ORIGINAL RESEARCH

The Role of Expert Opinion in Projecting Long-Term Survival Outcomes Beyond the Horizon of a Clinical Trial

Bartholomeus J. A. Willigers · Mario Ouwens · Andrew Briggs ⓘ ·

Hiddo J. L. Heerspink · Carol Pollock ⓘ · Roberto Pecoits-Filho ⓘ ·

Navdeep Tangri ⓘ · Csaba P. Kovesdy ⓘ · David C. Wheeler ⓘ ·

Juan Jose Garcia Sanchez ⓘ

ABSTRACT

Introduction: Clinical trials often have short follow-ups, and long-term outcomes such as survival must be extrapolated. Current extrapolation methods often produce a wide range of survival values. To minimize uncertainty in projections, we developed a novel method that incorporates formally elicited expert opinion in a Bayesian analysis and used it to extrapolate survival in the placebo arm of DAPA-CKD, a phase 3 trial of dapagliflozin in patients with chronic kidney disease (NCT03036150).

Methods: A summary of mortality data from 13 studies that included DAPA-CKD-like populations and training on elicitation were provided to six experts. An elicitation survey was used to gather the experts' 10- and 20-year survival estimates for patients in the placebo arm of DAPA-CKD. These estimates were combined with DAPA-CKD mortality and general population mortality (GPM) data in a Bayesian analysis to extrapolate long-term survival using seven parametric distributions. Results were compared with those from standard frequentist approaches (with and without GPM data) that do not incorporate expert opinion.

B. J. A. Willigers (✉)
Decision Science, Statistical Innovation, Biopharmaceuticals R&D, AstraZeneca, Cambridge, UK
e-mail: bartholomeus.willigers@astrazeneca.com

M. Ouwens
Medical & Payer Evidence Statistics, Real World Science and Digital, Biopharmaceuticals Business Unit, AstraZeneca, Mölndal, Sweden

A. Briggs
London School of Hygiene & Tropical Medicine, London, UK

H. J. L. Heerspink
University of Groningen, Groningen, The Netherlands

C. Pollock
University of Sydney, Sydney, Australia

R. Pecoits-Filho
Pontifical Catholic University of Paraná, Curitiba, Brazil

N. Tangri
University of Manitoba, Winnipeg, MB, Canada

C. P. Kovesdy
University of Tennessee Health Science Center, Memphis, TN, USA

D. C. Wheeler
University College London, London, UK

J. J. Garcia Sanchez
Health Economics, Biopharmaceuticals Business Unit, AstraZeneca, Cambridge, UK

Results: The group expert-elicited estimate for 20-year survival was 31% (lower estimate, 10%; upper estimate, 40%). In the Bayesian analysis, the 20-year extrapolated survival across the seven distributions was 14.9–39.1%, a range that was 2.4- and 1.6-fold smaller than those produced by the frequentist methods (0.0–56.9% without and 0.0–39.2% with GPM data).

Conclusions: Using expert opinion in a Bayesian analysis provided a robust method for extrapolating long-term survival in the placebo arm of DAPA-CKD. The method could be applied to other populations with limited survival data.

Key Summary Points

Why carry out this study?

Health technology assessment agencies recommend standard parametric frequentist methods for extrapolating survival data from randomized controlled trials beyond the trial duration.

However, these methods can produce a range of survival values, particularly where limited data are available, as can be the case for patients with chronic diseases.

What was learned from the study?

We developed a method that combines patient data from clinical trials with a formal expert elicitation using Bayesian methods to estimate long-term survival in cases where limited survival data exists.

The method produces estimates that are plausible, robust, and aligned with informed clinical opinion.

The survival extrapolations produced by this method could provide additional evidence for use in various situations, including clinician–patient conversations and in regulatory and reimbursement decision-making, for many chronic disease areas.

INTRODUCTION

Most randomized controlled trials (RCTs) of new therapies have maximum follow-up durations of less than 5 years [1], and they do not provide data on outcomes beyond this period. These time frames may be considered short for the purposes of clinical, reimbursement, and policy decision-making. Therefore, to obtain information about the impact of new therapies on long-term outcomes, the short-term results of RCTs are often extrapolated. Real-world evidence (RWE) can supplement RCT data in these decision-making processes, but RWE studies can have their own drawbacks: they are generally retrospective; they often have heterogeneous patient populations; and they are prone to biases such as performance bias and selection bias.

One outcome with important clinical and regulatory implications is patient survival. Standard parametric frequentist methods are by far the most commonly used methods for extrapolating survival data from RCTs beyond the trial duration. However, these frequentist methods can produce a wide range of survival values for a given time point, particularly where limited data are available, as can be the case for patients with chronic diseases that have low mortality. This variability introduces uncertainty into the survival projections and expert opinion is often sought on which parametric frequentist survival model is the most realistic. Health technology assessment (HTA) agencies have also expressed the need for methods that improve the accuracy of long-term survival predictions, such as Bayesian methods that incorporate expert judgment [2].

