Evidence-based prior for estimating the treatment effect of phase III randomized trials in oncology

Sherry, A.D.; Msaouel, P.; Kupferman, G.S.; Lin, T.A.; Jaoude, J.A.; Kouzy, R.; ... ; Zwet, E. van

Evidenced-Based Prior for Estimating the Treatment Effect of Phase III Randomized Trials in Oncology

Alexander D. Sherry, MD[1]; Pavlos Msaouel, MD, PhD[2,3]; Gabrielle S. Kupferman, BA[1]; Timothy A. Lin, MD, MBA[4]; Joseph Abi Jaoude, MD[5]; Ramez Kouzy, MD[1]; Zachary R. McCaw, PhD[6,7]; Ethan B. Ludmir, MD[8,9]; and Erik van Zwet, PhD[10]

DOI https://doi.org/10.1200/PO.24.00363

ABSTRACT

PURPOSE The primary results of phase III oncology trials may be challenging to interpret, given that results are generally based on *P* value thresholds. The probability of whether a treatment is beneficial, although more intuitive, is not usually provided. Here, we developed and released a user-friendly tool that calculates the probability of treatment benefit using trial summary statistics.

METHODS We curated 415 phase III randomized trials enrolling 338,600 patients published between 2004 and 2020. A phase III prior probability distribution for the treatment effect was developed on the basis of a three-component zero-mean mixture distribution of the observed z-scores. Using this prior, we computed the probability of clinically meaningful benefit (hazard ratio [HR] <0.8). The distribution of signal-to-noise ratios and power of phase III oncology trials were compared with that of 23,551 randomized trials from the Cochrane Database.

RESULTS The signal-to-noise ratios of phase III oncology trials tended to be much larger than randomized trials from the Cochrane Database. Still, the median power of phase III oncology trials was only 49% (IQR, 14%-95%), and the power was <80% in 65% of trials. Using the phase III oncology-specific prior, only 53% of trials claiming superiority (114 of 216) had a ≥90% probability of clinically meaningful benefits. Conversely, the probability that the experimental arm was superior to the control arm (HR <1) exceeded 90% in 17% of trials interpreted as having no benefit (34 of 199).

CONCLUSION By enabling computation of contextual probabilities for the treatment effect from summary statistics, our robust, highly practical tool, now posted on a user-friendly webpage, can aid the wider oncology community in the interpretation of phase III trials.

ACCOMPANYING CONTENT

Data Supplement

Accepted August 28, 2024
Published October 2, 2024

JCO Precis Oncol 8:e2400363

INTRODUCTION

The interpretation of modern phase III randomized trials in oncology is a considerable challenge.[1] The standard approach for estimating comparative survival advantages is to compute hazard ratios (HRs) and their 95% CIs, with most trials declaring superiority of an experimental intervention on the basis of *P* value thresholds.[2] However, 95% CIs and *P* values are widely misinterpreted. 95% CIs are often misunderstood as having 95% probability of containing the true effect, and *P* values are often mistaken as the probability of no difference.[3-5] *P* value thresholds may lead to both overestimation and underestimation of effects, particularly in scenarios where power is lower than planned.[6-9] For example, a significant *P* value (eg, *P* <.05) in a trial designed with 80% power does not imply an 80% probability that the experimental treatment was beneficial. Therefore, novel tools to improve the interpretation of primary outcomes in oncology trials are sorely needed. Some have proposed directly computing the probability of benefit (eg, HR <1) using Bayesian approaches, because the probability of whether an intervention is helpful or harmful is more intuitive to oncologists and patients than interpreting *P* values.[10-13] However, such calculations require specification of prior knowledge, which may appear controversial because of apparent subjectivity even when guided by domain expertise.[11,14,15]

There is a considerable need for a straightforward, data-driven approach to estimating the probability of benefit in phase III trials, including in the absence of individual-level patient data, which are often difficult for clinicians to access. Here, we propose a user-friendly and evidence-based solution for estimating the probability of whether new

CONTEXT

Key Objective

We sought to derive an informative prior distribution for the treatment effect of phase III oncology trials to facilitate an objective means to calculate posterior probabilities for the primary outcomes of randomized oncology trials.

Knowledge Generated

Using the summary statistics of 415 phase III trials, we fit a mixture model for the true (unobserved) treatment effect distribution by deconvoluting the distribution of z-scores into the signal-to-noise ratio distribution followed by scaling of the SE. We used this mixture model as a prior distribution to estimate the probabilities of clinically relevant effect sizes from this data set of phase III trials.

Relevance

Oncologists, researchers, and other stakeholders may readily estimate the probability of treatment effects at various effect sizes of interest with the derived prior by means of a user-friendly webpage (shinyapps.io).

oncology treatments tested in phase III trials are effective using standard trial-level summary statistics. The purpose of the present meta-epidemiologic study was to develop an informative, oncology-specific default prior, derived from the distribution of the z-scores obtained from 415 contemporary phase III oncology trials. This default prior can be used by practicing oncologists to interpret historical, current, and future phase III oncology trials.

METHODS

Institutional review board approval for this meta-epidemiologic study was not needed because of the public availability of the trial data. Trials were screened from ClinicalTrials.gov in February 2020 using the advanced search terms "cancer," phase "phase three," study results "with results," and status "excluded: not yet recruiting." Trials were required to be phase III with two-arm, superiority designs that tested anticancer interventional strategies (Fig 1). Trials that had not published their primary end point were excluded, as well as trials that did not use time-to-event primary end points.

