The Importance of Matching Language to Type of Evidence: Avoiding the Pitfalls of Reporting Outcomes Data

Payal Kohli, MD and Christopher P. Cannon, MD

Cardiology Division (Kohli), University of California San Francisco, San Francisco, California;
Cardiovascular Division (Cannon), Brigham and Women's Hospital, Boston, Massachusetts

ABSTRACT

Results from different types of clinical research studies provide different types of evidence for evaluating the effects of a new drug or intervention. For this reason, it is important to recognize this phenomenon during reporting and to choose appropriate language to match the type of study that was done, because this can become critical to the interpretation and application of the results in clinical practice. In this article, we aim to highlight this issue through a series of examples and provide some guidance on what the appropriate language for different types of studies should be.

Background on Types of Studies

When investigating the effects of a novel intervention on a patient population, a spectrum of different study designs exists and can be chosen by the investigator, including but not limited to a (1) purely retrospective case study; (2) retrospective case-control study; (3) prospective cohort study (eg, a registry); and (4) prospective, placebo-controlled, double-blind, randomized, controlled trial. Each of these study designs and analytic techniques has advantages and disadvantages, and the choice of study type can be a complex function of many considerations, such as the scientific hypothesis of interest, patient population, ethical implications, and resources available. These different studies, therefore, are valuable and contribute important results to clinical practice.

The Problem: Matching Reporting Language to Type of Study

Because they are all very different, the same language cannot be used to describe the results from distinct types

of studies when drawing conclusions and characterizing the risk relationship between the intervention and the outcome.

When describing clinical outcomes data, authors and readers should be cognizant of subtle language differences, which can result in differences in interpretation of results and clinical application. Randomized trials, in which patients are randomly assigned to receive 1 treatment or another and the only difference between the 2 groups is the intervention being studied, have the potential to establish causality if they are performed rigorously and have comprehensive follow-up.^{1,2} For example, if patients randomized to receive a new drug have a lower risk of events, and the groups are well-balanced with appropriate trial conduct, then the only explanation to attribute the lower risk to is the randomized therapy. Therefore, randomized controlled trials can potentially establish the cause and effect relationship between the intervention and the results.

Observational studies help to monitor clinical practice, guideline adherence, and safety issues in large numbers of patients. Observational studies can also identify risk relationships between risk factors, interventions and outcomes, and establish new hypotheses for study, but may not be able to establish causality because many variables differ between the groups. In these studies, large cohorts of patients (such as in registry data) are analyzed, dividing patients into those who received the medication of interest and those who do not. They are then followed, and their event rates are compared; conclusions are drawn about the potential effect of the medication based on differences in these events. The 2 groups of patients—treated and

This editorial review was cowritten and developed as an expanded background piece to accompany the statement by the HEART Group journal editors, which is being published simultaneously with this article.

The authors have no funding, financial relationships, or conflicts of interest to disclose.

not—can be different however, and thus multivariate adjustment (eg, for baseline characteristics and differences in other treatments received) is often necessary when comparing these outcomes. If the event rate continues to be lower in the group treated with the medication after multivariate adjustment for baseline differences, then the authors usually conclude that the treatment may be associated with some apparent benefit.

Although all types of outcome studies provide valuable descriptive data, one must be cautious when describing all results from different analytic techniques using the same language. Inherent to observational studies is the concept that, despite multivariate adjustment, many baseline differences cannot be captured and adjusted for, potentially leading to persistent differences between the 2 groups (due to other factors, not the one under study). Thus, one cannot fully attribute the difference in clinical event rates to the intervention of interest, and therefore use of language such as intervention X “reduces clinical events” is not appropriate.

As an illustrative (and somewhat humorous) example, if the number of storks in Scotland and the number of babies born in Scotland both increased by 10% from 2010 to 2011 (an observational study), one could not conclude that the storks “resulted in” the increased number of babies, but instead it is more appropriate to conclude that the number of storks was “correlated with” or “associated with” an increase in the number of babies born.

The problem arises when results from an observational analysis are erroneously described using language that concludes that a treatment provides direct clinical benefit or harm, but these results cannot be validated in a subsequent randomized controlled trial. This creates confusion and misinformation among clinicians, and, in turn, patients.

Language Mismatch Creating Confusion

There are many examples of conflicting conclusions drawn from observational studies vs randomized trials regarding the effect of an intervention as a result of the language used to describe the results. The prototypic example comes from studies of hormone replacement therapy (HRT) on risk of death and cardiovascular (CV) outcomes in postmenopausal women. The relationship between estrogen replacement and CV risk emerged in 1985, with the results of a prospective observational study of 121 964 females followed for 4 years. After multivariable adjustment, the risk of myocardial infarction and fatal coronary disease remained 50% lower for women taking HRT.³ This finding was subsequently duplicated in other registries and cohorts.⁴ In this latter report, the results section read, “We observed a marked decrease in the risk of major coronary heart disease among women who took estrogen with progestin, as compared with the risk among women who did not use hormones.” The conclusion stated, “The addition of progestin does not appear to attenuate the cardioprotective effects of postmenopausal estrogen therapy.”⁴ Following these reports, it then became widely accepted that HRT for postmenopausal women “resulted” in a mortality benefit. However, when the Women’s Health Initiative subsequently

