Original Investigation

The Net Chance of a Longer Survival as a Patient-Oriented Measure of Treatment Benefit in Randomized Clinical Trials

Julien Péron, MD, PhD; Pascal Roy, MD, PhD; Brice Ozenne, PhD; Laurent Roche, PhD; Marc Buyse, ScD

IMPORTANCE Time to events, or survival end points, are common end points in randomized clinical trials. They are usually analyzed under the assumption of proportional hazards, and the treatment effect is reported as a hazard ratio, which is neither an intuitive measure nor a meaningful one if the assumption of proportional hazards is not met.

OBJECTIVE To demonstrate that a different measure of treatment effect, called *the net chance of a longer survival*, is a meaningful measure of treatment effect in clinical trials whether or not the assumption of proportional hazards is met.

DESIGN In this simulation study, the net chance of a longer survival by at least *m* months, where *m* months is considered clinically worthwhile and relevant to the patient, was calculated as the probability that a random patient in the treatment group has a longer survival by at least *m* months than does a random patient in the control group minus the probability of the opposite situation. The net chance of a longer survival is equal to zero if treatment does not differ from control and ranges from −100% if all patients in the control group fare better than all patients in the treatment group up to 100% in the opposite situation. We simulated data sets for realistic trials under various scenarios of proportional and nonproportional survival hazards and plotted the Kaplan-Meier survival curves as well as the net chance of a longer survival as a function of *m*. Data analysis was performed from August 14 to 18, 2015.

MAIN OUTCOMES AND MEASURES The net chance of a longer survival calculated for values of *m* ranging from 0 to 40 months.

RESULTS When hazards are proportional, the net chance of a longer survival approaches zero as *m* increases. The net chance of a longer survival (Δ) was 13% (95% CI, 6.5%-19.4%; $P < .001$) when any survival difference was considered clinically relevant ($m = 0$ months). When survival differences larger than 20 months were considered relevant ($m = 20$), the net chance of a longer survival was very close to zero (Δ[20] = 0.5%; 95% CI, −0.1% to 1.1%; $P = .09$). In contrast, when treatment effects are delayed or when some patients are cured by treatment, the net chance of a longer survival benefit remains high and tends to the cure rate. For crossing hazards, the Δ was negative (Δ = −6.9%; 95% CI, −14.0% to −0.5%; $P = .047$). However when large survival differences were considered ($m = 20$), the Δ(*m*) was positive (Δ[20] = 8.9%; 95% CI, 6.7%-11.1%; $P < .001$).

CONCLUSIONS AND RELEVANCE The net chance of a longer survival is useful whether or not the assumption of proportional hazards is met in the analysis of survival end points and may be helpful as a measure of treatment benefit that has direct relevance to patients and health care professionals.

Invited Commentary page 906

Supplemental content at jamaoncology.com

Author Affiliations: Department of Oncology, Hospices Civils de Lyon, Pierre-Benite, France (Péron); Department of Biostatistics, Hospices Civils de Lyon, Pierre-Benite, France (Péron, Roy, Ozenne, Roche); Biometry and Evolutive Biology Laboratory, Biostatistics-Health Team, University of Lyon 1, Villeurbanne, France (Péron, Roy, Ozenne, Roche); Department of Biostatistics, International Drug Development Institute, Cambridge, Massachusetts (Buyse); Interuniversity Institute for Biostatistics and Statistical Bioinformatics, Hasselt University, Hasselt, Belgium (Buyse).

Corresponding Author: Marc Buyse, ScD, International Drug Development Institute, 185 Alewife Brook Pkwy, Ste 410, Cambridge, MA 02138 (marc.buyse@iddi.com).

JAMA Oncol. 2016;2(7):901-905. doi:10.1001/jamaoncol.2015.6359
Published online April 28, 2016.

