VOLUME 34 · NUMBER 15 · MAY 20, 2016

JOURNAL OF CLINICAL ONCOLOGY

STATISTICS IN ONCOLOGY

Comparison of Treatment Effects Measured by the Hazard Ratio and by the Ratio of Restricted Mean Survival Times in Oncology Randomized Controlled Trials

Ludovic Trinquart, Justine Jacot, Sarah C. Conner, and Raphaël Porcher

Ludovic Trinquart, Justine Jacot, Sarah C. Conner, and Raphaël Porcher, Institut National de la Santé et de la Recherche Médicale U1153; Ludovic Trinquart and Raphaël Porcher, Université Paris Descartes; and Assistance Publique-Hôpitaux de Paris; and Ludovic Trinquart, Cochrane France, Paris, France.

Published online ahead of print at www.jco.org on February 16, 2016.

L.T. was supported by Cochrane France.

Authors' disclosures of potential conflicts of interest are found in the article online at www.jco.org. Author contributions are found at the end of this article.

Corresponding author: Ludovic Trinquart, Centre de Recherche Epidémiologie et Statistique Paris Sorbonne Cité, Institut National de la Santé et de la Recherche Médicale U1153, Centre d'Epidémiologie Clinique, Hôpital Hôtel-Dieu, 1 place du Parvis Notre Dame, 75004 Paris, France; e-mail: ludovic.trinquart@aphp.fr.

0732-183X/16/3415w-1813w/$20.00

DOI: 10.1200/JCO.2015.64.2488

ABSTRACT

Purpose
We aimed to compare empirically the treatment effects measured by the hazard ratio (HR) and by the difference (and ratio) of restricted mean survival times (RMST) in oncology randomized trials.

Methods
We selected oncology randomized controlled trials from five leading journals during the last 6 months of 2014. We reconstructed individual patient data for one time-to-event outcome from each trial, preferably the primary outcome. We reanalyzed each trial and compared the treatment effect estimated by the HR with that by the difference (and ratio) of RMST. We estimated an average ratio of the HR to the ratio of RMST; an average ratio less than one indicates more optimistic assessments with HRs.

Results
We analyzed 54 randomized controlled trials totaling 33,212 patients. The selected outcome was overall survival in 21 (39%) trials. There was evidence of nonproportionality of hazards in 13 (24%) trials. The HR and RMST-based measures were in agreement regarding the statistical significance of the effect, except in one case. The median HR was 0.84 (Q1 to Q3 range, 0.67 to 0.97) and the median difference in RMST was 1.12 months (range, 0.22 to 2.75 months). The average ratio of the HR to the ratio of RMST was 1.11 (95% CI, 1.07 to 1.15), with substantial between-trial variability (I^2 = 86%). Results were consistent by outcome type (overall survival *v* other outcomes) and whether the proportional hazard assumption held or not.

Conclusion
On average, the HR provided significantly larger treatment effect estimates than the ratio of RMST. The HR may seem large when the absolute effect is small. RMST-based measures should be routinely reported in randomized trials with time-to-event outcomes.

J Clin Oncol 34:1813-1819. © 2016 by American Society of Clinical Oncology

INTRODUCTION

An essential component of clinical decision making for physicians, patients, and regulatory agencies is the evidence generated by randomized trials. Treatment outcomes need to show clinically meaningful benefit for the patient—in particular, that patients live longer or better.[1,2] Oncology trialists increasingly use the hazard ratio (HR) as the preferred measure of the difference between survival curves.[3] However, interpreting the magnitude of survival benefit using HRs may be challenging. HRs are commonly interpreted as relative risks, which can lead to an incorrect assessment of the clinical relevance of the treatment effect.[4] Moreover, HRs frequently vary with time, which violates the Cox proportional hazards model assumption. In such situations, the (average) HR depends on follow-up duration and may result in erroneous conclusions.[5,6]

An alternative treatment outcome measure is the mean survival time to some prespecified time point t^*—the restricted mean survival time (RMST).[7-9] The treatment effect can then be measured as the difference (and ratio) of RMST. It is straightforward to interpret such measures from a clinical perspective. For instance, if t^* is 12 months, the RMST measures the average number of months survived over 1 year. Thus, the difference (and ratio) of RMST measures the gain in 1-year life expectancy associated with the experimental treatment. Moreover,

these treatment-effect measures do not need any model assumptions, such as proportional hazards.

In recent examples, reanalysis of randomized trials using the difference in RMST found different results compared with the original HR estimate.[10,11] Our objective was to accumulate empirical evidence for the comparison of treatment effects measured by the HR and by the difference (and ratio) of RMST in a sample of oncology randomized trials.

METHODS

We performed a systematic review of phase III parallel-group randomized trials in oncology published in five journals during the last 6 months of 2014. We reconstructed individual patient data (IPD) from one time-to-event outcome from each trial, preferably the primary outcome. We reanalyzed each trial and we compared the treatment effect estimated by the HR with that by the difference (and ratio) of RMST.

