VALUE IN HEALTH 22 (2019) 431–438

Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/jval

Methodology

Measuring Survival Benefit in Health Technology Assessment in the Presence of Nonproportional Hazards

Giles Monnickendam, MSc [1,*], *Mingshu Zhu, MSc* [1], *Jan McKendrick, MSc* [1,2], *Yun Su, MD* [3]

[1]*PRMA Consulting, Fleet, UK;* [2]*University of Technology Sydney, Sydney, New South Wales, Australia;* [3]*Pfizer Inc, New York, NY, USA*

ABSTRACT

Background: Proportional hazards (PH) is an assumption often made by researchers, despite evidence of nonproportionality in a significant proportion of clinical trials. In the presence of non-PH, the interpretation of hazard ratios, medians, and landmark survival as summary measures of treatment effect can become problematic. Several recent studies have recommended restricted mean survival time (RMST) as an alternative metric for survival analysis, particularly where non-PH may apply. **Objectives:** To determine the current approaches of health technology assessment (HTA) agencies to value assessment in the presence of non-PH, and the extent to which RMST is accepted as an alternative measure of treatment benefit. **Methods:** Methodological guidelines published by 10 HTA agencies were reviewed to establish recommended approaches for presenting survival benefit from clinical trials. Published HTA reports for 23 oncology agents approved by the US Food and Drug Administration and the European Medicines Agency since 2014 were reviewed to determine how guidelines are implemented in practice and identify instances where the PH assumption was tested and RMST analyses reported. **Results:** Testing for non-PH is not widely incorporated into HTA except by the UK National Institute for Health and Care Excellence. RMST is used infrequently but has been used in a number of countries, particularly by agencies that focus on cost effectiveness. **Conclusions:** HTA agencies vary in their approaches to non-PH. Most do not routinely check the PH assumption. RMST has played a role in assessing clinical benefit within HTA, although not consistently within countries (across drugs) or across countries (for the same drug).

Keywords: HTA, nonproportional hazards, RMST, survival

Introduction

Survival Analysis

The relative efficacy of oncology interventions is usually assessed through survival analysis, with Kaplan-Meier (KM) plots used to present estimates of the evolution of survival probability over the length of clinical trial follow-up. Metrics that summarize the between-group difference in survival experience represented by KM curves into a single figure are also reported. The most commonly used summary metrics in clinical trials are the difference in median survival and the hazard ratio (HR).

Median survival is attractive as a summary measure because it is broadly recognized, it is an absolute measure of survival (measured in months), and it is intuitively appealing as the expected survival time of an "average" patient. Median survival requires observations on time-to-event of the first 50% of the population to experience the event and no more.

In contrast to median survival, the HR summarizes all the information in the KM curves over the full duration of trial follow-up. The HR is measured on an index scale, centered on 1. This allows HRs to be compared as a measure of relative efficacy across different trial populations or indications. HRs are typically estimated using a Cox regression model, requiring an assumption of proportional hazards (PH). An advantage of the HR over the median is that it can be reported before median has been reached.

Landmark survival rates are the proportion of subjects still alive at notable milestones, taken directly from the KM curve.

Conflicts of interest: G. Monnickendam, M. Zhu, and J. McKendrick are employees of PRMA Consulting who were paid for being consultants to Pfizer in connection with the development of this article. Y. Su has no conflicts of interest with regard to the content of this article.

* *Address correspondence to:* Giles Monnickendam, MSc, PRMA Consulting, Linea House, Harvest Crescent, Ancells Business Park, Fleet GU51 2UZ, UK.

E-mail: gmonnickendam@prmaconsulting.com

1098-3015 - see front matter

https://doi.org/10.1016/j.jval.2019.01.005

Landmark survival rates may be used to summarize comparative clinical benefit and are sometimes reported in clinical trials, although rarely used as an evaluative endpoint.

Measuring Survival Benefit in HTA

Health technology assessment (HTA) agencies involved in the pricing and reimbursement of new treatments are concerned with how much incremental benefit a new treatment offers. Equivalence or an improvement in clinical benefit is considered proven (on the basis of regulatory approval), and the HTA agency's focus is on understanding the magnitude of any difference in efficacy between 2 products, to determine willingness to pay.

HTA agencies vary in the methods they use to assess the value of new treatments, but can be broadly grouped into 2 categories according to their principal decision-making criteria: clinical-benefit agencies and cost-effectiveness agencies.

We include agencies such as the German Federal Joint Committee (*Gemeinsamer Bundesausschuss*) and the French National Authority for Health (*Haute Autorité de Santé*) in the clinical-benefit archetype. Their approach relies on comparing clinical effectiveness, relying almost exclusively on evidence produced directly from clinical trials. For oncology, this results in a strong focus on metrics that reflect observed survival, such as median survival and HR.

We include agencies such as the UK National Institute for Health and Care Excellence (NICE), the Australian Pharmaceutical Benefits Advisory Committee (PBAC), and the Canadian Agency for Drugs and Technologies in Health (CADTH) in the cost-effectiveness archetype. These agencies evaluate new technologies in 2 steps: an initial assessment of comparative clinical effectiveness focusing on clinical trial metrics, followed by assessment of incremental cost effectiveness, which involves translating clinical evidence into nonclinical metrics, such as quality-adjusted life-years. In the usual case in which survival data are incomplete, the calculation of quality-adjusted life-years requires estimates of mean survival, generated from post hoc extrapolation of trial data for the period beyond trial follow-up.