Time to event end points, such as survival time or time to a clinically important event, are widely used in randomized clinical trials. The treatment effect in survival analysis is most commonly quantified and reported as the hazard ratio (HR), a relative measure of the difference between 2 survival curves. One of the assumptions for computing a meaningful HR is that hazard functions are proportional over time. When the assumption of proportional hazards is not met, the computed HR does not reliably reflect the treatment benefit because the true HR is changing over time.[1,2] There is also evidence that treatment benefits tend to be overestimated by physicians (and potentially by patients) when expressed on a relative scale.[3] Researchers have attempted to express treatment effects on absolute scales; for example, as the difference in the percentage of patients still alive or free of the event of interest at 1 year or some other specific time point.[4] However, this solution provides a limited measure of benefit since it ignores all of the events that occur after the chosen time point and the treatment benefit may be very sensitive to the selection of this time point.[5,6] The difference in median survival is also commonly reported, but it too has limitations and cannot be estimated if less than half of the patients have died or experienced the event of interest.[4] Other measures of treatment effect have been proposed, but none has been widely adopted or has a straightforward application to the individual patient.[7-11]

We describe a new measure of treatment effect that directly addresses a question patients might raise, that is, "What is my net chance of surviving longer with treatment than without?" or, by extension, "What is my net chance of surviving at least 6 months longer with treatment than without?" We call this measure *the net chance of a longer survival*. We investigate the net chance of a longer survival through simulated data sets for a randomized clinical trial under typical scenarios for the treatment effect, and we show that this measure is meaningful and intuitively appealing whether or not hazards are proportional.

Key Points

Question Can net chance of survival be used to interpret treatment benefit and survival differences in clinical trials when hazards are not proportional?

Findings The net chance of a longer survival by at least m months, is the probability that a random patient in the experimental arm has a survival longer by at least m months than a random control patient, minus the probability of the opposite situation. When proportional hazards hold, the net chance of a longer survival goes to zero as *m* increases, while it tends to the cure rate when some patients are cured by treatment.

Meaning This method is useful whether or not the assumption of proportional hazards is met in the analysis of survival end points.

Methods

Rationale for the Net Chance of a Longer Survival

The net chance of a longer survival, denoted as Δ, is defined as the probability that a random patient in the treatment group survives longer than a random patient in the control group minus the probability of the opposite situation.[12] The Δ is equal to zero if treatment does not differ from the control, it is positive if treatment is better than the control, and it would be equal to 100% if all patients in the treatment group fared better than all patients in the control group. Conversely, the Δ would be equal to -100% if all patients in the control group fared better than all patients in the treatment group. For instance, if the Δ were estimated to be equal to 0.10, a random patient in the treatment group would have a 10% higher probability of a longer survival than a patient in the control group. The magnitude of Δ is directly related to the HR in the simple case of proportional hazards without censoring; in such cases, the following formula links the 2 measures of treatment effect: $\Delta = (1 - HR)/(1+HR)$.[13]

We used a generalized form of Δ, the net chance of a longer survival by at least *m* months, denoted as Δ(*m*). The Δ(*m*) was defined as the probability that a random patient in the treatment group survives by at least *m* months longer than does a random patient in the control group minus the probability of the opposite situation. A confidence interval for Δ(*m*) and a test of statistical significance can be computed using a randomization test, which is briefly summarized in eAppendix 1 in the Supplement and has been described in detail elsewhere.[12]

Simulation of Randomized Trial Data Sets

We simulated 5 scenarios of typical survival differences, keeping the overall HR equal to 0.75 in all cases. In scenario 1, the hazards were proportional between the 2 treatment groups, with a constant HR of 0.75. In the other scenarios, the hazards were nonproportional. In scenario 2, the HR increased over time from 0.4 to 1 (early survival differences). This scenario is typical of the survival outcomes observed in most trials evaluating cytotoxic chemotherapy or some molecular-targeted therapies for metastatic solid cancers.[14] In scenario 3, the HR decreased over time (late survival differences). This scenario is typical of the survival outcomes observed with modern immunotherapies in solid cancers.[15] In scenario 4, a total of 10% of the patients were cured by the treatment and the other 90% had no effect from the treatment. This scenario is typical of the survival outcomes observed with allografts in pediatric trials.[16] Finally, in scenario 5, half the patients had a benefit from treatment and the other half had a detrimental effect. This scenario is typical of survival outcomes when molecular-targeted therapy is compared with cytotoxic chemotherapy among all patients and only 50% of the patients respond to the targeted therapy.[17]

For each scenario, a data set was generated including 2 treatment groups, each with 600 patients. The simulation parameters are summarized in eAppendix 2 in the Supplement. For each data set, the net chance of a longer survival was calculated and plotted for values of *m* ranging from 0 to 40 months.

Generalized pairwise comparisons were performed with the package BuyseTest in R statistical software (R Foundation for Statistical Computing), available from the corresponding author upon request. Data analysis was performed from August 14 to 18, 2015. No institutional review board approval was required as no actual patient data were analyzed.