Selection of Randomized Trials

We searched reports of oncology randomized trials published between July 1, 2014 and December 31, 2014 in three general medical journals (*New England Journal of Medicine*, *Lancet*, and *Journal of the American Medical Association*) and two oncology journals (*Journal of Clinical Oncology* and *Lancet Oncology*).[12] We searched MEDLINE using the Cochrane Highly Sensitive Search Strategy to identify randomized controlled trials (search equation is given in Data Supplement). Two review authors independently examined titles and abstracts to exclude irrelevant reports and examined full-text articles to determine eligibility. Disagreements were discussed to reach consensus.

Eligibility required reports of superiority and noninferiority parallel-group randomized trials, which included a Kaplan-Meier curve for a primary or secondary time-to-event outcome. In cases of multiple Kaplan-Meier curves, we selected the primary outcome; in cases of coprimary outcomes, we gave priority to progression-free survival, relapse-free survival, or failure-free survival.

We excluded phase I, II, or IV trials; multiple-arm trials; supportive care, palliative care, or prevention trials; meta-analyses, or analyses using pooled data from two or more trials; reports of secondary, subgroup, or follow-up analyses; and trials with specific designs (cross-over, factorial or cluster trials, and trials for predictive biomarker validation).

Data Extraction and Reconstruction of IPD

One reviewer extracted the data using a standardized form, and a second reviewer independently checked the extraction. We collected information about the trial population, the trial design, the sample size, and the source of funding. We recorded the unadjusted HR and 95% CI whenever available. We noted if the proportional hazards assumption was assessed.

For each trial, we reconstructed IPD for each group from published Kaplan-Meier curves.[13] We used the DigitizeIt software to measure the time and survival probability coordinates on the Kaplan-Meier curves.[14] We extracted the numbers of patients at risk and the total numbers of events, when available. The data were then input into an algorithm on the basis of iterative numerical methods to solve the inverted Kaplan-Meier equations. All data were reconstructed by one reviewer. We assessed the accuracy and reproducibility of the algorithm, both of which were high (Data Supplement).

Estimation of Treatment Effects According to HR and RMST

For each trial, we reanalyzed the reconstructed IPD. We tested for a null treatment effect by a log-rank test. We tested for nonproportional hazards using the Grambsch-Therneau test.[15] We estimated the HR and associated variance by a Cox proportional hazards regression model. An HR less than one favored the experimental intervention.

We assessed the RMST in the experimental and control groups at the time horizon t^*. We prespecified t^* as the minimum of the largest observed event time in each of the two groups.[16,17] The RMST was determined from the Kaplan-Meier estimate of the survival function. We estimated the difference in RMST, as well as the ratio of RMST. The associated variances were estimated by the delta method. A difference in RMST greater than zero or a ratio greater than one favored the experimental treatment. Last, we tested for a null treatment effect by comparing the difference (and ratio) of the RMST with its SE with a standard normal distribution.[10]

Comparison of Treatment Effects According to HR and RMSTs

The HR and the ratio of RMST are two distinct ways of quantifying the difference between two survival curves and do not have the same meaning. However, these two measures of the magnitude of treatment effect estimates are on the same relative scale. We aimed to assess if one was systematically further from the null than the other.

To evaluate the concordance between the two measures, we plotted the HRs against the differences and ratios of RMST in each trial, respectively. In subsequent analyses, the HR was transformed to its inverse so that both an HR and a ratio of RMST greater than one indicated superiority of the experimental treatment. We assessed in how many cases the HR yielded a larger (respectively smaller) treatment effect for the experimental treatment than the ratio of RMST. We also used an alternative rule: the HR had to be at least double or less than half of the ratio of RMST.

For each trial, we measured the difference in treatment effect estimates by the ratio of the HR to the ratio of RMST; we estimated the associated variance by the bootstrap method (Data Supplement).[18] A ratio between the HR and the ratio of RMST greater than one indicates that assessments on the basis of HR are more optimistic. We meta-analyzed the individual trial ratios with a random-effects model to estimate a mean ratio of the HR to the ratio of RMST (Data Supplement).[19] The heterogeneity in ratios across trials was assessed by the I^2 coefficient, which describes the percentage of the variability in estimated ratios due to heterogeneity rather than sampling error; an I^2 value 50% or greater represents substantial between-trial heterogeneity.[20]

We prespecified subgroup analyses to compare the ratios between the HR and the ratio of RMST for subgroups of trials according to the type of outcome (overall survival v other selected outcomes) and non-proportionality of hazards (Grambsch-Therneau test P value $< .10$ v $> .10$). We tested if the ratios differed between subgroups of trials by meta-regression analyses (Data Supplement).

Analyses involved use of R with the survRM2 package to derive the RMST.[17,21] A two-tailed $P < .05$ was considered statistically significant, except for the Grambsch-Therneau test, for which we set the significance level at $P < .10$.

RESULTS

Characteristics of the Selected Randomized Controlled Trials

A total of 54 trials met the selection criteria and were included in our analysis (references are given in the Data Supplement). Their characteristics are summarized in Table 1. The 54 trials reported data on 33,212 patients (median, 503 patients; min-max, 168-2,559 patients).

