An Evaluation of Survival Curve Extrapolation Techniques Using Long-Term Observational Cancer Data

Adrian Vickers, Emma Hawe

RTI Health Solutions, Manchester, United Kingdom

BACKGROUND

Survival analysis has become an important part of cost-effectiveness methods for health technology appraisals in oncology. Current health technology assessments usually require mean survival times to estimate the life-years gained or, for Markov models, transition probabilities per cycle. Mean survival times are typically derived from fitting parametric survival curves for the lifetime of patients and the integral of the fitted survival curve used to estimate mean survival.

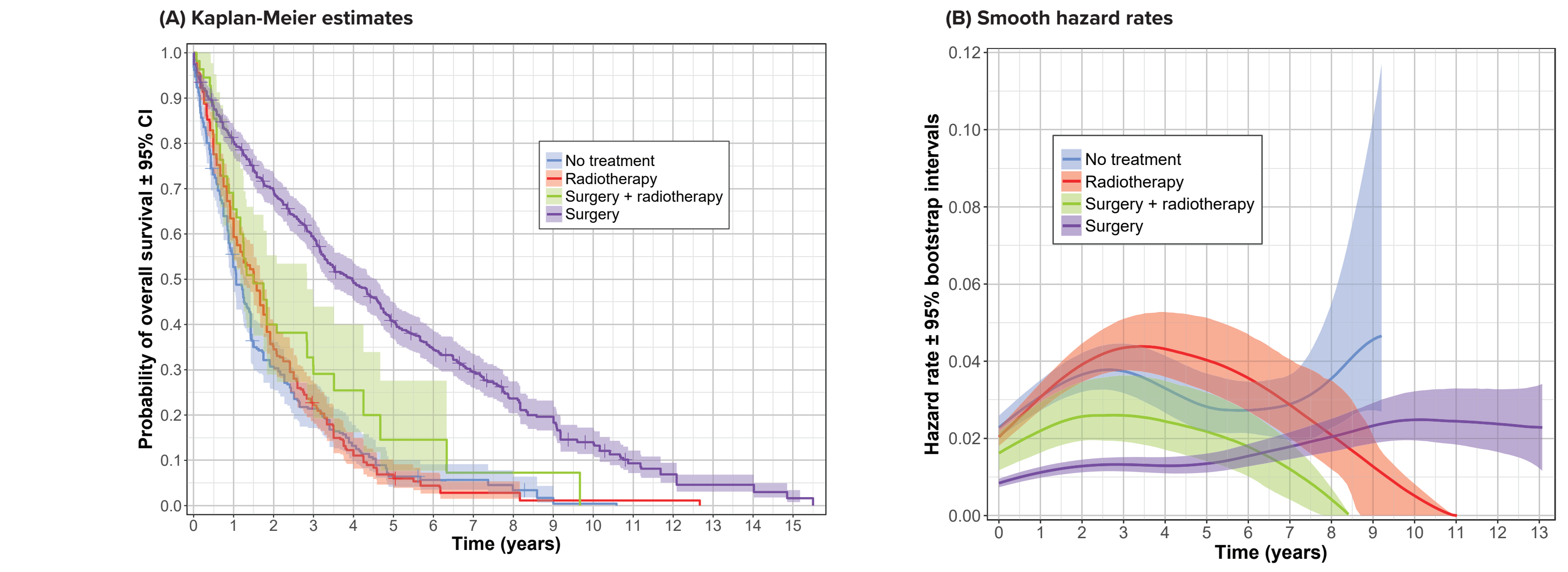
OBJECTIVES

The aim of this study was to recreate a common situation in which the control arm includes short-term trial data and long-term external data that, despite matching the disease, do not match the trial population exactly. Two long-term published data sources were identified for similar patient populations: one to be used as the external long-term data set and the other to generate pseudo short-term randomized controlled trial (RCT) data. Four different approaches were used to fit survival models and perform the extrapolation on the pseudo short-term data sets, two of which made direct use of long-term external data. The performance of these models to predict mean survival from the complete data set was evaluated. Finally, the findings and limitations of the approaches and these evaluation methods are discussed.

