



# Indication of Measures of Uncertainty for Statistical Significance in Abstracts of Published Oncology Trials

## A Systematic Review and Meta-analysis

Samuel M. Rubinstein, MD; Elizabeth A. Sigworth, BA; Shervin Etemad, BA; Richard L. Martin, MD, MPH; Qingxia Chen, PhD; Jeremy L. Warner, MD, MS

### Abstract

**IMPORTANCE** There is growing consensus that reliance on  $P$  values, particularly a cutoff level of .05 for statistical significance, is a factor in the challenges in scientific reproducibility. Despite this consensus, publications describing clinical trial results with  $P$  values near .05 anecdotally use declarative statements that do not express uncertainty.

**OBJECTIVES** To quantify uncertainty expression in abstracts describing the results of cancer randomized clinical trials (RCTs) with  $P$  values between .01 and .10 and examine whether trial features are associated with uncertainty expression.

**DATA SOURCES** A total of 5777 prospective trials indexed on HemOnc.org, as of September 15, 2019.

**STUDY SELECTION** Two-arm RCTs with a superiority end point with  $P$  values between .01 and .10.

**DATA EXTRACTION AND SYNTHESIS** Abstracts were evaluated based on an uncertainty expression algorithm. Ordinal logistic regression modeling with multiple imputation was performed to identify whether characteristics of study design, results, trial authors, and context were associated with uncertainty expression.  $P$  values were normalized by dividing by prespecified  $\alpha$  value.

**MAIN OUTCOMES AND MEASURES** Uncertainty expression in abstracts as determined by the algorithm and its association with trial and publication characteristics.

**RESULTS** Of 5777 trials screened, 556 met analysis criteria. Of these, 222 trials (39.9%) did not express uncertainty, 161 trials (29.0%) expressed some uncertainty, and 173 trials (31.1%) expressed full uncertainty. In ordinal logistic regression with multiple imputation, trial features with statistically significant associations with uncertainty expression included later year of publication (odds ratio [OR], 1.70; 95% CI, 1.24-2.32;  $P < .001$ ), normalized  $P$  value (OR, 1.36; 95% CI, 1.11-1.67;  $P = .003$ ), noncooperative group studies (OR, 1.72; 95% CI, 1.12-2.63;  $P = .01$ ), and reporting an end point other than overall survival (OR, 1.41; 95% CI, 1.01-1.96;  $P = .047$ ). Funding source, number of authors, journal impact tier, author nationality, study of unapproved drugs, abstract word count, whether the marginal end point was a primary or coprimary end point, and effect size (in subgroup analysis) did not have statistically significant associations with uncertainty expression.

**CONCLUSIONS AND RELEVANCE** Published oncology articles with marginally significant results may often incompletely convey uncertainty. Although it appears that more uncertainty is expressed in recent abstracts, full uncertainty expression remains uncommon, and seemingly is less common when reporting overall survival, results with  $P$  values lower than  $\alpha$  levels, and cooperative

(continued)

### Key Points

**Question** Do the abstracts of oncology randomized clinical trials with results of marginal statistical significance express uncertainty, and what characteristics are associated with uncertainty expression?

**Findings** This systematic review and meta-analysis of 556 phase 3 randomized clinical trials found that most of the abstracts do not fully express uncertainty, and that uncertainty expression decreases when the  $P$  value surrounding a result is below its prespecified  $\alpha$  level. The expression of uncertainty appears to be increasing over time.

**Meaning** It appears that a  $P$  value below the prespecified  $\alpha$  level is often treated as a discrete threshold for statistical significance in publication of randomized clinical trials.

+ [Invited Commentary](#)

+ [Supplemental content](#)

Author affiliations and article information are listed at the end of this article.

**Open Access.** This is an open access article distributed under the terms of the CC-BY License.

Abstract (continued)

group studies.

*JAMA Network Open.* 2019;2(12):e1917530.

Corrected on January 17, 2020. doi:10.1001/jamanetworkopen.2019.17530

---

## Introduction

Widely publicized clinical findings have often proven difficult to replicate in follow-up analyses.<sup>1,2</sup> One factor is the high type I error rate associated with the ubiquitous threshold for statistical significance, which is a *P* value less than .05. Owing in part to concerns that this threshold has raised reproducibility concerns, the American Statistical Association released a consensus statement advising caution in interpretation of *P* values for hypothesis testing, emphasizing that *P* values do not measure the effect size of a result, nor do they inform the likelihood that the null hypothesis is false.<sup>3</sup> However, it is common for publications to imply that a *P* value offers information on the probability of the null hypothesis (that a therapeutic intervention has no efficacy); instead, a *P* value provides the probability of observing the treatment effect assuming that the null hypothesis is valid.<sup>4</sup> In reality, a *P* value of .05 continues to serve as a discrete threshold for statistical significance.<sup>5</sup> The .05 threshold is so widely accepted as a threshold for publishable results that many collect data only until their results cross this threshold—a phenomenon known as “*P* hacking.”<sup>6</sup> In response to this controversy, some have advocated for lowering the threshold by an order of magnitude<sup>7</sup> or have questioned the validity of the concept of statistical significance.<sup>8</sup> Others have emphasized inclusion of a confidence interval in clinical trial reporting to highlight the uncertainty of the results.<sup>9</sup>

There also remains substantial room for improvement in the rigor that is applied to reporting of what are considered to be positive clinical trial results.<sup>1</sup> The magnitude of benefit reported as being clinically relevant declined between 1980 and 2010 in the reporting of clinical trials for non-small cell lung cancer, while at the same time reliance on surrogate end points increased.<sup>10</sup> An analysis of randomized clinical trials (RCTs) with statistically nonsignificant results (ie, results with a *P* value >.05) reported that the results were presented with “spin” (misrepresentation of study findings) in a significant number of trials; as such, findings were reported in a manner that was inconsistent with the results.<sup>9</sup> Spin has been shown to be associated with clinicians’ interpretation of trial results.<sup>11</sup> This style of reporting has implications for the practice of medical oncology in particular. Antineoplastic therapies generally have among the lowest therapeutic indices and highest costs in medicine, meaning that the adverse effects associated with these therapies can outweigh the benefits in many scenarios. If the data supporting a treatment’s efficacy are marginal, and therefore the benefit of a treatment is questionable, clinicians need to be aware of this uncertainty and convey it to patients before making an informed decision about treatment.