A common methodology has been developed for estimating mean survival in HTA.[1,2] The main advantage of estimating mean survival is that it produces a measure that represents the expected survival calculated over the full patient population and potential lifetime. Nevertheless, there are variations in approach and some degree of subjectivity involved in determining the optimal choice from these variations.[2–4] Consequently, although there are well-developed methods for calculating extrapolated mean survival, it is not used routinely to report clinical trial outcomes and it is accepted only by some HTA agencies as a valid approach to summarize clinical benefit.

Nonproportional Hazards

PH implies that the relationship between the hazard rates in 2 treatment groups being compared is consistent over the measurement period. It is not necessary that the hazard rates within each group are constant, but rather that when they change, they do so approximately at the same time as changes in the comparator group, and in a proportional manner over time.

Figure 1 provides examples of how PH and non-PH can manifest within KM curves for overall survival in oncology. Figure 1A, comparing trifluridine plus tipiracil with placebo for refractory metastatic colorectal cancer, shows a classic PH relationship, with a consistent HR over time demonstrated by curves first moving steadily apart and then converging asymptotically. Figure 1B, comparing panobinostat plus bortezomib plus dexamethasone with bortezomib plus dexamethasone alone in relapsed multiple myeloma, illustrates a case of non-PH in which HR initially decreases and then increases, demonstrated most clearly by initial opening followed by closing and crossing of survival curves at 40 months. Figure 1C, comparing nivolumab with dacarbazine in untreated melanoma, illustrates non-PH where the HR declines over time, with a plateau emerging in the curve for the experimental group while the curve for the control group continues to steadily decline. Figure 1D, comparing inotuzumab ozogamicin with investigator's choice of chemotherapy in acute lymphoblastic leukemia, illustrates a case of non-PH with the HR consistent for a period and then declining, with survival curves running approximately together before separating noticeably at 15 months.

The PH assumption is convenient and attractive for researchers seeking to quantify relative treatment effect, in particular because it is an underlying assumption for the calculation of the HR from the Cox model. Nevertheless, there is evidence of non-PH in a significant proportion of clinical trials,[5,6] yet the implications of this are rarely considered in the design of trials or discussed in the reporting of results.[6]

If hazards are not proportional, estimates of the HR from Cox models may no longer be a meaningful measure of the between-group difference in survival.[7] The HR from the Cox model cannot be considered as a simple average of the treatment effect over time if the PH assumption fails. Instead the HR becomes a weighted average, on the basis of the number of events. It is also questionable whether it is valid to represent the differences in survival experience by a single figure when the relative benefit provided by treatment can vary significantly depending on the patient's baseline expected survival time. The reported HR also becomes dependent on the length of follow-up and the pattern of censoring across the treatment groups, which means it is no longer a robust measure of survival differences.[8]

The introduction of time-dependent covariates can provide a solution to non-PH within the Cox model. There are, however, practical challenges for implementation, interpretation, and comparison,[9,10] and we do not see this approach used in either clinical trial reporting or HTA.

Non-PH also presents problems for the interpretation of median difference. If the difference in survival over the first half of the survival distribution (up to the point of the median estimate) cannot be assumed to be similar over the second half—which is the case with non-PH—then the median difference may not be informative of the expected treatment benefit across the full patient population.[11,12]

A range of methods exists for the estimation of mean survival in the presence of non-PH. Some parametric regression models can accommodate non-PH, and cost-effectiveness HTA agencies also use independent regression of each treatment arm to extrapolate survival curves, or implement complex piecewise modeling of survival functions to formally accommodate changing hazards over time within each treatment group. Nevertheless, use of these techniques requires non-PH to be identified first—testing for non-PH needs to be an integral part of the process for estimating mean survival.

PH may be a reasonable assumption for comparisons of treatments within the same class and with the same mechanism of action; nevertheless, when new classes of oncology therapy are introduced and evaluated against comparators from established classes, the routine assumption of PH is less appropriate.[3] Immuno-oncologic (IO) therapies in particular have produced atypical survival curves in comparison with cytotoxic chemotherapy drugs.[13–17] Delayed onset of treatment effect, "pseudo-progression", and subsets of long-term survivors with IO therapies are possible causes of non-PH in comparisons of IO and chemotherapy agents.[18–20]

