The Scientific Impact of Positive and Negative Phase 3 Cancer Clinical Trials

Joseph M. Unger, PhD; William E. Barlow, PhD; Scott D. Ramsey, MD; Michael LeBlanc, PhD; Charles D. Blanke, MD; Dawn L. Hershman, MD

IMPORTANCE Positive phase 3 cancer clinical trials are widely hailed, while trials with negative results are often interpreted as scientific failures. We hypothesized that these interpretations would be reflected in the scientific literature.

OBJECTIVE To compare the scientific impact of positive vs negative phase 3 cancer clinical treatment trials.

DESIGN, SETTING, AND PARTICIPANTS We examined the phase 3 trial history of SWOG, a national cancer clinical trials consortium, over a 30-year period (1985-2014). Scientific impact was assessed according to multiple publication and citation outcomes. Citation data were obtained using Google Scholar. Citation counts were compared using generalized estimating equations for Poisson regression. Any trial that was formally evaluated for the randomized treatment comparison was included for analysis of publication and citation outcomes. Trials were categorized as positive if they achieved a statistically significant result in favor of the new experimental treatment for the protocol-specified primary end point. Trials were categorized as negative if they achieved a statistically significant result in favor of standard therapy or a null result with no statistically significant benefit for either the experimental or standard therapy.

MAIN OUTCOMES AND MEASURES Impact factors for the journals publishing the primary trial results, and the number of citations for the primary trial articles and all secondary articles associated with the trials.

RESULTS Ninety-four studies enrolling n = 46,424 patients were analyzed. Twenty-eight percent of trials were positive (26 of 94). The primary publications from positive trials were published in journals with higher mean (SD) 2-year impact factors (28 [19] vs 18 [13]; P = .007) and were cited twice as often as negative trials (mean per year, 43 vs 21; relative risk, 2.0; 95% CI, 1.1-3.9; P = .03). However, the number of citations from all primary and secondary articles did not significantly differ between positive and negative trials (mean per year, 55 vs 45; relative risk, 1.2; 95% CI, 0.7-2.3; P = .53).

CONCLUSIONS AND RELEVANCE The scientific impact of the primary articles from positive phase 3 randomized cancer clinical trials was twice as great as for negative trials. But when all of the articles associated with the trials were considered, the scientific impact between positive and negative trials was similar. Positive trials indicate clinical advances, but negative trials also have a sizeable scientific impact by generating important scientific observations and new hypotheses and by showing what new treatments should not be used.
Phase 3 trials provide the highest level of evidence for showing the efficacy of new treatments or interventions. The phase 3 trial programs of the National Cancer Institute's (NCI's) National Clinical Trials Network and the NCI's Community Oncology Research Program (NCORP) are vital national resources and represent a substantial investment on the part of federal agencies. Given the size of the investment, negative trials—that is, those that fail to show that a new treatment is superior to standard treatment—may be incorrectly regarded as poor investments. However, negative trials are also important if they show that new treatments, which might otherwise be adopted into clinical practice, in fact do not work.

The rate at which trials are positive (showing that a new treatment is superior to standard treatment) has previously been examined, as has the relationship between trial results and publication rates in the context of publication bias. But the comparative scientific impact of positive vs negative trials using citation data has not been investigated. In this article, we use the phase 3 trial database of SWOG, a major national cooperative group, in combination with its trial publication database and citation data from Google Scholar, to compare the scientific impact of positive and negative cancer clinical trials.

Methods

SWOG is a member of the NCI's funded National Clinical Trials Network and NCORP programs. We surveyed the SWOG clinical trial database over the 30-year period 1985 through 2014 (inclusive). Any phase 3 clinical treatment trial that opened for enrollment during the period was included. Each trial was previously approved by an institutional review board.

SWOG maintains a comprehensive database of all publications associated with its trials. All SWOG trial publications are required to be vetted through the SWOG publications office for processing and data capture prior to submission. PubMed is periodically searched by the publications office staff to identify any SWOG-related publication that may have been missed. Publication lists are periodically reviewed by SWOG disease committee chairs and senior statisticians for completeness and accuracy.

This report is based on articles published in scientific journals.