Primary end point summary statistics (HRs and 95% CIs) were recorded. The control arm was taken as the reference for all comparisons, such that HR <1 always favored the experimental arm, and in cases where the experimental arm was set as the reference, reciprocals for the HR were used. For trials with multiple coprimary end points, the time-to-event primary end point with a reported 95% CI was used. If 95% CIs were used for all time-to-event co-primary end points, overall survival (OS) was used because of its intrinsic value and potential advantages compared with surrogate end points.[2,16,17]

In previous work, an informative prior for addressing treatment effect exaggeration was developed using 23,551 randomized trials in the Cochrane Database of Systematic Reviews.[9,18-21] In this study, we applied this methodology for the creation of a phase III, oncology-specific prior distribution of the treatment effect.[18-21] In brief, the z-score for each randomized clinical trial (RCT) was computed as the point estimate for the log HR (log HR) divided by the SE, computed by taking the difference between the log bounds of the 95% CI for the HR divided by twice the qnorm of 0.975.[20] Recall that the *P* value is <.05 when the z-score is >1.96 or <−1.96. Similar to the Cochrane-based prior distribution, we then fit a zero-mean three-component normal mixture to the z-scores obtained from the log HR.[21] We used three components because of the smaller number of phase III oncology trials compared with the four-component mixture used for the Cochrane-derived prior distribution. We chose a zero-mean mixture to set the prior probability of any benefit to be 50%-50%. In this way, neither treatment arm was favored in the prior distribution. Although the z-score is the ratio of the estimated treatment effect to the SE, the signal-to-noise ratio (SNR) is defined as the ratio of the true treatment effect to the SE. Because the z-score represents the sum of the SNR and a standard normal error term, the deconvolution trick was applied to obtain the distribution of the SNR by subtracting 1 from the variances of each mixture component.[19] It is quite remarkable that it is possible to obtain the distribution of the SNR because the true effect cannot be observed. The distribution of the power was then obtained as a transformation of the SNR.[9] Note that this is the power against the (unobserved) true effect, and not the so-called post hoc power, or the power against the effect that was assumed in the sample size calculation.[22] Finally, the prior for the treatment effect was obtained by scaling the distribution of the SNR by the SE.[18] The underlying data set and code for the development of the prior are provided in the Data Supplement.

On the basis of guidelines from the ASCO, the minimum clinically important difference (MCID) was defined by

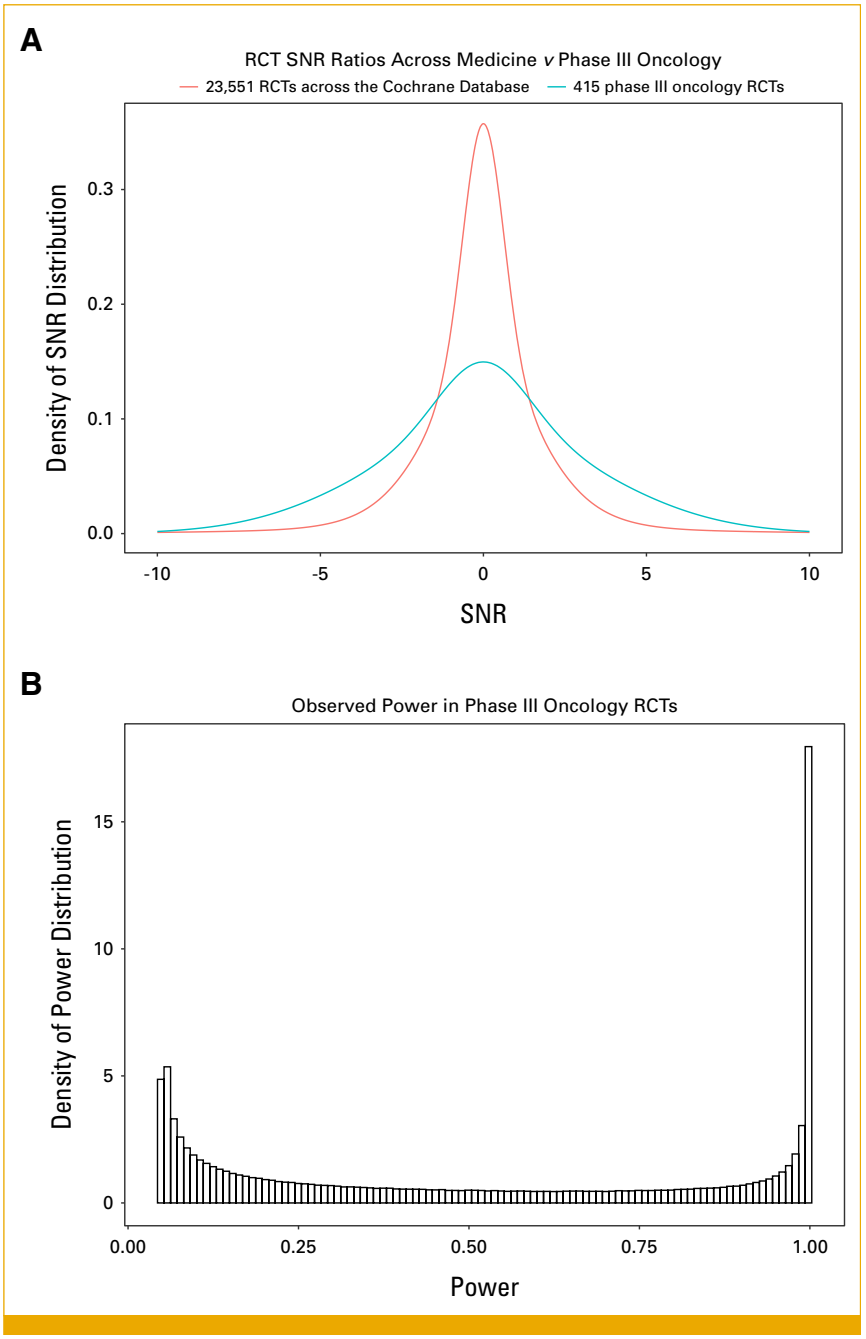

FIG 1. Distribution of the SNR and power of the primary outcomes of phase III RCTs in oncology. (A) The SNR in phase III oncology trials tended to be much larger than that of randomized trials in the Cochrane Database. (B) Estimated distribution of power against the true effect in phase III oncology RCTs. RCT, randomized clinical trial; SNR, signal-to-noise ratio.

