Supplementary Material

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

Bartholomeus J. A. Willigers[1], Mario Ouwens[2], Andrew Briggs[3], Hiddo J. L. Heerspink[4], Carol Pollock[5], Roberto Pecoits-Filho[6], Navdeep Tangri[7], Csaba P. Kovesdy[8], David C. Wheeler[9], Juan Jose Garcia Sanchez[10]

[1]Decision Science, Statistical Innovation, Biopharmaceuticals R&D, AstraZeneca, Cambridge, UK; [2]Medical & Payer Evidence Statistics, Real World Science and Digital, Biopharmaceuticals Business Unit, AstraZeneca, Mölndal, Sweden; [3]London School of Hygiene & Tropical Medicine, London, UK; [4]University of Groningen, Groningen, Netherlands; [5]University of Sydney, Sydney, Australia; [6]Pontifical Catholic University of Paraná, Curitiba, Brazil; [7]University of Manitoba, Winnipeg, MB, Canada; [8]University of Tennessee Health Science Center, Memphis, TN, USA; [9]University College London, London, UK; [10]Health Economics, Biopharmaceuticals Business Unit, AstraZeneca, Cambridge, UK

Corresponding author: Bartholomeus J. A. Willigers (bartholomeus.willigers@astrazeneca.com)

Contents

Supplementary methods

Literature search and mortality data extraction

Literature searches were performed to identify peer-reviewed articles published between 1990 and 2020 about CKD burden and treatment. The results of these literature searches were screened for articles reporting the results of randomized control trials (RCTs), observational cohort studies, or meta-analyses that reported all-cause mortality data for patients with characteristics similar to those of patients enrolled in DAPA-CKD (patients with non-dialysis-dependent CKD and elevated albuminuria, aged 18 years and older) (**Fig. S1**). These results were supplemented with national renal registry reports for Australia, Canada, China, France, Germany, Italy, Spain, the UK, and the USA, articles resulting from the CKDopps, CKD Prognosis Consortium, CKD-REIN, CPRD, CRIC, and DOPPS large-scale, longitudinal cohort studies and articles from landmark CKD studies.

The titles and abstracts of the articles identified by the literature searches were screened to identify articles that reported on the population of interest and were published in English. For articles retained after abstract screening, the full text was screened to identify articles that reported all-cause mortality incidence rate or Kaplan–Meier survival or all-cause mortality estimates, included patients aged ≥18 years with non-dialysis-dependent CKD and elevated albuminuria, had >500 patients per study arm and were published in English. References lists from relevant meta-analyses and systematic literature reviews were screened for additional papers using the same procedure and meeting the same criteria.

For relevant articles, study characteristics (including location, study date, duration of follow-up, type of study, and number of participants), patient characteristics (including age, proportion of female patients, and prevalence of comorbidities), relevant laboratory measurements (including estimated glomerular filtration rate and urine albumin-to-creatinine ratio), and all-cause mortality incidence rate were extracted and recorded in a standard Microsoft Excel spreadsheet. Kaplan–Meier survival estimates were extracted as .jpeg image files and digitized using a non-commercial solution that enables users to trace and scale digital images. They were then combined into a single figure.

Literature Kaplan–Meier survival estimates were also extrapolated to 20 years to provide a range of long-term survival projections to inform the experts' judgements. This was achieved 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) in Microsoft Excel [1].

The data were summarized in a data book, produced in Microsoft Word, which was provided to the participants of the expert elicitation to inform and support their judgements when providing survival estimates (**Table S1** and **Fig. S2**).

Expert elicitation survey

Expert elicitation is a well-established method for obtaining and synthesizing unbiased expert judgements that can provide valuable quantitative information when empirical data are lacking. Developed in the 1950s, early expert elicitations used the Delphi method, in which a panel of experts develop a consensus estimate for an uncertain parameter [2, 3]. As understanding of the cognitive errors that can occur during the estimation of uncertain parameters improved, so too did methods for expert elicitation [4-8].

Expert elicitation has recently received renewed attention for several reasons. These include an increased awareness of cognitive biases and their impact on expert opinions collected outside of formal elicitation processes; the development and increased use of Bayesian and Markov chain Monte Carlo (MCMC) methods in the design, monitoring, and analysis of RCTs [9-11], which require informative, robust, and unbiased a priori distributions [4, 12]; and endorsement of expert elicitation by agencies such as NICE and scientific advisory bodies such as the National Academies of Sciences, Engineering, and Medicine (NASEM) [13-15].

The expert elicitation process described here consisted of six steps: selection of experts, definition of the uncertainties to be assessed, creation of a data book, training of experts, elicitation of expert judgements, and aggregation of judgements.

Selection of experts

Six experts participated in the elicitation in this study: H.J.L. Heerspink (University of Groningen, Groningen, Netherlands), C.P. Kovesdy (University of Tennessee Health Science Center, Memphis, TN, USA), R. Pecoits Filho (Pontifical Catholic University of Paraná, Curitiba, Brazil), C. Pollock (University of Sydney, Sydney, Australia), N. Tangri (University of Manitoba, Winnipeg, MB, Canada), and D.C. Wheeler (University College London, London, UK). These experts are globally renown and recognized for their expertise in CKD. Several of the experts have co-authored studies included in the data book. Together, the

experts constitute a diverse group representing different countries and organizations. Expert responses were anonymized.

Definition of the uncertainties to be assessed

A set of 10 calibration questions about CKD and related medical topics were developed. The calibration questions have known answers taken from the scientific literature and were tailored to the area of expertise of the participating experts (**Table S2**). The calibration questions were based on the USA and the UK because of the wealth of studies and information on CKD in these countries. We followed recommendations from Cooke (1991) and Morgan and Henrion (1990) to use a total of 10 calibration questions and to ensure that sufficient background information was provided and that questions could not be misinterpreted [16, 17]. The results of the calibration questions were used to assign weights to the individual expert judgements, according to Cooke's classical method [16, 18]. A set of three survey questions about long-term survival of patients with CKD and elevated albuminuria were also developed. The questions were refined over a series of iterations involving several individuals who did not participate in the final expert elicitation. For each calibration question and survey question, participants were asked to provide low (P10), high (P90), and medium (P50) estimates for each parameter, where:

- P10 represents the value for which they are 90% confident that the true value is higher than this particular value
- P90 represents the value for which they are 90% confident that the true value is lower than this value
- and P50 represents the value for which they believe it is equally likely that the true value is either lower or higher than this value.