Treatment Effect Estimates On Selected Outcomes

The selected outcome was overall survival in 21 (39%) trials, progression-free survival in 23 (43%) trials, relapse-free survival in two (4%) trials, failure-free survival in two (4%) trials, disease-free survival in three (6%) trials, and event-free survival in three (6%)

Table 1. Characteristics of 54 Randomized Trials Included in Analysis

Feature	No. (%)*
Journal	
New England Journal of Medicine	11 (20)
Lancet	4 (7)
Journal of the American Medical Association	2 (4)
Journal of Clinical Oncology	22 (41)
Lancet Oncology	15 (28)
Funding source	
Industry	29 (54)
Nonindustry	17 (31)
Both	8 (15)
Cancer type	
Digestive/GI	11 (20)
Hematologic/blood	9 (17)
Breast	6 (11)
Endocrine and neuroendocrine	6 (11)
Respiratory/thoracic	6 (11)
Skin	6 (11)
Genitourinary	3 (6)
Gynecologic	2 (4)
Head and neck	2 (4)
Other	3 (6)
Type of experimental treatment	
Chemotherapy	8 (15)
Hormonal therapy	6 (11)
Targeted therapy	34 (63)
Chemotherapy plus targeted therapy	3 (6)
Immunotherapy	1 (2)
Radiation therapy	2 (4)
Type of control arm	
Placebo or best supportive care	36 (67)
Active comparator	18 (33)
Trial design	
Superiority	51 (94)
Noninferiority	3 (6)
Randomization ratio	
1:1	50 (93)
1:2	4 (7)
Sample size, median (Q1-Q3)	503 (296-764)
Trials stopped early	12 (22)†

*Except for the entry: sample size.
†Three trials were stopped early for efficacy, five trials for futility, and four trials for enrollment deficiencies.

trials. The selected outcome was the primary outcome in 49 (91%) trials. The reconstructed Kaplan-Meier curves are reported in the Data Supplement.

According to the log-rank test, there was evidence for a difference in survival curves between the experimental and control groups in 24 (44%) trials. The benefit always favored the experimental treatment, with HRs ranging from 0.19 (95% CI, 0.15 to 0.23) to 0.86 (95% CI, 0.75 to 0.98). In the other 30 trials, the HRs ranged from 0.52 (95% CI, 0.25 to 1.08) to 1.37 (95% CI, 0.87 to 2.14). In five (9%) trials, the authors reported that the proportionality assumption was checked. According to the Grambsch-Therneau test, evidence of nonproportionality of hazards was found in 13 (24%) other trials.

Across the 54 trials, the time horizons t^* defining RMST ranged from 11 months to 10 years (median [Q1 to Q3], 28.3 months [20.4 to 46.4 months]). The differences in RMST across the 54 trials ranged from 2.85 months in favor of the control treatment (95% CI, −3.89 to 9.60; $t^* = 77.2$ months) to 10.1 months in favor of the experimental treatment (95% CI, 0.6 to 19.6; $t^* = 111.6$ months). The median (Q1 to Q3) difference in RMST was 1.12 months (0.22 to 2.75 months). The ratios of RMST across the 54 trials ranged from 0.90 (95% CI, 0.80 to 1.01; $t^* = 21.6$ months) to 1.87 (95% CI, 1.48 to 2.37; $t^* = 28.4$ months). The median (Q1 to Q3) ratio of RMST was 1.10 (1.01 to 1.20). The difference (and ratio) of RMST was not correlated with the time horizon t^* (Data Supplement).

Comparison of Treatment Effects

Figures 1 and 2 show the differences and ratios of RMST against the HR, respectively. In all cases, there was agreement between the difference (and ratio) of RMST and HR about the direction of treatment effect, except in four trials. Fig 3 shows side by side the ratio of RMST and the HR with their CIs for each of the 54 trials. In all cases, there was agreement regarding the statistical significance, except for one trial where the difference and ratio of RMST indicated that the experimental treatment was significantly superior, whereas the HR did not.

In 41 (76%) of the 54 trials, the HR showed a larger treatment effect for the experimental treatment than the ratio of RMST; the difference was statistically significant in 20 (37%) trials. In four (7%) trials, the relative magnitude of the treatment effect according to the HR was at least double that according to the ratio of RMST. Conversely, the ratio of RMST showed a larger treatment effect for the experimental treatment than the HR in 13 trials, but none of the differences were statistically significant.