Expert elicitation is a well-established method for obtaining and synthesizing unbiased expert judgments that can provide valuable quantitative information when empirical data are lacking. The aim of expert elicitation is to develop a probability distribution for an uncertain parameter by combining a set of probabilistic expert judgments. Initially developed in the 1950s [3, 4], expert elicitation has received renewed attention, partly owing to endorsement by HTA agencies like the National Institute for Health and Care Excellence (NICE)

and scientific advisory bodies like the National Academies of Sciences, Engineering, and Medicine [2, 5, 6]. It has previously been applied to a variety of health research areas, including the extrapolation of long-term survival [7–9].

We describe a method that uses the expert clinical judgment of practising physicians within a Bayesian framework. We hypothesized that long-term survival extrapolations derived through this methodology would provide more robust and unbiased results than current standard methods for extrapolating RCT survival data. To test this and provide an example, we applied the method to patients with chronic kidney disease (CKD) in the placebo arm of the DAPA-CKD trial (NCT03036150) [10]. Patients with chronic cardiometabolic diseases, like CKD, generally receive treatment over a much longer period of time than the duration of a typical RCT and, while RCTs can demonstrate the short-term benefits of such treatments, often little is known about their potential long-term benefits. Therefore, robust methods for extrapolating outcomes are needed for these populations to aid clinical decision-making and to reduce uncertainty in regulatory and cost-effectiveness decision-making. Additionally, a large proportion of patients in these populations remain alive at the end of trials. Expert elicitation can be used to supplement the short-term mortality data available for these populations.

DAPA-CKD was a randomized, double-blind, placebo-controlled, multicentre clinical trial that investigated the effects of the sodium–glucose cotransporter 2 inhibitor on a composite outcome of kidney function decline, progression to end-stage kidney disease, and death from renal or cardiovascular causes in patients with CKD and elevated albuminuria, with and without type 2 diabetes. Death from any cause was a secondary outcome. DAPA-CKD was the first RCT in patients with CKD to show a statistically significant improvement in all-cause mortality compared with placebo [11]. After a regular review meeting, the independent Data Monitoring Committee recommended that the trial be discontinued because of clear efficacy, on the basis of 408 primary outcome events [10].

We performed an expert elicitation to obtain long-term survival estimates for patients in the placebo arm of DAPA-CKD. These estimates were used with general population mortality (GPM) data and survival data from DAPA-CKD in a Bayesian analysis to estimate long-term survival. The results of this method were compared with those from standard frequentist methods.

METHODS

General Approach

The generalizable method for projecting long-term survival had three steps (Fig. 1). First, literature survival data for populations similar to the population of interest were collated in a data book. Second, survival estimates for the population of interest were gathered using an expert elicitation. Finally, RCT survival data, survival estimates from the experts, and GPM data were combined in a Bayesian analysis. This approach was applied to the placebo arm of DAPA-CKD. See the Supplementary Material for full methodology. Ethics committee approval, consent to participate, consent for publication, and accordance with the Helsinki Declaration of 1964 were not required for this study because it did not involve human participants.

