Clinical Oncology 30 (2018) e75–e80

Contents lists available at ScienceDirect

Clinical Oncology

journal homepage: www.clinicaloncologyonline.net

Original Article

Cancer Biology and Survival Analysis in Cancer Trials: Restricted Mean Survival Time Analysis versus Hazard Ratios

R.P. A'Hern

London, UK

Received 9 December 2017; received in revised form 16 March 2018; accepted 13 April 2018

Abstract

Hazard ratios are commonly used when comparing survival between two groups and make the assumption that the relative event rates do not change markedly during follow-up, i.e. that event rates are proportional. However, there is currently debate about the use of the proportional hazards assumption to summarise the treatment effect in survival analysis compared with restricted mean survival time (RMST) analysis, particularly in cancer trials. In many situations it is unrealistic to assume that relative event rates in two groups will be proportional throughout follow-up and, hence, RMST analysis, which does not make this assumption, may be preferable. Several benefits of the latter approach have been identified but the biological perspective is not often discussed. Biological features such as the patterns of tumour growth and response can also contribute to assessing the relative merit of these two methods; such biological considerations are the subject of this paper. The observation that the most commonly observed approximation to the underlying distribution of time to event data, the lognormal distribution, does not reliably show proportional hazards in the comparison of two groups, lends weight to a statistical approach that is not based on proportional hazards. The proportional hazards assumption should be viewed more critically when estimating treatment effects. An optimum approach may be to include both proportional hazards analysis and RMST analysis when comparing time to event endpoints.

Key words: Cancer biology; Cox regression; proportional hazards; restricted mean survival time analysis; survival analysis

Introduction

There is currently stimulating discussion about the most appropriate method of survival analysis in randomised trials in oncology, particularly focussing on the common use of the proportional hazards assumption to summarise treatment effects compared with restricted mean survival time (RMST) analysis [1–5]. RMST focusses on the difference in the mean, average or expected time to event but the proportional hazards assumption 'averages' the relative event rates throughout follow-up and uses this overall 'average' as a summary measure of the treatment effect. RMST can be most simply thought of as the area under the survival curve.

The salient features of the debate can be illustrated by considering two specific trials. Figure 1A shows overall survival in the two arms of GOG-111 [6,7], a randomised trial that illustrates changing hazards over follow-up. In GOG-111, 410 women with advanced ovarian cancer and residual masses larger than 1 cm after initial surgery were randomly assigned to receive cisplatin (75 mg/m^2) with either cyclophosphamide (750 mg/m^2) or paclitaxel (135 mg/m^2 over 24 h). The overall hazard ratio was 0.71, but it was not constant, varying from 0.53 in year 1 to 1.0 in year 8. The fitted dashed curves in Figure 1A were calculated assuming that survival in the two arms follows a lognormal distribution, the ubiquity of this survival distribution, which is rarely associated with proportional hazards, is discussed below. Royston and Parmar [6] found the power of RMST analysis was on average 90% compared with 83% for the Logrank test when hazards were non-proportional in the context they considered. In GOG-111 the hazard ratio decays (moves towards unity) as follow-up extends but there are also situations that show the reverse pattern, the initial hazard ratio is close to 1 and the later hazard ratio decreases, showing a treatment effect. Trials of immunotherapy agents provide examples of this phenomenon in which proportional hazards may not hold because sufficient time is required for the initial induction of an immune

Address for correspondence: 81 Hillier Road, Battersea, London SW11 6AX, UK.

E-mail address: roger.ahern@icloud.com.