In this article, we describe a systematic evaluation of RCT abstracts in oncology with a marginal *P* value result, which we define as between .01 and .10, to center our analysis on results near the widely accepted threshold of .05.<sup>12,13</sup> We focused on whether uncertainty is expressed in the abstracts of published articles and examined factors potentially associated with uncertainty expression. We also examined the degree to which trials reporting these marginally significant results in the oncology literature expressed uncertainty and whether any secondary factors were associated with uncertainty expression.

---

## Methods

### Data Source

We derived our list of RCTs from HemOnc.org, a collaborative web-based knowledge base that aims to be a central repository of information regarding chemotherapy drugs and regimens that are used

or have been used in routine cancer care.<sup>14</sup> HemOnc.org has been continuously updated since 2011 and is the largest publicly available online repository of chemotherapy regimen information. Regimens included on HemOnc.org are discovered primarily through review of guidelines and the primary literature. The process of building HemOnc.org included systematic review of all American Society of Clinical Oncology, European Society of Medical Oncology, and National Comprehensive Cancer Network guidelines through December 31, 2018; all Cochrane Database of Systematic Reviews labeled as cancer specific ( $n = 685$ ); and all *Lancet*, *JAMA*, and *New England Journal of Medicine* tables of contents between 1946 and December 31, 2018. In addition, the citations of any included material are hand-searched for additional citations. All RCT publications have been manually reviewed for comparative efficacy of the primary and key secondary end points, and the findings have been labeled in a structured format on the HemOnc.org website.

This study followed the Preferred Reporting Items for Systematic Reviews and Meta-analyses (PRISMA) reporting guideline where applicable.<sup>15</sup> Although this is a meta-analysis of the expression of uncertainty in the oncology literature, the outcome of interest for this study is not a clinical outcome but rather the published text of clinical trial reporting, such that there is no intervention and comparison under examination. Typical features of meta-analyses delineated by the PRISMA guidelines, including aggregate measures of a clinical outcome or summary metric and assessment of bias in primary outcomes of the studies examined, are outside the scope of this analysis.

Efficacy is labeled following the recommendations by Pocock and Ware<sup>13</sup> to translate statistical findings into plain English. As uncertainty may be meaningfully lower for results with  $P$  values less than .01 than between .01 and .05, and many journals simply report  $P$  values as less than .01 rather than reporting the precise  $P$  value, these guidelines emphasize the importance of conveying the uncertainty of results with  $P$  values greater than .01.<sup>12,13</sup> With this justification, a regimen that shows an overall survival (OS) benefit above its comparator with a  $P$  value between .01 and .05 is labeled on the site as “seems to have superior OS,” whereas a regimen that shows an OS benefit less than .01 is labeled as having “superior OS.” To our knowledge, there are no other publicly available data sources that have similar labeling.

Efficacy labeling on HemOnc.org is first based on the primary outcome as reported by the clinical trial publication. If the primary outcome is a surrogate outcome and is positive with a  $P$  value less than .05, the least surrogate secondary outcome with  $P$  value of .10 or less is labeled instead. If the primary outcome is negative ( $P > .05$ ), the trial is labeled as negative even if a secondary outcome is reported as positive. Representative examples are reported in eTable 1 in the Supplement. In addition, end points are relabeled if an interim update is published with new findings, such as loss of statistical significance for a primary end point; in these cases, the label on HemOnc.org is marked with an asterisk and a note of the format “reported efficacy is based on the YYYY update” is added.

In this study, a frozen version of the HemOnc.org text content from September 15, 2019, was programmatically screened for 2-arm RCTs with a seems to have label or an asterisk indicating an interim update to efficacy. This was a 3-step process (Figure 1). The first step used an automated-parsing algorithm to exclude trials not suitable for this analysis, such as noncancer trials; the second involved a manual review of candidate abstracts for eligibility; and the third involved a manual screen of regimens with complicated comparative designs not parsed in step 1. Trials that involved more than 2 arms, with or without factorial designs, were excluded from the analysis, as the primary discussions of such articles often do not focus on the specific result that is statistically marginal. The specific statement in the abstract results section that reported the statistically marginal result, as well as the concluding statement of the abstract, were analyzed for expression of uncertainty. Abstracts of updates to clinical trials were included in this analysis. In some cases, the category of statistical significance for the initial and updated analyses differed; the process for selecting which abstract to include in the analysis is summarized in eTable 2 in the Supplement.

### Uncertainty Scoring Algorithm

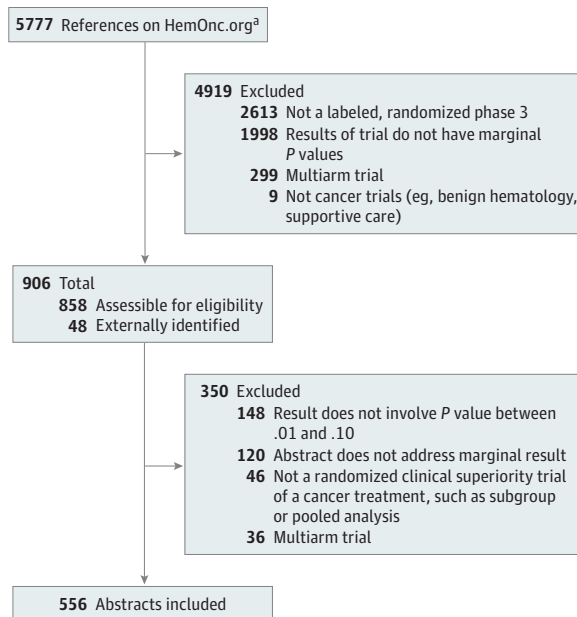
To our knowledge, no consensus guidelines currently exist for best practices in uncertainty expression in the medical literature, although the imperative for increased uncertainty expression in the medical literature has been recognized.<sup>16</sup> We therefore have proposed and developed a systematic approach for evaluating uncertainty expression based on 3 factors: 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 (**Figure 2**). The logic behind the algorithm is described more fully in the eAppendix in the [Supplement](#). We applied this algorithm to evaluate the abstracts for this study. The same approach was applied whether the marginal end point was OS or a surrogate of OS.

