization and methods;
	State if blinding is used or open-label
	Use of a placebo
Descriptive statistics	The effective sample size for outcomes and covariate
	Central tendency (mean and median)
	Dispersion (standard deviation, min, max, and inter-quartile range)

Sample size estimation and power analysis. A well-designed study should meet the sample size estimated at the design stage. The investigators/authors should state how the sample size was determined and what methods and parameters are used. In most cases, a range of sample sizes need to be provided under variable number of parameters, such as different effect sizes, thresholds for significance level, and levels of statistical power. When a study is completed, the actual sample size may have deviated from the original design (e.g., fewer samples). In such a case, a post hoc power analysis should be performed according to the actual sample size and effect size.

Statistics. Descriptive statistics should first be reported with sufficient details. The needed details usually include actual sample size, effective sample size for each outcome or covariates by treatment group, central tendency and

dispersion. A frequency distribution should be reported for a binary or categorical variable. This is usually reported along with the definition and coding for possible values of each individual variable. Reporting more than one measure of central tendency or dispersion is generally recommended. This will help provide a quick assessment for a reader that the collected data do or do not meet the assumptions for a specific statistical method. For example, in a metabolic syndrome study where multiple measures are collected and analyzed (19), one can tell if an assumption such as normality for a continuous variable are severely violated by merely comparing the mean and median.

Estimation of effect size should be reported with enough details to accord with the type of outcome and statistical methods used. Besides p-value, estimates of coefficients or least square mean, with their standard errors or 95% confidence limits, should be reported for a continuous variable; whereas odds ratio (OR) or hazard ratio (HR), or relative risk (RR) should be reported, with their 95% confidence limits. In addition, it is important to note that if authors choose to report statistics in graph or plot, the related statistics should also be included. This will be helpful for any future use in meta-analyses. The plot is more intuitive for presentation but does not wholly replace the role of statistical details in scientific publications.

Analysis of outcome with repeated measurement

With regard to the analysis of an outcome with repeated measurements, some care is needed in reporting of results. In clinical trials, time to the occurrence of certain events such as time to relapse (20), treatment failure (21), or discontinuation of medication(11,18), are often considered as a primary outcome, particularly in cancer research. However, continuous outcomes are collected at multiple time points during a period of clinical trials, for example, psychopathological symptoms and cognition in psychiatric research (22, 23). Repeated measures by design allow investigators to examine the timing and trajectory of a treatment effect. It is a powerful approach, but also creates some concerns about consistency, such as when and where data are collected and performed the clinical evaluation to collect study data.

Missing values due to dropout is a common problem in an analysis of outcome with repeated measurement. In a study with scheduled follow-up at multiple time points, subjects may have missing values due to skipping one of the scheduled follow-up visits or subjects may be lost to follow-up before the end of a study. A standard method to deal with missing values due to dropout is the Last-Observation-Carried-Forward (LOCF). LOCF assumes that the measurement of an outcome variable at one follow-up time can be replicated as the presumed observed value at later missing time points. This causes a biased estimate of treatment effect and reduced estimates of standard errors due to the increase in the number of constant observations within an individual (24). The US National Academy of Science has made a recommendation to the Food Drug Administration (FDA) against the use of LOCF in clinical trials and recommends an alternative approach such as generalized estimating equation (GEE) to deal with the missing value for repeated measurement (25). In addition, the random-effect model has recently gained support to be used as a primary methodology for analysis of outcome with repeated measurement. However, this type of modeling may require additional care in order to produce appropriate estimates.

Statistical modeling

Randomization is used to reduce potential confounding that is caused by the unequal distribution of an independent variable across treatment groups. In theory, a randomized trial should not have a significant difference in independent variables at baseline. However, due to the occurrence of an adverse event, a tolerability problem, ineffectiveness of treatment in some individuals, or compliance issues, treatment groups may consist of subgroups of patients who are unequally distributed and have divergence in some of their key characteristics. Multiple regression models are recommended for validating final estimates of parameters. If a treatment effect and a covariate are both significant, then the potential interaction between the treatment and covariates should be assessed. Finally, a model diagnosis should be carried out to make sure that a model is well fitted. With regression modeling, one can also perform stratification or sensitivity analysis to examine the internal validity of estimates across centers.

