INTRODUCTION

In several disease areas, approximately 30 percent of clinical trials fail, even though these trials are powered at 60-80 percent. Many methodological approaches for increasing signal detection have been proposed, including increased sample size, post hoc analyses, case report forms, and centralized assessments. Questions remain about the efficiency (or lack thereof) of these methods to increase signal detection. Reliability in reporting results of studies aimed at determining the efficacy of these novel methods makes it difficult to interpret, evaluate, and compare findings. Standardizing methods of reporting across studies and methodologies would alleviate these limitations and reduce reporting bias.

METHOD

While the Consolidated Standards of Reporting Trials (CONSORT)2 developed a set of guidelines for standardizing reporting of randomized controlled trials (RCTs), no similar set of guidelines exist for the reporting of results from studies of the efficacy of clinical trial methodologies. We propose a set of seven such guidelines, which recognize the need for each, and where appropriate, provide a detailed illustration of how misuse or omission can influence the interpretation of study results.

RESULT

1. Report inter-rater reliability (IRR). IRR (with a 95 percent confidence interval) should be reported in all studies with multiple raters and multiple observations. Low IRR will reduce study power and the ability to detect drug placebo separation.

2. Use appropriate statistical tests. Statistical tests should be appropriate for the type of variable (i.e., continuous, categorical, etc.) being analyzed.

Moreover, primary analyses should be adequately powered (i.e., 80 percent) for a statistical analysis plan, much like efficacy analyses. Usually, a primary analysis should be reported if no prior results have been reported. Typically, post hoc analyses are performed in small subsets of the sample. Failing to report these results can lead to over-interpreta-

3. Acknowledge and correct for multiple comparisons. If multiple comparisons are performed on a single sample, all analyses should be reported, whether or not they are published. Appropriate multiplicity adjustments must be made (e.g., Bonferroni or Hochberg) in order to avoid inflating false positives. Reporting a significant result on a subset of data without indicating the total number of comparisons made across the entire data set may lead to over interpretation of false positives. For example, a researcher might report isolation of a statistically significant p-value (i.e., a t-test comparing drug-placebo separation between two methodologies). This would appear to indicate that one methodology was statistically significantly superior to the other.

PROPOSED CHECKLIST FOR REPORTING CLINICAL TRIALS METHODOLOGY RESEARCH

1. Inter-rater reliability
   a. When assessed?
   b. How many raters?
   c. How many observations?
   d. How were ratings obtained (i.e., scoring videos, joint interviews, independent interviews)?
   e. What statistic was calculated?

2. Inferential statistics
   a. What inferential statistics were calculated?
   b. What significance tests are appropriate?

3. Analyses a priori vs. post-hoc
   a. What was primary analysis?
   b. Were analyses reported post-hoc?

4. Multiple comparisons
   a. How many comparisons were made across entire sample?
   b. What multiplicity adjustment were made? Why?

5. Appropriateness of statistical test
   a. Test appropriate for variable type?

6. Effect size for all analyses regardless of significance
   a. Effect size statistic reported?
   b. Interpretation of magnitude of effect?

7. Interpretation of null hypothesis testing (NHT)
   a. What does p-value indicate?
   b. Is interpretation of significance affected by magnitude of effect size?

6. Include inferential statistics for means comparisons.

CONCLUSION

We demonstrate how adherence to proposed guidelines for standardization reporting on outcomes evaluating clinical trial methodologies can reduce reporting bias. Empirical research evaluating the effectiveness of these methods is important, not only for clinical trial methodology but also for future drug development decisions being sponsored and regulated.