THE DATA

The long-term data sets identified included elderly patients with early stage non-small cell lung cancer (NSCLC), contained 15.5 years of follow-up, and represented complete survival estimates for four treatments. Figure 1 presents reconstructed data from Kaplan-Meier estimates and smoothed hazard rates derived from Surveillance, Epidemiology, and End Results data presented by Ganti et al.[1] for stages I and II NSCLC in an elderly population (≥ 80 years old) diagnosed between 1998 and 2007. These data were reconstructed, and RCT-like data sets were created as if patients had started a trial at random times during the first year, with a follow-up of 4 years. The external data were reconstructed from data presented by Bach et al.[2] from patients with stage I or II NSCLC who were ≥ 65 years old and were diagnosed between 1985 and 1993 (n = 2,589). General population data were also reconstructed. All data were reconstructed using the methods described by Guyot et al.[3] The smoothed hazard rates estimated from the complete data set appeared to show that hazard rates changed over time and were not consistent across treatments.

Figure 1. Data Presented by Ganti et al.[1] for Early Stage NSCLC in Elderly Patients



Number at risk

No treatment	286	149	86	60	37	18	15	15	9	4	1	0	0	0	0	0
Radiotherapy	339	215	117	75	37	23	14	9	5	2	2	2	2	0	0	0
Surgery + radiotherapy	55	36	22	16	14	8	8	4	4	4	0	0	0	0	0	0
Surgery	585	466	401	338	280	230	194	164	129	105	73	47	33	22	20	6

CI = confidence interval.

METHODS: EXTRAPOLATION AND EVALUATION

Extrapolation of Standard Parametric Models

This method was based on the method described by Latimer.[4] Standard parametric models were fitted and evaluated in terms of plausibility and fit. Models included standard parametric models, stratified parametric models, and Royston and Parmar[5] spline-based models with 1-3 knots, which assumed proportional hazards and time varying hazard ratios (HRs).