Real-world evidence in clinical research

While efficacy and safety from clinical trials have been acepted as standard evidence for approval of a drug for marketing, post-approval studies are commonly required. The core concern is usually that the drug approval was based on a few studies in well-characterized samples of recruited patients and that such study samples are often subject to some degree of recruiting bias, and thus may not represent the broader populations where the drug will be marketed or used. Real-world evidence (RWE) refers to "the output of real-world data (RWD) analysis that is used to generate insights, using appropriate study design and scientific methods, to inform decision-making by health care stakeholders."(26). Due to the wide-spread use of mobile devices and electronic medical or health records, collecting all kinds of data about patients from the real world is feasible, and analysis of RWD is anticipated to help satisfy post-approval study needs in many instances.

RWD may help generate evidence for a new indication for a previously approved drug. In the original clinical trials that generated the evidence for drug approval, potential interaction of the drug with other factors may not have been assessed due to the lack of prior evidence to collect those data, or data such as adverse events have been collected but lacked the power to generate statistically meaningful evidence. One of the most promising aspects of RWD is in its large sample sizes and rich variables, which could allow detection of high-order interactions and generate new hypotheses for further studies or approval of a drug for a new indication. For example, specific drug-food interactions that may have been known at the time of conduct of pivotal clinical trials(27) but might have been

excluded in the primary study protocols. Note that the FDA currently allows the use of RWE for monitoring drug safety and for drug approval for rare diseases (28). Analysis of RWD may require additional techniques such as machining learning and advanced data analyses for high dimensional data. For example, variables in RWD are not equally important; some of them may be more informative than the others. In addition, with highly diverse data, application of machine-learning techniques could lead to the selection of an extreme min or max case, which might, in turn, be difficult to replicate consistently or be fundamentally misleading. Therefore, findings generated from RWD may have to be subsequently validated by conducting additional rigorously designed prospective studies.

CONFLICT OF INTERESTS

The authors declare that there are no conflicts of interest regarding the publication of this paper.