For each survey question (**Table S3**), participants were asked to provide P10, P90, and P50 estimates for the survival percentages for patients in the placebo arm of the DAPA-CKD trial at:

- 10 years
- 20 years assuming the survival percentage at 10 years was 40%
- 20 years assuming the survival percentage at 10 years was 70%.

Participants in the expert elicitation were asked about the placebo arm of DAPA-CKD because they were not expected to have experience of long-term dapagliflozin use, and were

therefore not expected to be able to predict long-term survival for patients in the dapagliflozin arm of the trial.

Creation of a data book

A data book was created from the results of the literature review in order to present a summary of relevant data to the experts to help them complete the elicitation survey and to ensure they provided informed judgements. Study and patient characteristics and relevant mortality data from 13 studies were gathered in a table (**Table S1**). Kaplan–Meier survival estimates were summarized in a single plot (**Fig. S2**). In addition to data on the population of interest, mortality data for the general population was also included. Combining the data from several studies in a single data book gave participants a comprehensive summary of the relevant data and allowed them to compare and contrast the information as they made their judgements on long-term survival. This approach aimed to reduce the impact of the availability bias by presenting the participants with the results of studies that they may not have been familiar with alongside those they may have pre-existing knowledge of. This was intended to emphasise that the results of any familiar studies are a part of a broader collection of results.

Training of experts

Before the experts were invited to provide their judgements, they attended a one-hour training session via video conference. The training consisted of a presentation on the basic theory of expert elicitation, the electronic elicitation tool, the impacts of common cognitive biases on decision-making, and the need for judgements to be both informative and accurate [8]. Three prevalent cognitive biases were discussed: anchoring (over-reliance on an initial piece of information); availability bias (if something can be recalled easily, it must be important); and overconfidence (subjective confidence exceeds objective accuracy). Experts were also reminded of the statistical significance of the P10, P50, and P90 estimates, and guidance was provided on how to make these assessment (e.g. ‘the P50 signifies the value for which you believe the true value is equally likely to be higher or lower’). This helped to minimize the overconfidence bias. Several examples of non-informative and inaccurate judgements were shown to the experts.

The data book was provided to the participants 3 days before the training session to allow them to familiarize themselves with the contents. During the training session, the data book was reviewed, and participants confirmed that they understood the summary provided and that no significant pieces of evidence had been omitted. The studies in the data book were presented in a consistent manner that did not emphasize or highlight the results of any one study, which further helped to minimize the impact of the availability bias. No recommendations were made about whether participants should or should not consult other data sources.

Elicitation of expert judgements

The expert judgements were captured using a Microsoft Excel-based electronic elicitation tool that could be completed remotely and at a time convenient to each participant (**Fig. S3**). The elicitation survey was sent to the participants after the training session took place and all completed surveys were returned within approximately 1 month. The tool guides the participants through the questions by directing participants to complete each question in the correct order; participants were prompted to assess the percentiles in the tails (P10 and P90) first before assessing the central value (P50). This approach helps to minimize the impact of the overconfidence bias, which can arise when individuals begin with a central value and then adjust away from it, but not sufficiently far, consistent with the 'anchoring and adjustment' heuristic of Tversky and Kahneman (1974) [19]. The survey was completed independently by each participant, which helped to mitigate the anchoring bias by reducing the possibility that individuals' responses were influenced by those of the other participants. Error messages were displayed when participants provided answers with logical errors (e.g. an upper value being lower than the central value) or when answers were omitted. The tool also displayed a summary of the data contained in the data book (**Fig. S2**) as well as baseline characteristics of the patients in the population of interest for participants to refer to throughout the survey.

Aggregation of judgements

The participants responses to the calibration questions were used to assess the performance of each participant on the three survey questions and to assign a performance-based weight to the participant's responses for use when combining the judgements of the individual experts. Individual answers to the calibration questions were assessed for accuracy and for information. Accuracy reflects the likelihood that the realizations of the calibration questions

correspond, in a statistical sense, with an expert's assessments (i.e. the expectation is that 80% of the P10–P90 ranges as defined by an expert contain the actual, known, value of the calibration questions; overall, the experts achieved an accuracy of 7 out of the 10 ranges containing the true answer and the extremes were 6 [two of the experts] and 9 out of ten). An expert providing an extremely large interval can achieve an apparently excellent calibration. However, such judgement will not be informative. Information captures the expert's ability to articulate that some values more likely than others. Informativeness is measured using the Kullback–Leibler divergence between the probability distribution as defined by the expert and the least informative distribution, a uniform distribution across the range of opinion.[2] A combined score, representing the expert's ability to make good judgements, is determined by taking the product of the accuracy and information score. An expert specific weight was determined by normalizing the individual combined score. No fixed threshold was used to exclude experts based on their performance on the calibration questions. This was done because experts were selected on the basis of their knowledge, not on their predicted ability to make good judgments (defined as achieving a good balance between accuracy and information in their responses to the calibration questions).

The P10, P90, and P50 values elicited from the individual experts were pooled using a specific weighting for each expert that reflects their performance on the calibration questions, according to Cooke's method (**Table S4**) [16, 18]. Individual estimates were also combined with equal weighting for comparison. None of the participants' responses were excluded from the analysis.

The range for survival at 20 years (**Fig. 2 in main text**) are derived by combining the results for 10-year survival and the two conditional 20-year survival. The result is obtained from a Monte Carlo simulation and applied the conditional logic as described in the Bayesian survival modelling section below.

Survival extrapolation

Statistical analysis was performed using R (version 3.5.1). The frequentist analysis was performed using flexsurv and the Bayesian analysis was performed using RStan (version 2.17.3).

Bayesian survival modelling

In each MCMC iteration, a survival percentage was sampled for 10 and for 20 years by several steps.

The participants in the expert elicitation survey provided estimates of:

- S10, survival percentage for 10 years
- $S20_{S10=40\%}$, survival percentage for 20 years, assuming the survival percentage for 10 years was 40%
- $S20_{S10=70\%}$, survival percentage for 20 years, assuming the survival percentage for 10 years was 70%.