HR <0.8.[23-26] The probability of a detectable effect was defined by HR <1. The probability for both hypotheses was computed for each trial. Analyses and plots were completed in R v.4.3.2 (Vienna, Austria) and Prism v10 (La Jolla, CA).[27]

RESULTS

After screening 785 phase III randomized trials from ClinicalTrials.gov, we included 415 two-arm, superiority-design, interventional, therapeutic, time-to-event, phase III trials (Data Supplement, Fig S1). Publication dates of the primary end point ranged from 2004 to 2020, with a total of 338,600 patients enrolled. Most trials studied metastatic solid tumors (n = 263, 63%), and most used surrogate primary end points (n = 250, 60%; Table 1). Superiority was claimed for the experimental arm in 216 trials (52%) and was not claimed in 191 trials (46%); in eight trials, inferiority of the experimental arm was claimed (2%).

TABLE 1. Characteristics of Trials Included in the Analysis

Characteristic	No. (%)
Total trials	415
Disease stage	
Solid, nonmetastatic	87 (21)
Solid, metastatic	263 (63)
Hematologic	65 (16)
Disease site	
Breast	76 (18)
GI	66 (16)
Genitourinary	58 (14)
Hematologic	65 (16)
Thoracic	79 (19)
Other[a]	71 (17)
Treatment modality	
Systemic therapy	404 (97)
Local therapy	11 (3)
Cooperative group study	88 (21)
Industry sponsored	361 (87)
Enrolled patients, No., median (IQR)	596 (377-903)
Publication year, median (IQR)	2015 (2012-2017)
Primary end point	
OS	165 (40)
Surrogate	250 (60)
Primary outcome	
Superiority shown for the experimental arm	216 (52)
Superiority not shown for the experimental arm	191 (46)
Inferiority of the experimental arm	8 (2)

[a]Other disease sites included CNS, endocrine, gynecologic, head and neck, pediatric, sarcoma, and skin.
Abbreviation: OS, overall survival.

The absolute z-scores of phase III oncology trials tended to be much larger than those of 23,551 RCTs from the Cochrane Database ($P < .0001$, Mann-Whitney test; Data Supplement, Fig S2). This implies that both the SNR and the power of phase III oncology RCTs also tend to be much larger than that of general RCTs (Fig 1A). Z-scores appeared to be stable over time, where time was defined as the trial publication year (Data Supplement Fig S3). A zero-mean mixture of three normal distributions provided a reasonable fit to the z-scores of the phase III trials (Data Supplement, Fig S4). The proportions and variances of each subcomponent are reported in the Data Supplement (Table S1). We derived a zero-mean mixture of three normal distributions for the SNR by subtracting one from each of the variances (Data Supplement, Table S1). From the distribution of the SNR, we derived the distribution of the power. We find that the median power of phase III oncology RCTs was only 49% (IQR, 14% to 95%), with an average power of 52% (Fig 1B). An estimated 65% of trials had power <80%, with 71% of trials <90%. The power was >95% in an estimated 25% of trials. As previously reported, the power of RCTs from the Cochrane Database tended to be even lower (median power: 13%, with 78% of trials <80% power).[9] The SNR distribution was scaled by the observed SE to derive the prior for the treatment effect in a particular trial (Data Supplement, Fig S5). Examples of resulting priors are plotted in the Data Supplement (Fig S6).

The probabilities of a detectable effect (HR <1) exceeded 90% for all 216 trials that claimed superiority (Fig 2). However, only 53% of trials with superiority claims (114 of 216) had ≥90% probability of achieving the MCID (Table 2). Conversely, for the 199 trials that did not claim superiority, the median probability that the experimental arm had an HR <1 was 63% (IQR, 32%-86%; Fig 2). In 17% of trials that did not claim superiority (34 of 199), the probability that HR was <1 exceeded 90% (Table 2). Consistent with the differences in the SNR between phase III oncology RCTs and RCTs from the Cochrane Database, posterior probabilities computed by the Cochrane Database prior appeared to be overcorrected compared with the phase III oncology-specific prior (Fig 3).

A webpage has been created for users to compute hypothesis probabilities at a given level of HR on the basis of the trial's summary statistics (shinyapps.io),[28] with an illustration shown in Figure 4.

DISCUSSION

Using the distribution of the z-scores of the primary end points of 415 phase III oncology randomized controlled trials, to our knowledge, this study is the first to compute an evidenced-based default prior specifically designed to estimate the effects of phase III oncology trials. This prior has been deployed in a standalone webpage application that allows users to input the summary Cox regression statistics of a phase III trial and compute the posterior probabilities of benefit at any level of HR. By providing oncologists with the means to directly compute probabilities of interest, this study provides a robust, highly practical tool to immediately enhance the interpretation of phase III trials throughout the wider oncology community.

Consistent with previous work, we found that the actual power of most phase III trials is low relative to the power specified during trial design.[9,29] Lower power increases the risk of false-negative findings. When an underpowered trial does reach statistical significance, the effect is usually overestimated and leads to replication failure. Directly computing the probability of benefit using our phase III prior provides a more intuitive method of understanding and interpreting the uncertainty associated with underpowered trials. The consequences of relying on *P* value thresholds in underpowered trials are directly manifested in our finding that the experimental arms of 17% of trials interpreted as negative or inconclusive had >90% probability of superiority to the control arm.