REFERENCES

1. Dwan K, Altman DG, Clarke M, Gamble C, Higgins JP, Sterne JA, et al. Evidence for the selective reporting of analyses and discrepancies in clinical trials: a systematic review of cohort studies of clinical trials. PLoS Med. 2014;11(6):e1001666.
2. Al-Marzouki S, Roberts I, Marshall T, Evans S. The effect of scientific misconduct on the results of clinical trials: a Delphi survey. Contemp Clin Trials. 2005;26(3):331-7.
3. Rankin J, Ross A, Baker J, O'Brien M, Scheckel C, Vassar M. Selective outcome reporting in obesity clinical trials: a cross-sectional review. Clin Obes. 2017;7(4):245-54.
4. Bourgeois FT, Murthy S, Mandl KD. Outcome reporting among drug trials registered in ClinicalTrials.gov. Ann Intern Med. 2010;153(3):158-66.
5. Landis SC, Amara SG, Asadullah K, Austin CP, Blumenstein R, Bradley EW, et al. A call for transparent reporting to optimize the predictive value of preclinical research. Nature. 2012; 490(7419):187-91.
6. Journals unite for reproducibility. Nature. 2014;515(7525):7.
7. McNutt M. Journals unite for reproducibility. Science. 2014;346(6210):679.
8. Leek JT, Peng RD. Statistics: P values are just the tip of the iceberg. Nature. 2015;520(7549):612.
9. MacKinnon DP, Lockwood CM, Brown CH, Wang W, Hoffman JM. The intermediate endpoint effect in logistic and probit regression. Clin Trials. 2007;4(5):499-513.
10. Coxe S, West SG, Aiken LS. The analysis of count data: a gentle introduction to poisson regression and its alternatives. J Pers Assess. 2009;91(2):121-36.
11. Lieberman JA, Stroup TS, McEvoy JP, Swartz MS, Rosenheck RA, Perkins DO, et al. Effectiveness of antipsychotic drugs in patients with chronic schizophrenia. N Engl J Med. 2005; 353(12):1209-23.
12. Austin PC. A Tutorial on Multilevel Survival Analysis: Methods, Models and Applications. Int Stat Rev. 2017; 85(2): 185-203.
13. Ioannidis JP. Why most published research findings are false. PLoS Med. 2005;2(8):e124.
14. Nuzzo R. Scientific method: statistical errors. Nature. 2014; 506(7487):150-2.
15. Ioannidis JPA. The Proposal to Lower P Value Thresholds to .005. JAMA. 2018;319(14):1429-30.
16. Hougaard P. Frailty models for survival data. Lifetime Data Anal. 1995;1(3):255-73.
17. Short S, Zhang F. Use of maternal health services in rural China. Population Studies. 2004;58(1):3-19.
18. Zhang F, Tsui A, Suchindran C. The determinants of contraceptive discontinuation in Northern India: A multilevel analysis of calander data. . The MEASURE Evaluation Working Paper WP-99-15 1999.
19. Wu RR, Zhang FY, Gao KM, Ou JJ, Shao P, Jin H, et al. Metformin treatment of antipsychotic-induced dyslipidemia: an analysis of two randomized, placebo-controlled trials. Mol Psychiatry. 2016;21(11):1537-44.
20. Csernansky JG, Mahmoud R, Brenner R, Risperidone USASG. A comparison of risperidone and haloperidol for the prevention of relapse in patients with schizophrenia. N Engl J Med. 2002;346(1):16-22.
21. Boudreau JE, Giglio F, Gooley TA, Stevenson PA, Le Luduec JB, Shaffer BC, et al. KIR3DL1/ HL A-B Subtypes Govern Acute Myelogenous Leukemia Relapse After Hematopoietic Cell Transplantation. J Clin Oncol. 2017;35(20):2268-78.
22. Apud JA, Zhang F, Decot H, Bigos KL, Weinberger DR. Genetic variation in KCNH2 associated with expression in the brain of a unique hERG isoform modulates treatment response in patients with schizophrenia. Am J Psychiatry. 2012; 169 (7): 725-34.
23. Jenkins A, Apud JA, Zhang F, Decot H, Weinberger DR, Law AJ. Identification of candidate single-nucleotide polymorphisms in NRXN1 related to antipsychotic treatment response in patients with schizophrenia. Neuropsychopharmacology. 2014;39(9):2170-8.
24. Molnar FJ, Hutton B, Fergusson D. Does analysis using "last observation carried forward" introduce bias in dementia research? CMAJ. 2008;179(8):751-3.
25. Panel NRCU. The Prevention and Treatment of Missing Data in Clinical Trials. Washington (DC)2010.
26. Swift B, Jain L, White C, Chandrasekaran V, Bhandari A, Hughes DA, et al. Innovation at the Intersection of Clinical Trials and Real-World Data Science to Advance Patient Care. Clin Transl Sci. 2018;11(5):450-60.
27. Shader RI. Pharmacovigilance and Some Thoughts About What We Eat. Clin Ther. 2018;40(12):1957-61.
28. Jarow JP, LaVange L, Woodcock J. Multidimensional Evidence Generation and FDA Regulatory Decision Making: Defining and Using "Real-World" Data. JAMA. 2017;318(8):703-4.

How to cite this article:

Zhang F and Hughes C. Reporting Standards for Clinical and Translational Research. Glob Clin Transl Res. 2019; 1(2): 69-73. DOI:10.36316/gcatr.01.0010.