Extrapolating Survival Data Using Historical Trial-Based a Priori Distributions

Soikkeli, Fanni; Hashim, Mahmoud; Ouwens, Mario; Postma, Maarten; Heeg, Bart

Published in:
Value in Health

DOI:
10.1016/j.jval.2019.03.017

ScienceDirect

Contents lists available at sciencedirect.com
Journal homepage: www.elsevier.com/locate/jval

Methodology

Extrapolating Survival Data Using Historical Trial–Based a Priori Distributions

Fanni Soikkeli, MSc,[1,*] Mahmoud Hashim, MPH,[1] Mario Ouwens, PhD,[2] Maarten Postma, PhD,[3] Bart Heeg, PhD[1]

[1]Ingress Health, Rotterdam, The Netherlands; [2]AstraZeneca R&D, Mölndal, Sweden; [3]Department of Health Sciences, University Medical Center Groningen, University of Groningen, Groningen, The Netherlands

ABSTRACT

Objectives: To show how clinical trial data can be extrapolated using historical trial data–based a priori distributions.

Methods: Extrapolations based on 30-month pivotal multiple myeloma trial data were compared with 75-month data from the same trial. The 30-month data represent a typical decision-making scenario where early results from a clinical trial are extrapolated. Mature historical trial data with the same comparator as in the pivotal trial were incorporated in 2 stages. First, the parametric distribution selection was based on the historical trial data. Second, the shape parameter estimate of the historical trial was used to define an informative a priori distribution for the shape of the 30-month pivotal trial data. The method was compared with standard approaches, fitting parametric distributions to the 30-month data with noninformative prior. The predicted survival of each method was compared with the observed survival (ΔAUC) in the 75-month trial data.

Results: The Weibull had the best fit to the historical trial and the log-normal to the 30-month pivotal trial data. The ΔAUC of the Weibull with informative priors was considerably smaller compared with the standard Weibull. Also, the predicted median survival based on the Weibull with informative priors was more accurate (melphalan and prednisone [MP] 40 months, and bortezomib [V] combined with MP [VMP] 62 months) than based on the standard Weibull (MP 45 months and VMP 72 months) when compared with the observed median (MP 41.3 months and VMP 56.4 months).

Conclusions: Extrapolation of clinical trial data is improved by using historical trial data–based informative a priori distributions.

Keywords: Bayesian statistics, multiple myeloma, oncology, overall survival, survival analysis

VALUE HEALTH. 2019; 22(9):1012–1017

Introduction

Survival extrapolations beyond the clinical trial overall survival (OS) data are often needed for cost-effectiveness analysis submitted to health technology assessment (HTA) agents. The extrapolations are generally done based on standard parametric distributions. Standard parametric distributions might result in clinically implausible survival predictions and to a considerable uncertainty around the survival estimations hampering the HTA decision making.[1] As such, better utilization of available external information may lead to more accurate survival estimates.

According to the HTA guidelines, external data can be used either to assess the plausibility of extrapolations or to inform long-term survival estimates.[1] These data can be derived from general population survival, disease registries, historical clinical trials, or clinical expert opinion.[2] Few methodological studies incorporated external data in mean survival estimations.[3-6] This was usually done by either replacing the control arm of the trial by the external data and applying the relative efficacy from the trial to the external data or by incorporating the external data in the parametric fit in either frequentist or Bayesian way.

The current study presents a Bayesian method using historical trial data. In the method, mature historical trial data guide the parametric curve selection, and a priori distributions are used to inform the shape parameter of the parametric distributions fitted on the immature pivotal trial data. The OS data from the VISTA study in multiple myeloma (MM) is used by applying the methodology to the 30-month data-cut and validating the extrapolations with the 75-month follow-up data.

Methods

Data Sources and Extraction

The VISTA clinical trial[7,8] is a randomized controlled trial (RCT) comparing bortezomib (V) combined with melphalan and

* Address correspondence to: Fanni Soikkeli, MSc, Hofplein 20, 3032 AC Rotterdam, The Netherlands. Email: fanni.soikkeli@ingress-health.com

https://doi.org/10.1016/j.jval.2019.03.017

Table 1. Comparison of baseline characteristics of the MP patients in the VISTA and historical trials.

