








# Development and Validation of a New Hierarchical Composite End Point for Clinical Trials of Kidney Disease Progression

Hiddo L. Heerspink <sup>1,2</sup> Niels Jongs <sup>1</sup> Patrick Schloemer,<sup>3</sup> Dustin J. Little <sup>4</sup>,  
Meike Brinker <sup>5</sup> Christoph Tasto,<sup>5</sup> Martin Karpefors <sup>6</sup> David C. Wheeler <sup>2,7</sup>,  
George Bakris <sup>8</sup> Vlado Perkovic,<sup>2,9</sup> Richard Nkulikiyinka,<sup>3</sup> Jerome Rossert,<sup>4</sup>  
and Samvel B. Gasparyan <sup>6</sup>

Due to the number of contributing authors, the affiliations are listed at the end of this article.

## ABSTRACT

**Background** The established composite kidney end point in clinical trials combines clinical events with sustained large changes in GFR. However, the statistical method does not weigh the relative clinical importance of the end point components. A HCE accounts for the clinical importance of the end point components and enables combining dichotomous outcomes with continuous measures.

**Methods** We developed and validated a new HCE for kidney disease progression, performing *post hoc* analyses of seven major Phase 3 placebo-controlled trials that assessed the effects of canagliflozin, dapagliflozin, finerenone, atrasentan, losartan, irbesartan, and aliskiren in patients with CKD. We calculated the win odds (WOs) for treatment effects on a kidney HCE, defined as a hierarchical composite of all-cause mortality; kidney failure; sustained 57%, 50%, and 40% GFR declines from baseline; and GFR slope. The WO describes the odds of a more favorable outcome for receiving the active compared with the control. We compared the WO with the hazard ratio (HR) of the primary kidney outcome of the original trials.

**Results** In all trials, treatment effects calculated with the WO reflected a similar direction and magnitude of the treatment effect compared with the HR. Clinical trials incorporating the HCE would achieve increased statistical power compared with the established composite end point at equivalent sample sizes.

**Conclusions** In seven major kidney clinical trials, the WO and HR provided similar direction of treatment effect estimates with smaller HRs associated with larger WOs. The prioritization of clinical outcomes and inclusion of broader composite end points makes the HCE an attractive alternative to the established kidney end point.

JASN 00: 1–14, 2023. doi: <https://doi.org/10.1681/ASN.0000000000000243>

This is an open access article distributed under the terms of the [Creative Commons Attribution-Non Commercial-No Derivatives License 4.0 \(CCBY-NC-ND\)](https://creativecommons.org/licenses/by-nc-nd/4.0/), where it is permissible to download and share the work provided it is properly cited. The work cannot be changed in any way or used commercially without permission from the journal.

## INTRODUCTION

Kidney failure, which requires dialysis or kidney transplantation, is the most significant long-term complication of CKD for clinicians, patients, and their caregivers. As such, clinical trials aiming to develop new therapies for CKD have traditionally used kidney failure as a component of a composite

**Received:** April 19, 2023 **Accepted:** September 20, 2023.

**Published Online Ahead of Print:** October 24, 2023.

**Correspondence:** Prof. Hiddo L. Heerspink, Department of Clinical Pharmacy and Pharmacology, University of Groningen, Hanzeplein 1, PO Box 30 001 Groningen, The Netherlands. Email: [h.j.lambers.heerspink@umcg.nl](mailto:h.j.lambers.heerspink@umcg.nl)

Copyright © 2023 The Author(s). Published by Wolters Kluwer Health, Inc. on behalf of the American Society of Nephrology.

end point together with doubling of serum creatinine, equivalent to 57% GFR decline, which reflects a large decline in kidney function.<sup>1,2</sup> These end points only occur after a prolonged disease course that may extend 10–20 years, a time-frame which is not feasible for clinical trials. Surrogate end points that reliably reflect established disease outcomes could facilitate the conduct of clinical trials with smaller sample size and shorter duration. Progress in the validation of surrogate end points has led to the inclusion of smaller declines in GFR than 57% as a component of a composite kidney end point and the use of the rate of GFR decline (GFR slope) in some settings for full drug approval.<sup>3–5</sup>