To assess interrater variability, after the primary abstractor (S.M.R.) reviewed the entire data set, the algorithm was given to a second abstractor (R.L.M.) along with a 50-abstract tutorial set with answers provided. Following review of this material, the second abstractor was given an additional set of 50 trials to review and evaluate based on that individual's understanding of the algorithm. Interrater reliability was assessed using Fleiss  $\kappa$  and was computed using the R irr package, version 0.84.1 (R Foundation).<sup>17,18</sup>

### Data Collected for Analysis

The following metadata for each included RCT were recorded for analysis: (1) year of publication; (2) whether a for-profit pharmaceutical company funded the study; (3) the journal impact tier of publication (eTable 3 in the [Supplement](#)); (4) the total number of authors; (5) a score of 95% CI expression (eTable 4 in the [Supplement](#)); (6) the *P* value of the marginal result divided by prespecified  $\alpha$  level; (7) the nationality of the authors' affiliations (United States, non-United States, or collaborations between US and non-US authors); (8) whether the studies included a drug that, at the time of the RCT publication, was unapproved by the US Food and Drug Administration (FDA) or similar international regulatory bodies; (9) whether the study was conducted by a non-industry-funded cooperative group; (10) where reported, a point estimate for hazard ratio concerning the marginal end point; (11) the 95% CI surrounding the hazard ratio; (12) whether the marginal end point

Figure 1. Selection of Abstracts for Analysis



<sup>a</sup> References included in the analysis are from published prospective trials with results only; abstract-only references, trial protocol descriptions, meta-analyses, reviews (systematic or otherwise), case reports, and case series were excluded.

was a prespecified primary or coprimary end point; (13) whether the marginal end point was OS or a surrogate end point; and (14) abstract word count.

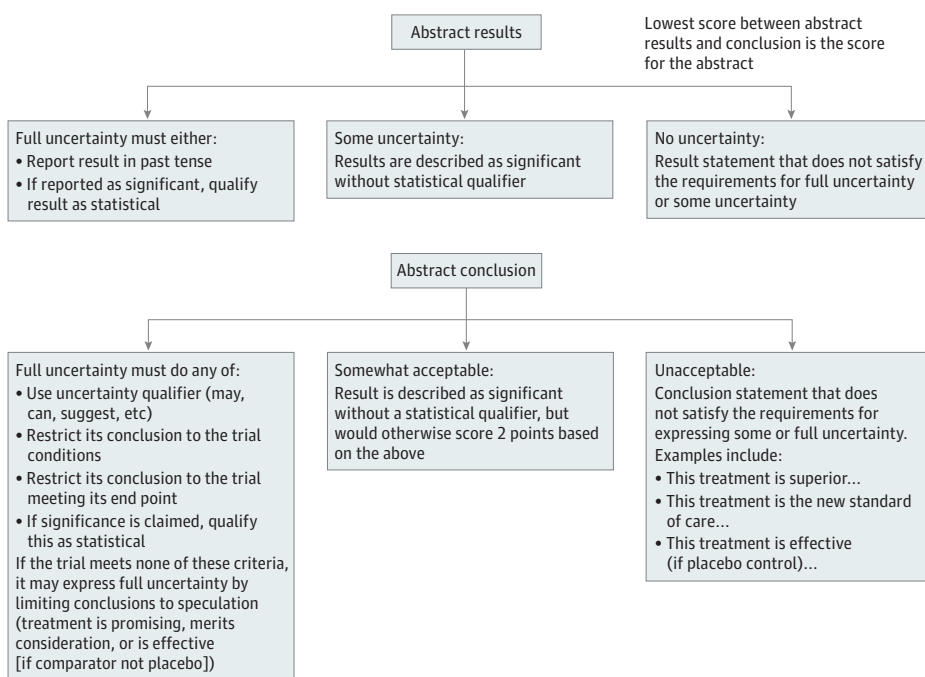
We examined results reporting for the presence of a 95% CI around the effect size to examine whether 95% CI expression was associated with uncertainty expression and scored articles on the degree to which a 95% CI was included (eTable 4 in the [Supplement](#)). Funding source information was extracted from the body of the article when available. If this information was not present in the body of the article, the metadata were searched, as well as the ClinicalTrials.gov record, when available.

Data regarding whether a trial involved an unapproved drug and whether a trial was conducted by a cooperative group are contained within the HemOnc.org ontologic framework and were abstracted from this database programmatically.<sup>19</sup> The FDA approval dates of all drugs included on HemOnc.org as well as the publication dates are concepts within the ontologic framework. If the FDA approval date was after or the same year as the RCT publication (as in the latter case, the trial must have been designed and conducted before approval), the trial was determined to have been conducted before FDA approval of all drugs. Trials that contained unapproved drugs were manually screened for approval by international agencies and, if so, the first date of international approval was used.

### Statistical Analysis

The outcome of interest was the ordinal uncertainty expression as evaluated by the 3-level algorithm (full uncertainty, some uncertainty, no uncertainty). Continuous variables, such as year and number of authors, were summarized as median and interquartile range as well as mean (SD) for each level of outcome variable and compared using the Kruskal-Wallis test.<sup>20</sup> Categorical variables were summarized as frequencies and counts. An ordinal logistic regression model or multivariable proportional odds model (eAppendix in the [Supplement](#)) was used to simultaneously study the associations of the covariates with uncertainty expression, allowing for nonlinear associations for year from 1974 and number of authors with restricted cubic spline functions with 5 knots.<sup>21</sup> With 15 *df* per parameter in the model in complete case analysis, no variable selection procedure was

Figure 2. Algorithmic Approach for Evaluation of Uncertainty



Graphical representation of the algorithm used to systematically evaluate the abstracts for uncertainty. A rationale is provided in the eAppendix in the [Supplement](#).

implemented in the analysis. Multiple imputation with 150 imputations, assuming missingness at random for funding resource and journal tier, was used.<sup>22</sup> Two-sided *P* values less than .05 were considered statistically significant. Ordinal logistic regression models and multiple imputation were performed with Hmisc and rms packages in R, version 3.4.4.<sup>23,24</sup>

Using the `popower` function from the Hmisc package, we analyzed our power to detect a range of odds ratios (ORs) given our current sample size, when binary predictors were in a 2:1 ratio (funding source, study group) or heavily imbalanced (10:1, whether experimental or control preferred). We found that in the 2:1 case, our current sample size gave us 80% power to detect an OR of 0.62 or lower (or  $\geq 1.6$  in the other direction); in the 10:1 imbalanced group case, we had 80% power to detect an OR of 0.48 or lower (or  $\geq 2.08$  in the other direction).