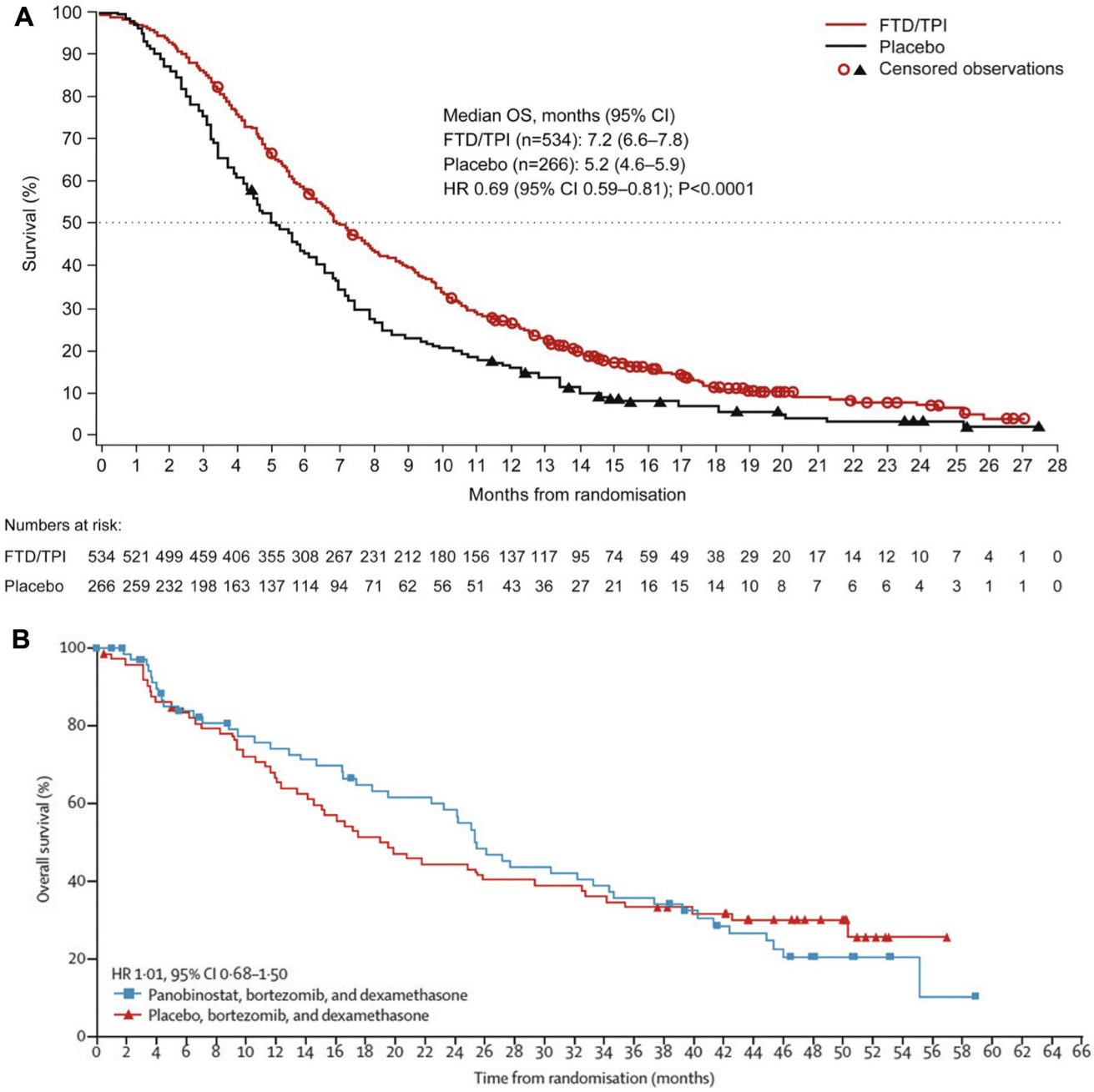

Fig. 1 – (A) PH, overall survival for FTD plus tipiracil vs placebo in refractory metastatic colorectal cancer.[33] (B) Non-PH, overall survival for panobinostat for subgroup with more than 2 previous therapies in relapsed multiple myeloma.[34] (C) Non-PH, overall survival for nivolumab vs dacarbazine in untreated melanoma.[16] (D) Non-PH, overall survival for inotuzumab ozogamicin vs standard therapy in acute lymphoblastic leukemia.[15] CI indicates confidence interval; FTD, trifluridine; HR, hazard ratio; OS, overall survival; PH proportional hazards; TPI, tipiracil.

Restricted Mean Survival Time

Restricted mean survival time (RMST) has been recommended as an alternative summary measure of treatment effect that remains meaningful in the presence of non-PH.[21] RMST is an established but not widely used survival metric—we are aware of only 1 occasion where it has been specified as an endpoint in an oncology trial.[22]

RMST is the average duration of survival for the trial population up to a given time point, t^*, typically the end of trial follow-up for the group with the shortest follow-up, such that the measurement period is the same for all groups compared. RMST is calculated as the area under the KM curve up to t^*, and is measured in units of time. The difference in RMST between treatment and control groups represents the treatment effect in absolute terms; the ratio of RMST can also be used as a relative