Data Book Creation: Literature Searches and Data Extraction

Literature searches were performed to identify peer-reviewed articles published during 1990–2020 in English that reported the results of RCTs, observational cohort studies, meta-analyses, and national renal registry reports in populations similar to that of DAPA-CKD (Fig. S1 in the Supplementary Material). The resulting articles were screened for those that reported all-cause mortality incidence rates or Kaplan–Meier (KM) estimates for survival or all-cause mortality, included patients at least 18 years of age with non-dialysis-dependent CKD and elevated albuminuria, and had more than 500 patients per study arm. Studies with
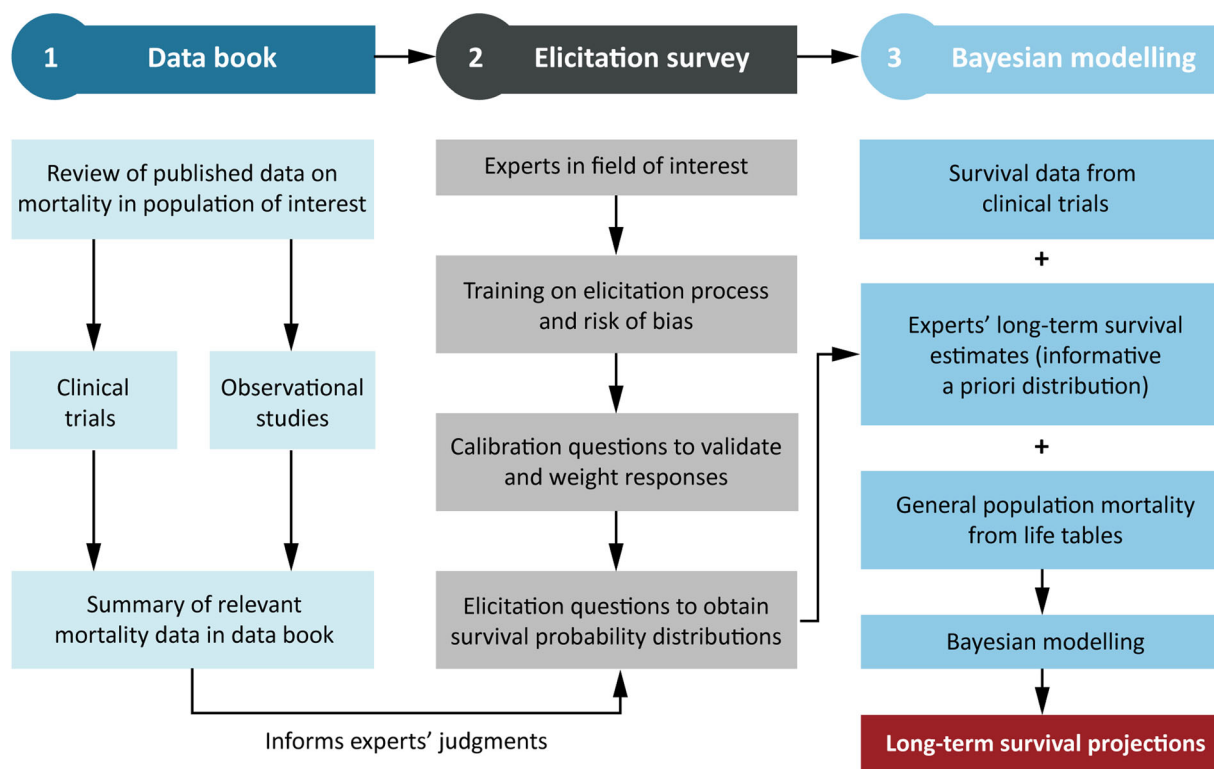
Fig. 1 Summary of the novel and generalizable method for projecting long-term survival using Bayesian methodology with randomized controlled trial survival data, expert-elicited values, and general population mortality data

fewer than 500 patients per study arm were excluded, prioritising larger, landmark studies with bigger sample sizes and therefore more accurate outcome measures.

For relevant articles, study and patient characteristics and all-cause mortality incidence rates were extracted and recorded in a standard form. KM estimates were extracted as JPEG image files and digitized. A model was fitted to the individual level data, which was then used to produce extrapolations to 20 years by calculating standard mortality ratios (SMRs) using age- and sex-adjusted general-population life table data (United States Life Tables 2017, US Department of Health and Human Services) and the internal additive hazards approach of van Oostrum et al. [12, 13]. Extracted data were summarized in a data book, which was provided to the participants of the expert elicitation before the elicitation to inform and support their judgments when providing survival estimates (Table S1 and Fig. S2 in the Supplementary Material).

Expert Elicitation Survey

A formal expert elicitation was used to gather long-term survival estimates for the population of interest from six leading nephrologists using an Excel-based elicitation survey (Fig. S3 in the Supplementary Material). Participants were trained on how to complete the survey and the impacts of common cognitive biases on judgment [14]. Participants then received the survey to be completed independently and at a convenient time. The survey consisted of 10 calibration questions about CKD and related medical topics with known answers from the scientific literature (Table S2 in the Supplementary Material) and three survey questions about the parameters of interest (10- and 20-year survival of patients in the DAPA-CKD placebo arm), which participants answered using their expertise and knowledge of the field, with support from the data book (Table S3 in the Supplementary Material).

Each participant's responses to the calibration questions were used to assess their performance on the survey questions (based on accuracy and information) and to assign performance-based weights to their responses for use when combining the individual judgments (Table S4 in the Supplementary Material) [14–17]. For all questions, participants provided low (P10), high (P90), and medium (P50) estimates for each parameter, where P10 represents the value they are 90% confident that the true value is higher than, P90 represents the value they are 90% confident that the true value is lower than, and P50 represents the value they believe it is equally likely that the true value is either lower or higher than.