	VISTA MP arm, *n* = 344	Historical trial MP arm, *n* = 234
Age, years		
Median	71.2	71
Range	57-90	41.4-90.6
Sex		
Male, %	51	43
Type of myeloma, %		
IgG	64	62
IgA	24	28
IgD	1	1
IgM	1	0
Light chain	8	9
Biclonal	2	0
>1 class	0	3

MP indicates melphalan and prednisone; Ig, immunoglobulin; VISTA, Velcade as Initial Standard Therapy in Multiple Myeloma.

prednisone (MP) versus MP alone for previously untreated transplant ineligible patients with MM. The OS data from the first data-cut of the VISTA trial (follow-up data up to 30 months)[7] were used to create a typical decision-making scenario in which early results from a clinical trial are extrapolated. For validation of the presented methodology, the 75-month data of the VISTA trial were used.[8] To inform extrapolations from the 30-month VISTA data, we used the study of Shustik et al,[9] which was an RCT of MP versus melphalan with dexamethasone (M-Dex) with 8 years of follow-up. This trial included patients with similar baseline characteristics compared with those contained in the VISTA trial. Table 1 compares the key patients' characteristics in the MP arms of the VISTA and historical clinical trial.[9] The similarity of the patient population and clinical trial design of historical and VISTA trials was seen sufficient that parameter estimates based on the historical trial can be used for the creation of informative a priori distributions for the parameters for the VISTA trial.

Individual patient-level data (IPD) were not available for the applied studies. Therefore, the OS Kaplan-Meier (KM) curves were reconstructed based on the published data. The reported KM curves were extracted by digitizing the data with Engauge Digitizer (version 9.2) software. The number of patients at risk at each time interval was retrieved and used together with the extracted KM data to construct IPD. The construction of the IPD was done with a published and validated algorithm in R.[10] The reconstructed KM curves were obtained from the IPD (Fig. 1).

Standard Methods (Bayesian Method With Noninformative Priors)

The selection of the parametric distribution for extrapolations was assessed with the Akaike information criterion (AIC) and Bayesian information criterion (BIC).[11] From the standard parametric distributions, the Weibull, log-normal, log-logistic, exponential, and generalized gamma were assessed. Also, Weibull, log-normal, and log-logistic spline models with one knot placed at median uncensored survival were tested. The estimated AIC and BIC scores are presented in the Appendix (see Appendix Table 2 in Supplemental Materials found at https://doi.org/10.1016/j.jval.2019.03.017).

All the analyses were carried out by Bayesian Markov Chain Monte Carlo simulation using BUGS software.[12] Three chains were run with 150 000 iterations. The chain convergence was determined by the plot of quantiles. Two approaches to select the parametric distribution were tested. Also, in these approaches, the historical trial data were not used to influence the estimation of the shape of the 30-month VISTA data. In the first approach, the selection of the parametric distribution to extrapolate the 30-month VISTA data was based on the historical trial data. In the

Figure 1. Overall survival (OS) trial data used for the analyses.