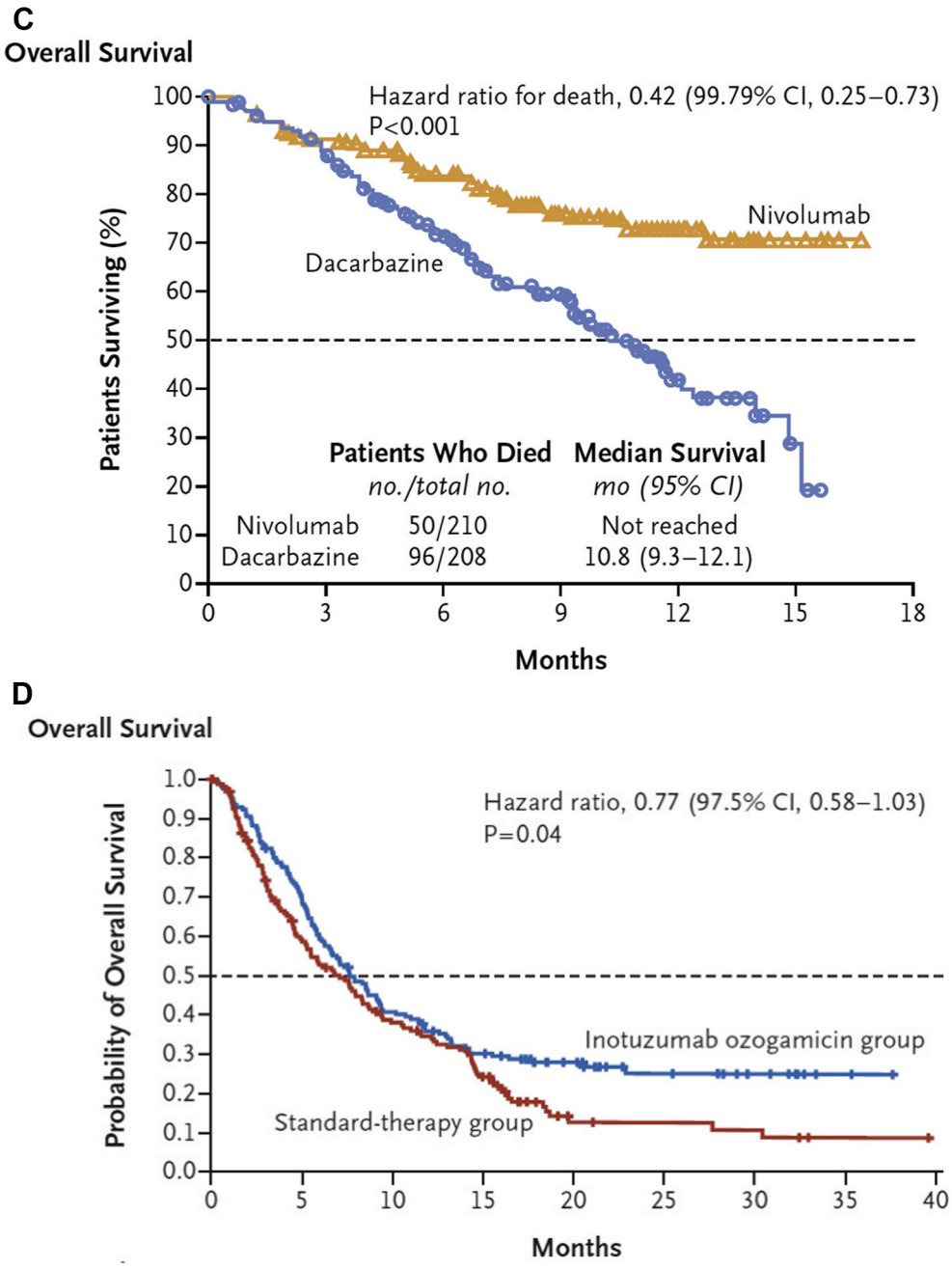

Fig. 1 – *Continued*

measure of efficacy. RMST summarizes KM curves over the full duration of follow-up, but unlike the HR it is fully nonparametric—no assumptions are required for calculation or interpretation. RMST is meaningful and robust when calculated over a range of follow-up times and so can be used to facilitate greater understanding of the evolution of comparative survival benefit over time. Unlike landmark survival rates, RMST does not represent a snapshot at a single point in time, because it integrates the survival function up to t^*.

A number of studies have considered the potential role of RMST in oncology trials. A'Hern[23] regards the HR and RMST as complementary techniques for summarizing treatment effect and argues that RMST should always be considered, given that PH cannot routinely be assumed in oncology trials.[23] Uno et al[8] recommend using the ratio of RMST estimates, with an appropriate prespecified value of t^*, as a model-free measure of treatment effect with clinical and analytic interpretability. Huang and Kuan[24] note that with the emergence of novel cancer treatments, including immunotherapies, observations of substantial departure from PH in clinical trials are not uncommon; they recommend including RMST analysis in studies in which it is suspected that non-PH may occur.

In an analysis by Trinquart et al,[5] comparison of the HR with the ratio of RMST showed that, on average, the HR provided a significantly larger estimate of treatment effect. Although the RMST and HR were usually consistent regarding the statistical

significance of the treatment effect, measures based on the RMST yielded more conservative estimates. The study concluded that RMST-based measures should routinely be reported in trials with time-to-event outcomes.[5]

Methods

The objective of this study was to identify whether HTA agencies routinely assess PH, the methods used when non-PH are observed in evaluations of oncology drugs, and the extent to which RMST is accepted as a measure of treatment benefit in these circumstances. We reviewed the methods of HTA agencies in major developed countries, as documented in published guidelines, and their actual practices in evaluations of oncology drugs.

HTA Guidelines

To establish the recommended approaches to presenting survival benefit from clinical trials, we reviewed methodological guidelines published by 10 HTA agencies in 8 countries: Australia (PBAC), Canada (CADTH), France (National Authority for Health and the Transparency Committee [*Commission de Transparence*, TC]), Germany (German Federal Joint Committee and the Institute for Quality and Efficiency in Healthcare [*Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen*, IQWiG]), Italy (regional HTA agencies in Emilia-Romagna and Veneto), Spain (Group for the Evaluation of Innovations, Standardization, and Research in Drug Selection [*Grupo de Evaluación de Novedades, Estandarización e Investigación en Selección de Medicamentos*, GENESIS]), Sweden (Dental and Pharmaceutical Benefits Agency [*Tandvårds- & läkemedelsförmånsverket, TLV*]), and the United Kingdom (NICE and the Scottish Medicines Consortium [SMC]). We also reviewed published material from key influencing organizations in the United States, where no formal HTA agencies exist; these organizations were the Institute for Clinical and Economic Review (ICER), the Academy of Managed Care Pharmacy, the Agency for Healthcare Research and Quality, the International Society for Pharmacoeconomics and Outcomes Research, and the Memorial Sloan Kettering Cancer Center.