The MCMC iterations used $CS20_{S10=40\%}$ and $CS20_{S10=70\%}$, where $CS20_{S10=40\%} = S20_{S10=40\%}/S10$ and $CS20_{S10=70\%} = S20_{S10=70\%}/S10$. Values for S10, for $CS20_{S10=40\%}$ and $CS20_{S10=70\%}$ are sampled per MCMC iteration using uniform distributions (with 0 as smallest value and 1 as largest). If S10 ≤40%, the 20 years survival is sampled to be $S20 = S10 \times CS20_{S10=40\%}$. If S10 ≥70%, the 20 years survival is sampled to be $S20 = S10 \times CS20_{S10=70\%}$. If S10 is in between, a weighted average of $CS20_{S10=40\%}$ and $CS20_{S10=70\%}$ is computed depending on how close S10 is to 40% compared to how close it is to 70% (Tables S7 and S8).

In the model with adjustment for general population mortality, patients can die from all-cause mortality as well as from the disease of interest at 10 and 20 years after initiation of trial participation. S10 is therefore written as S10 = GPM10×PSD10, where GPM10 is the percentage of patients alive at 10 years according to general population mortality information and PSD10 is the percentage of patients alive at 10 years according to one of the seven parametric survival distributions (exponential, gamma, generalized gamma, Gompertz, loglogistic, lognormal, or Weibull). Likewise, S20 = GPM20×PSD20. Values for which PSD20 is larger than 100% are ignored (Tables S7 and S8).

For the Weibull, lognormal, and loglogistic distributions, the values for the two parameters of these distributions are analytically determined per iteration using PSD10 and PSD20 and used to evaluate the corresponding likelihood.

For the more complex Gompertz and gamma two-parameter distributions, more sophisticated programming was necessary in which the two parameters were chosen to minimize the

difference between S10 and S20 and the modelled survival percentages PSD10×GPM10 and PSD20×GPM20, which is for the right parameter values equal to 0.

For the exponential distribution, the rate is set to the mean of the rate from baseline to 10 years and the rate from 10 years to 20 years based on PSD10 and PSD20 (adjusted for general population mortality).

For the generalized gamma distribution, a third parameter Q is sampled using a uniform distribution on [−1, 2]. This was needed because of convergence problems for larger absolute values of Q where the generalized gamma likelihood becomes very flat. The remainder is the same as for the gamma distribution.

Survival distribution functions and parameters used in the frequentist, frequentist accounting for general population mortality and Bayesian analyses are summarized in **Table S8**.

Supplementary tables

Table S1 Data summary adapted from the data book. Study and patient characteristics and all-cause mortality data for non-dialysis-dependent patients with elevated albuminuria

Study ID	Study characteristics					Patient characteristics								Mortality data	
	Study type	Location	Median follow-up duration, years	Trial arms	n	Age, years	eGFR, mL/min/ 1.73 m^2	Female, %	Albuminuria, %	Anaemia, %	Diabetes, %	CVD, %	HF, %	All-cause mortality incidence rate, per 100 patient-years	Kaplan–Meier figure reference
RENAAL Brenner 2001 [20]	RCT	28 countries	3.4[a]	Losartan Placebo	751 762	60.0 (7.0) 60.0 (7.0)	NR	38.5 35.2	NR	NR	100.0 100.0	NR	0.0	6.8 6.6	NR
IDNT Lewis 2001 [21]	RCT	USA	2.6	Irbesartan Amlodipine Placebo	579 567 569	59.3 (7.1) 59.1 (7.9) 58.3 (8.2)	NR	35.0 37.0 29.0	100.0	NR	100.0	27.0 30.0 29.0	NR	NR	**Fig. S2**
AASK Wright 2002 [22]	RCT	USA	3.0–6.4	Ramipril Amlodipine Metoprolol	436 217 441	54.4 (10.9) 54.5 (10.7) 54.9 (10.4)	45.4 (12.8) 45.8 (12.9) 45.8 (13.4)	38.5 39.6 38.6	NR	NR	0.0 0.0 0.0	NR	0.0 0.0 0.0	1.5 1.7 2.0	NR
PROMIS Levin 2008 [23]	Obs.	Canada	2.6	NA	4,231	66.8 (14.5)	<30.0	44.0	NR	NR	33.0	NR	NR	4.5	**Fig. S2**
BEACON de Zeeuw 2013 [24]	RCT	Australia, Canada, EU, Israel, Mexico, USA	0.75	Bardoxolone methyl Placebo	1,088 1,097	68.9 (9.7) 68.2 (9.4)	22.4 (4.3) 22.5 (4.6)	42.0 43.0	NR	NR	100.0 100.0	56.0 56.0	NR	NR	**Fig. S2**
Sunnybrook Sud 2014 [25]	Obs.	Canada	3.0	CKD stage 3A CKD stage 3B CKD stage 4 CKD stage 5	940 1,252 881 200	67.0 (14.0) 72.0 (13.0) 73.0 (14.0) 71.0 (15.0)	52.0 (4.0) 37.0 (4.0) 23.0 (4.0) 11.0 (2.0)	40.0 42.0 48.0 52.0	NR	NR	44.0 52.0 54.0 46.0	36.0 46.0 43.0 33.0	14.0 21.0 27.0 24.0	2.2 4.4 8.0 9.4	**Fig. S2**
CRIC Orlandi 2018 [26]	Obs.	USA	7.3	With haematuria Without haematuria	1,145 2,127	55.0 (12.0) 59.0 (10.0)	40.0 (15.0) 45.0 (16.0)	41.0 44.0	NR	NR	56.0 48.0	34.0 35.0	10.0 11.0	2.8 2.3	NR