## Results

A total of 5777 articles were analyzed with a total of 3218 efficacy labels in the context of superiority trials (eTable 5 in the [Supplement](#)). Of these trials, 4919 were programmatically excluded, leaving 858 abstracts available for analysis. An additional 48 abstracts were identified external to the pruning algorithm for review, and in total, 906 were manually reviewed for inclusion in the analysis. Of these, 556 abstracts were suitable for inclusion (Figure 1). The PubMed identification numbers and titles of the 556 abstracts analyzed are reported in eTable 6 in the [Supplement](#). The 2 abstractors of clinical trial uncertainty (S.M.R., R.L.M.) gave the same assessment on 45 of 50 (90.0%) mutually assessed abstracts (Fleiss  $\kappa$ , 0.85).

Characteristics of the included abstracts are reported in eTable 5 in the [Supplement](#). Median year of publication of the abstracts was 2009 (range, 1974-2019). The median number of authors was 15 (range, 2-54). Three-hundred thirty-two abstracts (59.7%) had exclusively non-US authors, 96 abstracts (17.3%) had exclusively US authors, and 128 abstracts (23.0%) were of trials with collaborations between US and non-US authors. One hundred ninety-one trials (34.4%) were funded exclusively by grants, and 286 trials (51.4%) were at least partially funded by industry; funding source was missing for 79 (14.2%) of the RCTs. Of the 556 abstracts evaluated, 222 reported trials (39.9%) that did not express uncertainty, 161 abstracts (29.0%) expressed some uncertainty, and 173 abstracts (31.1%) expressed full uncertainty.

Trial features with statistically significant associations with uncertainty expression after multiple imputation are shown in **Figure 3**: later year of publication (OR, 1.70; 95% CI, 1.24-2.32;  $P < .001$ ), lower normalized *P* value (OR, 1.36; 95% CI, 1.11-1.67;  $P = .003$ ), noncooperative group studies (OR, 1.72; 95% CI, 1.12-2.63;  $P = .01$ ), and reporting an end point other than OS (OR, 1.41; 95% CI, 1.01-1.96;  $P = .047$ ). Funding source, publication before regulatory drug approval, author nationality, confidence interval expression score, abstract word count, whether the marginal end point was a coprimary end point, journal tier, and number of authors were not associated with uncertainty expression after multiple imputation. Regression results are shown in **Figure 4** and eFigure 1 in the [Supplement](#).

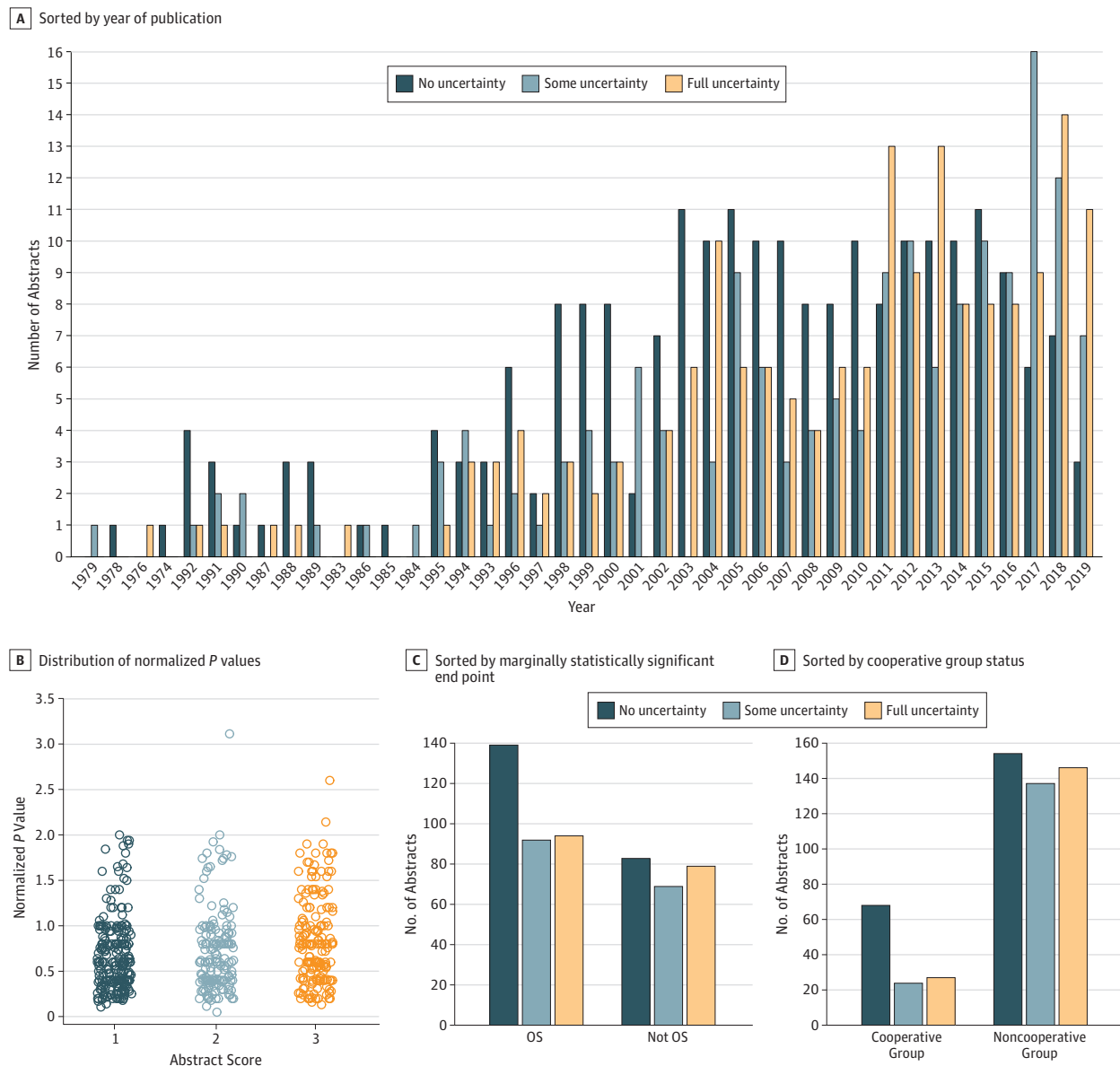
In the subgroup analysis to evaluate the association of effect size point estimate with uncertainty expression, there was greater uncertainty expression with lower hazard ratios, although this association was not statistically significant (OR, 1.22; 95% CI, 0.98-1.54;  $P = .08$ ). As in the complete model, publication year, normalized *P* value, noncooperative studies, and abstracts reporting end points other than OS expressed greater uncertainty. In this subanalysis, author nationality had a statistically significant association with uncertainty expression, with collaborations between US and non-US authors expressing more uncertainty than abstracts written by only US authors (OR, 2.38; 95% CI 1.14-5.00;  $P = .02$ ) and only non-US authors (OR, 2.08; 95% CI, 1.26-3.33;  $P = .003$ ). The subgroup analysis is summarized in eFigure 2 and eFigure 3 in the [Supplement](#).