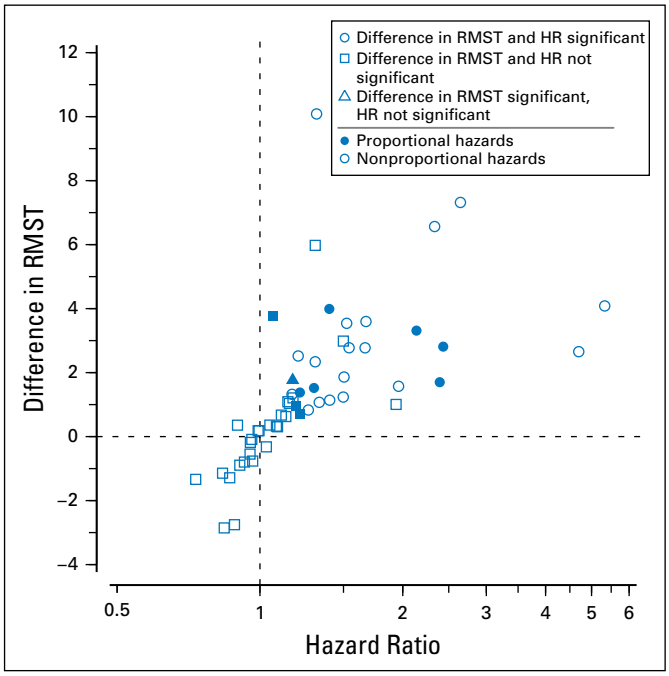

Fig 1. Estimates of treatment effect according to hazard ratio (HR) and difference in restricted mean survival times (RMST). The HR was transformed so that both an HR and a ratio of RMST greater than one indicated superiority of the experimental treatment. The dashed horizontal and vertical lines correspond to the null effect according to the difference in RMST and the HR, respectively. In all cases, there was agreement between the difference in RMST and the HR about the direction of treatment effect, except in four trials. In three trials, the HR favored the control treatment but the difference in RMST favored the experimental treatment; in one trial, the HR favored the experimental treatment but the difference in RMST favored the control treatment.

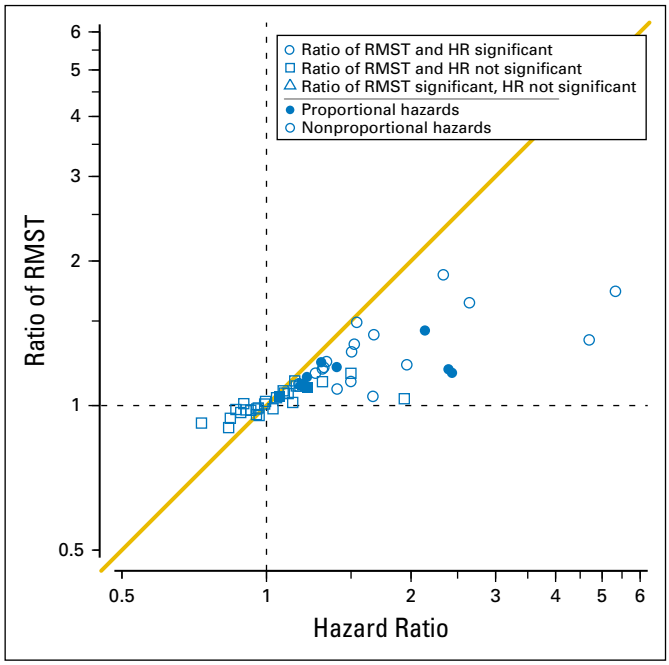

Fig 2. Estimates of treatment effect according to hazard ratio (HR) and ratio of restricted mean survival times (RMST). The HR was transformed so that both an HR and a ratio of RMST greater than one indicated superiority of the experimental treatment. The dashed horizontal and vertical lines correspond to the null effect according to the ratio of RMST and the HR, respectively. In all cases, there was agreement between the ratio of RMST and the HR about the direction of treatment effect, except in four trials. In three trials, the HR favored the control treatment but the ratio of RMST favored the experimental treatment; in one trial, the HR favored the experimental treatment but the ratio of RMST favored the control treatment.

Across the 54 trials, the average ratio of the HR to the ratio of RMST was 1.11 (95% CI, 1.07 to 1.15), suggesting that assessments on the basis of HRs are systematically more optimistic. The between-trial variability of discrepancies was substantial (I^2 = 86%; Fig 4).

We found no evidence of difference by type of outcome: average ratio 1.04 (95% CI, 1.00 to 1.07) across trials with overall survival (n = 21 trials) and 1.17 (95% CI, 1.10 to 1.23) across trials with other outcomes (n = 33 trials; *P* value for difference between subgroups = .08). There was no evidence of difference by nonproportionality of hazards: average ratio 1.12 (95% CI, 1.06 to 1.19) across trials with evidence of nonproportional hazards (n = 13 trials) and 1.10 (95% CI, 1.06 to 1.15) across trials without evidence of nonproportional hazards (n = 41 trials; *P* value for difference between subgroups = .55).

DISCUSSION

Although there was agreement between RMST and HR about the statistical significance of the treatment effect, we provided empirical evidence that treatment-effect measures on the basis of RMST yielded more conservative estimates than HRs. This finding was consistent for overall survival and other time-to-event outcomes and whether evidence of nonproportionality of hazards was identified or not. It was also observed in post hoc subgroup analyses according to trial size and follow-up duration (data not shown).

Our findings should be interpreted according to the understanding that treatment-effect measures on the basis of HRs and RMST are not interchangeable. The risk that HRs lead to more optimistic interpretations may arise from the fact that assessing the practical clinical significance on the basis of the magnitude of the treatment effect remains subjective. In practice, clinicians may apply a similar standard when interpreting the magnitude of HRs and ratios of RMST, despite differences in the meanings of the two relative measures. Given this practice, there is the potential for misinterpretation if HR is indeed systematically larger than the ratio of RMST.