Study ID	Study characteristics					Patient characteristics								Mortality data	
	Study type	Location	Median follow-up duration, years	Trial arms	n	Age, years	eGFR, mL/min/ 1.73 m^2	Female, %	Albuminuria, %	Anaemia, %	Diabetes, %	CVD, %	HF, %	All-cause mortality incidence rate, per 100 patient-years	Kaplan–Meier figure reference
CREDENCE Perkovic 2019 [27]	RCT	34 countries	2.6	Canagliflozin Placebo	2,202 2,199	62.9 (9.2) 63.2 (9.2)	56.3 (18.2) 56.0 (18.3)	34.6 33.3	NR	NR	100 100	50.5 50.3	NR	2.9 3.5	**Fig. S2**
FIDELIO-DKD Bakris 2020 [28]	RCT	48 countries	2.6	Finerenone Placebo	2,833 2,841	65.4 (8.9) 65.7 (9.2)	44.4 (12.5) 44.3 (12.6)	31.1 28.5	NR	NR	100 100	100.0 100.0	NR	2.90 3.23	**Fig. S2**
CRIC Correa 2020 [29]	Obs.	USA	5.1	MPO < 79 pmol/L MPO 79–109 pmol/L MPO 109–155 pmol/L MPO > 155 pmol/L	968 968 968 968	57.7 (11.3) 58.6 (10.4) 58.2 (10.9) 58.3 (11.3)	46.2 (35.8–57.5) 44.1 (33.9–53.9) 41.8 (32.6–52.7) 40.4 (31.1–51.0)	59.9 56.9 52.1 50.7	NR	43.1 47.4 50.3 49.0	43.0 46.0 55.5 50.5	18.4 19.8 23.1 25.6	5.7 8.3 11.5 13.5	7.4[b] 10.6[b] 13.9[b] 17.8[b]	**Fig. S2**
CRIC Hu 2020 [30]	Obs.	USA	12.0	HEI-2015 tertile 1 HEI-2015 tertile 3 AHEI-2010 tertile 1 AHEI-2010 tertile 3 aMed tertile 1 aMed tertile 3 DASH tertile 1 DASH tertile 3	801 801 801 801 870 682 912 795	55.0 (12.0) 60.0 (10.0) 56.0 (12.0) 59.0 (10.0) 55.0 (12.0) 56.0 (12.0) 55.0 (12.0) 60.0 (10.0)	45.0 (17.0) 48.0 (17.0) 44.0 (16.0) 49.0 (18.0) 45.0 (17.0) 44.0 (16.0) 45.0 (17.0) 48.0 (17.0)	39.0 57.0 43.0 53.0 48.0 49.0 37.0 59.0	NR	NR	40.0 43.0 38.0 47.0 41.0 44.0 37.0 49.0	30.0 30.0 31.0 29.0 31.0 30.0 30.0 30.0	NR	3.6 2.6 3.6 2.6 3.6 2.5 3.5 2.7	NR
CRIC Ku 2020 [31]	Obs.	USA	7.1	White Black	1,638 1,650	58.0 (11.0) 58.0 (11.0)	46.2 (14.7) 43.7 (14.9)	40.0 51.1	NR	NR	39.6 51.4	NR	7.1 13.2	NR	**Fig. S2**
DAPA-CKD Heerspink 2020 [32]	RCT	21 countries	2.4	Dapagliflozin Placebo	2,152 2,152	61.8 (12.1) 61.9 (12.1)	43.2 (12.3) 43.0 (12.4)	32.9 33.3	NR	NR	67.6 67.4	37.8 37.0	10.9 10.8	2.2 3.1	**Fig. S2**

Data are mean (SD) unless otherwise stated.

[a]Mean value.

[b]5-year cumulative incidence for death.

AHEI-2010, Alternative Healthy Eating Index-2010; aMed, Alternate Mediterranean diet; CKD, chronic kidney disease; CVD, cardiovascular disease; DASH, Dietary Approaches to Stop Hypertension; eGFR, estimated glomerular filtration rate; HEI-2015, Healthy Eating Index-2015; HF, heart failure; MPO, myeloperoxidase; NA, not applicable; NR, not reported; Obs, observational study; RCT, randomized controlled trial.

Table S2 Calibration questions used in the expert elicitation, and true answers and individual and group mean expert P50 responses to the calibration questions

Calibration questions		Individual responses						Group mean response	True answer	Reference
		Expert 1	Expert 2	Expert 3	Expert 4	Expert 5	Expert 6			
1	In the US, what is the expected remaining years of life for a male ESRD patient aged between 45–49 (2018)?	10	14	10	4	10	10	9.7	9.4	[33]
2	In the US, what percentage of new ESRD patients had a primary diagnosis of diabetes?	50	50	40	40	40	50	45	47	[34]
3	In the US, what percentage of adults have hypertension (applying the criteria from the American College of Cardiology (ACC) and American Heart Association (AHA) 2017)?	27	55	10	18	30	30	28.3	45	[35]
4	In the US, what is the prevalence (%) of obesity amongst non-Hispanic Asian adults (2017–2018)?	10	12	30	30	20	10	18.7	17.4	[36]
5	In the US, how many people were living with a working transplanted kidney in 2017?	210,000	200,000	200,000	70,000	500,000	220,000	233,333.3	222,848	[37]
6	In the US, how many people received a kidney transplant in 2019?	3,200	20,000	15,000	20,000	10,000	2,300	11,750	23,400	[37]
7	In the US, what is the expected remaining years of life for a female ESRD patient aged between 75–79 (2018)?	2	3	3	4	5	2	3.2	3.6	[33]
8	In the UK in 2009–2010, what is the diagnosed prevalence (%) of CKD (stages 3–5) amongst adults?	7.5	10	7	9	4	7.5	7.5	4.3	[38]
9	In the UK, how many adult patients were receiving RRT for ESRD (2018)?	65,000	70,000	600,000	80,000	30,000	65,000	151,667	66,612	[39]
10	In the UK, what proportion (%) of RRT patients are male (2018)?	60	58	55	60	50	60	57.2	61.1	[39]

CKD, chronic kidney disease; ESRD, end-stage renal disease; RRT, renal replacement therapy.

Table S3 Survey questions used in the expert elicitation

Survey questions	
1	What is the 10 years survival percentage of placebo patients in the DAPA-CKD trial (in 10 years mean age is 72)?
2	Given that the 10 years survival percentage is 40%, what is the 20 years survival percentage of placebo patients on the DAPA-CKD trial (in 20 years mean age is 82)?
3	Given that the 10 years survival percentage is 70%, what is the 20 years survival percentage of placebo patients on the DAPA-CKD trial (in 20 years mean age is 82)?