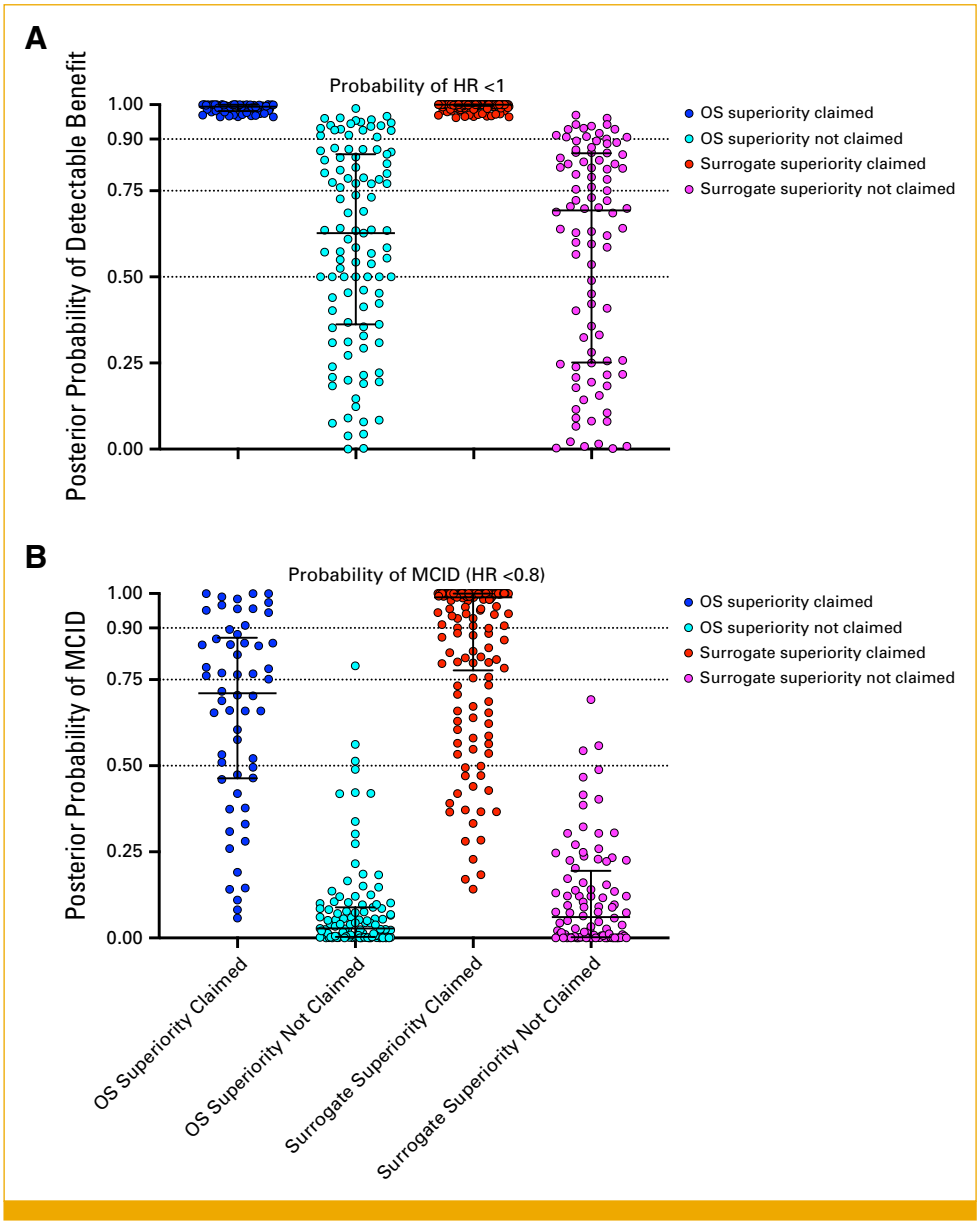

FIG 2. Posterior probabilities of primary end points of 415 phase III trials computed using the phase III oncology prior. Probabilities are grouped according to end point type (OS or surrogate survival) and the trial result interpretation (claim for superiority or not). (A) The median and interquartile range are shown for each group. Probability that the HR is <1 in favor of the experimental arm. (B) Probability that the experimental arm shows superiority according to the MCID, defined as HR <0.8. HR, hazard ratio; MCID, minimum clinically important difference; OS, overall survival.

Rather than specifying the beliefs of an effect for a single unique treatment, this study offers an objective approach to specifying the prior probability distribution by using SNRs that have been observed in phase III oncology trials. In the absence of specific beliefs regarding a unique treatment, this prior probability distribution for the treatment effect may serve as a reasonable default prior. The utility of this prior may be especially pertinent when Bayesian posteriors are not computed in the trial publication, which is the case for the majority of trials at present. Furthermore, this prior may be particularly helpful when individual patient-level data are not made readily available for reanalysis. In a separate study, we manually reconstructed individual patient-level data for the primary outcomes of 230 trials and found that the posteriors computed from individual patient-level data using conventional priors had a high degree of concordance with those computed using the prior proposed in this study.[30] Notwithstanding these advantages, in some cases or in sensitivity analyses, using priors unique to a certain treatment, rather than a default prior as proposed in this study, may be desirable. However, the selection of a prior under such circumstances has inherent subjectivity, representing

TABLE 2. Probabilities of a Detectable Effect (HR <1) and Achieving a MCID (HR <0.8) in Phase III Oncology RCTs, Computed by a Phase III, Oncology-Specific Prior

	RCTs Grouped by Trial Result Interpretation, No. (%)	
Posterior Probability	Superiority Claimed for the Experimental Arm (n = 216)	Superiority Not Claimed for the Experimental Arm (n = 199)
MCID (HR <0.8)		
≥90%	114 (53)	0 (0)
≥75%	149 (69)	1 (0.5)
≥50%	180 (83)	6 (3)
Detectable effect (HR <1)		
≥90%	216 (100)	34 (17)
≥75%	216 (100)	82 (41)
≥50%	216 (100)	130 (65)

Abbreviations: HR, hazard ratio; MCID, minimum clinically important difference; RCT, randomized clinical trial.

one of the primary criticisms of Bayesian methods, and may be less robust or less informative compared with the prior proposed in this study, which was derived from the data of 415 trials. Nonetheless, it is important to note that current tools using the methods of Wijeysundera et al are available to facilitate this evaluation with published trial summary statistics.[31,32] Taken together, this study represents a considerable advance and adds robust, objective, and unique value to the interpretative tools available to oncologists.