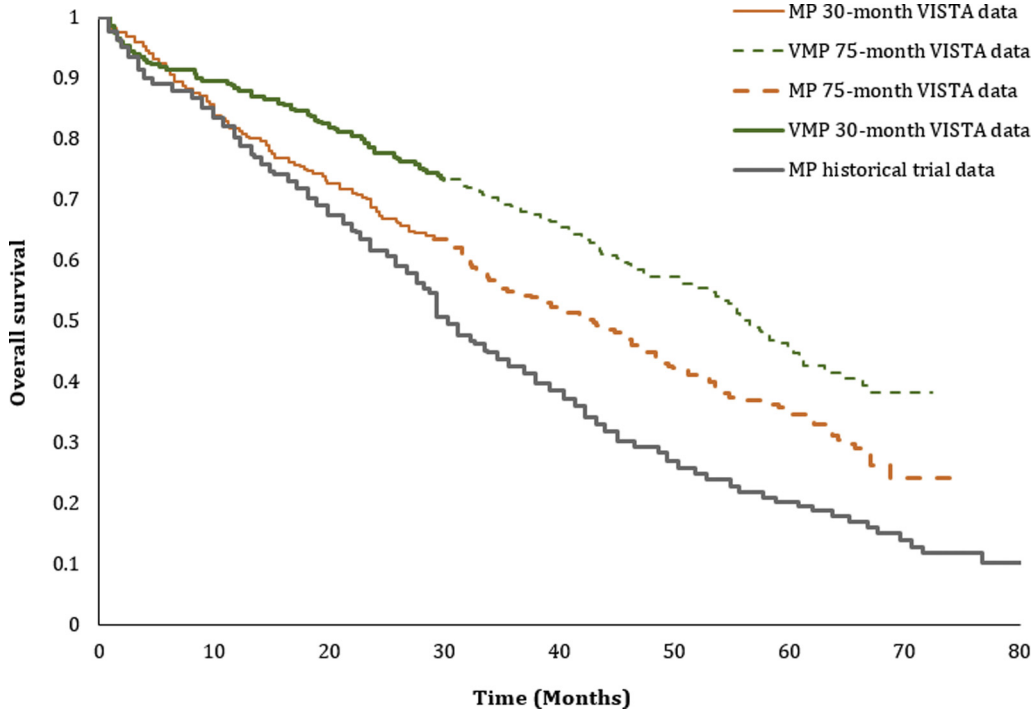
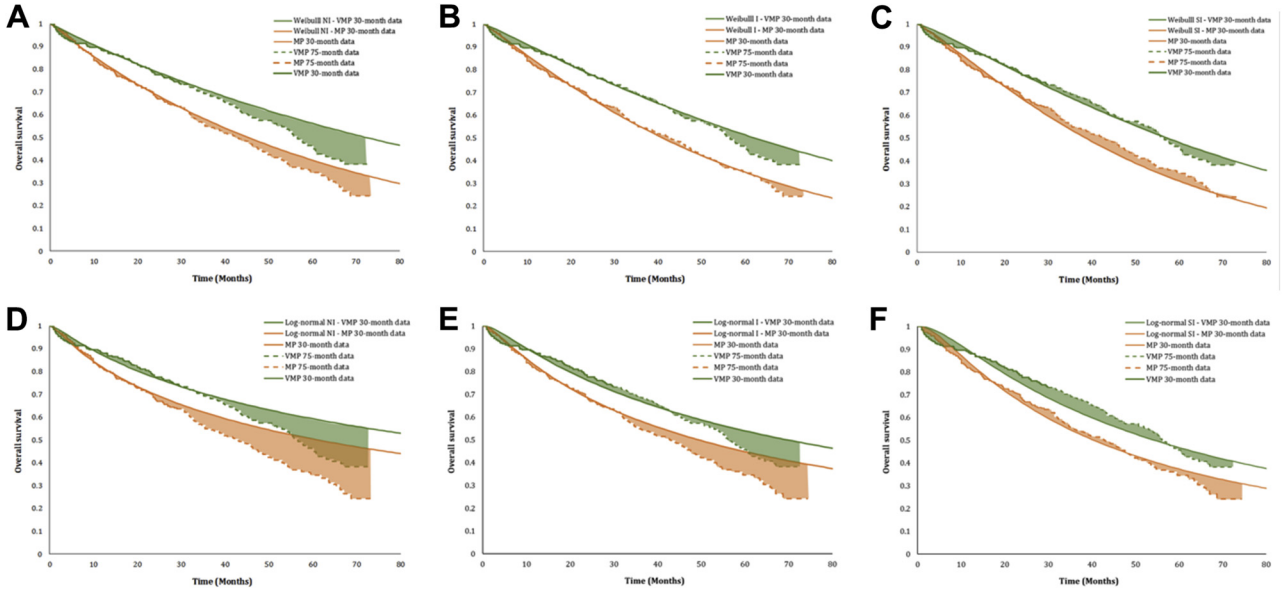


Figure 2. Observed and predicted overall survival (OS) when (A) Weibull NI, (B) Weibull I, (C) Weibull SI, (D) log-normal NI, (E) log-normal I, and (F) log-normal SI was applied.