Table S4 Accuracy and information scores and weights for the individual experts derived from the 10 calibration questions

Expert	Accuracy score	Information score	Combined score	Weight
1	0.526	1.818	0.956	0.185
2	0.109	2.930	0.263	0.051
3	0.109	2.563	0.256	0.050
4	0.371	2.715	0.576	0.112
5	0.526	2.894	2.145	0.415
6	0.526	1.861	0.972	0.188

Table S5 Estimates made by the six experts, followed by the unweighted and weighted group estimates for survival percentage at 10 years and at 20 years conditional on survival at 10 years being equal to 40% or 70% for patients in the placebo arm of the DAPA-CKD Trial [32]

	10-year survival (%)			20-year survival (%) conditional on 10-year survival being equal to 40%			20-year survival (%) conditional on 10-year survival being equal to 70%		
	P10	**P50**	**P90**	**P10**	**P50**	**P90**	**P10**	**P50**	**P90**
Expert 1	50	60	75	10	15	25	20	30	40
Expert 2	50	60	75	10	15	30	25	30	70
Expert 3	35	50	65	5	10	15	15	20	25
Expert 4	50	60	75	5	15	30	20	40	60
Expert 5	50	65	80	10	20	30	20	40	50
Expert 6	40	50	70	7	12	25	25	45	55
Unweighted group average	46	58	73	7	15	26	21	34	50
Weighted group average	**47**	**59**	**75**	**8**	**15**	**27**	**19**	**35**	**47**

Table S6 Estimated median survival for Bayesian and frequentist methods, for SMR extrapolations, and for general population life table data

Population	Distribution/source	Median survival (years)		
		Frequentist	Frequentist + GPM	Bayesian
Placebo arm of the DAPA-CKD trial [32]	**Exponential**	22	17	17
	Gamma	12	11	11
	Generalized gamma	10	9	11
	Gompertz	6	6	12
	Loglogistic	14	9	11
	Lognormal	27	14	14
	Weibull	11	10	11
	SMR extrapolation	13		
General population	**Life table data** [1]	22		

CKD, chronic kidney disease; GPM, general population mortality; SMR, standard mortality ratio.

Table S7 Definitions of parameters used in the survival extrapolation

Parameter	Definition
S10	Survival percentage at 10 years
S20	Survival percentage at 20 years
$S20_{S10=40\%}$	Survival percentage at 20 years, assuming the survival percentage at 10 years was 40%
$S20_{S10=70\%}$	Survival percentage at 20 years, assuming the survival percentage at 10 years was 70%
$CS20_{S10=40\%}$	$CS20_{S10=40\%} = S20_{S10=40\%}/S10$
$CS20_{S10=70\%}$	$CS20_{S10=70\%} = S20_{S10=70\%}/S10$
GPM10	Percentage of patients alive at 10 years according to general population mortality information
GPM20	Percentage of patients alive at 20 years according to general population mortality information
PSD10	Percentage of patients alive at 10 years according to one of the seven parametric survival distributions (exponential, gamma, generalized gamma, Gompertz, loglogistic, lognormal, or Weibull)
PSD20	Percentage of patients alive at 20 years according to one of the seven parametric survival distributions (exponential, gamma, generalized gamma, Gompertz, loglogistic, lognormal, or Weibull)

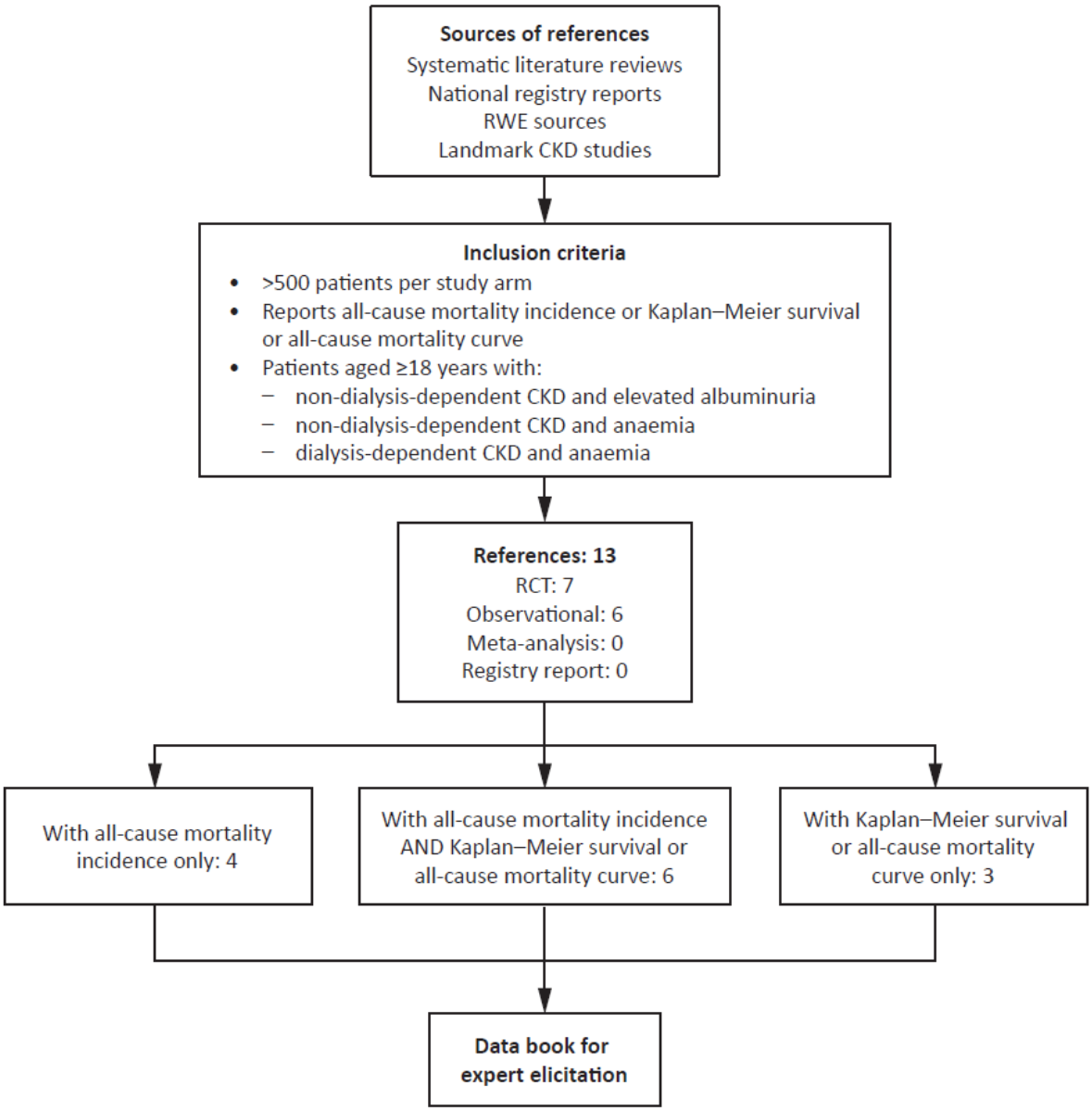
Table S8 Summary of survival distribution functions and parameters used in frequentist, frequentist accounting for general population mortality and Bayesian analyses