Results

The **Figure** shows the 5 simulated scenarios. In scenario 1 (proportional hazards), the survival curves separated harmoniously. Median survival was 9.3 months in the control group and 10.6 months in the treatment group. The net chance of a longer survival (Δ) was 13% (95% CI, 6.5%-19.4%; $P < .001$) when any survival difference was considered clinically relevant ($m = 0$ months), which means that a random patient in the treatment group would have a 13% higher chance of a longer survival compared with a random patient in the control group. However, the net chance of a longer survival decreased when long-term survival differences were evaluated. When survival differences larger than 20 months were considered relevant ($m = 20$), the net chance of a longer survival was very close to zero (Δ[20] = 0.5%; 95% CI, -0.1% to 1.1%; $P = .09$). In scenario 2 (early survival differences), the net chance of a longer survival was 23% (95% CI, 16.8%-28.9%; $P < .001$) when any survival benefit was considered clinically relevant ($m = 0$ months), but it decreased even more quickly than in scenario 1 and was close to zero (Δ[20] = 0.6%; 95% CI, −0.1% to 1.0%; $P = .06$) when survival differences larger than 20 months were considered relevant ($m = 20$). In scenario 3 (delayed survival differences), the Δ(m) remained about constant when m was less than 20 months and decreased slowly thereafter. In scenario 4 (curable disease), the Δ(m) remained stable at 10% (95% CI, 7.5%-12.5%; $P < .001$) regardless of m, which is exactly as expected given the assumption of a 10% cure rate. In scenario 5 (crossing hazards), the survival curves crossed near the eleventh month of follow-up. When any survival difference was considered relevant, the Δ was negative (Δ = -6.9%; 95% CI, -14.0% to -0.5%; $P = .047$). However when large survival differences were considered ($m = 20$), the Δ(m) was positive (Δ[20] = 8.9%; 95% CI, 6.7%-11.1%; $P < .001$).