Previously, an informative prior, on the basis of thousands of clinical trials in the Cochrane Database, was proposed.[18-21] However, we found that the distribution of SNRs of RCTs in the Cochrane Database tend to be lower than that of phase III oncology trials, consistent with the observation that phase III oncology trials are often larger than general medical RCTs, leading to more power. The use of a phase III–specific, oncology-specific prior improves the robustness of computed posterior probabilities by reducing the risk of overcorrection, as suggested by our comparison of posterior probabilities for the MCID. These findings thus support the stated rationale for using a separate, dedicated phase III oncology prior, namely that phase III trials, compared with general medical RCTs, are more typically large-scale, multicenter, multinational, and most likely to change practice and lead to regulatory approvals. That being said, low SNRs were observed even for phase III oncology trials, which can lead to upward bias in the estimate of HR.[21] Application of the phase III oncology prior distribution of treatment effect to trials, especially those with low SNR, may partially reduce this bias, and posterior mean estimates for HR are computed as part of the provided webpage.

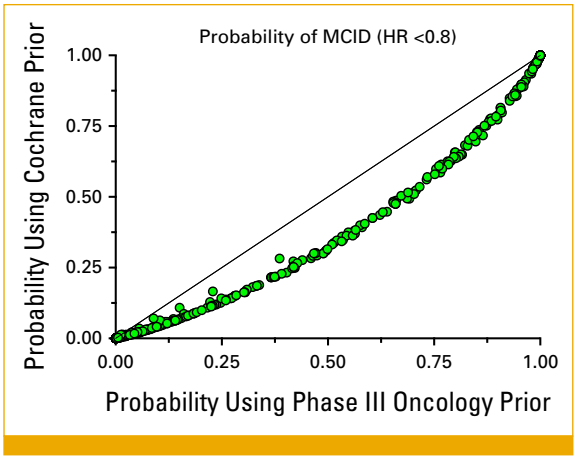

FIG 3. Comparison of posterior probabilities computed by the phase III oncology-specific prior versus the prior from the Cochrane Database of RCTs for the MCID (HR <0.8). HR, hazard ratio; MCID, minimum clinically important difference; RCT, randomized clinical trial.

To illustrate the potential value of estimating the probability of benefit, consider the results of two example phase III trials, the GEMPAX trial and CALGB 30610 (Alliance)/RTOG 0538, neither of which was included in the development of the phase III prior as both trials were recently published.[33,34] The GEMPAX trial compared second-line gemcitabine with paclitaxel versus gemcitabine alone for patients with metastatic pancreatic ductal adenocarcinoma.[33] GEMPAX showed an improvement in progression-free survival (HR, 0.64 [95% CI, 0.47 to 0.89]) and overall response rate, but interpreted the primary end point of OS as statistically negative (HR, 0.87 [95% CI, 0.63 to 1.20]) on the basis of a large *P* value (.41). Importantly, large *P* values do not support the null hypothesis and, in fact, provide little information.[4] Using the phase III prior developed in this study, the probability that gemcitabine plus paclitaxel is associated with better OS (HR <1) than gemcitabine alone is 78%, and is similar to the 75% probability that there is no clinically meaningful difference (0.8< HR <1.25) between the two treatments (Fig 4). This highlights that the OS results lacked the power and precision to make reliable assertions regarding treatment efficacy or lack thereof. Conversely, the CALGB 30610 (Alliance)/RTOG 0538 RCT noted a HR of 0.94 (95% CI, 0.76 to 1.17, *P* = .594) for OS among patients with limited-stage small cell lung cancer who underwent once-daily radiation compared with those who underwent