carried out 2 large-scale, randomized, controlled trials of healthy postmenopausal women, randomized to either estrogen plus progestin vs placebo⁵ or estrogen vs placebo⁶ to validate this interpretation, the results were alarming. Both trials were halted early due to an increase in risks associated with HRT. In the combination-therapy study, there was a higher risk of breast cancer, coronary heart disease, strokes, and pulmonary emboli.⁵ Here, the observational data had suggested a causal relationship between estrogen and protection from cardiovascular events, and the randomized controlled trial (RCT) was not able to validate this. As such, the language used in the observational registry report, citing “a marked decrease in the risk” and a “cardioprotective effect”⁴ appears to have overstated the strength of the evidence, which may have led to misinterpretation of the observational data.

Disparities between observational and randomized data are not just limited to reporting of results of drugs. The identical pattern has emerged with respect to devices, reflecting a universal need for attention to this issue when describing the findings of any type of clinical outcome study. For example, in a large national registry in Sweden (Swedish Coronary Angiography and Angioplasty Registry [SCAAR]), researchers reported in 2007 that patients treated with a drug-eluting stent (DES) had an increased relative risk of mortality than those treated with a bare-metal stent (BMS).⁷ In 2009, results from the same registry reported no difference,⁸ and in 2012, investigators reported exactly the opposite results—that those treated with a newer-generation DES had a lower mortality than those treated with a BMS.⁹ Although each of these studies provide incremental value in understanding the risk relationship between DES/BMS and mortality, this type of conflicting evidence can be very confusing if language used to draw causal conclusions about risk (or benefit) is not carefully selected. Indeed, when meta-analysis was conducted using patient-level pooled datasets of all DES vs BMS randomized trials were analyzed, absolutely no difference in mortality emerged.¹⁰ Similarly, although there continues to be some variability in the results, large randomized trials have generally found no difference in mortality between a first generation DES and BMS.^{11,12} Therefore, it is important to be mindful of reporting these results with clear, accurate, and consistent language that reflects the evidence being cited: for the registry data, “associated with” or “relative risk ratio” are appropriate, whereas for the randomized data, “risk reduction” is preferred.

A Caution: Pitfalls of Reporting Different Types of Studies

Observational analyses can provide very valuable information and are an effective means for uncovering epidemiological trends and characterizing potential risk relationships. Therefore, they should continue to be used to answer questions. Similarly, RCTs are valuable for investigating the effect of an intervention within a specific patient population within a closely regulated context. We are suggesting careful attention be paid toward reporting data from such observational studies and not using the same declarative language as one might use in an RCT, because these results can often be conflicting, and these types of trials offer different types of evidence.

Table 1. Suggested Language Based on Study Type

	Randomized Trial	Observational Study
Type of language		
Descriptive statements	“Reduced the risk by”	“A lower risk was observed,” “there is a relationship,” “there is an association”
Descriptive nouns	“Relative risk reduction,” “benefit”	“Difference in risk,” “risk ratio”
Verbs	“Affected,” “caused,” “modulated risk,” “treatment resulted in,” “reduced hazard”	“Correlates with,” “is associated with”
Incorrect terms/avoid using		“Reduced risk” (active verb), “lowered risk” (active verb), “benefitted”

A Solution: Matching Language to Type of Study

We therefore urge investigators and editors alike to carefully select the language used when presenting results, such that it is clear what type of study the conclusions are based on.

As an illustrative example, *The New England Journal of Medicine* published a case-control study on the risk of colorectal cancer and statins in 2005.¹³ Although the authors appropriately used the term “associated” when describing the results, they then concluded: “We found that the use of statins is associated with a 47 percent relative reduction in the risk of colorectal cancer after adjustment for other known risk factors and is specific to this class of lipid-lowering agents.” Here, the use of the active term “relative reduction in risk” implies causality. Instead, the authors should have written: “the use of statins is associated with a 47 percent lower risk of colorectal cancer. . .”, which more accurately reports the data based on the methodology of the evidence. Unless the authors had conducted an RCT, where subjects were randomized to statin vs placebo and then followed for incident colorectal cancer, this type of conclusion is misleading and incorrect. Multiple RCTs and meta-analysis of all 26 large trials have demonstrated no difference in the risk of incident cancer based on statin therapy.^{14–18}

Thus, a responsibility falls to both authors and editors to edit the manuscript not only for scientific accuracy and statistical rigor, but also for language that is appropriately suited to the study (Table 1). These differences in language may seem subtle, but the implications are enormous and critical to how the results are applied in practice. In all types of observational studies, the authors should report the difference in outcomes between 2 groups of patients descriptively; they cannot make conclusions about “reductions” or “increases” from this type of study.