The methods recommended for detecting and taking account of non-PH were identified when applicable. References to, and recommendations on, the use of RMST as a metric for assessment of survival benefit were identified.

Oncology Appraisals

To determine how guidelines are implemented in practice, we selected oncology drugs approved by the US Food and Drug Administration (FDA) and the European Medicines Agency (EMA) between 2014 and 2016 and reviewed published HTA reports from the 10 agencies for the relevant drug/indication combinations.

New oncology drug approvals by the FDA were identified from the Hematology/Oncology Approvals and Safety Notifications list, published on the FDA website, for the period January 1, 2014, to December 31, 2016. Approvals based on single-arm trials were excluded as were those for products not new to an indication (eg, biosimilars, companion diagnostics, or new methods of administration). A total of 75 records were identified, which reduced to 52 after applying exclusion criteria.

In the United Kingdom, NICE selects new drugs for assessment before EMA approval, and appraisal consultation documents (ACDs) are published at or very soon after approval. To identify drugs that had gained EMA approval and entered into HTA processes, we excluded drugs for which a NICE ACD had not been published by December 31, 2016. Of the 52 drugs selected from the FDA list, 22 had published NICE ACDs by this date.

For the 22 drug/indication combinations, we searched across the 10 scope agencies for public assessment documents published up to June 30, 2017; a total of 132 HTAs were identified. We reviewed the published documents to identify instances where the PH assumption was examined or formally tested, and where RMST analysis was reported, either within manufacturer's submissions or within HTA agency analysis.

We also reviewed published ICER reports on multiple myeloma and non–small cell lung cancer (NSCLC) to determine the approach used by the agency in its assessment process.

Results

HTA agencies vary in their approaches to non-PH, as shown in HTA guidelines (Table 1) and in practice (Table 2).

Three cost-effectiveness HTAs—NICE, PBAC, and CADTH—recognize the importance of the PH assumption in their methodological guidelines and recommend testing. SMC guidelines do not specifically mention the PH assumption, but indicate in broad terms that NICE methods are applied for SMC assessments. Of the clinical-benefit HTA agencies, only IQWiG referred to non-PH within guidelines. IQWiG uses the HR as its preferred measure of clinical benefit but recognizes that this cannot be interpreted meaningfully when non-PH applies. In these circumstances, IQWiG appears to recommend the use of landmark analysis alongside the HR. GENESIS does not specifically mention non-PH, but recognizes that the median difference may not be an appropriate measure of survival benefit if the relationship between survival curves changes in the latter part of the distribution.

The PH assumption is not mentioned in materials published by any of the US influencing organizations, with the exception of ICER, which has acknowledged its importance and also tested PH in both the HTA reports reviewed.

In practice, the PH assumption was not tested consistently within the assessment process of any agency. PH was assessed in 15 of the 22 NICE appraisals reviewed—formal statistical methods were used in each case, but varied in the type of test used. We found evidence of the assessment of PH in only 3 of 15 assessments by PBAC and 1 of 12 by CADTH, with formal statistical tests applied only once by each agency. SMC documents indicate that PH were assessed in 7 of 16 assessments, although given the summary nature of the reports, it was not possible to identify the extent to which formal statistical tests were used. We found evidence of the assessment of PH in only 1 clinical-benefit HTA.[25] The TC (France) identified and extensively discussed non-PH in the assessment of nivolumab for second-line treatment of nonsquamous NSCLC. It recognized that standard methods based on median difference and HR were insufficient to evaluate the incremental survival benefit of nivolumab, and that new methodological approaches were required for IO therapies. This was a significant evolution of the TC's position—in an earlier assessment of nivolumab for second-line treatment of squamous NSCLC,[26] non-PH were present but not discussed in those terms, and although it was acknowledged that median difference would not capture the strong survival benefit observed in the latter part of the KM curve, the HR was accepted as an appropriate measure of treatment effect over the full duration of follow-up.

The guidelines of half the HTA agencies included in the review referred to RMST; with the exception of GENESIS, these are all cost-effectiveness agencies. The TLV in Sweden is the only cost-effectiveness agency not to refer to RMST in its guidelines. For cost-effectiveness agencies, extrapolated mean survival time is always the preferred approach for assessment of cost

Table 1 – Recommendations for PH testing and use of RMST.

Country	Agency	Agency type	Non-PH recognized?	PH testing recommended?	RMST in guidelines?
Australia	PBAC	CE	Yes	Yes	Yes
Canada	CADTH	CE	Yes	Yes	Yes*
France	TC (HAS)	CB	No	No	No
Germany	IQWiG (G-BA)	CB	Yes	No	No
Italy	Emilia-Romagna	CB	No	No	No
	Veneto	CB	No	No	No
Spain	GENESIS	CB/CE†	No	No	Yes*
Sweden	TLV	CE	No	No	No
United Kingdom	NICE	CE	Yes	Yes	Yes
	SMC	CE	No	No	Yes*

CADTH indicates Canadian Agency for Drugs and Technologies in Health; CB, clinical benefit; CE, cost effectiveness; G-BA, *Gemeinsamer Bundesausschuss*; GENESIS, *Grupo de Evaluación de Novedades, Estandarización e Investigación en Selección de Medicamentos*; HAS, *Haute Autorité de Santé*; IQWiG, *Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen*; NICE, National Institute for Health and Care Excellence; PBAC, Australian Pharmaceutical Benefits Advisory Committee; PH, proportional hazards; RMST, restricted mean survival time; SMC, Scottish Medicines Consortium; TC, Transparency Committee; TLV, *Tandvårds- & läkemedelsförmånsverket*.