I indicates informative prior; NI, noninformative prior; SI, strong prior.

second approach, the selection of the parametric distribution was based on the 30-month VISTA trial data. Because the parameterization of some parametric distributions differs in BUGS and R,[6] the Bayesian method with noninformative priors was used. In general, Bayesian approaches with noninformative priors provide similar results as standard frequentist approaches.

Bayesian Method With Informative Priors

In this study, the Bayesian method with the historical trial data is used in 2 stages:

1 selection of the best fitting parametric distribution for survival extrapolations from the VISTA 30-month data and
2 creation of a priori distribution for the shape parameter of the 30-month VISTA data

The Weibull distribution, $W(\rho, \lambda_i)$, is taken here for illustration and fitted to the historical data to estimate priors for the shape parameter. The survival function of the Weibull takes the following form:

$$S(t_i) = e^{\left(-\lambda_1 t_i^{\rho 1}\right)}$$

Table 2. Overview of results of the Bayesian methods with the corresponding uncertainty.

		75-month VISTA data	
		The Bayesian method with noninformative prior (NI)	
		MP	VMP
Weibull distribution	ΔAUC		
	% difference in ΔAUC		
	Median OS	41	56
	Mean OS (95% CrI)	55.05 (48.30-62.82)	75.05 (66.01-87.83)
	Incremental OS (95% CrI)	20.00 (9.12-34.23)	
log-normal distribution	ΔAUC		
	% difference in ΔAUC		
	Median OS	39	59
	Mean OS (95% CrI)	75.57 (66.93-84.87)	93.51 (83.65-103.89)
	Incremental OS (95% CrI)	17.95 (5.09-29.88)	

AUC indicates area under the curve; CrI, credible interval; OS, overall survival.

where $\lambda_1 = e^{\beta_0}$ is the scale parameter of the historical trial and ρ_1 is the shape parameter of the historical trial, t denotes time, and i refers to an individual patient in the historical trial. For a priori distributions, we used gamma (α, β) distribution and normal (μ, τ) distribution, where μ stands for the mean and τ for precision (τ = 1/variance). To fit the historical trial data, a noninformative gamma distribution ($1.0E^{-3}$, $1.0E^{-3}$) was used for the shape ρ_1 and a noninformative normal distribution (0, $1.0E^{-9}$) for the scale λ_1. The mean (ρ_1) and the standard deviation (SD) of the posterior distribution of the shape ρ_1 were converted to parameters for a gamma distribution defined by (α, β) with $\beta = \frac{\rho_1}{SD^2}$ and $\alpha = \rho_1 \beta$. This gamma distribution was used as a priori distribution for the shape of the VISTA trial. For the scale of the MP of 30-month VISTA data, a noninformative normal (μ, τ) distribution was applied (0, $1.0E^{-9}$). This distribution is also used for the relative treatment effect of VMP versus MP (being the reference).

The posterior distribution of the parameters for all the tested methods was implemented into a survival partition framework. OS was extrapolated over time using a 30-year time horizon. The uncertainty around the point estimate of the mean was derived by taking the last 1000 iterations of the posterior mean of the shape and scale parameters. In the partitioned survival model, the survival extrapolations were restricted with the general population survival because patients with MM are expected to have a higher risk of death owing to their disease compared with the general population. It was assumed that in case the hazards in the estimated survival become lower than the hazards in the matched age-sex U.K. population,[13] the hazards of the U.K. general population were to be taken.

The performance of each method was assessed by comparing the extrapolated survival against the observed survival in the 75-month VISTA data-cut. This was both visually assessed and evaluated by examining the difference in the area between the predicted curve and the 75-month–based KM data (ΔAUC). At monthly intervals, the absolute difference between the predicted and observed survival curve was estimated and summed up to 75 months. The best performing method had the lowest ΔAUC up to 75-month follow-up.

Several summary statistics were derived from the tested methods, including median OS and mean OS, with the corresponding credible intervals for both MP and VMP arms in the VISTA trial. Additionally, the incremental difference in the mean OS between the VMP and MP arm was calculated. These outcomes were compared with the predicted mean and incremental mean OS of the 75-month VISTA trial data.

