Rubinstein et al\(^1\) propose an algorithm that measures uncertainty in abstracts reported from randomized clinical trials. Eligible trials were those in oncology that were designed as superiority trials with 2 arms in which \(P\) values ranged between .01 and .10 and were published from 1974 to 2017. The developed algorithm evaluates "whether reporting is restricted to the conditions of the trial, whether speculative language was used, and whether the significance of results is qualified as statistical."\(^1\) The algorithm was applied to the overall survival end point as well as other surrogate end points of overall survival. It categorizes an abstract as fully uncertain when it uses an "uncertainty qualifier" such as *may*, *can*, or *suggest*, restricts its conclusion to the trial conditions or to the trial meeting its end point, or, if significance is claimed, qualifies the significance as statistical. In contrast, it categorizes an abstract as somewhat uncertain when the results of the trial are described as significant without making a distinction between statistical and clinical significance. When neither of these conditions is met, the abstract is categorized as not uncertain.

The authors screened 5777 abstracts and identified 556 phase 3 randomized trials in which the \(P\) values ranged between .01 and .10 (that being the region of \(P\) values that they considered as not clearly indicating either statistically significant or not statistically significant). Of the 556 abstracts, 31% expressed full uncertainty, 29% expressed some uncertainty, and 40% did not express any uncertainty. Statistically significant factors associated with uncertainty expression were year of publication, normalized \(P\) values, noncooperative group trials, and abstracts reporting on end points other than overall survival. Of interest, funding source was not statistically associated with uncertainty expression.

The article by Rubenstein et al\(^1\) presents an important message for clinical trialists. There are already several best practice guidelines for reporting results for clinical trials.\(^2,3\) The Consolidated Standards of Reporting Clinical Trials (CONSORT) statement was first published in *JAMA* in 1996. It was developed to overcome issues of inadequate reporting of clinical trial results. It is considered a living document, and the CONSORT statement was last updated in 2010. Journals should consider implementing these guidelines, as currently authors do not always adhere to them.

In 2004, one of us (S.H.) led the development of biostatistical guidelines for reporting results from randomized trials to be used by investigators submitting abstracts for the American Society of Clinical Oncology (ASCO) annual meeting.\(^4\) The guidelines for phase 3 trials clearly describe the key elements of study design and results to report. For example, the planned sample size per arm, including operating characteristics (power and type I error rate, whether the type I error rate is 1-sided or 2-sided), the null and alternative hypotheses, and the magnitude of expected change are required. If the analysis that is presented in an abstract is prior to the trial's final analysis, it should clearly indicate whether the analysis was a planned interim analysis and the reason for early reporting. These guidelines\(^4\) have been updated and are available online. In an analysis\(^5\) of 500 abstracts presenting results from large randomized trials at the ASCO meetings, several abstracts were missing essential information on study design and analysis, and the manner in which the abstracts reported results was poor.

The analysis by Rubinstein et al\(^1\) is far from comprehensive, as they excluded multiple-arm trials (including factorial trial), noninferiority trials, and randomized phase 2 trials. We anticipate that these trials are challenging in terms of communicating results in abstracts. Still, we are not surprised by the quality of abstracts reported by Rubinstein et al.\(^1\)
We support the concept that uncertainty in levels of statistical significance should be clearly communicated in an abstract. An abstract is a brief summary of the background, primary goal of the trial, patient population, methods used in analyzing the data, key results, and conclusions. Usually, the content of the full article is condensed to only the most important aspects of the trial. The design and results sections in an abstract should be consistent with the details provided in the article. After reading an abstract, a person should comprehend the most salient results and messages of the trial, as well as be able to assess its design, conduct, primary end points, and, thereby, its overall credibility. While abstracts serve a vital role in condensing and communicating research findings, in and of themselves they are insufficient to draw meaningful conclusions and can even be misleading.

While we recognize that Rubinstein et al are focusing on abstracts, in which word count is always at a premium, we believe it is critical that readers understand that there are far broader issues that affect the reliability of study conclusions beyond simply the level of statistical significance (and any uncertainty around that). Without knowing important details of study design and execution, the level of statistical significance (however reliable the $P$ value may be) is of little value in deciding whether stated conclusions from a study are reliable. With regard to statistical testing, we believe that the analysis of data from a randomized clinical trial should follow the advice published by the American Statistical Association (ASA)$^6$ and by Chavalarias et al in JAMA.$^7$ The ASA statement on statistical significance and $P$ values "is intended to steer research into a 'post $p < 0.05$ era.'"$^6$ This action is necessary because "Over time it appears the $P$ value has become a gatekeeper for whether work is publishable...."$^6$ The ASA statement concludes, "...Good statistical practice ... emphasizes ... a variety of numerical and graphical summaries of data..., complete reporting and proper logical and quantitative understanding of what data summaries mean. No single index should substitute for scientific reasoning."$^6$ The JAMA article$^7$ reported on the improper use of the $P$ value in biomedical research and came to similar a conclusion: "Rather than reporting isolated $P$ values, articles should include effect sizes and uncertainty metrics."$^7$

In summary, while we agree that reporting uncertainty in $P$ values and levels of statistical significance is important, particularly in cases in which statistical significance is not clear cut, we caution about considering this a sufficient description of uncertainty. There are many aspects of study design, conduct, execution, analysis, and reporting that $P$ values do not capture but that are critical to the reliability of the reported conclusions. We caution clinical trialists and clinicians on relying on abstracts alone for the reasons cited here.

**ARTICLE INFORMATION**

**Published:** December 13, 2019. doi:10.1001/jamanetworkopen.2019.17543

**Open Access:** This is an open access article distributed under the terms of the CC-BY License. © 2019 Halabi S et al. JAMA Network Open.

**Corresponding Author:** Susan Halabi, PhD, Duke University Medical Center, 2424 Erwin Rd, Ste 11088, Durham, NC 27710 (susan.halabi@duke.edu).

**Author Affiliations:** Department of Biostatistics and Bioinformatics, Duke University, Durham, North Carolina (Halabi); Clinical Trials Consulting & Training Limited, Buckingham, England (Day).

**Conflict of Interest Disclosures:** None reported.

**REFERENCES**