* RMST was mentioned without using the exact term (eg, referred to as "area under the curve" or "truncated survival benefit").

† GENESIS is considered a hybrid agency—clinical-benefit and cost-effectiveness evidence are presented without specific criteria for decision making.

Table 2 – Assessment of PH and use of RMST in HTAs.

Country	Agency	Agency type	HTAs with PH assessment	HTAs with RMST reported
Australia	PBAC	CE	3 of 15 (20%)	1 of 15 (7%)
Canada	CADTH	CE	1 of 12 (8%)	1 of 12 (8%)
France	TC (HAS)	CB	1 of 14 (7%)	0 of 14 (0%)
Germany	IQWiG (G-BA)	CB	0 of 18 (0%)	2 of 18 (11%)
Italy	Emilia-Romagna	CB	0 of 8 (0%)	0 of 8 (0%)
	Veneto	CB	0 of 4 (0%)	0 of 4 (0%)
Spain	GENESIS	CB/CE*	1 of 9 (11%)	2 of 9 (22%)
Sweden	TLV	CE	2 of 14 (14%)	0 of 14 (0%)
United Kingdom	NICE	CE	15 of 22 (68%)	10 of 22 (45%)
	SMC	CE	7 of 16 (44%)	1 of 16 (6%)

CADTH indicates Canadian Agency for Drugs and Technologies in Health; CB, clinical benefit; CE, cost effectiveness; G-BA, *Gemeinsamer Bundesausschuss*; GENESIS, *Grupo de Evaluación de Novedades, Estandarización e Investigación en Selección de Medicamentos*; HAS, *Haute Autorité de Santé*; HTA, health technology assessment; IQWiG, *Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen*; NICE, National Institute for Health and Care Excellence; PBAC, Australian Pharmaceutical Benefits Advisory Committee; PH, proportional hazards; RMST, restricted mean survival time; SMC, Scottish Medicines Consortium; TC, Transparency Committee; TLV, *Tandvårds- & läkemedelsförmånsverket*.

* GENESIS is considered a hybrid agency—clinical-benefit and cost-effectiveness evidence are presented without indicating criteria for decision making.

effectiveness, but RMST is recommended for scenario analysis or for use in the baseline assessment in specific circumstances (eg, when data are almost complete). GENESIS is notable for recommending plurality in methods to assess survival benefit, suggesting that the median, RMST, and HR should be analyzed together and any difference in signal understood.

Overall, RMST was used infrequently in HTAs, but with broad use across countries: RMST was reported in at least 1 assessment by 6 of the 10 HTA agencies reviewed. It was much more commonly used in assessments by cost-effectiveness agencies, reported in 13 of 79 assessments (16%), compared with 4 of 53 assessments by clinical-benefit agencies (8%).

RMST was used most frequently by NICE. It was used in a number of assessments to validate estimates of extrapolated mean survival in cost-utility analyses; we did not find any instances where it was used directly for the primary cost-effectiveness analysis. In 1 NICE assessment, RMST was used as the measure of treatment effect to address issues of non-PH in an indirect treatment comparison.[27] In 1 case, pembrolizumab for the treatment of NSCLC, RMST was used in the manufacturer's submission to demonstrate the impact of non-PH and provide evidence of treatment effect when median difference was not statistically significant.[28] Non-PH were also identified in the CADTH assessment of pembrolizumab for NSCLC, which used RMST in a scenario analysis in cost-utility modeling.[29]

In the assessment of nivolumab for nonsquamous NSCLC by PBAC (Australia), the PH assumption was tested and found to be invalid.[30] The agency recommended reporting RMST, but the manufacturer did not include it in the submission.

In the GENESIS assessment of ramucirumab for treatment of gastric adenocarcinoma, RMST was used to estimate survival benefit, although referred to as the "area under the curve" method of estimating survival. In the IQWiG assessment of ramucirumab for the same indication, RMST was reported in tables alongside other efficacy data but not discussed in the analysis.

Discussion

Our research indicates that the attention given by HTA agencies to non-PH can be significant, particularly those that focus on cost effectiveness. Nevertheless, testing for PH and use of appropriate methods when these tests fail remain inconsistent.

It is particularly concerning that among clinical-benefit HTA agencies—which assess value directly from clinical trial summary measures—understanding of the implications of non-PH appears more limited and testing is neither recommended nor routinely conducted. Conventional clinical trial measures such as the median and HR may be insufficient to describe and evaluate differences in survival with new oncology drugs. The complementary use of RMST may be beneficial because it is a single summary measure, using only observed data with no assumptions required, and is robust and meaningful irrespective of whether PH applies.

The TC assessments of nivolumab in squamous and nonsquamous NSCLC suggest that with increasing experience of evaluating IOs, awareness of the challenges of non-PH for evaluating survival benefit may increase among clinical-benefit HTAs. This may lead to the development of new methodological approaches, which could potentially include more widespread use of RMST, alongside the median and HR.