Our analysis also highlights how the difference in RMST provides a clinically meaningful summary of evidence. It allows for quantifying the absolute survival difference and grading the magnitude of clinical benefit.[22] In turn, one can assess if the studied treatment exceeds a predefined threshold of direct clinical benefit (minimally effective treatment).[23]

Also, RMST-based treatment-effect measures do not rely on any specific assumption. It is poor practice to report a single HR when nonproportionality of hazards is present because the parameter being estimated in that case is particularly difficult to interpret.[24] In our sample, we found evidence of nonproportionality of hazards in 24% of trials. It has been suggested that the difference in RMST should be considered as a standard treatment-effect measure under nonproportionality of hazards.[9,10] Our findings suggest that treatment effects seem systematically more beneficial when measured with HRs rather than RMST, whether the proportional hazards assumption holds or not. Thus, we extend the previous recommendation to systematically report RMST-based measures of treatment effects in any trial with time-to-event outcomes.

An additional advantage of RMST-based measures is that they are more efficient than HRs. When the number of events is small, the CI for the true HR is wide, whether the proportional hazards assumption holds or not. RMST-based measures do not suffer from this limitation and would be a better measure in this regard, especially in the setting of noninferiority trials.[25,26]

Previous studies provided empirical evidence regarding the comparison between measures of time-to-event outcomes in cancer trials. Seruga and colleagues[27,28] compared the difference in survival probabilities at specific time points, or in median survival times, with the difference in RMST; the absolute benefits were larger and more variable with the former than the latter. Cortés et al[29] compared the HR to the ratio of median survival times and did not find evidence of difference, on average, between the two. A limitation is that both survival probabilities at specific time points and median survival times do not reflect the entire survival history; on the contrary, RMST takes the whole survival distribution up to t^* into account. Moreover, the median survival time is frequently inestimable (as was the case in 30% [16 of 54] of trials in our sample), whereas RMST is always estimable. To our knowledge, our study is the first to compare empirically RMST-based measures with HRs, and the results need to be replicated.

Our study has some limitations. First, it was on the basis of reconstructed IPD and we cannot claim the level of standard that would be provided by the original IPD for each trial. However, it would be impractical to obtain them for all trials in such a methodological study. Moreover, we used a validated algorithm; the accuracy and reproducibility of the reconstruction was excellent, as demonstrated in the seminal work of Guyot et al.[13] Also, in our sample, reconstructed data provided estimates close to the HRs or