Returning to the example of DES vs BMS, if future observational studies on stent type and outcomes are performed, it is important to note that the patients who received a next-generation DES may be quite different from those who received a BMS previously. Thus, any comparison between the groups should merely report differences in outcomes without ascribing causality. So, if patients treated with a DES had a lower mortality within the registry, the publication should read: “Patients treated with a DES had a lower mortality than those treated with a BMS, HR = X.XX”

(the preferred way); or “Use of a DES was associated with a lower mortality rate than use of a BMS.” These statements are descriptive and accurate. However, it would be incorrect to report that “the use of a DES reduced mortality by x-fold.”

In conclusion, we would like to make a plea to all investigators and editors to carefully select language used (descriptive or declarative) during reporting to match the type of study conducted. As Oliver Wendell Holmes so accurately stated, “carve every word before you let it fall.”

References

1. Moher D, Hopewell S, Schulz KF, et al. CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials. *BMJ*. 2010;340:c869.
2. Schulz KF, Altman DG, Moher D. CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. *BMJ*. 2010;340:c332.
3. Stampfer MJ, Willett WC, Colditz GA, et al. A prospective study of postmenopausal estrogen therapy and coronary heart disease. *N Engl J Med*. 1985;313:1044–1049.
4. Grodstein F, Stampfer MJ, Manson JE, et al. Postmenopausal estrogen and progestin use and the risk of cardiovascular disease. *N Engl J Med*. 1996;335:453–461.
5. Rossouw JE, Anderson GL, Prentice RL, et al. Risks and benefits of estrogen plus progestin in healthy postmenopausal women: principal results from the Women's Health Initiative randomized controlled trial. *JAMA*. 2002;288:321–333.
6. Anderson GL, Limacher M, Assaf AR, et al. Effects of conjugated equine estrogen in postmenopausal women with hysterectomy: the Women's Health Initiative randomized controlled trial. *JAMA*. 2004;291:1701–1712.
7. Lagerqvist B, James SK, Stenestrand U, et al. Long-term outcomes with drug-eluting stents versus bare-metal stents in Sweden. *N Engl J Med*. 2007;356:1009–1019.
8. James SK, Stenestrand U, Lindback J, et al. Long-term safety and efficacy of drug-eluting versus bare-metal stents in Sweden. *N Engl J Med*. 2009;360:1933–1945.
9. Sarno G, Lagerqvist B, Frobert O, et al. Lower risk of stent thrombosis and restenosis with unrestricted use of 'new-generation' drug-eluting stents: a report from the nationwide Swedish Coronary Angiography and Angioplasty Registry (SCAAR). *Eur Heart J*. 2012;33:606–613.
10. Kastrati A, Dibra A, Spaulding C, et al. Meta-analysis of randomized trials on drug-eluting stents vs. bare-metal stents in patients with acute myocardial infarction. *Eur Heart J*. 2007;28:2706–2713.
11. Vink MA, Dirksen MT, Suttrop MJ, et al. 5-year follow-up after primary percutaneous coronary intervention with a paclitaxel-eluting stent versus a bare-metal stent in acute ST-segment elevation myocardial infarction: a follow-up study of the PASSION

- (Paclitaxel-Eluting Versus Conventional Stent in Myocardial Infarction with ST-Segment Elevation) trial. *JACC Cardiovasc Interv.* 2011;4:24–29.
12. Di Lorenzo E, De Luca G, Sauro R, et al. The PASEO (Paclitaxel or Sirolimus-Eluting Stent Versus Bare Metal Stent in Primary Angioplasty) Randomized Trial. *JACC Cardiovasc Interv.* 2009;2:515–523.
 13. Poynter JN, Gruber SB, Higgins PD, et al. Statins and the risk of colorectal cancer. *N Engl J Med.* 2005;352:2184–2192.
 14. Baigent C, Blackwell L, Emberson J, et al. Efficacy and safety of more intensive lowering of LDL cholesterol: a meta-analysis of data from 170,000 participants in 26 randomised trials. *Lancet.* 2010;376:1670–1681.
 15. Ford I, Murray H, Packard CJ, et al. Long-term follow-up of the West of Scotland Coronary Prevention Study. *N Engl J Med.* 2007;357:1477–1486.
 16. Bulbulia R, Bowman L, Wallendszus K, et al. Effects on 11-year mortality and morbidity of lowering LDL cholesterol with simvastatin for about 5 years in 20,536 high-risk individuals: a randomised controlled trial. *Lancet.* 2011;378:2013–2020.
 17. Sever PS, Chang CL, Gupta AK, Whitehouse A, Poulter NR. The Anglo-Scandinavian Cardiac Outcomes Trial: 11-year mortality follow-up of the lipid-lowering arm in the UK. *Eur Heart J.* 2011;32:2525–2532.
 18. Kohli P, Cannon CP. Statins and safety: can we finally be reassured? *Lancet.* 2011;378:1980–1981.