Survival Extrapolation

For all analyses, parametric survival estimates were generated using exponential, gamma, generalized gamma, Gompertz, loglogistic, lognormal, and Weibull survival distributions, as recommended by NICE, the Pharmaceutical Benefits Advisory Committee (PBAC), and the Canadian Agency for Drugs and Technologies in Health (CADTH) [2, 6, 18, 19]. Survival estimates to 40 years, point estimates for survival at 20 years, and median survival were determined for each distribution. Where relevant, GPM was accounted for using age- and sex-adjusted general-population life table data [12], according to the internal additive hazards method of van Oostrum et al. [13].

Bayesian Analysis
Elicited survival estimates and GPM data were used in a Bayesian method to extrapolate survival in the placebo arm of DAPA-CKD (Tables S7 and S8 in the Supplementary Material). Bayesian statistical methods determine the probability of an event based on both data and previously held beliefs about the event or conditions associated with the event; the probability of an event occurring can be updated as more evidence is obtained. Parametric Bayesian analysis was used to extrapolate the DAPA-CKD placebo arm KM survival estimate using the

elicited survival estimates and accounting for GPM.

Frequentist Analysis
We hypothesized that survival extrapolations derived through the Bayesian methodology would be more robust and unbiased than those derived through current standard methods. Therefore, for comparison, survival was extrapolated using frequentist methods, as is currently recommended by HTA agencies for estimating long-term survival. In frequentist methods, the probability of an event occurring is determined using the frequency of that event in a repeatable, objective test, and model parameters and hypotheses are considered fixed. Frequentist analyses were used to extrapolate the DAPA-CKD placebo arm KM survival estimate. Two analyses were run, one accounting for GPM and one not accounting for GPM. Elicited survival estimates were not used in this analysis.

RESULTS

Expert Elicitation: Calibration Results

Results from the calibration questions showed that all experts provided responses that are accurate (i.e. the realizations of the calibration questions are likely to correspond statistically with an expert's assessments) and informative (i.e. the expert can articulate that some values are more likely than others) (Tables S2 and S4 in the Supplementary Material). The group average responses to the calibration questions demonstrated greater accuracy than the individual responses (Table S2). Calibration results were used to weight participant responses to the survey questions (Table S4).

Expert Elicitation: Survival of Patients in a DAPA-CKD-Like Population

The weighted P50, P10, and P90 10-year survival predictions were 59%, 47%, and 75%, respectively for patients in a DAPA-CKD-like population. The 20-year survival predictions were 31%, 10%, and 40% (Fig. 2 and Table S5 in the
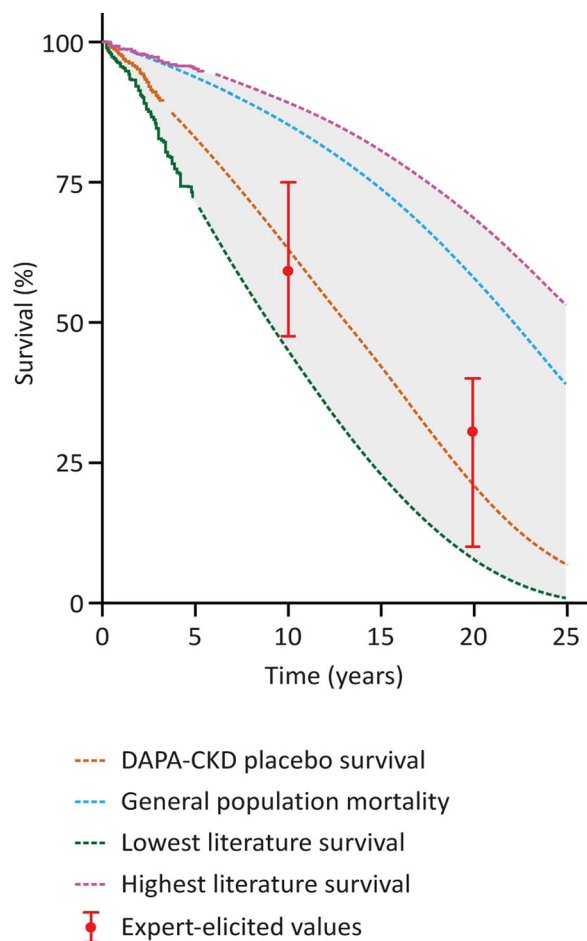
Fig. 2 Group weighted estimates for survival percentage at 10 and 20 years for patients in the placebo arm of DAPA-CKD compared to KM survival estimates from DAPA-CKD placebo and active arms. Trial data and highest and lowest literature KM estimates (solid portions of lines) were extrapolated (dashed portions of lines) using SMRs using age- and sex-adjusted general-population life tables. *KM* Kaplan–Meier, *SMR* standard mortality ratio

Supplementary Material). The predictions were in line with the highest and lowest extrapolated literature KM estimates (Fig. 2).