RMST may also have a significant role to play in cost-effectiveness HTA. In some circumstances, it may be used as the measure of survival benefit in the primary cost-effectiveness analysis (eg, where data are almost complete, or no further survival benefit is expected beyond the trial period). Nevertheless, these occasions are limited—in one instance, NICE chose extrapolated mean survival over RMST when only 1 patient remained alive.[31] RMST has the potential for a more substantial role in secondary analysis, including in the validation of models for extrapolated mean survival, meta-analyses, and indirect treatment comparisons in the presence of non-PH. Finally, in the initial assessment of comparative clinical effectiveness, where the median and HR remain the key metrics used, RMST assessed over a range of follow-up times can be a valuable tool to analyze the nature of nonproportionality and develop a better understanding of the evolution of treatment benefit over time.

Although RMST is an established measure of survival benefit, it is not widely used and awareness is limited. Lack of familiarity with RMST is likely to limit its acceptance by HTA agencies, as is the fact that RMST analyses are often post hoc rather than prespecified in statistical analysis plans. This will be a particular concern for clinical-benefit HTA agencies, which often regard such analyses as purely supportive and carrying little weight in their own right.

The use of RMST involves some methodological challenges, in particular the appropriate choice of the time point t^* at which to measure survival. Solutions to methodological issues have been proposed,[32] but a consensus has not been established.

A further challenge to the use of RMST concerns the implicit value it places on different parts of the survival distribution, compared with other survival metrics. In a conventional clinical trial, where survival data are not fully mature, the numbers at risk in the tail of the survival distribution are relatively low, and so estimates of survival here are more uncertain. The difference in medians takes account of this by discounting entirely the latter part of the survival distribution; similarly, the HR from the Cox model is influenced less strongly by parts of the survival curve with few events. Nevertheless, the difference in RMST, when reported as a single summary figure for the difference in survival experience over the course of a trial, does not take account of the increasing uncertainty of observed survival at later stages of follow-up. RMST may therefore be perceived as a metric that places undue value on the more uncertain tail of the survival distribution, although this can be addressed with the appropriate use of confidence intervals.

Although there are potential benefits to using RMST, further empirical studies of the behavior of RMST and its performance compared with other survival metrics, particularly the median and HR, may be needed to increase understanding of RMST, develop consensus on its appropriate application within survival analysis, and clarify the potential implications of use in decision making by both regulators and HTA agencies.

Limitations of Research

We rely on the guidelines and individual assessment reports published by HTA agencies. Some methods may be routinely applied, but not formally documented in guidelines. For those agencies that publish more limited information, non-PH may have been assessed or RMST used without reference in published materials.

Conclusions

As treatment for cancer expands to new drug classes and new indications, non-PH may become increasingly common in clinical trials. In these circumstances, routine testing for non-PH and the selection of appropriate metrics and techniques to analyze survival differences are important to ensure that new treatments are valued appropriately. Our analysis found considerable variability in approaches to non-PH, both across and within HTA agencies. The methodological guidelines of cost-effectiveness agencies demonstrate awareness of the importance of the PH assumption, but thorough assessment of PH is not evident in practice. No agency adopted a consistent approach, even when guidelines specifically mentioned non-PH, and most agencies do not routinely test the validity of the PH assumption. The importance of the PH assumption is not well recognized by clinical-benefit agencies. This is of particular concern given their reliance on clinical trial metrics, which are problematic for interpretation in the presence of non-PH. RMST has played a role in the assessment of clinical effectiveness by both cost-effectiveness and clinical-benefit HTA agencies; nevertheless, use has been much more limited among the latter, which also do not recommend or conduct PH testing. Among the agencies that use RMST, NICE in particular has shown increasing awareness and use of RMST, especially in appraisals of cancer immunotherapies.

Acknowledgments

We thank Max Pruetzel-Thomas of PRMA Consulting who provided medical writing support funded by Pfizer Inc.

Source of financial support: This study was sponsored by Pfizer Inc.

REFERENCES

1. Latimer N. *NICE DSU Technical Support Document 14: Survival Analysis for Economic Evaluations Alongside Clinical Trials—Extrapolation With Patient-Level Data*. London: National Institute for Health and Care Excellence; 2011.
2. Latimer NR. Survival analysis for economic evaluations alongside clinical trials—extrapolation with patient-level data: inconsistencies, limitations, and a practical guide. *Med Decis Making*. 2013;33(6):743–754.
3. Bagust A, Beale S. Survival analysis and extrapolation modeling of time-to-event clinical trial data for economic evaluation: an alternative approach. *Med Decis Making*. 2014;34(3):343–351.
4. Latimer NR. Response to "survival analysis and extrapolation modeling of time-to-event clinical trial data for economic evaluation: an