https://doi.org/10.1016/j.clon.2018.04.011

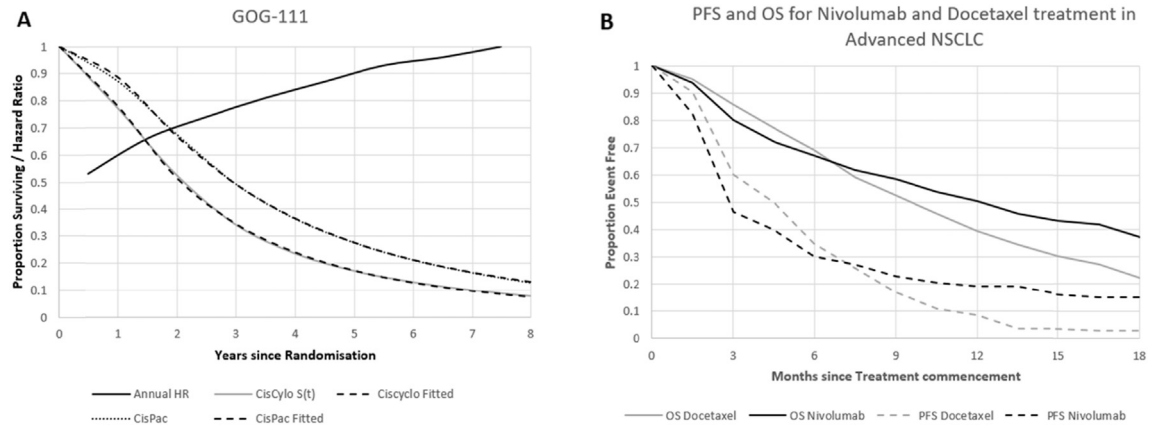

Fig 1. (A) Overall survival curves, together with the annual hazard ratio, for the two arms of GOG-111, a trial comparing two chemotherapy regimens in advanced ovarian cancer. (B) Progression-free and overall survival curves from a trial of nivolumab versus docetaxel in advanced non-small cell lung cancer. CisCyclo, cisplatin plus cyclophosphamide; CisPac, cisplatin plus paclitaxel; Fitted, fitted lognormal survival curve; HR, hazard ratio; OS, overall survival; PFS, progression-free survival.

response that is effective against the tumour. Figure 1B, for example, shows progression-free survival (PFS) and survival curves extracted from a publication of an immunotherapy trial that compared treatment with nivolumab versus docetaxel in advanced non-small cell lung cancer [8] ($n = 582$). Both PFS and overall survival only begin to show a beneficial effect for nivolumab after about 6 months. The hazard ratios were 0.92 (nivolumab: docetaxel, $P =$ ns) and 0.73 ($P = 0.002$) for PFS and overall survival, respectively, but the difference in RMST estimated from the curves was more similar, 0.6 months for PFS and 0.9 months for overall survival (both estimated up to 18 months). Statistical methods are detailed in Supplementary Appendix 1.

Three benefits of calculating mean survival will be briefly described. First, overall survival can be split into two periods when considering advanced cancer: the time to disease progression or PFS and the time from progression to death (survival post-progression; SPP). The simple equality PFS + SPP = overall survival can be best considered statistically in terms of means, i.e. mean PFS + mean SPP = mean overall survival [1]. This would imply that if mean SPP is similar in both arms of a trial, i.e. the biology of the progressive disease is similar in the two arms, the difference in mean PFS in the trial will tend to be similar to the difference in mean overall survival, a pattern consistent with that noted by a several authors in a number of different cancer types [1]. The hypothesis of similar mean SPP in the trial arms, although clearly not guaranteed in any specific trial, forms a valuable hypothesis when extrapolating differences in PFS to realistic differences in overall survival because a novel treatment may extend PFS, but there are no a priori grounds to conclude that disease that is resistant to treatment and leads to progression will lead to different survival in the two trial arms. Saad *et al.* [9] found that the median overall survival is typically three times the median PFS in advanced breast cancer trials. If this observation is applied uncritically to a trial in which the median PFSs in two arms are 7 months versus 9 months then it might be anticipated that the median overall survivals would be 21 months versus 27 months, a 6 month difference. However, the broad pattern across advanced breast cancer suggests the true overall survival difference is more likely to be similar to the difference in median PFS [10] (i.e. 21 months versus 23 months).