Survival Extrapolation for Patients in a DAPA-CKD-Like Population

The 20-year survival values for patients in a DAPA-CKD-like population for the seven parametric distributions were 0.0–56.9% in the frequentist analysis, and 0.0–39.2% in the frequentist analysis accounting for GPM

(Fig. 3a, b, Table 1). Median survival was 6–-27 years and 6–17 years, respectively (Table S6 in the Supplementary Material).

The Bayesian analysis using survival data from DAPA-CKD, expert-elicited survival estimates, and GPM resulted in good agreement with both the expert-elicited values and the DAPA-CKD KM survival estimate for all seven distributions (Fig. 3c, Table 1). The 20-year survival values for patients in the placebo arm of DAPA-CKD for the seven distributions were 14.9–39.1%, which was in line with the expert-elicited 20-year survival estimates of 10–40% (Figs. 2, 3c). Median survival was 11–17 years (Table S6 in the Supplementary Material).

Of the three extrapolation methods, the frequentist analysis without GPM produced the greatest variability across the seven distributions over time relative to the SMR-extrapolated KM survival estimate from the DAPA-CKD placebo arm, followed by the frequentist analysis accounting for GPM, followed by the Bayesian analysis (Fig. S4 in the Supplementary Material).

DISCUSSION

We describe an innovative and generalizable method that uses expert opinion to inform long-term survival extrapolations in patient populations for which only limited survival data from RCTs or RWE studies are available. It uses an expert elicitation method that incorporates a data book summarizing data relating to the uncertain parameters of interest, participant training on the potential effects of cognitive biases on their responses, and a remote elicitation survey. The method was applied to data from patients in the placebo arm of DAPA-CKD to generate long-term survival estimates for this population, for which little is known about long-term survival. While only the placebo arm was considered here, in the future treatment effects could be dealt with as indicated in NICE, PBAC, and CADTH guidelines [6, 18, 19].

Survival data from DAPA-CKD were initially extrapolated using a frequentist approach, as recommended by various HTA agencies [2, 18, 19]. This method produced a wide range
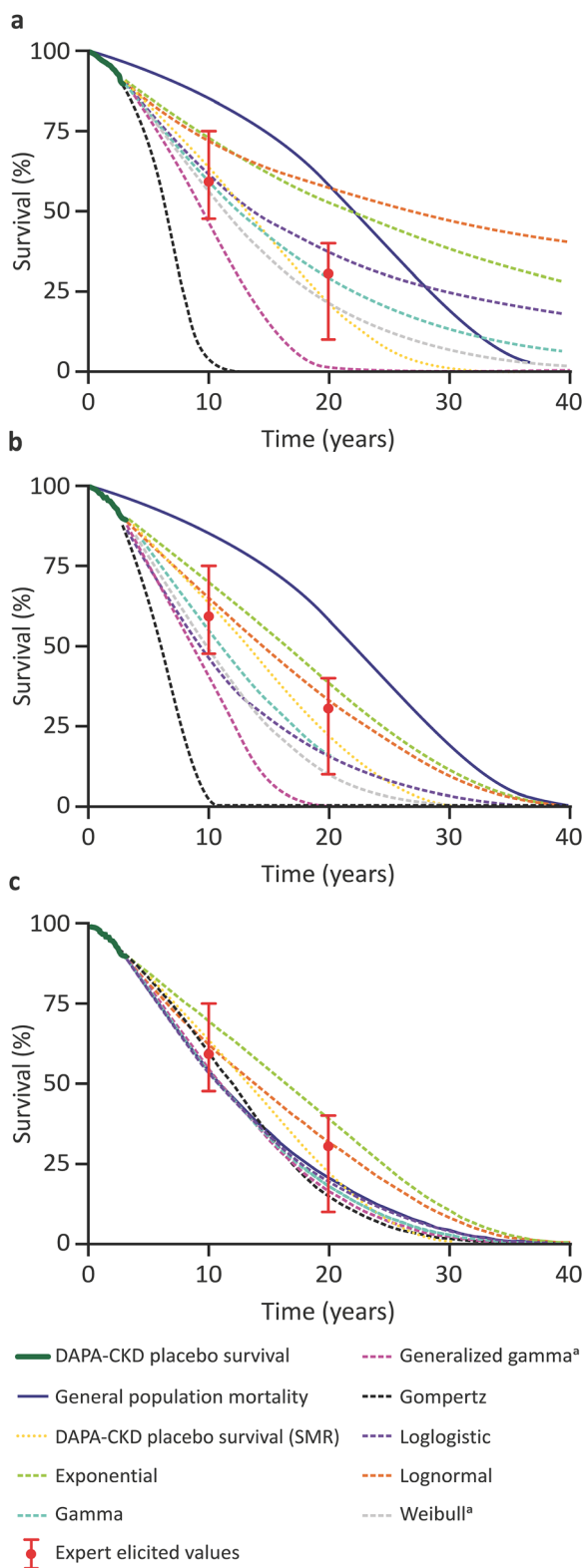
◄**Fig. 3** Long-term survival extrapolations for patients in the placebo arm of DAPA-CKD. Results for **a** frequentist methods, **b** frequentist methods accounting for GPM, and **c** Bayesian methods are presented for seven distributions (exponential, gamma, generalized gamma, Gompertz, loglogistic, lognormal, and Weibull). The DAPA-CKD placebo arm KM survival estimate, SMR extrapolation, GPM, and expert-elicited values are also presented in **a–c** for reference. *GPM* general population mortality, *KM* Kaplan–Meier, *SMR* standard mortality ratio. [a]Results for the Weibull and generalized gamma distributions for the Bayesian analysis overlap

