CLINICAL INVESTIGATION

EDITORIAL

Clin. Invest. (2012) 2(5), 447–449

Over-reporting of p-values in the medical literature should be discouraged

Eric P Brass

"The proper use of p-values may be well understood in principle, but the cited examples and many others in recent publications suggest that misuse persists in practice. All editors should take steps to discourage the over-reporting of p-values in the literature."

The reporting of clinical trial data increasingly includes p-values that are not based on prospective hypothesis testing. These p-values often add little to the utility of the data presentation and have the potential to facilitate misinterpretation by the reader. This trend and the associated potential problems are illustrated in recent papers published in the literature. This over-reporting of p-values should be discouraged.

Rigorous statistical methodologies have provided critical tools for designing, analyzing and interpreting clinical research. Effective communication of clinical trial results, particularly to the non-statistician, is an important responsibility of clinical investigators. The calculation of p-values for predefined end points using prespecified methodologies greatly enhances the utility of trial results for the reader. However, there has been a substantial increase in the reporting of p-values for comparisons that lack the rigor and robustness required for hypothesis testing. As highlighted by Fleming, the reporting of these p-values adds little information to the data presentation and risks misleading the reader with respect to the proper interpretation of the data [1]. Recent publications from several major journals illustrate this problem.

Ensuring the comparability between study arms in a clinical trial is important. Thus, Lee *et al.* list baseline characteristics of the study groups in Table 1 of their paper on statin effects in patients with acute myocardial infarctions [2]. This table appropriately provides values and indices of variability for each cohort. The table also provides p-values for each characteristic – a total of 33 p-values in this one table! How is the reader to interpret these values which range from <0.001 to 0.995? Is the p-value more informative about potential imbalances then the values themselves which are so clearly provided? No adjustment was made to these p-values for multiple comparisons, and thus the reader who focuses on an individual low p-value for a characteristic of interest may attach 'significance' to what could be a chance finding. If clinical features have the potential to impact the effect of an intervention there are robust strategies for dealing with this concern. In this case, Lee *et al.* used a propensity score analysis to adjust for potential confounders [2]. The p-values in Table 1 of their paper thus cannot add to the reader's interpretation of the study and may detract form the more formal propensity score analysis. Similarly, Ogedegbe *et al.* report 20 p-values in the table of baseline characteristics in their interesting paper on interventions to improve adherence in patients with hypertension, thus risking inappropriate inference by the reader [3].

Redon *et al.* report on patient characteristics based on outcomes and

Department of Medicine, Harbor-UCLA Medical Center, 1124 W Carson Street, Torrance, CA, USA
E-mail: ebrass@ucla.edu

Keywords: clinical trial reporting • p-values • statistics

subgroups in Table 3 of their interesting paper on blood pressure effects in high-risk diabetic patients [4]. This table contains 74 p-values without adjustment. The 'Methods' section indicates that a $p < 0.05$ should be considered significant for these analyses. The typical reader might be quite challenged to try and understand the relevance of these p-values and may naively over-interpret their importance in understanding the dataset, to the detriment of the impact of the dataset itself. Vakil *et al.* in their study of arbaclofen placarbil's effect on gastroesophogeal reflux disease symptoms forthrightly state that the statistical analysis of secondary end points "should be considered as exploratory" and that "nominal p-values are presented" as no correction was done for multiple comparisons [5]. Nonetheless, Table 2 of their paper includes 48 p-values. The busy reader may review the results in this table without appreciating the disclaimer related to the p-values presented. Zesiewiez and colleagues also clearly state that the 15 p-values in their Table 2 includes no correction for multiple comparisons, highlighting the study's small sample size as a rationale [6]. However, this caveat may be lost on a reader viewing the table, and the sample size could easily provide a rationale for omitting the p-values to avoid over-interpretation of the small study.

"The proliferation of p-values in publications diminishes the impact of those p-values that are the result of prespecified, hypothesis-testing analyses..."

Few studies are adequately powered to make meaningful statistical inferences about adverse event rates, yet the reporting of these data is of obvious import. Suh *et al.* reported on safety end points in Table 6 of their paper on anticoagulation after percutaneuous coronary interventions [7]. This included an analysis of bleeding events, of which there were three events in one arm and one event in the other arm. The table reports these values and provides a p-value (0.511), as it does for other assessments in the table. This p-value may be read as implying that a meaningful conclusion can be drawn, and the text reinforces this by stating "bleeding complications did not differ". A more appropriate conclusion is that no inferences can be made based on the small number of events observed, and that clinically meaningful differences could not be excluded. Similarly, Diletti *et al.* report on adverse events in Table 3 of their study of bioresorbable everolimus-eluting vascular scaffolding placement. The 'Methods' section states explicitly that the p-values were calculated for 'descriptive purposes' [8]. However, reporting a p-value of 0.5645 when listing two myocardial infarctions in one arm and one myocardial infarction in the other arm adds little useful description and again risks a false-confidence in the meaning of the observations. Importantly, approaches for analysis of critical safety data have been suggested [9,10], which provides a more meaningful context than possible with simple group comparisons and p-value reporting.