The parametric distributions were fitted combinedly to the MP and VMP data, implying that the MP data were used as a reference curve and VMP was determined by a treatment coefficient. The combined fits can be used in case the proportional hazards (PH) assumption is met.[2] This was assessed based on log-cumulative hazard plots and was confirmed with the Schoenfeld residuals test (see Appendix Table 1, Fig. 1, and Fig. 2 in Supplemental Materials found at https://doi.org/10.1016/j.jval.2019.03.017). The log-cumulative hazards plot showed that the curves cross at the time point of 5 months and then become parallel, which might indicate that after 5 months the assumption holds. This was confirmed by the Schoenfeld test, which was not significant.

Scenario Analysis

In the Bayesian analyses, priors are considered subjective. As such, Bayesian guidelines recommend[14] that sensitivity analysis on the selected prior distribution should be conducted. As a scenario analysis, the shape of the historical trial was taken as true by reducing the standard deviation of the mean shape from the historical trial to the point that the posterior shape for the VISTA trial was identical to the shape in the historical trial. This implies that a broad range of possible priors was covered; the standard approach relies on a noninformative prior, whereas the scenario with strong priors takes the historical trial shape as true. We believe that all realistic priors would be in between the noninformative prior and the strong prior.

Although the assumption of the PH was met based on a visual assessment and Schoenfeld test, the KM curves in the VISTA trial cross. When parametric distributions are fitted individually to each treatment arm, there is no assumption on the proportionality of the hazards between the treatment arms. Thus, the parametric distributions were fitted independently to each treatment arm. In this case, the shape of the historical trial was used as an informative prior for the shape parameter of the VMP fit and separately to the shape parameter of the MP fit.

Results

The first step of the 2-stage approach was the selection of the parametric distribution to extrapolate 30-month VISTA data. The

Table 2. Continued

30-month VISTA data					
The Bayesian method with noninformative prior (NI)		**The Bayesian method with informative prior (I)**		**Scenario: The Bayesian method with strong informative prior (SI)**	
MP	**VMP**	**MP**	**VMP**	**MP**	**VMP**
2.25	3.26	0.87	1.69	1.94	1.37
		−62%	−48%	−14%	−58%
45	72	40	62	38	57
65.89 (48.91-91.53)	98.89 (70.03-127.23)	55.69 (44.27-71.33)	83.13 (63.28-108.34)	50.32 (41.70-59.37)	74.25 (60.19-92.45)
33.00 (7.93-53.47)		27.44 (5.82-50.23)		23.93 (7.60-43.51)	
7.86	5.27	3.58	2.78	2.08	3.81
		−54%	−47%	−74%	−28%
72	108	47	65	37	48
110.9 (94.14-125.46)	127.35 (109.24-142.02)	87.24 (74.07-100.67)	103.31 (88.32-119.12)	67.20 (62.9-79.8)	80.69 (78.2-98.6)
16.92 (0.04-32.97)		16.07 (0.2-32.61)		13.49 (3.9-29.9)	

Weibull distribution had the best statistical fit to the historical trial data. The log-normal distribution had the best statistical fit to the 30-month VISTA data (see Appendix Table 2 in Supplemental Materials found at https://doi.org/10.1016/j.jval.2019.03.017).

Figure 2A,D shows the observed 75-month VISTA data and the predicted survival based on the Weibull with non-informative prior (Weibull NI) and log-normal with non-informative prior (log-normal NI) fitted to the 30-month VISTA data. The marked areas in orange (MP) and green (VMP) represent the AUC between the predicted and observed survival. Visually, the Weibull NI predicts 75-month survival better than the log-normal NI.

The observed median OS in the VISTA trial[6] was 41.3 months in the MP group and 56.4 months in the VMP group. The median OS based on the Weibull NI was 45 for MP (vs 41.3 months observed) and 72 months for VMP (vs observed 56.4 months) (Table 2). The log-normal NI–based median OS was 72 (vs observed 41.3) and 108 (vs observed 56.4) months for MP and VMP, respectively (Table 2). The ΔAUC for the Weibull NI (2.25 and 3.26 for MP and VMP, respectively) was lower than the ΔAUC for the log-normal NI (7.86 and 5.27 for MP and VMP groups, respectively). Therefore, the choice for the Weibull based on the historical trial outperforms the choice for the log-normal based on the VISTA 30-month data.