## Discussion

Publication of RCT results is the primary means by which knowledge of new and potentially effective therapies is disseminated. A definitive trial result is usually a factor in some or all of the following: new drug approvals, uptake of new combination therapies by practicing clinicians, insurance coverage decisions for new therapies, and inclusion in guidelines and compendia. The phrasing of the findings has particular importance to many constituents: clinicians, approval agencies, pharmacy benefit managers, and, by extension, the patients who receive these therapies.

To our knowledge, this is the largest analysis of uncertainty phrasing in oncology RCTs with statistically marginal results to date. We found that nearly 70% of the abstracts presented findings in a way that could be construed as definitive, even when the reported *P* value suggests a substantial

**Figure 3. Abstract Uncertainty Expression by Year of Publication, Normalized *P* Values, Marginally Statistically Significant End Point, and Cooperative Group Status**



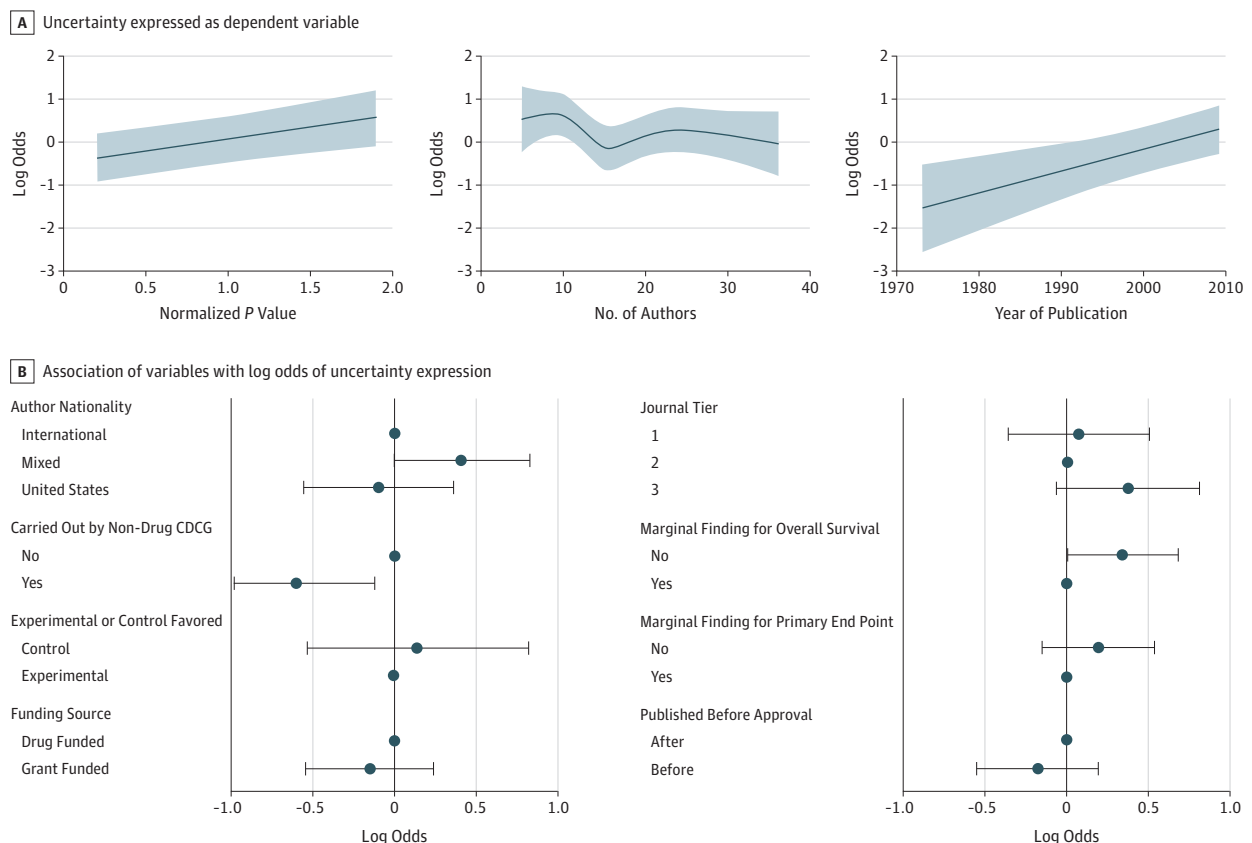
A, Abstract uncertainty expression sorted by year of publication. B, Distribution of normalized *P* values for abstracts with each degree of uncertainty expression. C, Abstract uncertainty expression sorted by the marginally statistically significant end point (overall

survival [OS], or a surrogate of OS). D, Abstract uncertainty expression sorted by cooperative group status.