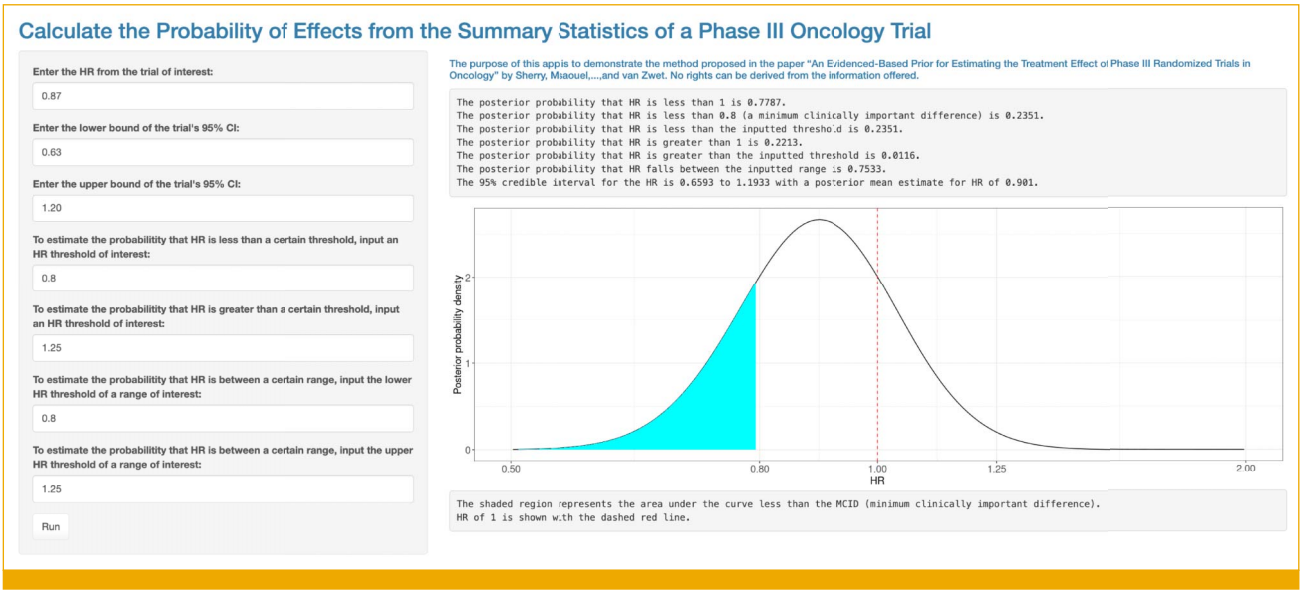

FIG 4. Illustration of the standalone webpage[28] facilitating a user-friendly approach to estimating the posterior distribution from a phase III oncology trial of interest using the proposed informative prior. After entering the summary statistics from the published trial's Cox regression (HR and its 95% CI), the user can calculate the posterior mean estimate and its 95% credible interval, visualize the posterior distribution, and estimate the posterior probability of effects above or below a threshold of interest as well as between a range of interest. Posterior probabilities for HR <1, HR >1, and HR <0.8 (which is often considered as a MCID) are provided as a default. For the posterior distribution plot, a dashed line indicates HR = 1, and the AUC for HR <0.8 is shaded in blue. HR, hazard ratio; MCID, minimum clinically important difference.

standard-of-care twice-daily radiation.[34] Using our phase III prior, we can see that the probability that the two treatments yielded no clinically meaningful difference (0.8< HR <1.25) was 95%. Therefore, although both RCTs yielded a similar *P* value, only RTOG 0538 had adequate power to conclusively determine no clinically meaningful difference. Notably, the SE of the natural logarithm of HR for RTOG 0538, 0.11, was less than that of GEMPAX, 0.16. A trial with a smaller SE is expected to have a narrower prior for the treatment effect because treatment effect and SE are inversely related. For example, to obtain 80% power in a trial, the treatment effect must be 2.8 times the SE. Accordingly, in this example, GEMPAX was powered for a larger treatment effect (HR of 0.625) compared with RTOG 0538 (HR of 0.77).[33,34] Trials expecting small treatment effects and choosing large sample sizes to reduce the SE will result in narrower priors for the treatment effect, because the trialists themselves suspected smaller treatment effects. This distinction, made manifest in the computation of posterior probabilities by our prior distribution, highlights an example of the importance of comprehensively evaluating RCT results using our provided webtool. Importantly, probabilities of benefit do not represent rules for decision making.[1] Inference, obtained from data such as this study, must be applied by the oncologist to each clinical scenario in light of the risks, alternatives, patient characteristics, and patient values.[25,35] Nonetheless, the additional information provided by posterior computation can facilitate a more informed and data-driven approach to clinical care.

There are some important limitations to consider in this study. First, we assumed that the Cox regressions that formed the basis of the primary analysis of each phase III trial met their underlying statistical assumptions, including proportional hazards; however, this may not have always been the case.[36,37] In general, we suggest that trials should report summary statistics that are interpretable and make as few assumptions as possible.[38] Second, there may have been underlying variation in the approach of each trial toward computing the primary end point; Cox models fit with prognostic covariates are likely more precise than univariable models and this may have influenced the resultant SNRs that were used to create the prior.[39,41] Third, phase III trials that were not published because of nonsignificant results, which was estimated in one study to be as high as 7% of trials, may have resulted in a file drawer effect and influenced our treatment effect distribution, as only published studies were included.[40] Fourth, our prior was specifically fit to the primary end points of phase III trials, which is both a strength and limitation. Importantly, this prior does not assume homogeneity of treatment effects, but represents general information about the SNRs of phase III trials of oncology conducted between 2004 and 2020. Although the z-score distribution appeared stable during the time period these trials were published, if the SNRs of phase III oncology trials were to meaningfully change in the coming decades, the robustness of the prior may be lessened. However, the foundation provided by this study, including public availability of the key data and methods, facilitates ready updates

to the distribution of the informative prior if the underlying basis of the prior were to meaningfully change, consistent with the general principles of Bayesian epistemology. Furthermore, the distribution of SNRs of phase III oncology trials appeared meaningfully different from the distribution of SNRs from general medical RCTs in the Cochrane Database, as well as subsequent posterior probability estimations, thus establishing the need for a separate, phase III oncology-specific prior. Consequently, other trials or end points, such as phase II trials or even secondary end points of phase III trials, are expected to have lower SNRs. Our prior is not appropriate for these cases because they lack exchangeability with the trials included in our analysis. Similarly, the prior is not as relevant to phase III trials conducted in other fields of medicine because it was derived and developed exclusively from oncology trials. Fifth, although we defined the MCID as HR <0.8 as an illustration of the output of this tool, the MCID may be also be context-specific.[24] Using the provided webpage, any interested user can compute the probability of MCID according to their own definition, or likewise compute the probability of a detectable effect on the basis of an alternative definition (such as MCID/2, rather than HR <1). Finally, although this tool allows any user to obtain inferences from published trial data, we encourage consideration of these limitations and the fact that inference and decision making are distinct.[25,35] Bayesian methodology is not strictly superior to frequentism in all settings, but rather complementary. An understanding of the advantages and tradeoffs of both approaches is critical towards the optimal interpretations of landmark trials. For trial primary end point analysis, clinical trial biostatisticians and principal investigators are the best equipped to select statistical approaches for their trial, including frequentist approaches and/or priors on the basis of the specific beliefs unique to their proposed treatment for posterior calculation based on the complete underlying individual patient-level data.

In summary, this study provides an evidence-based off-the-shelf prior that can be used to improve trial interpretation by enabling computation of the probabilities of any benefit and a clinically meaningful benefit on the basis of summary statistics from published Cox regression analyses. This tool is freely available online, without the need for coding. Practicing oncologists, patients, scientists, and students may find estimation of posterior probabilities to be valuable for placing trial results in context. We encourage clinical trial principal investigators, regulators, and other stakeholders to consider computing and reporting the probability of whether a treatment provides a clinically meaningful benefit, and not just the *P* value, when weighing the merits and drawbacks of a new treatment.