As the second step, an informative priori distribution for the shape parameter of the 30-month VISTA data was created. Figure 2B,E shows the observed 75-month VISTA data and the predicted survival based on the informative prior distributions. Visually, the Weibull with informative prior (Weibull I) and the log-normal with informative prior (log-normal I) predicted the 75-month survival better than the Weibull NI and log-normal NI.

Of the tested methods, Weibull I predicted the observed median OS in the VISTA trial[6] most accurately (40 months Weibull I vs 41.3 months observed for the MP group, and 62 months Weibull I vs 56.4 months for the VMP group). The Weibull I also showed the lowest ΔAUC of 0.87 for MP group and 1.69 for VMP group when compared with the corresponding results of the Weibull NI and log-normal I (Table 2).

In the cost-effectiveness analysis the accurate estimation of incremental survival between the VMP and MP groups over a lifetime time horizon is a crucial estimate, as is the uncertainty around the estimated incremental survival. The estimated mean OS and incremental OS of different methods are presented in Table 2. The estimated mean OS with the Weibull NI was 65.89 months (95% credible interval [95% CrI]: 48.91-91.53) for MP and 98.89 months (95% CrI: 70.03-127.23) for VMP, resulting in the incremental survival of 33.00 months (95% CrI: 7.93-53.47). The corresponding results with the Weibull I was 55.69 months (95% CrI: 44.27-71.33) for MP and 83.13 months (95% CrI: 63.28-108.34) for VMP months, resulting in the incremental survival of 27.44 months (95% CrI: 5.82-50.23). The figures presenting the extrapolations over the lifetime from the tested methods are provided in the Appendix (see Appendix Figs. 3 and 4 in Supplemental Materials found at https://doi.org/10.1016/j.jval.2019.03.017).

The scenario analyses with strong priors show that the Weibull with strong prior (Weibull SI) improved the predicted survival compared with the Weibull NI by having a ΔAUC of 1.94 for MP group and 1.37 for VMP group (Table 2). The improvement in the predicted survival can also be seen in Figure 2C. Similar results were found with the log-normal with strong prior (log-normal SI; Table 2 and Fig. 2F). The results with individually fitted distributions were similar to the combined fits to the 30-month VISTA data. The results of that scenario are provided in the Appendix (see Appendix Table 5 in Supplemental Materials found at https://doi.org/10.1016/j.jval.2019.03.017).

Discussion

In this study, we presented a Bayesian method in which historical trial data were used to inform survival extrapolations. The historical trial was applied in 2 stages, by first guiding the selection of the parametric distribution and then by creating a priori distribution to inform the shape parameter of the pivotal clinical trial. The method was illustrated with an application in the first-line treatments to MM. First, the 75-month VISTA data were more accurately predicted by the Weibull NI, which fitted best the historical trial data, compared with the log-normal NI, which fitted best the 30-month VISTA data. Second, the use of informative a priori distribution to the shape parameter of the 30-month VISTA trial improved the predictions compared with the noninformative approach. The conducted scenario analyses supported these findings.

In this application, the patient characteristics in the VISTA trial and historical trial were similar. In case the patient characteristics between the trials differ, population-adjusted indirect comparison[15] methods can be used to correct for these differences. In those methods, the patient characteristics and corresponding survival may be adjusted to reflect the historical trial data. Nevertheless, in case the patient-level data of both the pivotal and historical trial are available, the historical trial data can be included as a third treatment group in the parametric curve fitting.[5] This approach assumes that the shape of the hazard is the same in both clinical trials, whereas in this study the historical data are guiding the shape. Another alternative is the approach proposed by Bagust and Beale,[16] which does not rely on external data but assumes that an exponential distribution can be applied to the remaining data after left truncation of the data at the point when the constant hazards are found.

There are some limitations in the conducted analyses. First, such Bayesian method is the most appropriate when proportionality of the hazards holds. Generally, when this assumption does not hold, the parametric distributions are fitted individually to each treatment group rather than using a combined fit. This can be done with the presented Bayesian method, but the stronger the prior on the shape parameter, the more likely that both treatment groups will have a similar posterior shape parameter. This equals the combined fit method where the shape parameter is assumed to be a shared parameter, and the scale parameter explains the differences between the treatment groups. A second limitation of the study relates to assessing the performance based on the ΔAUC because it does not consider the impact of decreasing patient numbers at risk over time.