A second benefit of using mean survival is that it represents average duration in a specific disease state, e.g. average months progression free, hence it is of direct relevance in calculating quality-adjusted life years (QALYS), an important end point used in cost analysis to establish the cost/benefit balance of a therapy, for example by the National Institute for Health and Care Excellence (NICE) on behalf of the (UK) National Health Service [11]. Finally, in contrast to the hazard ratio, which does not have a straightforward interpretation in terms of duration of survival, RMST analysis gives a simple quantitative measure of the improvement in survival in months or years summarised over the whole period of follow-up over which it is calculated and can give a better measure of treatment effect [5].

It is relevant to ask whether biological considerations can contribute to the debate surrounding these two types of analysis. Robust estimation of treatment effects in small biologically defined subgroups of patients is necessary to accomplish the aims of personalised medicine and proportional hazards analysis may not be the optimum method to estimate such effects [12].

Tumour Growth and Long-term Outcome

The seminal model of tumour growth and treatment response, the log-kill model, was developed by Skipper *et al.* [33] at the Southern Research Institute, Birmingham, Alabama, USA and was founded on the empirical observation that leukaemia L1210 cells in mice grow exponentially up to a fatal size (10^9 cells). This model can be used to demonstrate some important properties of tumour growth and regression in response to treatment, many of which are also

pertinent to more complex growth models [13] (this model is briefly described in Supplementary Appendix 2). Figure 2A shows this exponential growth model as a simplified diagram on a semi-log plot, illustrating scenarios of no treatment, treatment of the primary tumour only and treatment of both primary and recurrent disease. If one considers treatment of advanced disease, i.e. a scenario in which there is no cure, then as implied above, an interesting implication of this model is that the difference in the time to progression (PFS) will tend to be similar to the difference in overall survival if SPP (i.e. the nett growth rate) is similar in both arms of a trial (Figure 2B). Wilkerson *et al.* [14], for example, found that estimated exponential tumour response and growth rates were independent of each other in a study involving individual patient data from chemotherapy trials in advanced prostate cancer ($n = 2353$). Note that generally the growth rate post-progression does not have to be the same but just similar in the two groups for the improvement in PFS to be similar to the improvement in overall survival. This same schema is shown extended to the commonly used and more sophisticated Gompertz growth model, in which the exponential tumour growth rate decreases with increasing tumour volume (Figure 2C). However, note that the tendency for the difference in PFS to