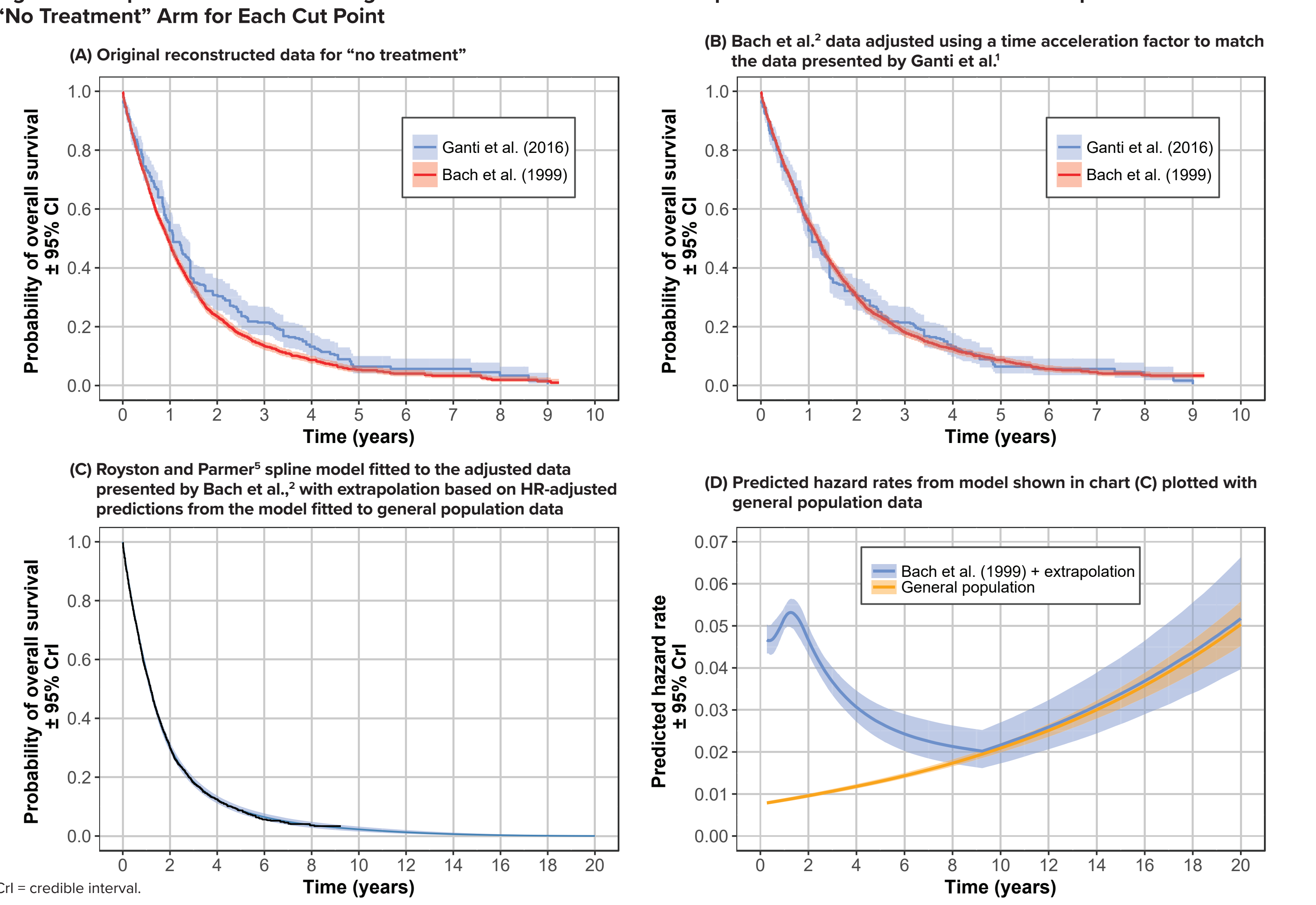
Bootstrapped Hybrid Model

This method was based on the method described by Gelber et al.[6] and Bagust and Beale.[7] A Chow-break test[8] was used to cut Kaplan-Meier estimates in two. Kaplan-Meier estimates were calculated for the first part of the curve, and parametric models (exponential, Weibull, log-normal, and log-logistic) were fitted to the tail of the distribution. The whole procedure was bootstrapped. The most plausible model was chosen.

Extrapolation Parametric Models With Direct Use of Long-Term Data

This method was based on the method described by Jackson et al.[9] Parametric models were fitted to the pseudo RCT data, and, after follow-up, predicted hazards were used from external data adjusted to match the RCT data. Model averaging was used to give an ensemble of model predictions based on mean Akaike's information criterion and Bayesian information criterion for the models fitted to the RCT data. HR adjusted predictions from general population data were used after the follow-up of the external data. Figure 2 presents the steps involved in using the external data to perform the extrapolation. HR tapering was used to obtain predicted survival estimates for radiotherapy, surgery plus radiotherapy, and surgery. It was assumed that hazard rates would eventually equal that from the "no treatment" arm. It was assumed that the time to an HR of 1 was a distribution of 10 years with a standard deviation of 2, which gives a range of approximately 3 to 17 years after the start of the trial, and that HRs went to 1 in a linear way. This distribution reflected a belief that the treatment effect may continue for a long time after patients received treatment but that the upper limit of the distribution could not exceed the plausible limit imposed by age.

Figure 2. Steps Involved in Estimating the Hazard Rates From Disease-Specific External Data and General Population Data for the "No Treatment" Arm for Each Cut Point



CrI = credible interval.

Bayesian Simultaneous Spline Model of RCT and Long-Term Data

This method was based on the method described by Guyot et al.[10] A Bayesian spline-based model was simultaneously fitted to the RCT and predicted survival from the external data, which included general population data. A simplified 1-knot model was needed to achieve convergence, which contained a treated parameter (vs. "no treatment") for the intercept and other parameters allowed to vary with each treatment. Priors were used in the same way as in Goyet et al.[10] for the treatment effect after follow-up and assumed that HRs equaled 1 at 10 years. This model was fitted using JAGS.[11] All other analyses were conducted in R.[12]

Reference Model

Mean survival times were estimated from the complete data set. They were based on Kaplan-Meier estimates from bootstrap samples and the area under the curve used to give distributions of mean survival times for each treatment.