Discussion

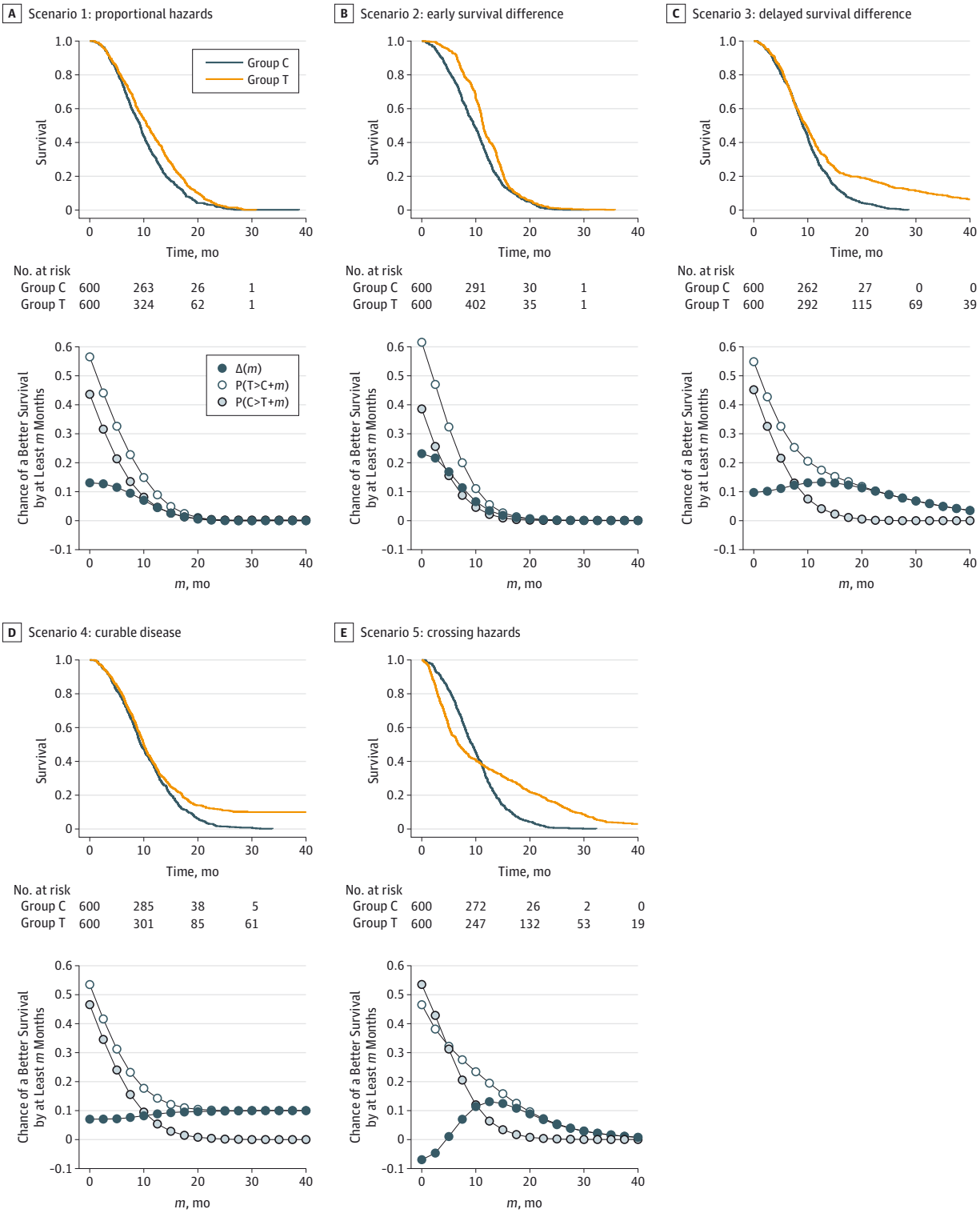
In this study, we attempt to address the question of treatment benefit from the point of view of a patient asking, "What is my chance of surviving longer with treatment than without?" We call this measure of treatment effect *the net chance of a longer survival*. The Δ(m) has a probabilistic interpretation: it is the difference between the probability that a random patient in the treatment group has a survival longer by at least m months than a random patient in the control group and the probability of the opposite situation, with m being a specified minimal, clinically relevant difference in survival, such as 6 months. An intuitively appealing assessment of the treatment effect may be derived from the Figure representing the Δ(m) as a function of the minimal clinically relevant difference m. For any value of m, the Figure provides the probabilities that the new treatment prolongs or shortens the survival time for a patient by at least m months and the difference between these probabilities. The graphs of the Δ(m) as a function of m show strikingly different patterns in the 5 scenarios considered and, as such, can be useful to describe potential survival benefits in situations of nonproportional hazards. For instance, if there is a cure rate, the Δ(m) trends, as it should, to the proportion of patients who experience a long-term benefit of the treatment (or, in the best-case scenario, to the cure rate).

The Δ extends the previously proposed probabilistic index, a general nonparametric measure of the effect size.[18] The association between the probabilistic index and the HR was investigated by Moser and McCann[19] and by Buyse.[13] The advantages and limitations of the probabilistic index have been debated extensively.[20,21] In the absence of any treatment effect, the probabilistic index is equal to 0.5 while the net chance of a longer survival is equal to zero, which is more intuitively appealing. The Δ also has a direct association with other measures of treatment effect.[12] Tan and Murphy[22] proposed to use the "average duration of life gained" to summarize the difference between 2 survival curves. Royston and Parmar[4] proposed to use the difference in "restricted mean survival," which is estimated by the area under the survival curve up to that point. As it turns out, this statistic is also equal to the area under the curve of the Δ(m) for all values of m, restricted to the same time point as the area under the survival curve.[10,21-25]

The Δ has intuitive and descriptive appeal, but it also leads to a test of significance through a randomization test, which is asymptotically equivalent to Gehan's generalized Wilcoxon rank sum test.[12,26,27] When the treatment effect is tested for a single value of m (the most common case being $m = 0$), no adjustment for multiplicity is needed. When multiple values of m are of interest, several tests can be carried out, with an appropriate adjustment for multiplicity.[12,28,29]