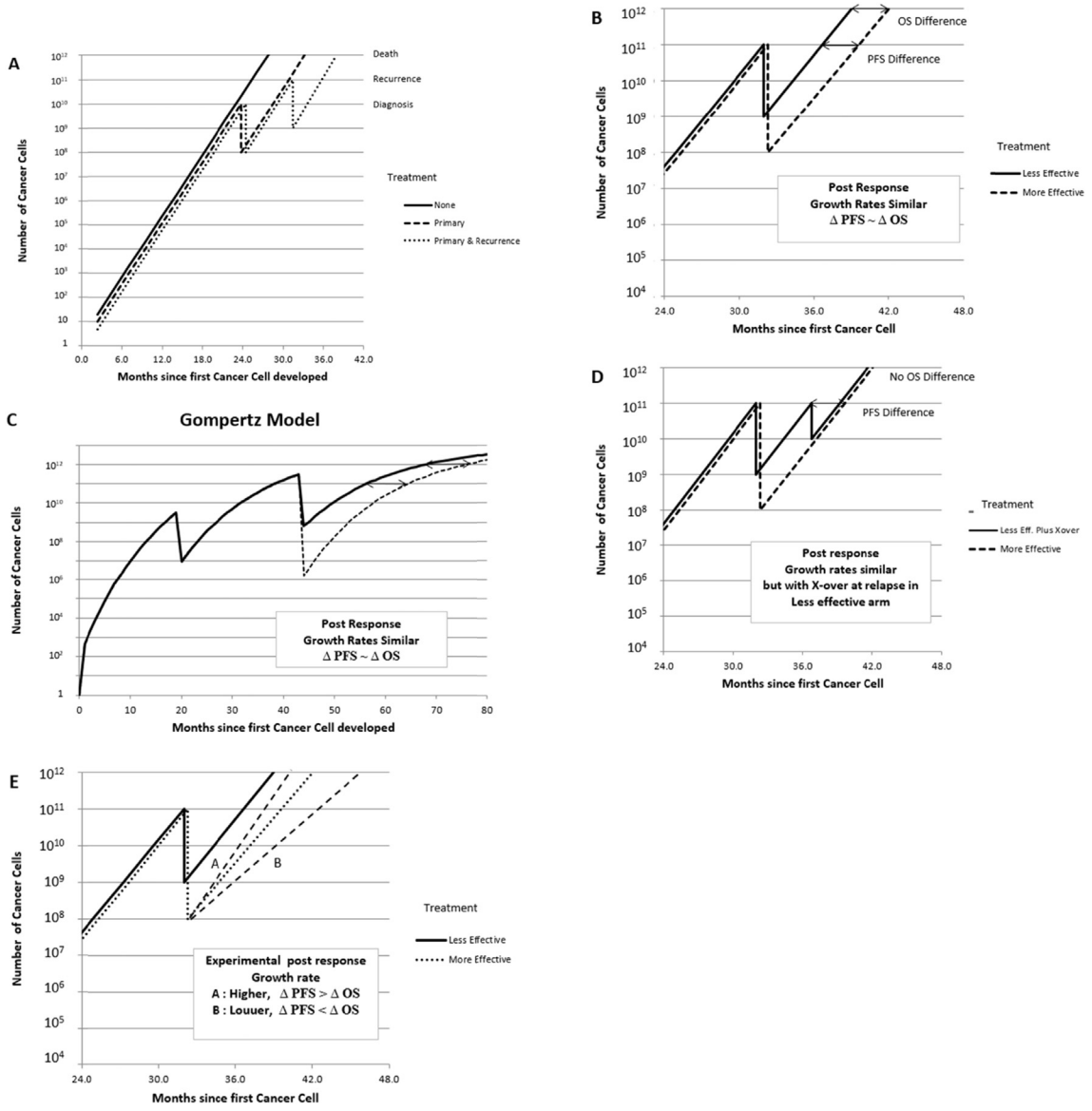

Fig 2. (A) Simplified diagram of exponential tumour growth, illustrating scenarios of no treatment, treatment of the primary tumour only and treatment of both the primary and recurrent disease. (B) Trial in advanced disease in which treatments have a 2 log and 3 log response, an interesting implication is that the difference in the time to progression (progression-free survival; PFS) will tend to be similar to the difference in overall survival if survival post-progression (i.e. the net growth rate) is similar in both arms. (C)The same schema is extended to the more sophisticated Gompertz model, in which the exponential tumour growth rate decreases with increasing tumour volume. (D) This effect will be counteracted if patients in the arm with the inferior log-kill rate are crossed over on to the other treatment at some point and show a similar log-kill. (E) A greater effect on overall survival is seen if the growth rates of the resistant clones that persist after progression are smaller in the experimental arm (curve B), but the effect on overall survival will be diminished if patients in the arm with the superior log-kill rate have a greater post-progression growth rate (curve A).

equal the difference in overall survival will be counteracted if patients in the arm with the inferior log-kill rate are crossed over on to the other treatment at some point and benefit from a similar log-kill (Figure 2D); such crossover can be a confounding factor in randomised trials in advanced disease. Piccart-Gebhart *et al.* [15], for example, observed crossover rates of 24–57% to treatment with taxanes in the control arms of trials of taxanes alone or in combination with anthracyclines as first-line therapy of patients with metastatic breast cancer.

The response to therapy given second line rather than first line may not be the same if the tumour characteristics have changed in the intervening time, e.g. the tumour could have become more resistant to treatment. Finally, note that a greater effect on overall survival will be seen if the growth rate is lower in the resistant clones that persist after progression in the experimental arm (Figure 2E, curve B) and the effect on overall survival will be diminished if the tumours of the patients in the arm with the superior log-kill rate have a greater post-progression growth rate (Figure 2E, curve A). The former scenario is clearly the most desirable outcome of treatment. General practical examples of the phenomena illustrated in Figure 2 for three cancer types are presented in Supplementary Appendix 3.