AFFILIATIONS

[1]Department of Radiation Oncology, Division of Radiation Oncology, The University of Texas MD Anderson Cancer Center, Houston, TX
[2]Department of Genitourinary Medical Oncology, Division of Cancer Medicine, The University of Texas MD Anderson Cancer Center, Houston, TX
[3]Department of Translational Molecular Pathology, The University of Texas MD Anderson Cancer Center, Houston, TX
[4]Department of Radiation Oncology and Molecular Radiation Sciences, Johns Hopkins University School of Medicine, Baltimore, MD
[5]Department of Radiation Oncology, Stanford University, Stanford, CA
[6]Insitro, South San Francisco, CA
[7]Department of Biomedical Informatics, University of North Carolina at Chapel Hill, Chapel Hill, NC
[8]Department of Gastrointestinal Radiation Oncology, Division of Radiation Oncology, The University of Texas MD Anderson Cancer Center, Houston, TX
[9]Department of Biostatistics, The University of Texas MD Anderson Cancer Center, Houston, TX
[10]Department of Biomedical Data Sciences, Leiden University Medical Center, Leiden, the Netherlands

PREPRINT VERSION

Preprint version available on https://www.medrxiv.org/content/10.1101/2024.04.01.24305158v1.

CORRESPONDING AUTHOR

Erik van Zwet, PhD; e-mail: E.W.van_Zwet@lumc.nl.

SUPPORT

Supported in part by Cancer Center Support (Core) grant P30CA016672 from the National Cancer Institute to The University of Texas MD Anderson Cancer Center and by the Sabin Family Fellowship Foundation (E.B.L. and P.M.).

DATA SHARING STATEMENT

The underlying data set, code for the development of the prior, and code for computing posterior probability are provided in the Data Supplement. A freely available webpage has been created for users to compute the probabilities at a given level of hazard ratio on the basis of the trial's summary statistics: https://alexandersherry.shinyapps.io/shinyapp/.

AUTHOR CONTRIBUTIONS

Conception and design: Alexander D. Sherry, Pavlos Msaouel, Joseph Abi Jaoude, Zachary R. McCaw, Ethan B. Ludmir, Erik van Zwet
Administrative support: Ethan B. Ludmir
Provision of study materials or patients: Ethan B. Ludmir
Collection and assembly of data: Alexander D. Sherry, Pavlos Msaouel, Gabrielle S. Kupferman, Timothy A. Lin, Joseph Abi Jaoude, Ramez Kouzy, Ethan B. Ludmir
Data analysis and interpretation: Alexander D. Sherry, Pavlos Msaouel, Timothy A. Lin, Joseph Abi Jaoude, Ramez Kouzy, Zachary R. McCaw, Ethan B. Ludmir, Erik van Zwet
Manuscript writing: All authors
Final approval of manuscript: All authors
Accountable for all aspects of the work: All authors

AUTHORS' DISCLOSURES OF POTENTIAL CONFLICTS OF INTEREST

The following represents disclosure information provided by authors of this manuscript. All relationships are considered compensated unless otherwise noted. Relationships are self-held unless noted. I = Immediate Family Member, Inst = My Institution. Relationships may not relate to the subject matter of this manuscript. For more information about ASCO's conflict of interest policy, please refer to www.asco.org/rwc or ascopubs.org/po/author-center.

Open Payments is a public database containing information reported by companies about payments made to US-licensed physicians (Open Payments).

Alexander D. Sherry
Employment: MD Anderson Cancer Center
Honoraria: Sermo

Pavlos Msaouel
Honoraria: Exelixis, Mirati Therapeutics, Pfizer
Consulting or Advisory Role: Axiom Healthcare Strategies
Research Funding: Mirati Therapeutics (Inst), Takeda (Inst), Bristol Myers Squibb (Inst), Gateway for Cancer Research (Inst)

Timothy A. Lin
Employment: Johns Hopkins Hospital

Joseph Abi Jaoude
Employment: Stanford Health Care

Ramez Kouzy
Employment: MD Anderson Cancer Center

Zachary R. McCaw
Employment: Insitro

Ethan B. Ludmir
Employment: Alaunos Therapeutics
Consulting or Advisory Role: Xerient

No other potential conflicts of interest were reported.

ACKNOWLEDGMENT

The authors thank Erica Goodoff, Senior Scientific Editor in the Research Medical Library at The University of Texas MD Anderson Cancer Center, for editing this article.