Distribution	Functional form	Scale	Shape	Q
Exponential				
Frequentist	Exp(-scale t)	0.032	NA	NA
Frequentist with GPM	Exp(-scale t)	0.020	NA	NA
Bayesian[a]	Exp(-scale t)	0.020/0.020	NA	NA
Gamma				
Frequentist[b]	(1 - pgamma(t, scale, sigma))	0.09	1.45	NA
Frequentist with GPM[b]	(1 - pgamma(t, scale, sigma)) × GP(t)	0.11	1.76	NA
Bayesian[a,b]	(1 - gamma_cdf(t, scale, sigma)) × GP(t)	0.10/0.11	1.75/1.77	NA
Generalized Gamma				
Frequentist[b]	1 - pgengamma(t, scale, sigma, Q)	2.57	0.37	1.94
Frequentist with GPM[b]	(1 - pgengamma(t, scale, sigma, Q)) × GP(t)	2.54	0.30	1.99
Bayesian[a,b]	(1 - pgengamma(t, scale, sigma, Q)) × GP(t)	2.86/2.83	0.67/0.65	1.03/0.96
Gompertz				
Frequentist	exp(-scale/shape (exp(shape t) - 1))	0.018	0.45	NA
Frequentist with GPM	exp(-scale/shape (exp(shape t) - 1)) × GP(t)	0.018	0.45	NA
Bayesian[a]	exp(-scale/shape (exp(shape t) - 1)) × GP(t)	0.017/0.017	0.12/0.12	NA
Loglogistic				
Frequentist	$1/(1 + (t/scale)^{shape})$	13.8	1.43	NA
Frequentist with GPM	$(1/(1 + (t/scale)^{shape})) \times GP(t)$	11.1	1.68	NA
Bayesian[a]	$(1/(1 + (t/scale)^{shape})) \times GP(t)$	14.5/13.7	1.70/1.71	NA
Lognormal				
Frequentist	1 - Φ((ln t - ln scale)/shape)	26.7	1.67	NA
Frequentist with GPM	(1 - Φ((ln t - ln scale)/shape)) × GP(t)	26.8	1.45	NA
Bayesian[a]	(1 - Φ((ln t - ln scale)/shape)) × GP(t)	24.8/24.8	1.4/1.4	NA
Weibull				
Frequentist	$S = \exp(-(t/scale)^{shape})$	14.7	1.41	NA
Frequentist with GPM[c]	$S = \exp(-scale\ t^{shape}) \times GP(t)$	0.011	1.69	NA
Bayesian[a]	$S = \exp(-(t/scale)^{shape}) \times GP(t)$	17.6/17.0	1.54/1.54	NA

[a]Bayesian analyses: instead of using the parameters, figures are based on mean survival percentages over time across MCMC iterations. Provided parameters are the mean parameters/median parameters.
[b]pgamma and pgengama are built in functions for R.
[c]WeibullPH used.
GPM, general population mortality; MCMC, Markov chain Monte Carlo; NA, not applicable.

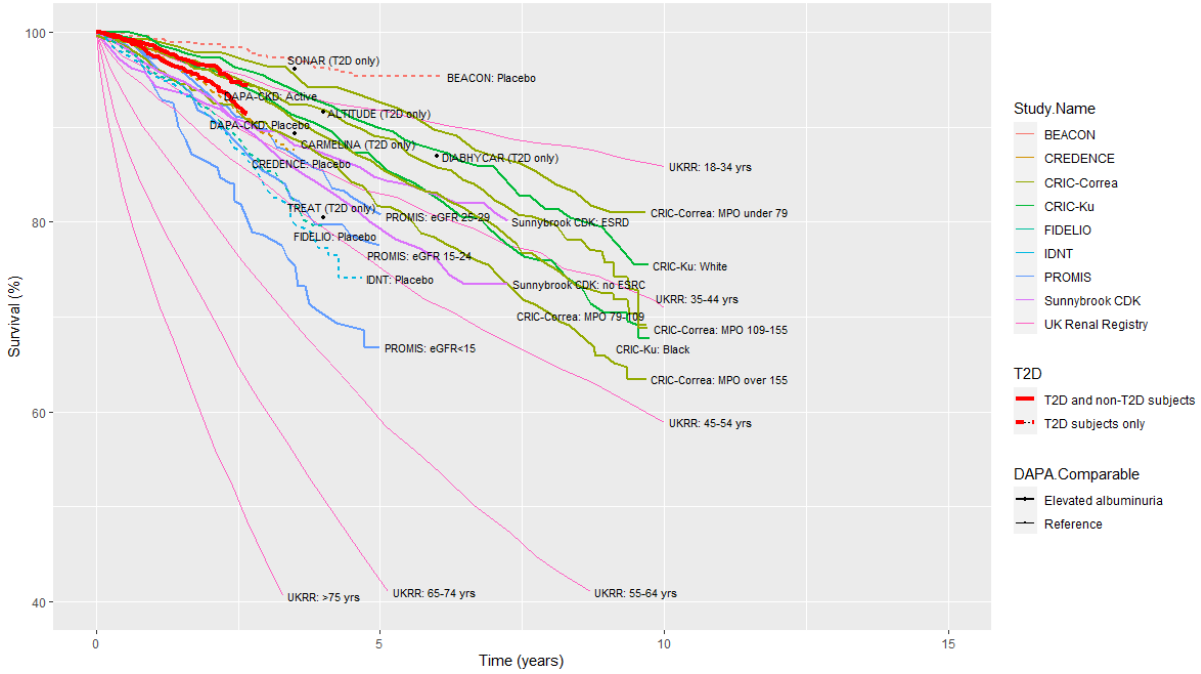
Supplementary figures

Fig. S1 Summary of literature search and screening process



CKD, chronic kidney disease; RCT, randomized control trial; RWE, real-world evidence.

Fig. S2 Kaplan–Meier estimate figure from the data book: Kaplan–Meier estimates from studies involving patients with non-dialysis-dependent CKD with elevated albuminuria. Results from the UK Renal Registry were included for reference [39]



CKD, chronic kidney disease; eGFR, estimated glomerular filtration rate (mL/min/1.73 m^2); MPO, myeloperoxidase (pmol/L); T2D, type 2 diabetes; UKRR, UK Renal Registry; yrs, years.

Fig. S3 Screen capture of the Microsoft Excel-based elicitation tool

What is the 10 years survival percentage of placebo patients in the DAPA-CKD trial (in 10 years mean age is 72)?