The Distribution of Tumour Growth and Long-term Outcome

In any population of patients there will be a wide variation in growth rates and treatment response and it is interesting to ask how these will be reflected in the pattern of long-term outcome. The lognormal distribution has a fundamental role to play in this context. Tumour growth rates estimated in breast cancer [16,17], prostate cancer [14] and renal cancer [18] (Supplementary Appendix 4) have been observed to have an approximate lognormal distribution, which suggests that the times to long-term outcome will also tend to have an approximate lognormal distribution if the schemas in Figure 1 are broadly representative. On the basis of extensive experience, it is also common to assume tumour growth rates are lognormally distributed in the modelling of tumour growth in animal xenograft experiments [19]. Evidence to support this in terms of overall survival of patients, looking across a wide range of cancer types, can also be found in a study of the Surveillance, Epidemiology and End Results (SEER) database [20] and a study of stage IV clinical outcomes undertaken by Qazi *et al.* [21] in three tumour types (pancreatic, breast and colon cancers); see Figure 3. More information about these two studies is given in Supplementary Appendix 5. Other studies examining single cancer types have also observed similar lognormal patterns with respect to disease progression and survival end points: in the adjuvant treatment of colorectal cancer, examining time to recurrence, disease-free survival and overall survival [12,22,23]; in trials of the treatment of advanced ovarian cancer, examining PFS and overall survival [24]; and in early breast cancer, examining disease-free survival [25,26]. It has also been noted above (Figure 1A) that survival in GOG-111 follows an approximate lognormal distribution. However, there are good reasons to suspect that intervening factors (including

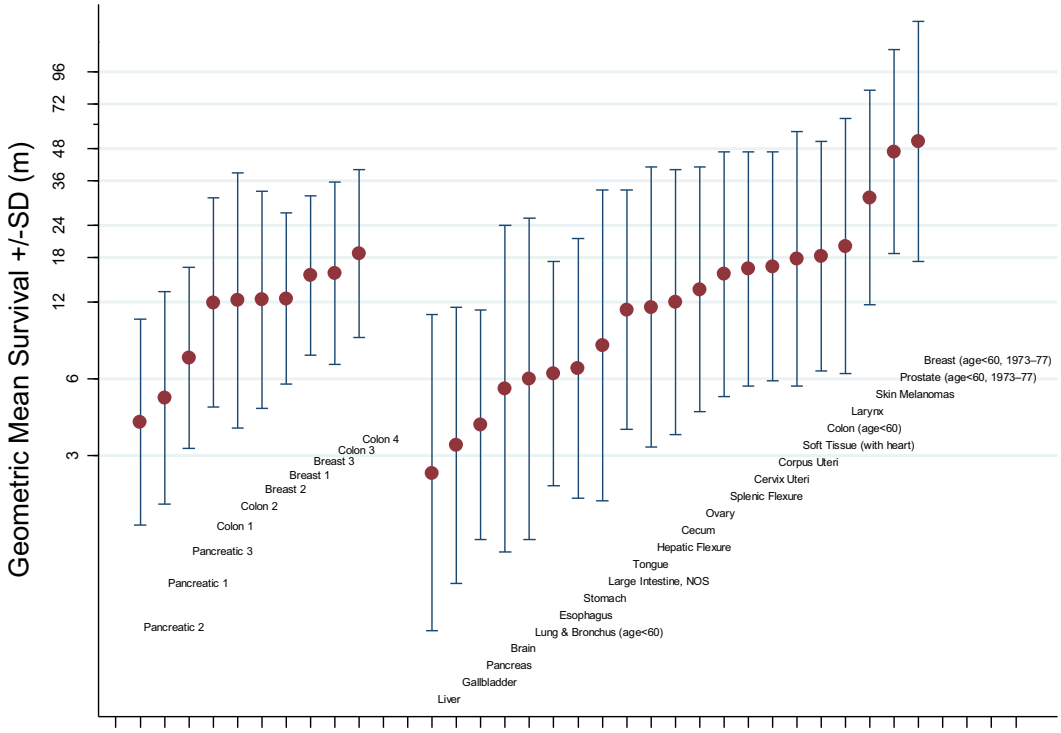

Fig 3. (A) Parameters for the lognormal survival distributions observed by Qazi *et al.* [21] and by the SEER study [20], Supplementary Appendix 5, the latter only including cancer diagnoses with >400 patients.