Study Outcome Definitions

A trial was defined as “completed” if it was closed to accrual. Any trial that was analyzed in the context of a formal interim analysis, or after achieving at least 50% accrual, was considered “evaluated.” An evaluated trial was defined as “full accrual” if at least 90% of protocol-specified enrollment was achieved; “futility” if the trial was closed early following a recommendation by the Data and Safety Monitoring Committee that the experimental therapy would never achieve its protocol-specified end point or was too toxic; “efficacy” if the trial was closed early following a recommendation by the Data and Safety Monitoring Committee that the trial had shown the benefit of the experimental treatment; and “sufficient” if the trial did not reach full accrual but was analyzed for the treatment comparison. Trials that were not evaluated were either closed early as a result of poor accrual or for other reasons (such as withdrawal of NCI or pharmaceutical industry support) and were counted as negative trials.

Definition of Positive and Negative Trials

To be categorized as a positive trial, trial results must have achieved a statistically significant finding in favor of the new, experimental therapy for a designated primary end point according to the statistical design prespecified in the study protocol. A trial with a negative result indicated that there was a statistically significant finding in favor of standard therapy. A trial with a null result indicated that there was no statistically significant benefit for either the experimental or standard therapy. (Of note, a null result does not equate to proving that the treatments were the same, only that there was no evidence that they were different.) For presentation purposes, both null and negative trials are regarded as failed experiments and are referred to, collectively, as negative trials.

We also considered the rate of positive trials based on an assessment of net benefit. In this calculation, trials that were positive according to the protocol-specified primary end point, but in which the toxicity for the new therapy was too high to recommend the new treatment, were considered negative. Conversely, trials that did not achieve their protocol-specified primary end point for the new treatment, but that had other substantial advantages (ie, lower cost, reduced toxicity) leading the authors to consider the new treatment an advance over standard care, were considered positive.
due to poor accrual) was unknown, so could not be categorized as positive or negative. The “primary” article was the article reporting the results of the analysis for the primary protocol-specified end point by randomized treatment arm. A “secondary” article was any other article that relied on data from a given phase 3 trial, such as subset or meta-analyses. Letters to the editor were excluded.

Scientific impact was evaluated according to multiple article publication end points. To assess the potential for publication bias, we calculated the rate at which primary articles were published in any journal. We also calculated the total number of primary and secondary articles associated with each trial. To account for time, for each trial, we calculated the publication rate as the number of publications divided by the number of years from accrual closure date through December 31, 2014. Finally, we calculated the mean impact factor related to the primary article. Because historical impact factors were not fully available, we used current (2015) scientific journal 2-year and 5-year impact factor levels.

Citation Analysis
We also examined scientific impact using citation analysis, a bibliometric tool.6,7 We created a database of the total number of articles in the biomedical literature that cited the SWOG articles in each year following its initial publication using Google Scholar, a web-based scholarly search engine. Citation data were obtained in July 2015 for citations that occurred through December 31, 2014. The Web of Science search engine was also used to conduct a sensitivity analysis for primary article citation data.

Statistical Considerations
The total number of primary and secondary publications associated with each trial and the mean impact factor were compared between positive and negative trials using t tests. Secondary articles (and their citation counts) associated with multiple trials were ascribed a weight equal to the inverse of the number of associated trials.

The citation rates for positive and negative trials are a function of the number of trials in each category and the number of years since publication for each trial article. Also, yearly citation counts are likely correlated; that is, an article is more likely to be cited if it has previously been cited. For these reasons, we modeled citation counts using generalized estimating equations (GEEs) for Poisson regression.8,9 The GEE models fit the number of citations by study over time (in yearly intervals). To account for correlation between citation counts, we specified an independence working correlation using robust standard errors. An article was considered “at risk” of being cited in the literature starting the calendar year in which it was published until the end of the period on December 31, 2014. For unpublished trials, the primary manuscript date was set at the year when enrollment completed, and citation counts were assigned as zero for all at-risk years. Citation counts up to 20 years after the primary article publication were analyzed. Model-based expected citation counts for positive vs negative trials are provided, as are relative risk (RR) estimates and their 95% confidence intervals.