Evaluation

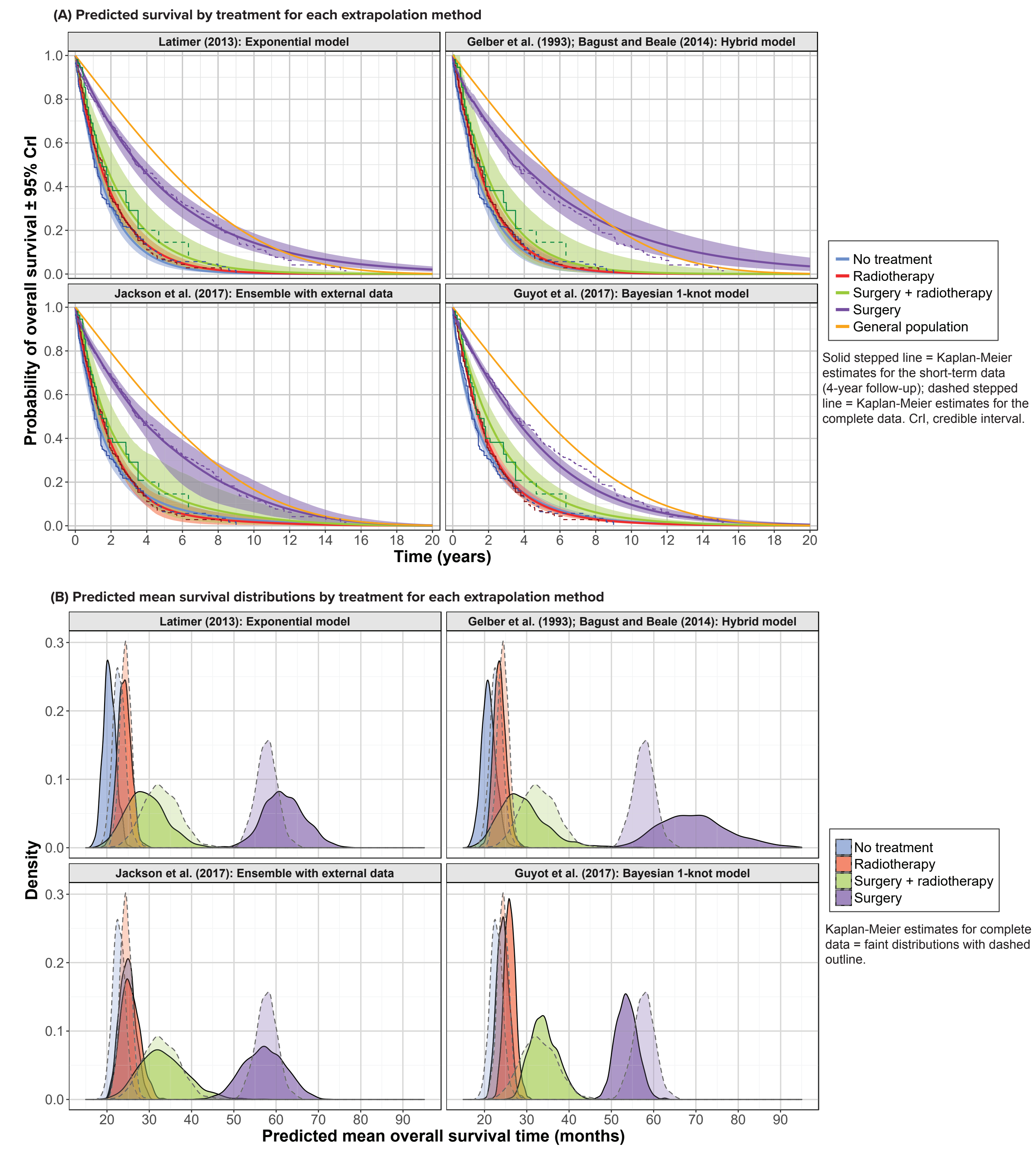
Survival curves and predicted mean overall survival (OS) from the above models were compared with those from the reference model. Plausibility of standard parametric models was based on visual fit to the long-term external data.

RESULTS

Figure 3 presents the predicted survival curves and predicted mean survival distributions from the models tested, with a follow-up of 4 years, compared with the reference model. Only the results from the most plausible models are shown for the parametric approach[4] and the hybrid model.[6,7] This chart shows the following:

- The simple parametric model and hybrid exponential models were unable to produce predictions that remained below those derived from the general population data and were therefore not considered to give plausible predictions.
- The methods that directly used external data appeared to perform better. However, the Bayesian model appeared to underestimate the error as a high proportion of the estimates from the reference model fall outside the predicted 95% credible intervals for the extrapolated part of the curve.