of long-term survival estimates, potentially because approximately 90% of the trial population were still alive at the end of DAPA-CKD, and these survival extrapolations are therefore based on few instances of death. Furthermore, at later time points, survival for the DAPA-CKD-like population was predicted to be higher than that of the general population for the lognormal, exponential, loglogistic, and gamma distributions, which is clinically implausible. Also, some of the distributions were not aligned with the expert elicitation—the generalized gamma, Gompertz, lognormal, and exponential distributions did not fall within the credibility intervals defined by the elicited survival estimates, indicating that they would be considered unlikely by the experts.

Incorporating GPM into the frequentist analysis, as indicated by NICE and implemented by van Oostrum et al. [2, 13], reduced the variability in the survival estimates and made the projections more plausible because survival of the disease population could, by definition, no longer be higher than that of the general population. However, this approach still produced a wide range of long-term survival estimates. The generalized gamma and Gompertz distributions were not aligned with expert opinion and predicted higher mortality than expected based on the expert-elicited survival estimates.

As expected with the Bayesian approach that incorporated expert opinion, the values for survival at 20 years from the seven distributions converged and the method provided more

Table 1 Long-term survival extrapolations for patients in the placebo arm of DAPA-CKD

Population	Distribution/source	Estimated survival at 20 years (%) (95% CI)		
		Frequentist	Frequentist + GPM	Bayesian
Placebo arm of DAPA-CKD [10]	Exponential	52.8 (46.9–58.2)	39.2 (35.6–42.3)	39.1 (35.3–43.0)
	Gamma	28.9 (16.2–41.6)	17.5 (5.0–29.0)	18.1 (11.3–29.7)
	Generalized gamma	1.0 (0.0–96.3)	0.0 (0.0–58.2)	16.5 (3.4–33.1)
	Gompertz	0.0 (0.0–0.0)	0.0 (0.0–0.1)	14.9 (0.2–39.5)
	Loglogistic	37.1 (25.5–47.2)	15.8 (9.4–23.0)	20.8 (13.9–30.5)
	Lognormal	56.9 (47.4–63.7)	33.8 (21.1–40.6)	32.0 (21.6–39.9)
	Weibull	21.3 (7.1–35.8)	9.9 (0.2–27.4)	16.7 (9.2–28.7)
	SMR extrapolation	21.1		
General population	Life table data [12]	58.2		

Estimated survival percentages at 20 years are presented for Bayesian, and frequentist methods and frequentist methods accounting for GPM for seven distributions (exponential, gamma, generalized gamma, Gompertz, loglogistic, lognormal, and Weibull). Estimated survival percentages at 20 years are also presented for SMR extrapolations and for general-population life table data

CI confidence interval, *GPM* general population mortality, *SMR* standard mortality ratio

consistent survival estimates across a lifetime horizon than the frequentist analyses.

The estimates produced using the Bayesian approach can be used by patients, clinicians, and those making decisions about the regulation of new therapies. These estimates could lead to reduced uncertainty for patients and healthcare professionals (HCPs) and help with HCP–patient conversations about life expectancy, disease trajectory, and long-term outcomes. HCPs also require accurate survival estimates to help them make decisions about patient care and evaluate the long-term risk/benefit profiles of newly approved therapies. Reliable survival estimates may also provide encouragement to patients to continue treatment by demonstrating long-term positive outcomes associated with treatment compliance, thereby optimizing patient management. Additionally, validated long-term survival projections with reduced uncertainty are fundamental to HTA agencies for making balanced assessments of the long-term benefits of new treatments. Clinical trial sponsors should

consider factoring this analysis approach into primary statistical analysis plans.