treatment) may contribute to obscuring a clear lognormal relationship when two groups with different prognosis are being compared, as discussed below. Cure can also occur, with the result that some patients are no longer at risk of an event due to disease. However, these patients typically remain in trial survival analyses that are undertaken on an intention to treat basis, obscuring a pattern that will exist only in patients still at risk of relapse (Supplementary Appendix 6). These phenomena, and the general complexity of estimating survival models, may explain why a wide variety of distributional models for survival times have been found to be suitable in different practical situations. Further information on the lognormal distribution, including some mathematical background, is presented in Supplementary Appendix 7.

Patterns of Hazard Ratios

It is instructive to consider the pattern of hazards and hazard ratios generated in survival analysis of lognormal distributed time to event data. Hazard ratios are not likely to be proportional across time, see, for example, the survival pattern observed in GOG-111 (Figure 1A). One interpretation of the lognormal model is that there is a consistent treatment effect on the mean survival (e.g. increase), with no confounding of this treatment effect caused by the impact of other extraneous factors with increasing follow-up time. For example, it could be inferred that if breast cancer growth rates immediately after treatment follow a lognormal distribution [16,17] then the time to recurrence and death will as a result follow a similar distribution, which is determined only by this initial growth rate pattern. However, random heritable genetic events are likely to occur in cancers throughout follow-up and there is also ongoing evolutionary selection between multiple within tumour clones [27]; these effects might be expected to make the prognosis in the two arms of a trial more similar with increasing follow-up than would be implied by a simple shift in the mean. Such a phenomenon may, for example, contribute to late events in oestrogen receptor-positive early breast cancer (Supplementary Appendix 8).

Conclusion

Biological considerations should have a greater influence on the choice of the methods of survival analysis used in cancer trials and arguments support the use of analytical methods such as RMST analysis that do not rely on the proportional hazards assumption. It has long been recognised that there is no general survival distribution that is universally applicable for survival analysis but the lognormal distribution has been observed to have relevance in a number of contexts. It is not reliably associated with proportional hazards and it is likely that biological factors such as the effect of genetic instability and clonal selection through follow-up also cause deviations from both a proportional hazards and lognormal pattern. The proportional hazards assumption should therefore be viewed more critically when estimating treatment effects. One solution would be to benchmark proportional hazards analysis against RMST analysis. Statistical programmes for the analysis of RMST are available [28] and RMST has a clear association with the estimation of QALYS, an end point that is widely used in the assessment of cost-effectiveness. The observation that hazard ratios frequently change with the duration of follow-up also suggests there may be confounding of interim analyses and their associated stopping rules if these are based on the assumption of constant hazard ratios. Statisticians have long argued for the need for trials that can detect realistic effect sizes. The simple equation mean overall survival = mean PFS + mean SPP offers a method of extrapolating a realistic survival effect size from an effect on PFS in the advanced disease setting and can also be extended to the early disease setting by incorporating cure. The above observations also suggest that tumour growth rates post-relapse should be a focus of interest in control subjects versus subjects receiving the experimental therapy, both in randomised trials of patients and in the study of animal models of cancer therapy. Note this paper has highlighted non-proportionality of treatment effects but the non-proportionality of hazards relating to prognostic factors in cancer has also long been recognised [26,29].

In practice there need not be a binary choice between RMST analysis and proportional hazards analysis, both of which have the common goal of describing the difference between two unknown event processes that have typically been estimated by undertaking a randomised controlled clinical trial. The two methods can be seen as complementary and their results will be correlated. Hence, it has been suggested that investigators can incorporate both approaches in the analysis of a specific trial with only a minor increase in the number of patients [30]. Both analysis methods show the parsimony that is necessary to efficiently test the hypothesis of the equality of survival curves by minimising the number of patients necessary to effectively compare interventions, typically achieved by testing a single primary hypothesis relating to the most important endpoint. It is pertinent to ask whether the effects of some cancer therapies have been misjudged because non-proportional treatment effects have been evaluated assuming a proportional hazards model. The consensus statement [31] describing guidelines for the content of statistical analysis plans for clinical trials recommends that details and the results of the methods used to check assumptions made in statistical analysis are routinely included in trial analysis. It would also be beneficial for the CONSORT [32] statement to include advice to publish information on the applicability of the proportional hazards assumption in situations where it is employed. This would yield a greater understanding of the frequency and impact of such potential model misspecification.