risk of a type I error. Our focus on abstracts was intentional, given that many clinicians rely on abstracts as the primary sources of information that may change their clinical practice, despite well-recognized limitations in fully conveying the results of a trial.<sup>25,26</sup> It would appear that clinicians increasingly rely on third-party knowledge bases that are often then integrated into clinical decision support systems.<sup>27,28</sup> If the literature is biased toward overly optimistic statistically inaccurate interpretation of results, this bias may be associated with the validity of the meta-analyses and knowledge summarization required to generate these products. Systematic biases in the phrasing of the literature curated by knowledge bases are likely to be associated with automated curation efforts, such as CancerLinQ and other learning health systems used in oncology.<sup>29,30</sup>

Our finding that uncertainty expression is far from universal in the oncology literature is in line with other analyses.<sup>31,32</sup> However, unlike previous analyses, our algorithm demonstrated high interobserver correlation without the need for post hoc resolution of disagreement. The finding that uncertainty expression is improving over time may be associated with an increase in published position articles and awareness regarding the aforementioned reproducibility concerns. There was a statistically significant tendency for uncertainty expression to decrease when the *P* value of a result is below the  $\alpha$  level. In subgroup analysis, normalized *P* values had a more statistically significant association with uncertainty expression than the effect size. This finding further supports the hypothesis that the oncology community treats a *P* value of .05 as a cutoff not only for publication

Figure 4. Multivariable Regression of Uncertainty Expression on Trial Predictors



Multivariable logistic regression of uncertainty expression on various trial covariates. These figures show estimated values of the log odds at each value of the continuous factor while assuming all other variables are at their set reference levels (either prespecified for categorical variables or taken as the median for continuous variables). A, Multivariable logistic regressions with uncertainty expression as the dependent variable; the left side demonstrates the association between normalized *P* value and uncertainty

expression, the middle examines the association between number of authors and uncertainty expression (not statistically significant), and the right side demonstrates the association between year of publication and uncertainty expression (statistically significant). B, Association of ordinal and categorical variables with the log odds of uncertainty expression. CDCG indicates Company-Driven Cooperative Group; error bars, 95% CIs for the estimate of the log odds.



but also for external validity of a result. We also found that marginal OS results were conveyed with less uncertainty than marginal results that are surrogates of OS. Owing to the high clinical relevance of OS as a trial end point, it is understandable that marginally positive OS results are conveyed with more enthusiasm, and therefore less uncertainty, than other marginal results. It is also possible that awareness of the limitations of surrogate end points is growing, resulting in increasing uncertainty expression for trials that only show superiority in a surrogate end point.<sup>33,34</sup>

In contrast to other published data on the topic of spin in the oncology literature, we did not find an association between funding source and uncertainty expression. This finding is somewhat at odds with the widely held view that pharmaceutical industry-funded trials are more likely to be biased.<sup>35-37</sup> However, funding sources are often difficult to discern from published material, especially in older studies, so we advise interpreting these results with caution.

### Limitations

This analysis has several limitations. First, HemOnc.org aims to capture all oncology RCTs but is not complete and disproportionately includes RCTs that are published in very high-impact journals or endorsed by guidelines. Second, the algorithm proposed for evaluating uncertainty expression is imperfect. Although the interobserver variability is low, there remain cases in which 2 individuals may interpret uncertainty phrasing differently based on the rules described herein. Subtle messaging, such as the use of tone, is not considered. Existing technologies for programmatically parsing text, such as natural language processing, are not currently capable of reliably evaluating tone in technical documents, and thus an automated approach is not plausible. Better algorithms may be developed in the future based on collaborative efforts, and we hope that this work sparks a discussion within the community regarding best practices for uncertainty expression. Third, we did not evaluate uncertainty expression of results with *P* values below .01. Although our analysis focused on uncertainty expression for results that marginally cleared the widely accepted *P* value threshold, uncertainty should be expressed for the vast majority of scientific results, including those for which the *P* value crosses .01. Future analyses will focus on how descriptions of these less-marginal results express uncertainty. Fourth, we did not analyze multiarm or factorial trials owing to concerns with the generalizability of the algorithm to these trials; this analysis is a planned focus of future work. Fifth, we did not evaluate whether each drug included was approved for the specific indication of the study, as opposed to approved at all. Our focus on the first date of drug approval was intentional because off-label drug use is common in oncology, but there may remain incentives to underexpress uncertainty if a study is aimed at obtaining a new indication for an already approved drug, and this question may merit additional study.<sup>38</sup> Sixth, we did not examine individual author's conflict of interest disclosures, which may represent a source of indirect funding; given the increased focus on this issue in the scientific literature, this source of support will be a focus of future work.<sup>38,39</sup>

### Conclusions

The results suggest that clinical trials commonly fail to convey uncertainty when describing results of marginal statistical significance. These results are often conveyed as definitively demonstrating that the null hypothesis of an experiment is false, which may be associated with reliance by the oncology community on a discrete threshold for statistical significance. Many prominent voices in the scientific community are advocating for a change in this culture, and we believe our data support their conclusion.<sup>40-42</sup>

#### ARTICLE INFORMATION

**Accepted for Publication:** October 25, 2019.

**Published:** December 13, 2019. doi:[10.1001/jamanetworkopen.2019.17530](https://doi.org/10.1001/jamanetworkopen.2019.17530)

**Correction:** This article was corrected on January 17, 2020, to fix a run-on sentence in the Abstract.

**Open Access:** This is an open access article distributed under the terms of the [CC-BY License](#). © 2019 Rubinstein SM et al. *JAMA Network Open*.

**Corresponding Author:** Jeremy L. Warner, MD, MS, Vanderbilt University, 2220 Pierce Ave, Preston Research Bldg 777, Nashville, TN 37232 ([jeremy.warner@vumc.org](mailto:jeremy.warner@vumc.org)).

**Author Affiliations:** Division of Hematology and Oncology, Department of Medicine, Vanderbilt University, Nashville, Tennessee (Rubinstein, Martin, Warner); Department of Biostatistics, Vanderbilt University, Nashville, Tennessee (Sigworth, Chen); Vanderbilt University School of Medicine, Nashville, Tennessee (Etemad); Department of Biomedical Informatics, Vanderbilt University, Nashville, Tennessee (Chen, Warner); Vanderbilt-Ingram Cancer Center, Vanderbilt University Medical Center, Nashville, Tennessee (Warner).