The expert elicitation method was designed to reduce the impacts of cognitive biases that can influence the formation of judgments. The data book aimed to mitigate the availability bias by presenting participants with studies they may not have been familiar with and by presenting the studies consistently without emphasizing the results of any one particular study. Training about cognitive biases was provided to the experts before the elicitation survey. Training about the statistical significance of the P10 and P90 estimates and guidance on making these assessments helped mitigate the overconfidence bias. Participants were instructed to consider extreme (P10 and P90) outcomes first, which helped mitigate the anchoring bias. Participants completed the survey independently, which also helped mitigate the anchoring bias by reducing the risk that individual responses were influenced by other participants.

Participants completed the survey remotely, avoiding potential scheduling and geographical constraints and allowing experts to be recruited

from across the world. Previous expert elicitations have required lengthy in-person interviews. However, in the approach described here, participants completed the remote survey successfully. Although face-to-face elicitations are generally preferred to virtual ones [7, 20], the average accuracy and information scores for the participants in this virtual elicitation were higher than those from some face-to-face elicitations [21–23], demonstrating that virtual elicitation can be highly effective.

Finally, in contrast to existing Bayesian methods, the use of conditional survival percentages in the method described here meant that 20-year survival could not be higher than 10-year survival in any Markov chain Monte Carlo iteration. As a result, the method is more clinically plausible than methods in which a higher survival percentage can be sampled at 20 years than at 10 years or, in other words, in which the analyzed patients are allowed to be resurrected.

This study has certain limitations. The literature review that was performed to create the data book was not systematic, which may have influenced the responses of some experts participating in the elicitation since some were also authors of the studies included in the data book. In addition, the literature review only included papers published in English. However, it included 13 relevant RCTs and observational studies. Furthermore, the populations of the trials presented in the data book were not identical to the population of DAPA-CKD. In the future, the data book could be enriched using data from electronic health records or from risk equations. The extrapolations of the literature survival data provided in the data book were intended to help the experts in their assessments; however, it is possible that the inclusion of these extrapolations may have biased the experts' assessments. We also did not account for differences in background mortality between countries represented in the literature survival data and the USA (which was used as the reference population to generate the extrapolation). Finally, the DAPA-CKD population may not accurately reflect the general population of patients with CKD, and survival estimates made for patients in the DAPA-CKD placebo arm may not be applicable to a broader population.

CONCLUSION

We describe a method for obtaining long-term survival estimates in cases where RCTs provide limited data. This method combines the results of an expert elicitation with short-term RCT mortality and GPM data in a Bayesian analysis. It was applied to extrapolate survival of patients in the placebo arm of DAPA-CKD to produce long-term survival estimates that were plausible, robust, and aligned with expert clinical opinion. The approach is versatile and generalizable. Beyond the initial application to CKD, we propose it could be applied to a wide range of patient populations in which a lack of long-term survival data makes the use of conventional statistical methods challenging, including chronic disease populations and populations expected to have long remaining lifespans, such as paediatric populations or those undergoing gene therapy. The method provides consistent long-term survival extrapolations that are in line with clinical opinion, which could provide additional evidence for use by HCPs, patients, cost-effectiveness and HTA decision-makers, and more broadly by policy, regulatory, and reimbursement agencies.

ACKNOWLEDGEMENTS

Author Contributions. All authors contributed to data acquisition and/or interpretation, all contributed to drafting and revising the manuscript and all approved the final version.

Bartholomeus JA Willigers, Mario Ouwens, and Juan Jose Garcia Sanchez designed the study. Bartholomeus JA Willigers and Mario Ouwens additionally contributed to methods development and data analysis.

Disclosures. Bartholomeus JA Willigers, Mario Ouwens, and Juan Jose Garcia Sanchez are employees of AstraZeneca. Andrew Briggs has an ongoing consultancy contract with AstraZeneca. Hiddo JL Heerspink has received research support from AstraZeneca and support from AstraZeneca to his institution for the DAPA-CKD trial. Carol Pollock has received consultancy fees from Astra-Zeneca. Roberto Pecoits-Filho has received honorarium (paid to his employer) from AstraZeneca to participate in advisory boards and educational activities. Navdeep Tangri has received consultancy fees from AstraZeneca. Csaba P Kovesdy has received honoraria from AstraZeneca for advisory boards, is a member of various clinical trial executive committees for AstraZeneca, and has received consultancy fees from Abbott, Akebia, Bayer, Boehringer Ingelheim, Cara Therapeutics, CSL Behring, GlaxoSmithKline, Rockwell, and Vifor. David C Wheeler has an ongoing consultancy contract with AstraZeneca and has received honoraria from Astellas, Amgen, Bayer, Boehringer Ingelheim, GlaxoSmithKline, Gilead, Janssen, Napp/Mundipharma, Merck Sharp and Dohme, Tricida, Vifor, and Zydus.