Results
One-hundred seven phase 3 trials were activated from 1985 through 2014. Of these, 6 finished enrollment but the data were still maturing, and 7 were still open for enrollment (eFigure 1 in the Supplement). Therefore, n = 94 studies involving 46,424 patients were examined. Trial characteristics are shown in eTable 1 in the Supplement. Breast, gastrointestinal, genitourinary, leukemia, lung, and lymphoma cancer trials were most common (≥10% each). Nearly all trials had superiority design (90 [96%]) and included evaluation of systemic therapy (90 [96%]). The majority were designed to assess either survival alone (35 [37%]) or some other time-to-event end point alone (26 [28%]). There were no observed differences between positive and negative trials in factors that might influence citation rates, such as the incidence of the cancer type or the nature of the treatment.

Accrual Outcomes
Among the 94 completed studies, nearly half reached full accrual (46 of 94 [49%]), 5 (5%) closed early because of an efficacy finding and 10 (11%) because of a futility finding, and 5 (5%) were analyzed after reaching sufficient accrual (eFigure 1 in the Supplement). Among the trials closed early for futility, 9 closed because the experimental treatment was worse—or would never be better—than standard treatment, and 1 because the experimental treatment was too toxic. In no case did a study reach full accrual and the experimental treatment was found to be statistically significantly worse than standard treatment. Twenty-six trials (28%) were closed early because of poor accrual.

Rate of Positive Phase 3 Trials
Overall, 26 of 94 completed trials (28%) were positive according to the protocol-specified primary end point analysis. Given the likelihood that trials that fail are more likely to fail quickly, we excluded 7 trials activated in the most recent 10 years (2005-2014); the rate of positive trials was 30% (26 of 87). Among the 66 evaluated trials, the rate of positive trials was 39% (26 of 66).

Two trials were positive according to the protocol-specified primary end point, but the new therapy was sufficiently toxic that the authors did not recommend it over standard therapy. Four other trials did not achieve their...
protocol-specified primary end point but were considered “positive” by the study authors based on other treatment benefits. Therefore, in 28 of 94 trials (30%), the overall risk-benefit profile was considered to be in favor of the new treatment.

Publication Outcomes
Among the n = 66 trials that were evaluated for their primary end point, 38 of 40 negative trials had an article publication (95%) and all 26 positive trials had an article publication (100%). There was no statistically significant evidence of publication bias (P = .25).

In total, 273 unique articles were published during the period, including 229 articles associated with a single trial and 44 articles associated with 2 or more trials. Among the 209 secondary articles, the most common types pertained to biomarker predictors of outcomes (52 [25%]); clinical, treatment, or demographic predictors of outcomes (52 [25%]); and investigations of cancer biological characteristics (43 [21%]). Twenty-four of the secondary articles (11%) were from collaborative databases or meta-analyses with other research groups.

Neither the mean (SD) number of publications (3.7 [4.2] vs 4.4 [5.8]; P = .57) nor the publication rate relative to the number of years since trial accrual completion date (0.23 [0.26] vs 0.35 [0.37] per year; P = .16) differed between positive vs negative trials (respectively).

The mean (SD) 2-year impact factor was higher for the primary article publications of positive vs negative trials (28 [19] vs 18 [13]; P = .007). Results were similar for the 5-year impact factor. Also, the proportion of studies published in very high impact factor journals (JAMA, Lancet, New England Journal of Medicine) was higher for positive trials (9 of 26 [35%]) than for negative trials (5 of 40 [13%]; P = .03).

Citation Analysis
For the 66 evaluated trials, there were 24,235 citations over 783 at-risk years for the primary articles, including 15,064 citations over 350 years for positive trials and 9171 citations over 433 years for negative trials (Figure 1A).

The model-adjusted citation rate per at-risk year was 43 for positive trials and 21 for negative trials. In multilevel Poisson regression analysis, this difference was statistically significant (RR, 2.0 [95% CI, 1.1-3.9]; P = .03). Ignoring autocorrelation does not change the effect size but does reduce the standard errors. In this setting, the strength of the difference was much greater (RR, 2.03 [95% CI, 1.98-2.09]; P < .001). Adjusting for the journal impact factor as covariate, results were consistent but weaker (RR, 1.5 [95% CI, 0.9-2.4]; P = .12). There was no evidence that the pattern of higher citation counts for positive trials differed according to whether the primary article was published in a high-impact journal (rates for positive vs negative trials: 91 vs 67 in high impact factor journals, and 24 vs 15 in low impact factor journals; interaction P = .75).