**Author Contributions:** Drs Rubinstein and Warner had full access to all of the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis.

*Concept and design:* Rubinstein, Warner.

*Acquisition, analysis, or interpretation of data:* All authors.

*Drafting of the manuscript:* Rubinstein, Sigworth, Etemad.

*Critical revision of the manuscript for important intellectual content:* Rubinstein, Sigworth, Martin, Chen, Warner.

*Statistical analysis:* Sigworth, Martin, Chen.

*Obtained funding:* Warner.

*Administrative, technical, or material support:* Rubinstein.

*Supervision:* Warner.

**Conflict of Interest Disclosures:** Dr Chen reported receiving grants from the National Institutes of Health during the conduct of the study. Dr Warner reported receiving grants from National Cancer Institute during the conduct of the study and consulting fees from Westat outside the submitted work; in addition, Dr Warner is cofounder of HemOnc.org LLC, the content holder of the HemOnc.org database. This is an uncompensated position and shares have no monetary value. No other disclosures were reported.

**Funding/Support:** The study was funded by National Cancer Institute grants CA068485 (Dr Warner) and CA194215 (Ms Sigworth and Drs Chen and Warner); National Human Genome Research Institute grant T32 HG008341 (Dr Rubinstein); and the Carol and Jim O'Hare Fund award from a donor at Vanderbilt University Medical Center (Dr Rubinstein).

**Role of the Funder/Sponsor:** The funding sources had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

**Additional Contributions:** All contributors and editors of HemOnc.org, in particular Peter C. Yang, MD (editor-in-chief, HemOnc.org), provided helpful commentary on the manuscript; Andrew Maly assisted with abstract word count determination; and Frank Harrell, PhD (Vanderbilt University and creator of hmisc), provided comments and suggestions. There was no financial compensation outside of salary.

## REFERENCES

1. Greenhalgh T, Howick J, Maskrey N; Evidence Based Medicine Renaissance Group. Evidence based medicine: a movement in crisis? *BMJ*. 2014;348:g3725. doi:10.1136/bmj.g3725
2. Ioannidis JP. Why most published research findings are false. *PLoS Med*. 2005;2(8):e124. doi:10.1371/journal.pmed.0020124
3. Wasserstein RL, Lazar NA. The ASA's statement on *P* values: context, process, and purpose. *Am Stat*. 2016;70(2):129-133. doi:10.1080/00031305.2016.1154108
4. Mark DB, Lee KL, Harrell FE Jr. Understanding the role of *P* values and hypothesis tests in clinical research. *JAMA Cardiol*. 2016;1(9):1048-1054. doi:10.1001/jamacardio.2016.3312
5. Grabowski B. "*P* < 0.05" might not mean what you think: American Statistical Association clarifies *P* values. *J Natl Cancer Inst*. 2016;108(8):djw194. doi:10.1093/jnci/djw194
6. Head ML, Holman L, Lanfear R, Kahn AT, Jennions MD. The extent and consequences of *p*-hacking in science. *PLoS Biol*. 2015;13(3):e1002106. doi:10.1371/journal.pbio.1002106
7. Benjamin DJ, Berger JO, Johannesson M, et al. Redefine statistical significance. *Nat Hum Behav*. 2018;2(1):6-10. doi:10.1038/s41562-017-0189-z
8. Amrhein V, Greenland S, McShane B. Scientists rise up against statistical significance. *Nature*. 2019;567(7748):305-307. doi:10.1038/d41586-019-00857-9