Compliance with Ethics Guidelines. Ethics committee approval, consent to participate, consent for publication, and accordance with the Helsinki Declaration of 1964 were not required for this study because it did not involve human participants.

Data Availability. All data generated or analyzed during this study are included in this published article/as supplementary information files.

Code Availability. The analysis code files are available in the electronic supplementary material.

Open Access. This article is licensed under a Creative Commons Attribution-NonCommercial

REFERENCES

1. Ioannidis JP. Effect of the statistical significance of results on the time to completion and publication of randomized efficacy trials. JAMA. 1998;279(4): 281–6.

2. Rutherford MJ, Lambert PC, Sweeting MJ, et al. Technical support document 21: flexible methods for survival analysis. National Institute for Health and Care Excellence Decision Support Unit. 2020.

3. Savage LJ. The foundations of statistics. New York: Wiley; 1954.

4. Dalkey NC. The Delphi method: an experimental study of group opinion. Santa Monica: RAND; 1969.

5. National Academies of Sciences, Engineering, and Medicine. Valuing climate damages: updating estimation of the social cost of carbon dioxide. Washington: The National Academies Press; 2017.

6. Latimer NR. Technical support document 14: survival analysis for economic evaluations alongside clinical trials—extrapolation with patient-level data. National Institute for Health and Care Excellence Decision Support Unit. 2011.

7. Dallow N, Best N, Montague TH. Better decision making in drug development through adoption of formal prior elicitation. Pharm Stat. 2018;17(4): 301–16.

8. Azzolina D, Berchialla P, Gregori D, et al. Prior elicitation for use in clinical trial design and analysis: a literature review. Int J Environ Res Public Health. 2021;18:1833.

9. Cope S, Ayers D, Zhang J, et al. Integrating expert opinion with clinical trial data to extrapolate long-term survival: a case study of CAR-T therapy for children and young adults with relapsed or refractory acute lymphoblastic leukemia. BMC Med Res Methodol. 2019;19(1):182.

10. Heerspink HJL, Stefansson BV, Correa-Rotter R, et al. Dapagliflozin in patients with chronic kidney disease. N Engl J Med. 2020;383(15):1436–46.

11. Garcia Sanchez JJ, Thompson J, Scott DA, et al. Treatments for chronic kidney disease: a systematic literature review of randomized controlled trials. Adv Ther. 2022;39:193–220.

12. Arias E, Xu J. United States life tables, 2017 vol 68 no 7. National Vital Statistics Reports. Hyattsville, MD: National Center for Health Statistics; 2019.

13. van Oostrum I, Ouwens M, Remiro-Azocar A, et al. Comparison of parametric survival extrapolation approaches incorporating general population mortality for adequate health technology assessment of new oncology drugs. Value Health. 2021;24(9):1294–301.

14. O'Hagan A. Expert knowledge elicitation: subjective but scientific. Am Stat. 2019;73:69–81.

15. Cooke RM. Experts in uncertainty—opinion and subjective probability in science. New York: Oxford University Press; 1991.

16. Colson AR, Cooke RM. Expert elicitation: using the classical model to validate experts' judgments. Rev Environ Econ Policy. 2018;12(1):113–32.

17. Morgan M, Henrion M. Uncertainty: a guide to dealing with uncertainty in quantitative risk and policy analysis. Cambridge: Cambridge University Press; 1990.

18. Pharmaceutical Benefits Advisory Committee guidelines: model transition probabilities or variables, transformation and extrapolation. Australian Government Department of Health. https://pbac.pbs.gov.au/section-3a/3a-4-model-transition-probabilities-or-variables-transformation-extrapolation.html. Accessed 25 Nov 2021.

19. Procedures for the CADTH common drug review and interim plasma protein product review. Canadian Agency for Drugs and Technologies in Health. https://www.cadth.ca/sites/default/files/cdr/process/Procedure_and_Guidelines_for_CADTH_CDR.pdf. Accessed 25 Nov 2021.

20. Cooke RM, Marti D, Mazzuchi T. Expert forecasting with and without uncertainty quantification and weighting: what do the data say? Int J Forecast. 2021;37(1):378–87.

21. Van der Fels-Klerx HJ, Cooke RM, Nauta MN, et al. A structured expert judgment study for a model of Campylobacter transmission during broiler-chicken processing. Risk Anal. 2005;25(1):109–24.

22. Rothlisberger JD, Finnoff DC, Cooke RM, et al. Ship-borne nonindigenous species diminish Great Lakes ecosystem services. Ecosystems. 2012;15(3):1–15.

23. Wittmann ME, Cooke RM, Rothlisberger JD, et al. Using structured expert judgment to assess invasive species prevention: Asian carp and the Mississippi-Great Lakes hydrologic connection. Environ Sci Technol. 2014;48(4):2150–6.