Of the 5 scenarios investigated in this study, the most intriguing was that of crossing hazards. The corresponding data set was simulated to closely resemble the reported progression-free survival (PFS) curves of the Iressa Pan-Asia Study.[17] In that study, previously untreated patients with advanced lung adenocarcinoma were randomized to receive either gefitinib (a tyrosine kinase inhibitor of the epidermal growth factor receptor [EGFR]) or a chemotherapy combination of carboplatin and paclitaxel. Activating *EGFR* mutations are now known to be predictive of benefit from gefitinib.[30] Although this predictive role was unknown when the Iressa Pan-Asia Study[17] was initiated, the frequency of such mutations was high (approximately half of all patients) because the study was carried out among Asians and a large proportion were nonsmokers or light smokers, 2 characteristics associated with activating *EGFR* mutations. In the corresponding simulated data set, the PFS Kaplan-Meier curves crossed, and the assumption of proportional hazards was obviously violated; hence, the overall HR of 0.75 reported in the trial was not meaningful as well as potentially misleading. The analysis of the net chance of a longer PFS identified a clear benefit in favor of the treatment (gefitinib) when long-term PFS differences were considered ($m > 20$ months). However, the control group had a better outcome when any PFS difference was considered relevant ($m = 0$). The mutation data provide a clear explanation for this pattern: patients who did not have an *EGFR* mutation had a poorer outcome when they received gefitinib because this targeted agent had no antitumor effect in them; hence, these patients did worse than patients receiving chemotherapy. In contrast, patients who had an *EGFR* mutation had

Figure. Survival Benefits in a Scenario of Proportional Hazards and 4 Scenarios of Nonproportional Hazards



Kaplan-Meier estimates and net chance of a longer survival by at least *m* months for proportional hazards (A), early survival difference (B), delayed survival difference (C), curable disease (D), and crossing hazards (E). *Δ(m)* Indicates net chance of a longer survival by at least *m* months. *Δ(m)* = P[*T* > *C* + *m*] – P[*C* > *T* + *m*], where P[*T* > *C* + *m*] is the probability that a random patient in the treatment group survives by at least *m* months longer than a random patient receiving the control intervention, while P[*C* > *T* + *m*] is the probability that a random patient in the control group survives by at least *m* months longer than a random patient in the treatment group. C indicates control group; P, probability; T, treatment group.

a better outcome when they received gefitinib and experienced much longer PFS times. Now that the predictive value of *EGFR* mutations is established, clinical trials would no longer include all patients, or an analysis stratified by *EGFR* mutation would be performed.

Conclusions

The net chance of a longer survival may be particularly helpful as a measure of treatment effect when hazards are not proportional. Such nonproportional hazards may result from the following 2 main mechanisms: interactions between the treatment effect and features of the patient or disease (eg, *EGFR* mutations in the Iressa Pan-Asia Study[17]), and variation of the treatment effect over time (eg, in trials comparing transplantation with nontransplantation strategies). In such cases, Kaplan-Meier curves often display unusual shapes, and standard comparison techniques may lead to erroneous conclusions.[23,31] Even in these cases, the net chance of a longer survival provides a heuristic interpretation of the treatment effects on the time scale, which is the scale most relevant to patients and health care professionals.

ARTICLE INFORMATION

Accepted for Publication: December 16, 2015.

Published Online: April 28, 2016. doi:10.1001/jamaoncol.2015.6359.

Author Contributions: Dr Péron had full access to all 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: Péron, Roy, Roche, Buyse.
Acquisition, analysis, or interpretation of data: All authors.
Drafting of the manuscript: Péron, Buyse.
Critical revision of the manuscript for important intellectual content: Roy, Ozenne, Roche, Buyse.
Statistical analysis: All authors.
Administrative, technical, or material support: Roy.
Study supervision: Roy, Buyse.

Conflict of Interest Disclosures: None reported.

Funding/Support: This study was supported by a grant from the Nuovo-Soldati research foundation (Dr Péron).

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

REFERENCES

1. Hattori S, Henmi M. Estimation of treatment effects based on possibly misspecified Cox regression. *Lifetime Data Anal*. 2012;18(4):408-433.

2. Schoenfeld DA. Sample-size formula for the proportional-hazards regression model. *Biometrics*. 1983;39(2):499-503.

3. Bobbio M, Demichelis B, Giustetto G. Completeness of reporting trial results. *Lancet*. 1994;343(8907):1209-1211.

4. Royston P, Parmar MKB. Restricted mean survival time. *BMC Med Res Methodol*. 2013;13:152.

5. Klein JP, Logan B, Harhoff M, Andersen PK. Analyzing survival curves at a fixed point in time. *Stat Med*. 2007;26(24):4505-4519.

6. Royston P, Parmar MKB. An approach to trial design and analysis in the era of non-proportional hazards of the treatment effect. *Trials*. 2014;15:314.

7. 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*. 2012;9(5):570-577.

8. Logan BR, Klein JP, Zhang MJ. Comparing treatments in the presence of crossing survival curves: an application to bone marrow transplantation. *Biometrics*. 2008;64(3):733-740.