alternative approach" by Bagust and Beale. *Med Decis Making*. 2014;34(3):279–282.
5. Trinquart L, Jacot J, Conner SC, et al. 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(15):1813–1819.
6. Batson S, Greenall G, Hudson P. Review of the reporting of survival analyses within randomised controlled trials and the implications for meta-analysis. *PLoS One*. 2016;11(5):e0154870.
7. Hasegawa T, Uno H, Wei LJ. Nivolumab in nonsquamous non–small-cell lung cancer. *N Engl J Med*. 2016;374(5):492–493.
8. 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.
9. Wolfe RA, Strawderman RL. Logical and statistical fallacies in the use of Cox regression models. *Am J Kidney Dis*. 1996;27(1):124–129.
10. Fisher LD, Lin DY. Time-dependent covariates in the Cox proportional-hazards regression model. *Annu Rev Public Health*. 1999;20:145–157.
11. Davies A, Briggs A, Schneider J, et al. The ends justify the mean: outcome measures for estimating the value of new cancer therapies. *Health Outcomes Res Med*. 2012;3(1):e25–e36.
12. Gould SJ. The median isn't the message. *Virtual Mentor*. 2013;15(1):77–81.
13. Borghaei H, Paz-Ares L, Horn L, et al. Nivolumab versus docetaxel in advanced nonsquamous non-small-cell lung cancer. *N Engl J Med*. 2015;373(17):1627–1639.
14. Kantarjian H, Stein A, Gokbuget N, et al. Blinatumomab versus chemotherapy for advanced acute lymphoblastic leukemia. *N Engl J Med*. 2017;376(9):836–847.
15. Kantarjian HM, DeAngelo DJ, Stelljes M, et al. Inotuzumab ozogamicin versus standard therapy for acute lymphoblastic leukemia. *N Engl J Med*. 2016;375(8):740–753.
16. Robert C, Long GV, Brady B, et al. Nivolumab in previously untreated melanoma without BRAF mutation. *N Engl J Med*. 2015;372(4):320–330.
17. Hamid O, Puzanov I, Dummer R, et al. Final OS analysis for KEYNOTE-002: pembrozliumab vs investigator-choice chemotherapy for ipilimumab-refractory melanoma. Copenhagen, Denmark: ESMO; 2016. https://oncologypro.esmo.org/Meeting-Resources/ESMO-2016/Final-overall-survival-for-KEYNOTE-002-pembrolizumab-pembro-versus-investigator-choice-chemotherapy-chemo-for-ipilimumab-ipi-refractory-melanoma.
18. Chen TT. Statistical issues and challenges in immuno-oncology. *J Immunother Cancer*. 2013;1:18.
19. Hoos A, Eggermont AM, Janetzki S, et al. Improved endpoints for cancer immunotherapy trials. *J Natl Cancer Inst*. 2010;102(18):1388–1397.
20. Wolchok JD, Hoos A, O'Day S, et al. Guidelines for the evaluation of immune therapy activity in solid tumors: immune-related response criteria. *Clin Cancer Res*. 2009;15(23):7412–7420.
21. 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*. 2011;30(19):2409–2421.
22. Perren TJ, Swart AM, Pfisterer J, et al. A phase 3 trial of bevacizumab in ovarian cancer. *N Engl J Med*. 2011;365(26):2484–2496.
23. A'Hern RP. Restricted mean survival time: an obligatory end point for time-to-event analysis in cancer trials? *J Clin Oncol*. 2016;34(28):3474–3476.
24. Huang B, Kuan PF. Comparison of the restricted mean survival time with the hazard ratio in superiority trials with a time-to-event end point. *Pharm Stat*. 2018;17(3):202–213.
25. Haute Autorité de Santé. Opdivo (CNBPC non épidermoïde)—Audition. 2017. https://www.has-sante.fr/portail/upload/docs/application/pdf/2017-06/dir71/opdivo_cbnpc_non_epidermoide_11012017_ct15261opdivo_cbnpc_non_epidermoide_11012017_transcription.pdf. Accessed January 4, 2017.
26. Haute Autorité de Santé. Opdivo—Audition. 2016. https://www.has-sante.fr/portail/upload/docs/application/pdf/2016-06/opdivo_03022016_transcription.pdf. Accessed January 4, 2017.
27. National Institute for Health and Care Excellence. *Nivolumab for Previously Treated Non-Squamous Non-Small-Cell Lung Cancer*. London: National Institute for Health and Care Excellence; 2017.
28. National Institute for Health and Care Excellence. *Pembrolizumab for Treating PD-L1-Positive Non-Small-Cell Lung Cancer After Chemotherapy*. London: National Institute for Health and Care Excellence; 2017.
29. Canadian Agency for Drugs and Technologies in Health. *Pembrolizumab (Keytruda) for Non-Small Cell Lung Cancer*. Toronto, ON: Canadian Agency for Drugs and Technologies in Health; 2016.
30. Australian Pharmaceutical Benefits Advisory Committee. Nivolumab, injection concentrate for I.V. infusion 40 mg in 4 mL, injection concentrate for I.V. infusion 100 mg in 10 mL, Opdivo®. Bristol-Myers Squibb Australia Pty Ltd; 2016. http://www.pbs.gov.au/industry/listing/elements/pbac-meetings/psd/2016-11/files/nivolumab-non-squamous-nsclc-psd-november-2016.pdf. Accessed January 16, 2017.
31. National Institute for Health and Care Excellence. *Pegylated Liposomal Irinotecan for Treating Pancreatic Cancer After Gemcitabine*. London: National Institute for Health and Care Excellence; 2017.
32. 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*. 2013;13:152.
33. Van Cutsem E, Mayer RJ, Laurent S, et al. The subgroups of the phase III RECOURSE trial of trifluridine/tipiracil (TAS-102) versus placebo with best supportive care in patients with metastatic colorectal cancer. *Eur J Cancer*. 2018;90:63–72.
34. San-Miguel JF, Hungria VT, Yoon SS, et al. Overall survival of patients with relapsed multiple myeloma treated with panobinostat or placebo plus bortezomib and dexamethasone (the PANORAMA 1 trial): a randomised, placebo-controlled, phase 3 trial. *Lancet Haematol*. 2016;3(11):e506–e515.