What value defines the 10th percentile, i.e. I am 90% confident that the true value is higher than this value

What value defines the 50th percentile, i.e. I am 50% confident that the true value is higher than this value

What value defines the 90th percentile, i.e. I am 90% confident that the true value is lower than this value

Inclusion criteria:
eGFR: 25–75 mL/min/1.73m2
UACR: 200–5000 mg/g
With and without T2D
Receiving ACEi/ARB
Exclusion criteria:
T1D
Polycystic kidney disease lupus nephritis, or antineutrophil cytoplasmic antibody-associated vasculitis
Key patient characteristics (dapagliflozin):
Female, %: 32.9
Mean age (SD), years: 61.8 (12.1)
Mean eGFR (SD), mL/min/1.73m2: 43.2 (12.3)
Median UACR (IQR), mg/g: 965 (472–1903)
With T2D, %: 67.6
History of CVD, %: 37.8

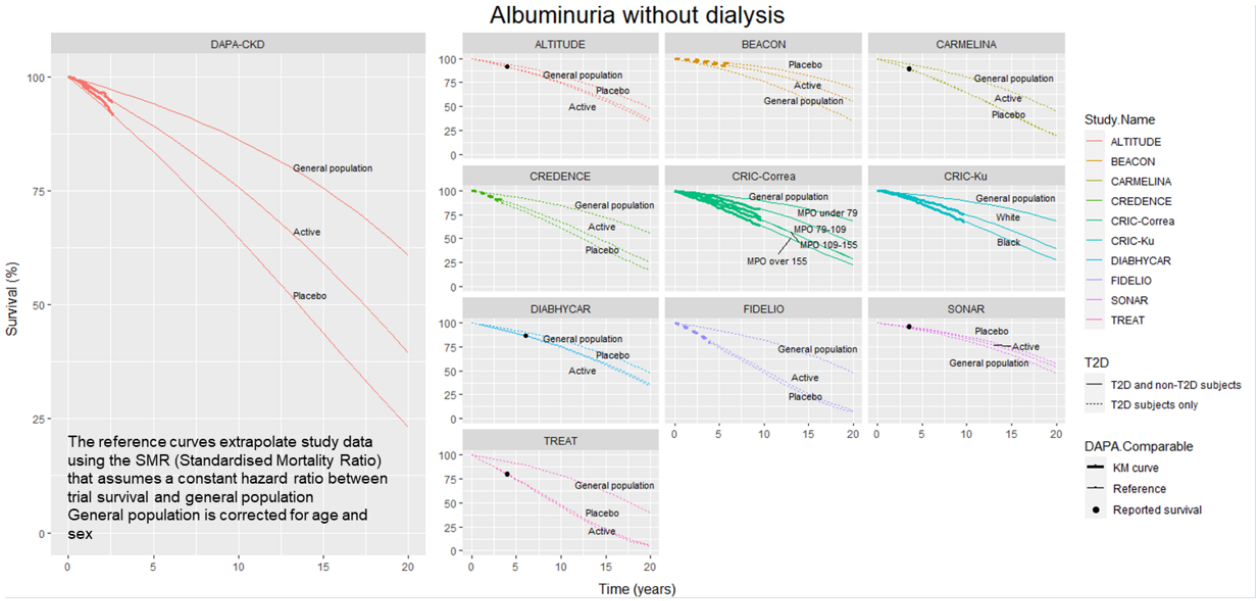
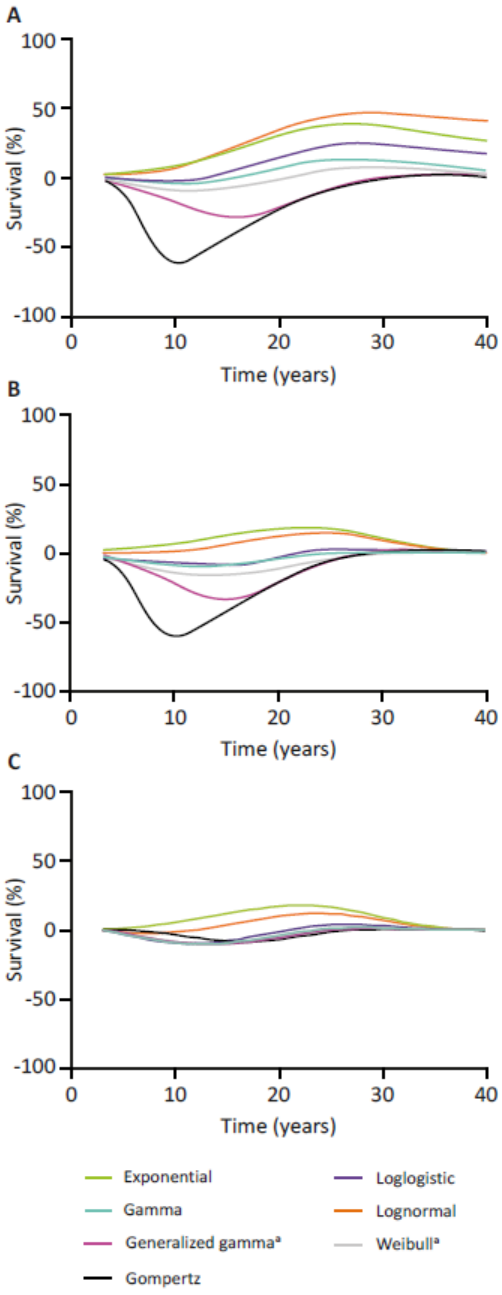


Fig. S4 Relative differences between exponential, gamma, generalized gamma, Gompertz, loglogistic, lognormal, and Weibull distributions for (A) frequentist, (B) frequentist accounting for general population mortality, and (C) Bayesian analyses



Relative difference estimates are calculated as the difference between each survival distribution and the Kaplan–Meier survival estimate from the DAPA-CKD placebo arm extrapolated by calculating standard mortality ratios using age- and sex-adjusted general-population life table data (United States Life Tables 2017, US Department of Health and Human Services) [1].

[a]Results for the generalized gamma and Weibull distributions for the Bayesian analysis overlap.