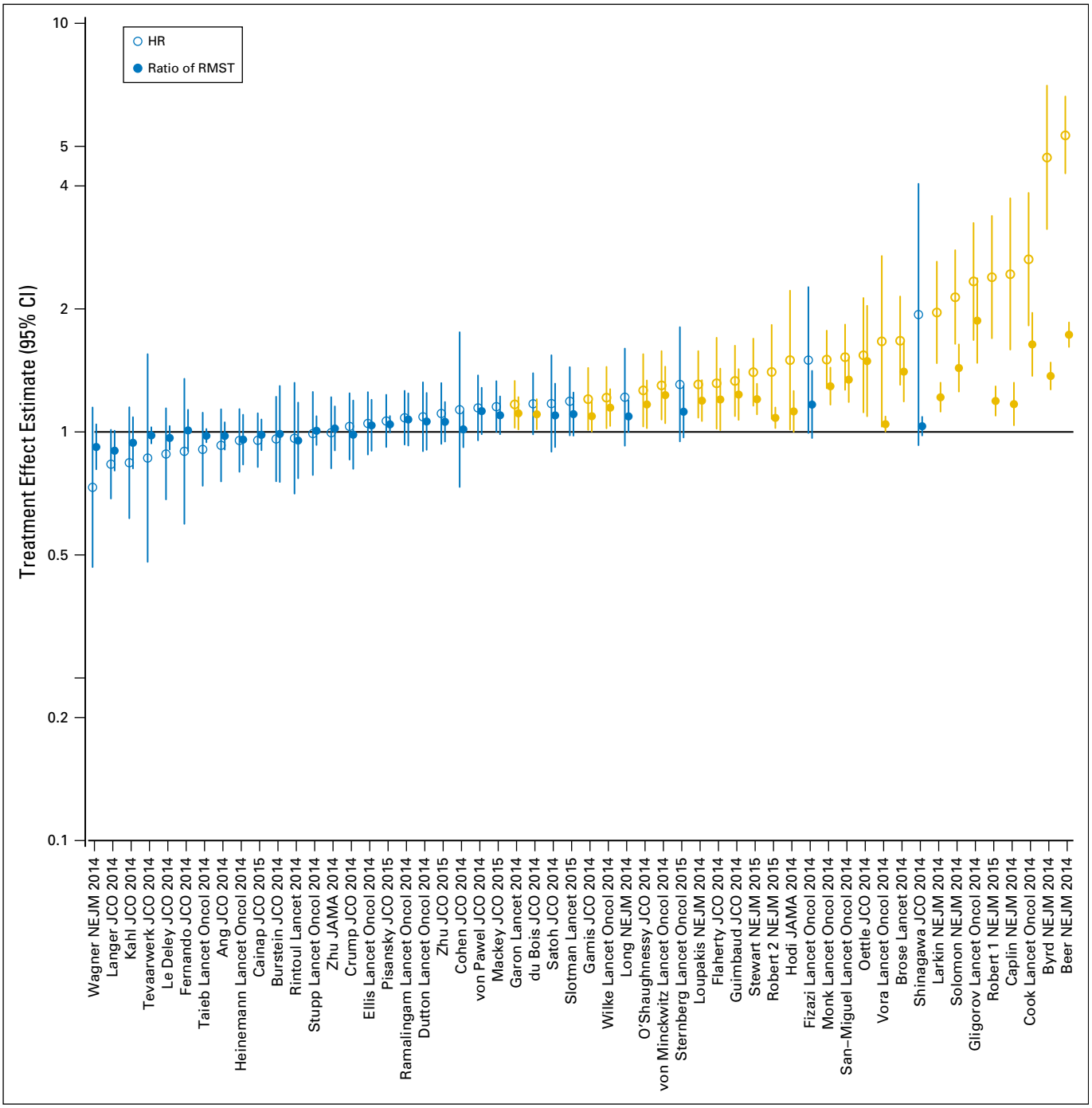

Fig 3. Comparison of the ratio of restricted mean survival times (RMST) with the hazard ratio (HR) and 95% CIs for the 54 trials. The HR was transformed so that both an HR and a ratio of RMST greater than one indicated superiority of the experimental treatment. Gold indicates a statistically significant HR or ratio of RMST. In all cases, there was agreement regarding the statistical significance, except for one trial where the ratio of RMST (as well as the difference in RMST) suggested that the experimental treatment was significantly superior, whereas the HR did not. JAMA; *Journal of the American Medical Association*; JCO, *Journal of Clinical Oncology*; Lancet Oncol; *Lancet Oncology*; NEJM, *New England Journal of Medicine*.

median survival times reported in the original articles. We could not re-estimate adjusted treatment effect estimates, as was done in some of the original reports. However, trialists can derive adjusted RMST.[30,31]

Second, we estimated the RMST by directly integrating the Kaplan-Meier survival curve. The Kaplan-Meier estimate becomes unstable when the number of patients at risk becomes small. This may happen at the tail of the curve, usually beyond the time horizon at which the RMST is estimated. Nonetheless, where this is a concern, more robust approaches to estimate the RMST have been recommended, such as flexible parametric models.[9,32] Third,