The best-fit model included both a linear and quadratic interaction of time and whether the study was positive or negative (eTable 2 in the Supplement). This can be interpreted as indicating that there was a quadratic relationship between citation counts and time that differed for the positive vs negative trials. For positive trials, mean citation counts increased for approximately 10 years after initial publication, after which they decreased, at approximately a similar rate, out to 20 years; for negative trials, citation counts also increased out to approximately 10 years, but at a more modest pace, after which mean citation counts slowly decreased (Figure 2A).

For secondary articles, citation counts were higher for negative trials (12,461 vs 6029) (Figure 1B), while for the combination of primary and secondary articles, citation counts were similar for positive vs negative trials (21,093 vs 21,632, respectively) (Figure 1C). The model-adjusted citation rate per at-risk year was 55 for positive trials and 45 for negative trials (relative risk, 1.2 [95% CI, 0.7-2.3]; P = .53) (Figure 2B). The absence of a statistically significant overall difference between positive and negative trials was due to the substantial number of citations from secondary publications for negative trials. As before, the best-fit model included a quadratic interaction of time and whether the study was positive or negative (eTable 2 in the Supplement), although the strength of the interaction was weaker than for the primary article citations.

Sensitivity and Outlier Analyses
Using the Web of Science search engine, the total absolute number of citations for primary articles was lower (n = 15,828), including 9749 for positive trials and 6079 for negative trials. The model-adjusted citation rate for positive trials was 28 and for negative trials was 14 (RR, 2.0 [95% CI, 1.0-3.8]; P = .04). As before, a quadratic interaction model best fit the data (data not shown).

We assessed the potential for very highly cited positive and negative trials to influence the analysis of total primary and secondary article citations by excluding the 2 top positive and negative trials with the greatest number of citations from all associated articles. The simple citation rate per at-risk year was 41 for positive trials and 36 for negative trials (RR, 1.1 [95% CI, 0.6-2.0]; P = .66) (eFigure 2 in the Supplement). The observed modest early increase in total citations for negative trials (prior to year 0) (Figure 2B) is no longer evident, suggesting that the early difference was due to a few negative trials with a large amount of well-cited secondary article publications.

Discussion
We found that the scientific impact of primary articles from positive trials was greater than for negative trials, as measured by publication in a very high impact journal and mean citation rates over 20 years. But negative trials also had a sizeable impact on the scientific literature. When the citation counts from both the primary and secondary publications were considered, the total scientific impact between positive and negative trials was roughly comparable.

Recent evaluations of NCI-sponsored cooperative group trials showed high rates of full publication of completed trials in journals (≥90%).11,28 Our findings were similar, showing an overall publication rate of 97%, which did not statistically significantly differ according to whether trial results were positive or negative. Earlier observations about the potential for publication bias in clinical trials, and the subsequent require-
ment that all National Institutes of Health–funded clinical trials be registered through http://www.clinicaltrials.gov, have likely helped to reduce publication bias in NCI-sponsored cooperative group trials.2,11,12

Figure 1. Number of Citations by Calendar Year

A Primary article citations

Positive trials: total citations = 15,064
Negative trials: total citations = 9,171

B Secondary article citations

Positive trials: total citations = 6,029
Negative trials: total citations = 12,461

C Primary and secondary article citations

Positive trials: total citations = 21,093
Negative trials: total citations = 21,632

Calendar year was truncated at 1995 for presentation purposes. The number of publications that were at risk of citation are shown below each panel.
Overall, randomized phase 3 trials generated positive results at a rate of 28%, consistent with prior evidence and consistent with the idea that important scientific questions should have equipoise.\textsuperscript{1,4,13} That is, a comparative phase 3 trial should only be conducted in an environment in which true uncertainty exists as to whether a new treatment is better than standard treatment. In this setting, the likelihood that an individual trial will be positive should lie within a certain range; if the likelihood of success is too high or too low, then patients and clinicians are unlikely to entrust their treatment choice to random assignment.\textsuperscript{13} Moreover, if the scientific question does have equipoise, then negative trial results reduce uncertainty about which treatments should direct guideline-based care. Thus, their publication serves a vital scientific and societal interest. It is also essential to publish negative trial results given the costliness of trials; the high probability that the clinical experiment is unlikely to be repeated (or, conversely, to reduce the risk that a similar trial is repeated due to lack of knowledge of a negative trial’s results); and importantly, in fulfillment of the obligation to the trial participants who volunteered to participate.