Appendix A. Supplementary data

Supplementary data related to this article can be found at https://doi.org/10.1016/j.clon.2018.04.011.

References

[1] A'Hern RP. Restricted mean survival time: an obligatory end point in cancer trials? *J Clin Oncol* 2016;34:3474–3476.

[2] Royston P, Parmar MKB. 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* 2011;30:2409–2421.

[3] Uno H, Claggett B, Tian L, Inoue E, Gallo P, Miyata T, *et al*. Moving beyond the hazard ratio in quantifying the between group difference in survival analysis. *J Clin Oncol* 2014;32: 2380–2385.

[4] Trinquart L, Jacot J, Conner SC, Porcher R. Comparison of treatment effects measured by the hazard ratio and by the ratio of restricted mean survival times in oncology randomized controlled trials. *J Clin Oncol* 2016;34:1813–1819.

[5] Lawrence J, Qui J, Bai S, Hung J. Comparison of hazard ratio and restricted mean survival analysis for cardiorenal trial analysis. Available at: https://ww2.amstat.org/meetings/biopharmworkshop/2017/onlineprogram/ViewPresentation.cfm?file=300553.pdf.

[6] Royston P, Parmar MKB. 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* 2013;13:152.

[7] McGuire WP, Hoskins WJ, Brady MF, Kucera PR, Partridge EE, Look KY, *et al*. Cyclophosphamide and cisplatin compared with paclitaxel and cisplatin in patients with stage III and stage IV ovarian cancer. *N Engl J Med* 1996;334:1–6.

[8] Borghaei H, Paz-Ares L, Horn L, Spigel DR, Steins M, Ready NE, *et al*. Nivolumab versus docetaxel in advanced nonsquamous non-small-cell lung cancer. *N Engl J Med* 2015;373(17): 1627–1639. https://doi.org/10.1056/NEJMoa1507643.

[9] Saad ED, Katz A, Buyse M. Overall survival and post-progression survival in advanced breast cancer: a review of recent randomized clinical trials. *J Clin Oncol* 2010;28: 1958–1962.

[10] Sherrill B, Amonkar M, Wu Y, Hirst C, Stein S, Walker M, *et al*. Relationship between effects on time-to-disease progression and overall survival in studies of metastatic breast cancer. *Br J Cancer* 2008;99:1572–1578.

[11] https://www.nice.org.uk/advice/lgb10/chapter/judging-the-cost-effectiveness-of-public-health-activities.

[12] Chapman JW, O'Callaghan CJ, Hu N, Ding K, Yothers GA, Catalano PJ, *et al*. Innovative estimation of survival using log-normal survival modelling on ACCENT database. *Br J Cancer* 2013;108:784–790.

[13] Gilewski TA, Dang C, Surbone A, Norton, Cytokinetics L. *Holland-Frei cancer medicine* [Chapter 38], 5th ed. 2000. ISBN-10: 1-55009-113-1.

[14] Wilkerson J, Abdallah K, Hugh-Jones C, Curt G, Rothenberg M, Simantov R, *et al*. Estimation of tumour regression and growth rates during treatment in patients with advanced prostate cancer: a retrospective analysis. *Lancet Oncol* 2017; 18(1):143–154.

[15] Piccart-Gebhart MJ, Burzykowski T, Buyse M, Sledge G, Carmichael J, Luck H-J, *et al*. Taxanes alone or in combination with anthracyclines as first-line therapy of patients with metastatic breast cancer. *J Clin Oncol* 2008;26(12):1980–1986. https://doi.org/10.1200/JCO.2007.10.8399.