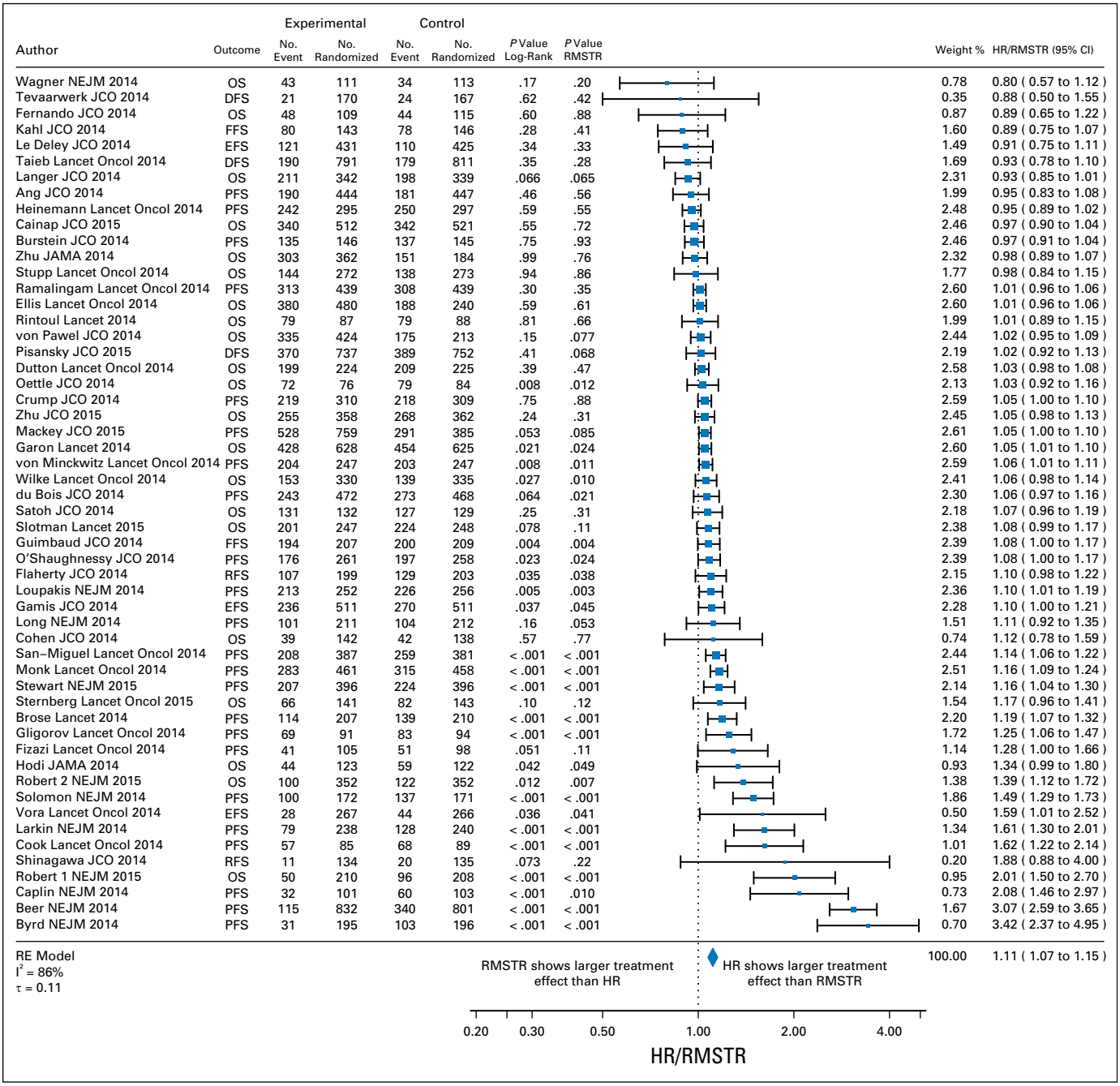

Fig 4. Forest plot of ratios between the hazard ratio (HR) and the ratio of restricted mean survival times (RMST). The HR was transformed so that both an HR and a ratio of RMST greater than one indicated superiority of the experimental treatment. A ratio between the HR and the ratio of RMST greater than one indicates that assessments on the basis of HR are more optimistic. We meta-analyzed the individual trial ratios with inverse variance methods using a random-effects model; the diamond shows the mean ratio between the HR and the ratio of RMST, and its associated CI. DFS, disease-free survival; EFS, event-free survival; FFS, failure-free survival; JAMA; *Journal of the American Medical Association*; JCO, *Journal of Clinical Oncology*; Lancet Oncol; *Lancet Oncology*; NEJM, *New England Journal of Medicine*; OS, overall survival; PFS, progression-free survival; RE, random effects; RFS, relapse-free survival; RMSTR, ratio of restricted mean survival time.

we analyzed oncology randomized trials published in high-impact journals and our findings may lack generalizability.

Fourth, for the sake of comparability, we transformed the HR so that both an HR and a ratio of RMST greater than one favored the experimental treatment. This could be avoided by using the restricted mean time lost (RMTL = $t^{\star}$ − RMST) instead of RMST. A ratio of RMTL less than one indicates the beneficial effect of the experimental treatment. Another possible measure is the relative difference in RMST, ie, the difference in RMST divided by the time horizon $t^{\star}$.[26] This measure somehow circumvents the fact that RMST must be interpreted relative to a time horizon $t^{\star}$.

Here, we used the same prespecified definition for all trials. In essence, taking the time horizon as the minimum of the largest observed event time in both groups comes down to going out as far as

possible along the Kaplan-Meier curves. One may select a t^* value on the basis of data maturity, or consider a curve showing the difference in RMST over time with a simultaneous confidence band, so that no single t^* needs to be selected.[10,33] However, in a given trial, the time horizon t^* may be instead prespecified at the design stage on the basis of clinical relevance rather than statistical considerations.[10,11]

Reporting specific treatment-effect measures has the potential to affect the interpretation of trial findings.[34] Relying on HRs to the exclusion of measures of absolute survival benefit may result in a gap between the apparently large relative effect and actually small absolute beneficial effect of the treatments studied. This problem may also affect meta-analyses of randomized trials, because using HRs is considered the most appropriate way to summarize time-to-event data.[35] However, it is possible to use RMST in meta-analyses of time-to-event outcomes, although the choice of t^* may be an issue due to variable follow-up times across trials.[36]

In conclusion, the HR and RMST-based measures were in agreement regarding the statistical significance of treatment effects. However, the HR provided treatment effect estimates that were, on average, significantly further from the null effect than the ratio of RMST. Thus, RMST-based measures should be routinely reported in randomized trials with time-to-event outcomes.

AUTHORS' DISCLOSURES OF POTENTIAL CONFLICTS OF INTEREST

Disclosures provided by the authors are available with this article at www.jco.org.

AUTHOR CONTRIBUTIONS

Conception and design: Ludovic Trinquart, Raphaël Porcher
Collection and assembly of data: All authors
Data analysis and interpretation: All authors
Manuscript writing: All authors
Final approval of manuscript: All authors