"Just because a p-value can be calculated does not mean it should be calculated, and if a p-value is calculated does not mean the p-value should be reported."

The problem of p-value over-reporting is not unique to the journals or papers cited. The papers referred to were selected in part because of the high quality of the journals in which they appeared, the study designs, the data reported and their potential impact. Each table cited would be clear in the absence of the reported p-values, and the p-values provided no incremental clarity while risking misinterpretation. Just because a p-value can be calculated does not mean it should be calculated, and if a p-value is calculated does not mean the p-value should be reported. Simple reporting of the point estimate and a measure of precision such as a confidence interval will provide the critical information with less risk of misinterpretation. Yet, personal experience suggests that editors contribute to the misuse of p-values (see Table VI in Brass *et al.* [11]).

The proliferation of p-values in publications diminishes the impact of those p-values that are the result of prespecified, hypothesis-testing analyses and increases the likelihood that the reader will make inappropriate inferences based on incidentally reported p-values. The proper use of p-values may be well understood in principle, but the cited examples and many others in recent publications suggest that misuse persists in practice. All editors should take steps to discourage the over-reporting of p-values in the literature.

Acknowledgements

The author thanks T Fleming for his suggestions on the manuscript.

Financial & competing interests disclosure

E Brass is a consultant to the pharmaceutical industry on issues related to clinical trial design and interpretation. The author

has no other relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript apart from those disclosed.

No writing assistance was utilized in the production of this manuscript.

References

1 Fleming TR. Clinical trials: discerning hype from substance. *Ann. Intern. Med.* 153(6), 400–406 (2010).

2 Lee KH, Jeong MH, Kim HM *et al.* Benefit of early statin therapy in patients with acute myocardial infarction who have extremely low low-density lipoprotein cholesterol. *J. Am. Coll. Cardiol.* 58(16), 1664–1671 (2011)

3 Ogedegbe GO, Boutin-Foster C, Wells MT *et al.* A randomized controlled trial of positive-affect intervention and medication adherence in hypertensive african americans. *Arch. Intern. Med.* 172(4), 322–326 (2012).

4 Redon J, Mancia G, Sleight P *et al.* Safety and efficacy of low blood pressures among patients with diabetes subgroup analyses from the ONTARGET (Ongoing Telmisartan Alone and in combination with Ramipril Global Endpoint Trial). *J. Am. Coll. Cardiol.* 59(1), 74–83 (2012)

5 Vakil NB, Huff FJ, Bian A, Jones DS, Stamler D. Arbaclofen placarbil in GERD: a randomized, double-blind, placebo-controlled study. *Am. J. Gastroenterol.* 106(8), 1427–1438 (2011).

6 Zesiewicz TA, Greenstein PE, Sullivan KL *et al.* A randomized trial of varenicline (Chantix) for the treatment of spinocerebellar ataxia type 3. *Neurology* 78(8), 545–550 (2012).

7 Suh JW, Lee SP, Park KW *et al.* Multicenter randomized trial evaluating the efficacy of cilostazol on ischemic vascular complications after drug-eluting stent implantation for coronary heart disease: results of the CILON-T (Influence of Cilostazol-based Triple Antiplatelet Therapy on Ischemic Complication after Drug-Eluting Stent Implantation) trial. *J. Am. Coll. Cardiol.* 57(3), 280–289 (2011).

8 Diletti R, Onuma Y, Farooq V *et al.* 6-month clinical outcomes following implantation of the bioresorbable everolimus-eluting vascular scaffold in vessels smaller or larger than 2.5 mm. *J. Am. Coll. Cardiol.* 58(3), 258–264 (2011).

9 Brass EP, Lewis RJ, Lipicky R, Murphy J, Hiatt WR. Risk assessment in drug development for symptomatic indications: a framework for the prospective exclusion of unacceptable cardiovascular risk. *Clin. Pharmacol. Ther.* 79(3), 165–172 (2006).

10 Fleming TR. Identifying and addressing safety signals in clinical trials. *N. Engl. J. Med.* 359(13), 1400–1402 (2008).

11 Brass EP, Cooper LT, Morgan RE, Hiatt WR. A Phase II dose-ranging study of the phosphodiesterase inhibitor K-134 in patients with peripheral artery disease and claudication. *J. Vasc. Surg.* 55(2), 381–389 (2012).