Supplementary references

1. Arias E, Xu J: United States life tables, 2017 vol 68 no 7. National Vital Statistics Reports. National Center for Health Statistics, Hyattsville, MD (2019).
2. Savage LJ. The Foundations of Statistics. New York: Wiley, 1954.
3. Dalkey NC. The Delphi Method: an Experimental Study of Group Opinion. Santa Monica, CA: RAND Corporation, 1969.
4. 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.
5. Spetzler CS, Staël Von Holstein CAS. Probability encoding in decision analysis. Manag Sci. 1975;22:340–358.
6. Cooke RM, Goossens LJH. Procedures Guide for Structured Expert Judgment. TU Delft: Nuclear Science and Technology, EUR 18820, 1999.
7. Gosling JP. SHELF: the Sheffield Elicitation Framework. Cham, Switzerland: Springer, 2018.
8. O'Hagan A. Expert knowledge elicitation: subjective but scientific. Am Stat. 2019;73:69–81.
9. Chevret S. Bayesian adaptive clinical trials: a dream for statisticians only? Stat Med. 2012;31(11–12):1002–1013.
10. Baldi I, Gregori D, Desideri A, et al. Accrual monitoring in cardiovascular trials. Open Heart. 2017;4(2):e000720.
11. Lesaffre E, Baio G, Boulanger B. Bayesian Methods in Pharmaceutical Research. Chapman and Hall, 2020.
12. Dallow N, Best N, Montague TH. Better decision making in drug development through adoption of formal prior elicitation. Pharm Stat. 2018;17(4):301–316.
13. National Academies of Sciences, Engineering, and Medicine. Valuing Climate Damages: Updating Estimation of the Social Cost of Carbon Dioxide. Washington, DC: The National Academies Press, 2017.
14. 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.
15. 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.
16. Cooke RM. Experts in Uncertainty – Opinion and Subjective Probability in Science. New York: Oxford University Press, 1991.
17. Morgan M, Henrion M. Uncertainty: a Guide to Dealing with Uncertainty in Quantitative Risk and Policy Analysis. Cambridge University Press, 1990.
18. Colson AR, Cooke RM. Expert elicitation: using the classical model to validate experts' judgments. Rev Environ Econ Policy. 2018;12(1):113–132.
19. Tversky A, Kahneman D. Judgment under uncertainty: heuristics and biases. Science. 1974;185:1124–1131.
20. Brenner BM, Cooper ME, de Zeeuw D, et al. Effects of losartan on renal and cardiovascular outcomes in patients with type 2 diabetes and nephropathy. N Engl J Med. 2001;345(12):861–869.
21. Lewis EJ, Hunsicker LG, Clarke WR, et al. Renoprotective effect of the angiotensin-receptor antagonist irbesartan in patients with nephropathy due to type 2 diabetes. N Engl J Med. 2001;345(12):851–860.

22. Wright JT, Jr., Bakris G, Greene T, et al. Effect of blood pressure lowering and antihypertensive drug class on progression of hypertensive kidney disease: results from the AASK trial. JAMA. 2002;288(19):2421–2431.
23. Levin A, Djurdjev O, Beaulieu M, et al. Variability and risk factors for kidney disease progression and death following attainment of stage 4 CKD in a referred cohort. Am J Kidney Dis. 2008;52(4):661–671.
24. de Zeeuw D, Akizawa T, Audhya P, et al. Bardoxolone methyl in type 2 diabetes and stage 4 chronic kidney disease. N Engl J Med. 2013;369(26):2492–2503.
25. Sud M, Tangri N, Levin A, et al. CKD stage at nephrology referral and factors influencing the risks of ESRD and death. Am J Kidney Dis. 2014;63(6):928–936.
26. Orlandi PF, Fujii N, Roy J, et al. Hematuria as a risk factor for progression of chronic kidney disease and death: findings from the Chronic Renal Insufficiency Cohort (CRIC) study. BMC Nephrol. 2018;19(1):150.
27. Perkovic V, Jardine MJ, Neal B, et al. Canagliflozin and renal outcomes in type 2 diabetes and nephropathy. N Engl J Med. 2019;380(24):2295–2306.
28. Bakris GL, Agarwal R, Anker SD, et al. Effect of finerenone on chronic kidney disease outcomes in type 2 diabetes. N Engl J Med. 2020;383:2219–2229.
29. Correa S, Pena-Esparragoza JK, Scovner KM, et al. Myeloperoxidase and the risk of CKD progression, cardiovascular disease, and death in the Chronic Renal Insufficiency Cohort (CRIC) Study. Am J Kidney Dis. 2020;76(1):32–41.
30. Hu EA, Coresh J, Anderson CAM, et al. Adherence to healthy dietary patterns and risk of CKD progression and all-cause mortality: findings from the CRIC (Chronic Renal Insufficiency Cohort) study. Am J Kidney Dis. 2021;77(2):235–244.
31. Ku E, Yang W, McCulloch CE, et al. Race and mortality in CKD and dialysis: findings from the Chronic Renal Insufficiency Cohort (CRIC) study. Am J Kidney Dis. 2020;75(3):394–403.
32. Heerspink HJL, Stefansson BV, Correa-Rotter R, et al. Dapagliflozin in Patients with Chronic Kidney Disease. N Engl J Med. 2020;383(15):1436–1446.
33. United States Renal Data System. 2020 USRDS Annual Data Report: Epidemiology of kidney disease in the United States. Bethesda, MD: National Institutes of Health, National Institute of Diabetes and Digestive and Kidney Diseases, 2020.
34. Kidney Disease: The Basics. National Kidney Foundation. Available from: https://www.kidney.org/news/newsroom/fsindex. Accessed 2 August 2021.
35. Estimated hypertension prevalence, treatment, and control among US adults. Million Hearts. Available from: https://millionhearts.hhs.gov/data-reports/hypertension-prevalence.html. Accessed 2 August 2021.
36. Adult Obesity Facts. Centers for Disease Control and Prevention. Available from: https://www.cdc.gov/obesity/data/adult.html. Accessed 2 August 2021.
37. Chronic Kidney Disease in the United States, 2021. Centers for Disease Control and Prevention. Available from: https://www.cdc.gov/kidneydisease/publications-resources/ckd-national-facts.html. Accessed 2 August 2021.
38. Chronic Kidney Disease in England: The Human and Financial Cost. NHS Kidney Care. Available from: https://www.england.nhs.uk/improvement-hub/publication/chronic-kidney-disease-in-england-the-human-and-financial-cost/. Accessed 2 August 2021.
39. UK Renal Registry. UK Renal Registry 22nd Annual Report – data to 31/12/2018. Bristol, UK: The Renal Association, 2020.