REFERENCES

1. Wilson MK, Collyar D, Chingos DT, et al: Outcomes and endpoints in cancer trials: Bridging the divide. Lancet Oncol 16:e43-e52, 2015
2. Wilson MK, Karakasis K, Oza AM: Outcomes and endpoints in trials of cancer treatment: The past, present, and future. Lancet Oncol 16:e32-e42, 2015
3. Blagoev KB, Wilkerson J, Fojo T: Hazard ratios in cancer clinical trials–a primer. Nat Rev Clin Oncol 9: 178-183, 2012
4. Case LD, Kimmick G, Paskett ED, et al: Interpreting measures of treatment effect in cancer clinical trials. Oncologist 7:181-187, 2002
5. Prentice RL, Pettinger M, Anderson GL: Statistical issues arising in the Women's Health Initiative. Biometrics 61:899-911, 2005; discussion 911-941
6. Hernán MA: The hazards of hazard ratios. Epidemiology 21:13-15, 2010
7. Irwin JO: The standard error of an estimate of expectation of life, with special reference to expectation of tumourless life in experiments with mice. J Hyg (Lond) 47:188-189, 1949
8. Karrison TG: Use of Irwin's restricted mean as an index for comparing survival in different treatment groups–interpretation and power considerations. Control Clin Trials 18:151-167, 1997
9. Royston P, Parmar MK: The use of restricted mean survival time to estimate the treatment effect in randomized clinical trials when the proportional hazards assumption is in doubt. Stat Med 30:2409-2421, 2011
10. Royston P, Parmar MK: Restricted mean survival time: An alternative to the hazard ratio for the design and analysis of randomized trials with a time-to-event outcome. BMC Med Res Methodol 13:152, 2013
11. Uno H, Claggett B, Tian L, et al: Moving beyond the hazard ratio in quantifying the between-group difference in survival analysis. J Clin Oncol 32: 2380-2385, 2014
12. ISI Web of Knowledge: 2013 journal citation reports science edition. http://admin-apps.webofknowledge.com/JCR/JCR
13. Guyot P, Ades AE, Ouwens MJ, et al: Enhanced secondary analysis of survival data: Reconstructing the data from published Kaplan-Meier survival curves. BMC Med Res Methodol 12:9, 2012
14. Bormann I: DigitizeIt. http://www.digitizeit.de/
15. Grambsch PM, Therneau TM: Proportional hazards tests and diagnostics based on weighted residuals. Biometrika 81:515-526, 1994
16. Pepe MS, Fleming T: Weighted Kaplan-Meier statistics: A class of distance tests for censored survival data. Biometrics 45:497-507, 1989
17. Uno H: Vignette for survRM2 package: Comparing two survival curves using the restricted mean survival time. https://cran.r-project.org/web/packages/survRM2/vignettes/survRM2-vignette3-1.pdf
18. Efron B: Better bootstrap confidence intervals. J Am Stat Assoc 82:171-185, 1987
19. DerSimonian R, Laird N: Meta-analysis in clinical trials. Control Clin Trials 7:177-188, 1986
20. Higgins JP, Thompson SG: Quantifying heterogeneity in a meta-analysis. Stat Med 21:1539-1558, 2002
21. R Core Team: R: A Language and Environment for Statistical Computing. Vienna, Austria, R Foundation for Statistical Computing, 2014
22. Schnipper LE, Davidson NE, Wollins DS, et al; American Society of Clinical Oncology: American Society of Clinical Oncology statement: A conceptual framework to assess the value of cancer treatment options. J Clin Oncol 33:2563-2577, 2015
23. Buyx AM, Friedrich DR, Schöne-Seifert B: Ethics and effectiveness: Rationing healthcare by thresholds of minimum effectiveness. BMJ 342:d54, 2011
24. Struthers CA, Kalbfleisch JD: Misspecified proportional hazard models. Biometrika 73:363-369, 1986
25. Uno H, Wittes J, Fu H, et al: Alternatives to hazard ratios for comparing the efficacy or safety of therapies in noninferiority studies. Ann Intern Med 163:127-134, 2015
26. Zhao L, Tian L, Uno H, et al: Utilizing the integrated difference of two survival functions to quantify the treatment contrast for designing, monitoring, and analyzing a comparative clinical study. Clin Trials 9:570-577, 2012
27. Seruga B, Pond GR, Hertz PC, et al: Comparison of absolute benefits of anticancer therapies determined by snapshot and area methods. Ann Oncol 23:2977-2982, 2012
28. Seruga B, Hertz PC, Wang L, et al: Absolute benefits of medical therapies in phase III clinical trials for breast and colorectal cancer. Ann Oncol 21: 1411-1418, 2010
29. Cortés J, González JA, Campbell MJ, et al: A hazard ratio was estimated by a ratio of median survival times, but with considerable uncertainty. J Clin Epidemiol 67:1172-1177, 2014
30. Andersen PK, Hansen MG, Klein JP: Regression analysis of restricted mean survival time based on pseudo-observations. Lifetime Data Anal 10: 335-350, 2004
31. Tian L, Zhao L, Wei LJ: Predicting the restricted mean event time with the subject's baseline covariates in survival analysis. Biostatistics 15:222-233, 2014
32. Royston P, Parmar MK: Flexible parametric proportional-hazards and proportional-odds models for censored survival data, with application to prognostic modelling and estimation of treatment effects. Stat Med 21:2175-2197, 2002
33. Zhao L, Claggett B, Tian L, et al: On the restricted mean survival time curve in survival analysis. Biometrics 10.1111/biom.12384 [epub ahead of print on August 24, 2015]
34. Boutron I, Altman DG, Hopewell S, et al: 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 32:4120-4126, 2014
35. Deeks J, Higgins J, Altman D: Analysing data and undertaking meta-analyses, in Higgins JPT, Green S (eds): Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0. http://handbook.cochrane.org/
36. Wei Y, Royston P, Tierney JF, et al: Meta-analysis of time-to-event outcomes from randomized trials using restricted mean survival time: Application to individual participant data. Stat Med 34:2881-2898, 2015

AUTHORS' DISCLOSURES OF POTENTIAL CONFLICTS OF INTEREST

Comparison of Treatment Effects Measured by the Hazard Ratio and by the Ratio of Restricted Mean Survival Times in Oncology Randomized Controlled Trials

The following represents disclosure information provided by authors of this manuscript. All relationships are considered compensated. 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 jco.ascopubs.org/site/ifc.

Ludovic Trinquart
No relationship to disclose

Justine Jacot
No relationship to disclose

Sarah C. Conner
No relationship to disclose

Raphaël Porcher
No relationship to disclose