[16] Kuroishi T, Tominaga S, Morimoto T, Tashiro H, Itoh S, Watanabe H, *et al*. Tumour growth rate and prognosis of breast cancer mainly detected by mass screening. *Jpn J Cancer Res* 1990;81:454–462.

[17] Weedon-Fekjær H, Lindqvist BH, Vatten LJ, Aalen OO, Tretli S. Breast cancer tumor growth estimated through mammography screening data. *Breast Cancer Res* 2008;10. R41.

[18] Stein WD, Figg WD, Dahut W. Tumor growth rates derived from data for patients in a clinical trial correlate strongly with patient purvival: a novel strategy for evaluation of clinical trial data. *Oncologist* 2008;13(10):1046–1054.

[19] Zhao L, Morgan MA, Parsels LA, Maybaum J, Lawrence TS, Normolle D, *et al*. Bayesian hierarchical changepoint methods in modelling the tumour growth profiles in xenograft experiments. *Clin Cancer Res* 2011;17(5):1057–1064.

[20] Tai P, Yu E, Cserni G, Vlastos G, Royce M, Kunkler I, *et al*. Minimum follow-up time required for the estimation of statistical cure of cancer patients: verification using data from 42 cancer sites in the SEER database. *BMC Cancer* 2005;5:48.

[21] Qazi S, DuMez D, Uckun FM. Meta-analysis of advanced cancer survival data using lognormal parametric fitting: a statistical method to identify effective treatment protocols. *Curr Pharm Des* 2007;13(15):1533–1544.

[22] Sargent DJ, Patiyil S, Yothers G, Haller DG, Gray R, Benedetti J, *et al*. End points for colon cancer adjuvant trials: observations and recommendations based on individual patient data from 20,898 patients enrolled onto 18 randomized trials from the ACCENT Group. *J Clin Oncol* 2007;25(29):4569–4574.

[23] Sargent DJ, Marsoni S, Monges G, Thibodeau SN, Labianca R, Hamilton SR, *et al*. Defective mismatch repair as a predictive marker for lack of efficacy of fluorouracil-based adjuvant therapy in colon cancer. *J Clin Oncol* 2010;28:3219–3226.

[24] Buyse M, Molenberghs G, Burzykowski T, Renard D, Geys H. The validation of surrogate endpoints in meta-analyses of randomized experiments. *Biostatistics* 2000;1(1):49–67.

[25] A'Hern R, Gillett C, Pinder S, Kalaitzaki E, Johnston S. Differing patterns of treatment effect in a trial assessing sequential docetaxel as adjuvant chemotherapy for early breast cancer (TACT). In: *San Antonio Breast Cancer Symposium* 2013 (Poster) P6-06-03.

[26] Royston P. The lognormal distribution as a model for survival time in cancer, with an emphasis on prognostic factors. *Stat Neerl* 2001;55(1):89–104.

[27] Gerlinger M, Rowan AJ, Horswell S, Larkin J, Endesfelder D, Gronroos E, *et al*. Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. *N Engl J Med* 2012;366:883–892.

[28] http://discovery.ucl.ac.uk/10041884/1/Royston_sjart_st0479.pdf.

[29] Gore SA, Pocock SJ, Kerr GR. Regression models and non-proportional hazards in the analysis of breast cancer survival. *Appl Stat* 1984;33:176–195.

[30] Royston P, Parmar MK. Augmenting the logrank test in the design of clinical trials in which nonproportional hazards of the treatment effect may be anticipated. *BMC Med Res Methodol* 2016;16:16.

[31] Gamble C, Krishan A, Stocken D, Lewis S, Juszczak E, Doré C, *et al*. Guidelines for the content of statistical analysis plans in clinical trials. *JAMA* 2017;318(23):2337–2343.

[32] http://www.consort-statement.org/

[33] Skipper HE, Schabel Jr FM, Wilcox WS. Experimental evaluation of potential anticancer agents XIII: on the criteria and kinetics associated with "curability" of experimental leukemia. *Cancer Chemother Rep* 1964;35:1.