The citation trends observed in this study point to patterns of advances in treatment discovery. Positive and negative trials achieved their maximum citation rates at approximately 10 years after publication. Thereafter, citation counts diminished, likely because new treatments were developed to replace prior standards. However, even 20 years after publication, both positive and negative trials remained frequently cited (15–20 per year), suggesting that successfully completed phase 3 trials—regardless of their outcome—set the stage for new treatment discovery for decades to come.

These findings also point to the tremendous secondary value of well-conducted phase 3 clinical trials. Fully 43% of total citations (18 490 of 42 725) in this set of trials were from secondary articles. Clinical trial participants are typically younger and healthier than patients who do not participate in trials.\textsuperscript{14,15} However, randomized phase 3 trials provide large samples of uniformly staged and treated patients with prospective data collection and long-term follow-up. In an age of translational medicine research, nearly all trials also collect biologic ancillary data to investigate and generate important hypotheses regarding prognostic and predictive biomarkers. Finally, patients who receive treatment on trials have access to health care in a protocol-specified treatment setting, so studies of health outcomes and health care use by demographic or socioeconomic groups using clinical trial cohorts are less subject to confounding by access to care.\textsuperscript{16} Indeed, in some cases a negative trial may actually be more advantageous as a data resource for secondary science, because the treatment effect can also be discounted and the full trial data set used. In total, trial data represent a powerful resource for well-designed secondary data analyses. The power of this data resource is increasingly recognized, as reflected by a recent Institute of Medicine report,\textsuperscript{17} by the mission of the new Cancer Care Delivery Research program of the NCI’s NCORP research base,\textsuperscript{18} and by the increasing efforts of pharmaceutical companies to advance their trial data sharing.\textsuperscript{19,20}

This study was potentially limited by the fact that citation data are likely incomplete, although Google Scholar has been estimated to be at least as comprehensive as other scholarly search engines.\textsuperscript{21,22} The use of Google Scholar for citation data also provided a different profile of citation data than Web of Science.\textsuperscript{23,24} Although this affected absolute citation counts, it did not affect the relative citation patterns between positive and negative trials. In addition, a given citation does not represent a uniform level of scientific impact, and citation analysis may not fully represent the scientific impact of a given trial. As such, we defined scientific impact as a multidimensional construct incorporating multiple publication end points.\textsuperscript{25,26} Other ways to assess scientific impact are also possible, such as usage log data.\textsuperscript{27} Finally, these data represent only a single cooperative group. Publication patterns for other cooperative cancer groups of the NCI or for noncooperative group trials may be different.
Conclusions

Positive phase 3 trials indicate clinical advances, and as such their scientific impact as reflected in the literature is substantial. But well-designed and conducted negative trials also have a sizeable scientific impact by generating important scientific observations and new hypotheses and by showing what new treatments should not be used. Inevitably, the trials with the least scientific impact are those that close early as a result of poor accrual or other issues. This suggests the importance of designing trials with strong support in the scientific and treatment communities. Once a trial is successfully completed, the totality of its scientific impact, considering both primary and secondary results, promises to be substantial, regardless of whether the trial results are positive or negative.

ARTICLE INFORMATION

Accepted for Publication: December 22, 2016.
Published Online: March 10, 2016. doi:10.1001/jamaoncol.2015.6487.

Author Contributions: Dr Unger had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis. Study concept and design: Unger, Barlow, Ramsey, LeBlanc, Hershman. Acquisition, analysis, or interpretation of data: Unger, Barlow, Blanke, Hershman. Drafting of the manuscript: Unger. Critical revision of the manuscript for important intellectual content: All authors. Statistical analysis: Unger, Barlow, LeBlanc. Obtained funding: Unger. Administrative, technical, or material support: Unger, Blanke. Study supervision: Unger, Ramsey, Hershman.

Conflict of Interest Disclosures: None reported.

Funding/Support: This research was supported in part by the Dr Charles A. Coltman, Jr, Fellowship Program of the Hope Foundation; and by the National Institutes of Health, National Cancer Institute, NCI Community Oncology Research Program (NCORP) Research Base grant U10CA189974-01.

Role of the Funder/Sponsor: The funding organizations 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.

REFERENCES