9. Royston P, Parmar MKB, Altman DG. Visualizing length of survival in time-to-event studies. *J Natl Cancer Inst*. 2008;100(2):92-97.

10. Seruga B, Pond GR, Hertz PC, Amir E, Ocana A, Tannock IF. Comparison of absolute benefits of anticancer therapies determined by snapshot and area methods. *Ann Oncol*. 2012;23(11):2977-2982.

11. Coory M, Lamb KE, Sorich M. Risk-difference curves can be used to communicate time-dependent effects of adjuvant therapies for early stage cancer. *J Clin Epidemiol*. 2014;67(9):966-972.

12. Buyse M. Generalized pairwise comparisons of prioritized outcomes in the two-sample problem. *Stat Med*. 2010;29(30):3245-3257.

13. Buyse M. Reformulating the hazard ratio to enhance communication with clinical investigators. *Clin Trials*. 2008;5(6):641-642.

14. Johnston S, Pippen J Jr, Pivot X, et al. Lapatinib combined with letrozole versus letrozole and placebo as first-line therapy for postmenopausal hormone receptor–positive metastatic breast cancer. *J Clin Oncol*. 2009;27(33):5538-5546.

15. Maio M, Grob J-J, Aamdal S, et al. Five-year survival rates for treatment-naive patients with advanced melanoma who received ipilimumab plus dacarbazine in a phase III trial. *J Clin Oncol*. 2015;33(10):1191-1196.

16. Oliansky DM, Rizzo JD, Aplan PD, et al. The role of cytotoxic therapy with hematopoietic stem cell transplantation in the therapy of acute myeloid leukemia in children: an evidence-based review. *Biol Blood Marrow Transplant*. 2007;13(1):1-25.

17. Mok TS, Wu YL, Thongprasert S, et al. Gefitinib or carboplatin–paclitaxel in pulmonary adenocarcinoma. *N Engl J Med*. 2009;361(10):947-957.

18. Acion L, Peterson JJ, Temple S, Arndt S. Probabilistic index. *Stat Med*. 2006;25(4):591-602.

19. Moser BK, McCann MH. Reformulating the hazard ratio to enhance communication with clinical investigators. *Clin Trials*. 2008;5(3):248-252.

20. Senn S. Probabilistic index: an intuitive non-parametric approach to measuring the size of treatment effects by L. Acion, J. J. Peterson, S. Temple and S. Arndt, *Statistics in Medicine* 2006; 25(4):591-602. *Stat Med*. 2006;25(22):3944-3946.

21. Acion L, Peterson JJ, Temple S, Arndt S. Authors' reply. *Stat Med*. 2006;25(22):3946-3948. doi:10.1002/sim.2586.

22. Tan LB, Murphy R. Shifts in mortality curves. *Lancet*. 1999;354(9187):1378-1381.

23. 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*. 2014;32(22):2380-2385.

24. Datta S. Estimating the mean life time using right censored data. *Stat Methodol*. 2005;2(1):65-69. doi:10.1016/j.stamet.2004.11.003.

25. Ajani JA. The area between the curves gets no respect. *J Clin Oncol*. 2007;25(34):5531.

26. Edgington ES, Onghena P. *Randomization Tests*. 4th ed. New York, NY: Chapman Hall/CRC Press; 2007.

27. Good P. Comparing two populations. In: Good P. *Resampling Methods: A Practical Guide to Data Analysis*. 3rd ed. New York, NY: Birkhauser; 2006:31-57.

28. Dmitrienko A, Tamhane AC, Bretz F. *Multiple Testing Problems in Pharmaceutical Statistics*. New York, NY: Chapman and Hall/CRC Press; 2009.

29. Lan KKG, DeMets DL. Discrete sequential boundaries for clinical trials. *Biometrika*. 1983;70(3):659-663. doi:10.1093/biomet/70.3.659.

30. Lynch TJ, Bell DW, Sordella R, et al. Activating mutations in the epidermal growth factor receptor underlying responsiveness of non-small-cell lung cancer to gefitinib. *N Engl J Med*. 2004;350(21):2129-2139.

31. Bellera CA, MacGrogan G, Debled M, de Lara CT, Brouste V, Mathoulin-Pélissier S. Variables with time-varying effects and the Cox model. *BMC Med Res Methodol*. 2010;16:10-20.