Figure 3. Results From the Cut Point of 4 Years Compared With the Model Fitted to the Complete Data Set



DISCUSSION

- The data used in this study were from elderly patients (≥ 80 years old) in whom general, age-related mortality was a factor, which was not detectable in the short-term data. This resulted in methods that did not use long-term external data and general population data that produced biased predictions. It would be interesting to see how these models perform in younger patient populations.
- For this study, the difference in study arms was assumed to be caused by the treatment effects. However, because the data were observational, the differences between treatments and duration of treatment effect may have been due to variation in patient populations that made them more likely to receive radiotherapy and/or surgery. This makes modelling observational data more difficult, as expert opinion may not be able to advise on what happens after follow-up.
- An accurate estimate of the duration of treatment effect after follow-up is needed to make accurate, long-term predictions. RCTs of systemic drugs are likely to have a shorter duration of treatment effect, on which clinical experts can give advice.

CONCLUSIONS

- The results from this study support the direct use of external data to extrapolate survival curves even when the external data may not be an exact match with the RCT data.
- At this stage, for this example, the method based on the Jackson et al.[9] approach performed better with less bias compared with the other models tested.
- This study demonstrates the value of including all sources of error within a single model. The method based on Jackson et al.[9] was the only model that included errors for the choice of distribution and the uncertainty of the treatment effect after follow-up. Sensitivity analyses can be used to address this to some extent in other models. However, presenting and making sense of numerous models for multiple approaches makes the results less transparent and harder to interpret.
- The methods described in this study based on the Jackson et al.[9] approach, which includes model averaging, using HR distributions and distribution for the time to HR of 1, can be implemented in a health economic model.

REFERENCES

1. Ganti Ak, Shostrom V, Alorabi M, Zhen W, Marr AS, Tujollo K, et al. Early stage non-small-cell lung cancer in octogenarian and older patients: A SEER database analysis. Clin Lung Cancer. 2016;17(4):285-91.
2. Bach PB, Laura MD, Cramer LD, Warren JL, Begg CB. Racial differences in the treatment of early stage cancer. N Engl J Med. 1999;341(16):1198-5.
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):1-13.
4. 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;33:743-54.
5. Royston P, Parmar M. Flexible parametric proportional-hazards and proportional-odds models for censored survival data, with application to prognostic modelling and estimation of treatment effects. Statistics Med. 2002;21(1):2175-97.
6. Gelber R, Goldhirsch A, Cole BF. Parametric extrapolation of survival estimates with applications to quality of life evaluation of treatments. Controlled Clin Trials. 1993;14:485-99.
7. 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):343-51.
8. Chow, GC. Tests of equality between sets of coefficients in two linear regressions. Econometrica. 1960;28(3):591-605.
9. Jackson C, Stevens J, Ren S, Latimer N, Bojke L, Manca A, et al. Extrapolating survival from randomized trials using external data: a review of methods. Med Decis Making. 2017; 37(4):377-90.
10. Guyot P, Ades AE, Beasley M, Lueza B, Pignon J-P, Welton NJ. Extrapolation of trial-based survival curves using external information. Med Decis Making. 2017;37(4):353-66.
11. Plummer M. JAGS: a program for analysis of Bayesian graphical models using Gibbs sampling. Proceedings of the Third International Workshop on Distributed Statistical Computing; Vienna, Austria. March 20-22, 2003. ISSN 1609-395X. http://www.r-project.org/conferences/DSC-2003/Proceedings/Plummer.pdf (2003).
12. R Development Core Team. R version 3.4.3: a language and environment for statistical computing. R Foundation for Statistical Computing; Vienna, Austria. 2017. ISBN 3-900051-07-0. 2017. Available at: http://CRAN.R-project.org.

CONTACT INFORMATION

Adrian Vickers PhD, MSc
Director, Data and Design Strategy
RTI Health Solutions

Presented at: ISPOR Europe 2018; 10-14 November 2018; Barcelona, Spain