In the future, the proposed Bayesian method could also be applied by using more flexible models such as spline models in case they provide a better fit to the data.[17] Also, external data from registries or real-world evidence could be used. In addition, the approach can be tested within a parametric NMA (network meta-analysis) framework.[18]

In conclusion, the presented Bayesian approach with informative a priori distributions for the shape parameter improved the survival extrapolations. Given that there is an external source of adequately similar historical evidence, this approach may be valuable for HTAs that deal with immature data.

Acknowledgments

The authors would like to thank the reviewers of the article for their valuable input in finalizing the article.

Supplemental Material

Supplementary data associated with this article can be found in the online version at https://doi.org/10.1016/j.jval.2019.03.017

REFERENCES

1. 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.
2. Jackson C, Stevens J, Ren S, et al. Extrapolating survival from randomized trials using external data: a review of methods. *Med Decis Making*. 2017;37(4):377–390.
3. Guyot P, Ades AE, Ouwens MJ, Welton NJ. Enhanced secondary analysis of survival data: reconstructing the data from published Kaplan-Meier survival curves. *BMC Med Res Methodol*. 2012;12:9.
4. Demiris N, Sharples LD. Bayesian evidence synthesis to extrapolate survival estimates in cost-effectiveness studies. *Stat Med*. 2006;25(11):1960–1975.
5. Pennington M, Grieve R, van der Meulen J, Hawkins N. Value of external data in the extrapolation of survival data: a study using the NJR data set. *Value Health*. 2018;21(7):822–829.
6. Jackson CH. Flexsurv: a platform for parametric survival. 2016;70(8):1–33.
7. San Miguel JF, Schlag R, Khuageva NK, et al. Bortezomib plus melphalan and prednisone for initial treatment of multiple myeloma. *N Engl J Med*. 2008;359(9):906–917.
8. San Miguel JF, Schlag R, Khuageva NK, et al. Persistent overall survival benefit and no increased risk of second malignancies with bortezomib-melphalan-prednisone versus melphalan-prednisone in patients with previously untreated multiple myeloma. *J Clin Oncol*. 2013;31(4):448–455.
9. Shustik C, Belch A, Robinson S, et al. A randomised comparison of melphalan with prednisone or dexamethasone as induction therapy and dexamethasone or observation as maintenance therapy in multiple myeloma: NCIC CTG MY.7. *Br J Haematol*. 2007;136(2):203–211.
10. Guyot P, Ades AE, Beasley M, Lueza B, Pignon J-P, Welton NJ. Extrapolation of survival curves from cancer trials using external information. *Med Decis Making*. 2017;37(4):353–366.
11. Collett D. *Modelling Survival Data in Medical Research. 2nd ed*. Boca Raton: Chapman and Hall/CRC; 2003.
12. Spiegelhalter D, Thomas A, Best N, Lunn D. *WinBUGS 1.4 User Manual*. Cambridge: University of Cambridge; 2004. https://www.mrc-bsu.cam.ac.uk/software/bugs/. Accessed May 25, 2018.
13. Office for National Statistics. National life tables: UK. https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/lifeexpectancies/datasets/nationallifetablesunitedkingdomreferencetables. Accessed May 22, 2018.
14. Sung L, Hayden J, Greenberg ML, Koren G, Feldman BM, Tomlinson GA. Seven items were identified for inclusion when reporting a Bayesian analysis of a clinical study. *J Clin Epidemiol*. 2005;58(3):261–268.
15. Signorovitch JE, Sikirica V, Erder MH, et al. Matching-adjusted indirect comparisons: a new tool for timely comparative effectiveness research. *Value Health*. 2012;15(6):940–947.
16. 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.
17. Royston P, Parmar MKB. Flexible parametric proportional-hazards and proportional-odds models for censored survival data, with application to prognostic modelling and estimation of treatment effects. *Stat Med*. 2002;21(15):2175–2197.
18. Ouwens MJ, Philips Z, Jansen JP. Network meta-analysis of parametric survival curves. *Res Synth Methods*. 2010;1(3-4):258–271.