REFERENCES

1. Msaouel P, Lee J, Thall PF: Interpreting randomized controlled trials. Cancers (Basel) 15:4674, 2023
2. Lin TA, Sherry AD, Ludmir EB: Challenges, complexities, and considerations in the design and interpretation of late-phase oncology trials. Semin Radiat Oncol 33:429-437, 2023
3. Greenland S, Senn SJ, Rothman KJ, et al: Statistical tests, *P* values, confidence intervals, and power: A guide to misinterpretations. Eur J Epidemiol 31:337-350, 2016
4. Rafi Z, Greenland S: Semantic and cognitive tools to aid statistical science: Replace confidence and significance by compatibility and surprise. BMC Med Res Methodol 20:244, 2020
5. Greenland S: Invited commentary: The need for cognitive science in methodology. Am J Epidemiol 186:639-645, 2017
6. Amrhein V, Greenland S, McShane B: Scientists rise up against statistical significance. Nature 567:305-307, 2019
7. Wasserstein RL, Lazar NA: The ASA statement on p-values: Context, process, and purpose. The Am Statistician 70:129-133, 2016
8. Wasserstein RL, Schirm AL, Lazar NA: Moving to a world beyond "*P* <.05." Am Statistician 73:1-19, 2019 (suppl 1)
9. van Zwet E, Gelman A, Greenland S, et al: A new look at P values for randomized clinical trials. NEJM Evid 3:EVIDoa2300003, 2024
10. Adamina M, Tomlinson G, Guller U: Bayesian statistics in oncology: A guide for the clinical investigator. Cancer 115:5371-5381, 2009
11. Good IJ: The Bayes/non-Bayes compromise: A brief review. J Am Stat Assoc 87:597-606, 1992
12. Senn S: You may believe you are a Bayesian but you are probably wrong. RMM 2:48-66, 2011
13. Siddique J: Bayesian (re)-analyses of clinical trial data. NEJM Evid 2:EVIDe2200297, 2023
14. Diamond GA, Kaul S: Prior convictions: Bayesian approaches to the analysis and interpretation of clinical megatrials. J Am Coll Cardiol 43:1929-1939, 2004
15. Fornacon-Wood I, Mistry H, Johnson-Hart C, et al: Understanding the differences between Bayesian and frequentist statistics. Int J Radiat Oncol Biol Phys 112:1076-1082, 2022
16. Booth CM, Eisenhauer EA, Gyawali B, et al: Progression-free survival should not Be used as a primary end point for registration of anticancer drugs. J Clin Oncol 41:4968-4972, 2023
17. Kemp R, Prasad V: Surrogate endpoints in oncology: When are they acceptable for regulatory and clinical decisions, and are they currently overused? BMC Med 15:134, 2017
18. Zwet E, Gelman A: A proposal for informative default priors scaled by the standard error of estimates. Am Statistician 76:1-9, 2022
19. van Zwet E, Schwab S, Greenland S: Addressing exaggeration of effects from single RCTs. Significance 18:16-21, 2021
20. van Zwet E, Schwab S, Senn S: The statistical properties of RCTs and a proposal for shrinkage. Stat Med 40:6107-6117, 2021
21. van Zwet E, Tian L, Tibshirani R: Evaluating a shrinkage estimator for the treatment effect in clinical trials. Stat Med 43:855-868, 2024
22. Hoenig JM, Heisey DM: The abuse of power. Am Statistician 55:19-24, 2001
23. Hahn AW, Dizman N, Msaouel P: Missing the trees for the forest: Most subgroup analyses using forest plots at the ASCO annual meeting are inconclusive. Ther Adv Med Oncol 14: 17588359221103199, 2022
24. 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 32:1277-1280, 2014
25. Msaouel P, Lee J, Thall PF: Making patient-specific treatment decisions using prognostic variables and utilities of clinical outcomes. Cancers (Basel) 13:2741, 2021
26. Jaeschke R, Singer J, Guyatt GH: Measurement of health status. Ascertaining the minimal clinically important difference. Control Clin Trials 10:407-415, 1989
27. R Core Team: R: A Language and Environment for Statistical Computing. Vienna, Austria, R Foundation for Statistical Computing, 2023
28. Sherry AD, Msaouel P, Kupferman GS, et al: Calculate the Probability of Effects from the Summary Statistics of a Phase III Oncology Trial, 2024. https://alexandersherry.shinyapps.io/shinyapp/
29. Shen C, Xu H: Randomized phase III oncology trials: A survey and empirical bayes inference. J Stat Theor Pract 13:49-13, 2019
30. Sherry AD, Msaouel P, Kupferman G, et al: Towards treatment effect interpretability: A bayesian Re-analysis of 194,129 patient outcomes across 230 oncology trials. medRxiv 2024.07.23.24310891
31. Wijeysundera DN, Austin PC, Hux JE, et al: Bayesian statistical inference enhances the interpretation of contemporary randomized controlled trials. J Clin Epidemiol 62:13-21.e5, 2009
32. Lane D, Andrew B: Bayesian re-analysis of clinical trials, 2021. https://benjamin-andrew.shinyapps.io/bayesian_trials/
33. Fouchardière CDL, Malka D, Cropet C, et al: Gemcitabine and paclitaxel versus gemcitabine alone after 5-Fluorouracil, Oxaliplatin, and irinotecan in metastatic pancreatic adenocarcinoma: A randomized phase III PRODIGE 65-UCGI 36-GEMPAX UNICANCER study. J Clin Oncol 0:1055-1066, 2024
34. Bogart J, Wang X, Masters G, et al: High-dose once-daily thoracic radiotherapy in limited-stage small-cell lung cancer: CALGB 30610 (Alliance)/RTOG 0538. J Clin Oncol 41:2394-2402, 2023
35. Msaouel P, Lee J, Karam JA, et al: A causal framework for making individualized treatment decisions in oncology. Cancers (Basel) 14:3923, 2022
36. Lin TA, McCaw ZR, Koong A, et al: Proportional hazards violations in phase 3 cancer clinical trials: A potential source of trial misinterpretation. Clin Cancer Res, 2024
37. Rahman R, Fell G, Ventz S, et al: Deviation from the proportional hazards assumption in randomized phase 3 clinical trials in oncology: Prevalence, associated factors, and implications. Clin Cancer Res 25:6339-6345, 2019
38. McCaw ZR, Tian L, Wei J, et al: Choosing clinically interpretable summary measures and robust analytic procedures for quantifying the treatment difference in comparative clinical studies. Stat Med 40:6235-6242, 2021
39. Senn S: Seven myths of randomisation in clinical trials. Stat Med 32:1439-1450, 2013
40. Pasalic D, Fuller CD, Mainwaring W, et al: Detecting the dark matter of unpublished clinical cancer studies: An analysis of phase 3 randomized controlled trials. Mayo Clin Proc 96:420-426, 2021
41. Sherry AD, Passy AH, McCaw ZR, et al: Increasing power in phase III oncology trials with multivariable regression: An empirical assessment of 535 primary end point analyses. JCO Clin Cancer Inform 10.1200/CCI.24.00102