9. Boutron I, Dutton S, Ravaud P, Altman DG. Reporting and interpretation of randomized controlled trials with statistically nonsignificant results for primary outcomes. *JAMA*. 2010;303(20):2058-2064. doi:10.1001/jama.2010.651
10. Sacher AG, Le LW, Leighl NB. Shifting patterns in the interpretation of phase III clinical trial outcomes in advanced non-small-cell lung cancer: the bar is dropping. *J Clin Oncol*. 2014;32(14):1407-1411. doi:10.1200/JCO.2013.52.7804
11. Boutron I, Altman DG, Hopewell S, Vera-Badillo F, Tannock I, Ravaud P. Impact of spin in the abstracts of articles reporting results of randomized controlled trials in the field of cancer: the SPIIN randomized controlled trial. *J Clin Oncol*. 2014;32(36):4120-4126. doi:10.1200/JCO.2014.56.7503
12. Goodman SN. Toward evidence-based medical statistics—1: the *P* value fallacy. *Ann Intern Med*. 1999;130(12):995-1004. doi:10.7326/0003-4819-130-12-199906150-00008
13. Pocock SJ, Ware JH. Translating statistical findings into plain English. *Lancet*. 2009;373(9679):1926-1928. doi:10.1016/S0140-6736(09)60499-2
14. Warner JL, Cowan AJ, Hall AC, Yang PC. HemOnc.org: a collaborative online knowledge platform for oncology professionals. *J Oncol Pract*. 2015;11(3):e336-e350. doi:10.1200/JOP.2014.001511
15. Shamseer L, Moher D, Clarke M, et al; PRISMA-P Group. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015: elaboration and explanation. *BMJ*. 2015;350:g7647. doi:10.1136/bmj.g7647
16. Bauchner H, Golub RM, Fontanarosa PB. Reporting and interpretation of randomized clinical trials. *JAMA*. 2019;322(8):732-735. doi:10.1001/jama.2019.12056
17. Fleiss JL. Measuring nominal scale agreement among many raters. *Psychol Bull*. 1971;76(5):378. doi:10.1037/h0031619
18. Gamer M. Package irr: various coefficients of interrater reliability and agreement. <https://cran.r-project.org/web/packages/irr/irr.pdf>. Published January 26, 2019. Accessed October 28, 2019.
19. Maly AM, Jain SK, Yang PC, Harvey K, Warner JL. Computerized approach to creating a systematic ontology of hematology/oncology regimens. *JCO Clin Cancer Inform*. 2018;2:1-11. doi:10.1200/CCI.17.00142
20. McKight PE, Najab J. Kruskal-Wallis test. In: *The Corsini Encyclopedia of Psychology*. Hoboken, NJ: John Wiley & Sons Inc; 2010:1-1.
21. Harrell FE. Ordinal logistic regression. In: *Regression Modeling Strategies*. Cham, Switzerland: Springer; 2015: 311-325. doi:10.1007/978-3-319-19425-7\_13
22. Little RJ, Rubin DB. *Statistical Analysis With Missing Data*. 3rd ed. Hoboken, NJ: Wiley; 2019.
23. Harrell FE Jr, Dupont MC. Hmisc: the design package. R Package. <http://ftp.uni-bayreuth.de/math/statlib/R/CRAN/doc/packages/Design.pdf>. Published June 18, 2007. Accessed November 7, 2019.
24. R Computing Team. R: a language and environment for statistical computing. <https://www.r-project.org/>. Published 2013. Accessed September 27, 2019.
25. Saint S, Christakis DA, Saha S, et al. Journal reading habits of internists. *J Gen Intern Med*. 2000;15(12):881-884. doi:10.1046/j.1525-1497.2000.00202.x
26. Berwanger O, Ribeiro RA, Finkelsztejn A, et al. The quality of reporting of trial abstracts is suboptimal: survey of major general medical journals. *J Clin Epidemiol*. 2009;62(4):387-392. doi:10.1016/j.jclinepi.2008.05.013
27. Malin JL. Envisioning Watson as a rapid-learning system for oncology. *J Oncol Pract*. 2013;9(3):155-157. doi:10.1200/JOP.2013.001021
28. Yu PP. Knowledge bases, clinical decision support systems, and rapid learning in oncology. *J Oncol Pract*. 2015;11(2):e206-e211. doi:10.1200/JOP.2014.000620
29. Sledge GW, Miller RS, Center JHO. Practice Management and Information Technology. CancerLinQ: a rapid learning system for oncology. <http://www.medsci.cn/webeditor/uploadfile/201306/20130602212623645.pdf>. Published 2013. Accessed July 7, 2019.
30. Abernethy AP, Etheredge LM, Ganz PA, et al. Rapid-learning system for cancer care. *J Clin Oncol*. 2010;28(27):4268-4274. doi:10.1200/JCO.2010.28.5478
31. Lazarus C, Haneef R, Ravaud P, Boutron I. Classification and prevalence of spin in abstracts of non-randomized studies evaluating an intervention. *BMC Med Res Methodol*. 2015;15(1):85. doi:10.1186/s12874-015-0079-x
32. Vera-Badillo FE, Shapiro R, Ocana A, Amir E, Tannock IF. Bias in reporting of end points of efficacy and toxicity in randomized, clinical trials for women with breast cancer. *Ann Oncol*. 2013;24(5):1238-1244. doi:10.1093/annonc/mds636

33. Kim C, Prasad V. Cancer drugs approved on the basis of a surrogate end point and subsequent overall survival: an analysis of 5 years of US Food and Drug Administration approvals. *JAMA Intern Med.* 2015;175(12):1992-1994. doi:10.1001/jamainternmed.2015.5868
34. Ellis LM, Bernstein DS, Voest EE, et al. American Society of Clinical Oncology perspective: raising the bar for clinical trials by defining clinically meaningful outcomes. *J Clin Oncol.* 2014;32(12):1277-1280. doi:10.1200/JCO.2013.53.8009
35. Djulbegovic B, Lacey M, Cantor A, et al. The uncertainty principle and industry-sponsored research. *Lancet.* 2000;356(9230):635-638. doi:10.1016/S0140-6736(00)02605-2
36. Lexchin J, Bero LA, Djulbegovic B, Clark O. Pharmaceutical industry sponsorship and research outcome and quality: systematic review. *BMJ.* 2003;326(7400):1167-1170. doi:10.1136/bmj.326.7400.1167
37. Zwierzyna M, Davies M, Hingorani AD, Hunter J. Clinical trial design and dissemination: comprehensive analysis of clinicaltrials.gov and PubMed data since 2005. *BMJ.* 2018;361:k2130. doi:10.1136/bmj.k2130
38. Saiyed MM, Ong PS, Chew L. Off-label drug use in oncology: a systematic review of literature. *J Clin Pharm Ther.* 2017;42(3):251-258. doi:10.1111/jcpt.12507
39. Conti RM, Bernstein AC, Villaflor VM, Schilsky RL, Rosenthal MB, Bach PB. Prevalence of off-label use and spending in 2010 among patent-protected chemotherapies in a population-based cohort of medical oncologists. *J Clin Oncol.* 2013;31(9):1134-1139. doi:10.1200/JCO.2012.42.7252
40. Boothby A, Wang R, Cetnar J, Prasad V. Effect of the American Society of Clinical Oncology's conflict of interest policy on information overload. *JAMA Oncol.* 2016;2(12):1653-1654. doi:10.1001/jamaoncol.2016.2706
41. Tao DL, Boothby A, McLouth J, Prasad V. Financial conflicts of interest among hematologist-oncologists on Twitter. *JAMA Intern Med.* 2017;177(3):425-427. doi:10.1001/jamainternmed.2016.8467
42. DeAngelis CD. The influence of money on medical science. *JAMA.* 2006;296(8):996-998. doi:10.1001/jama.296.8.jed60051

#### SUPPLEMENT.

**eAppendix.** Rationale of the Algorithm for Describing Uncertainty Expression and Explanation of a Proportional Odds Model

**eFigure 1.** Forest Plot for Correlates of Uncertainty Expression

**eFigure 2.** Visualization of Model for Subgroup Analysis

**eFigure 3.** Forest Plot for Correlates of Uncertainty Expression in Subgroup Analysis

**eTable 1.** Example Labeling of Comparative Efficacy of Superiority Trials on HemOnc.org

**eTable 2.** Inclusion of Marginal Initial and Subsequent Analyses

**eTable 3.** Journal Tiers as Defined in This Analysis

**eTable 4.** Scoring Algorithm for Confidence Intervals

**eTable 5.** Efficacy Labels for Trials on HemOnc.org

**eTable 6.** PMIDs and Titles of Examined Abstracts