The conventional method to assess treatment effects in clinical trials of CKD progression is to define the end point as the time to the first event of the composite outcomes without taking into account the severity or clinical importance of that first event (analyzed using Kaplan–Meier estimates, log-rank tests, or Cox proportional hazards models). This is particularly important when the effects on the different components vary or when the components of less clinical effect occur earlier. For example, a participant experiencing a 50% reduction in GFR decline after 9 months is considered to have reached the composite end point. The clinically more impactful event (*e.g.*, requirement for dialysis or kidney transplantation), which may occur later, is ignored in the primary analysis. Moreover, a participant reaching 50% GFR decline after 9 months is considered to have had a worse outcome than another participant reaching dialysis after 11 months. Thus, the components of the composite end point receive equal weight in the analysis, irrespective of their clinical importance. Another issue with the conventional kidney end points is that the estimated effect of an intervention is determined by the number of patients who reach the outcomes included in the composite end point. In clinical trials in nephrology, these are patients with a faster progression of kidney disease. Other patients who do not experience a sustained large (*e.g.*, 50%) decline in kidney function only contribute exposure and time at risk to the analysis. Ideally, an end point should capture the effect of the intervention in all trial participants. The GFR slope provides an estimate of the effect of an intervention in all participants, both fast and slow progressors, even when they do not experience events included in the conventional composite outcome. However, the conventional composite clinical end point cannot incorporate such a continuous quantitative measure.

To overcome some of these limitations, new approaches for the analysis of composite end points are emerging which take into account the prioritization of the severity of the components and combining dichotomous end points and quantitative (continuous) measures.<sup>6</sup> The flexibility of such new end points, in particular the combination of different types of outcomes and the hierarchical structure of the end point components, makes them an attractive alternative to the established kidney end point. We refer to the accompanying review for more background details of hierarchical

### Significance Statement

The established composite kidney end point in clinical trials combines clinical events with sustained large changes in GFR but does not weigh the relative clinical importance of the end point components. By contrast, a hierarchical composite end point (HCE) accounts for the clinical importance of the end point components. The authors developed and validated a kidney HCE that combines clinical kidney outcomes with longitudinal GFR changes (GFR slope). They demonstrate that in seven major placebo-controlled kidney outcome trials with different medications, treatment effect estimates on the HCE were consistently in similar directions and of similar magnitudes compared with treatment effects on the established kidney end point. The HCE's prioritization of clinical outcomes and ability to combine dichotomous outcomes with GFR slope make it an attractive alternative to the established kidney end point.

composite end point (HCE). In brief, the novel HCE is analyzed using win odds (WOs), which describes the odds of a patient receiving the active treatment of having a more favorable outcome compared with a patient treated with control. For both hazard ratio (HR) and WO, a value of 1 corresponds with the null hypothesis of no treatment effect. However, unlike for HR where a result of  $<1$  is indicative of a favorable treatment effect, a WO of  $>1$  corresponds with a treatment benefit, by signifying that treated patients are more likely to have a favorable outcome, compared with control patients. The aim of this study was to develop and validate a novel kidney HCE using a WO approach.

## METHODS

### Overall Study Design

In this *post hoc* analysis, we used data from completed Phase 3 placebo-controlled randomized clinical trials that assessed the efficacy and safety of two sodium–glucose cotransport 2 inhibitors, a nonsteroidal mineralocorticoid receptor antagonist, an endothelin receptor antagonist, two angiotensin receptor blockers, and a direct renin inhibitor on composite end points of kidney failure or death due to kidney disease with GFR decline thresholds of 40%, 50%, and 57% (as prespecified in each trial<sup>7–13</sup>). We selected these clinical trials because they recruited patients with CKD and demonstrated varying effects on the primary composite kidney end point. We included the Dapagliflozin and Prevention of Adverse Outcomes in CKD (DAPA-CKD) (ClinicalTrials.gov [NCT03036150](#)), Canagliflozin and Renal Events in Diabetes with Established Nephropathy Clinical Evaluation (CREDESCO) ([NCT02065791](#)), Finerenone in reducing kidney failure and disease progression in Diabetic Kidney Disease (FIDELIO-DKD) ([NCT02540993](#)), Study Of diabetic Nephropathy with AtRasentan (SONAR) ([NCT01858532](#)), Reduction of Endpoints in Non-insulin-dependent diabetes mellitus with the Angiotensin II Antagonist Losartan (RENAL)

(NCT00308347), Irbesartan Diabetic Nephropathy Trial (IDNT) (NCT00317915), and ALTITUDE (NCT00549757) trials.

### End Point Definitions

We compared treatment effects on the established kidney end point as defined in each trial with the hierarchical composite kidney end point. The definitions of the established kidney end points in each trial are shown in [Supplemental Table 1](#). The defined hierarchical composite kidney end point accounts for the clinical effect of events. We defined the HCE as a composite end point including seven components which we ranked in order of highest to lowest effect as (1) all-cause mortality; (2) kidney replacement therapy defined as dialysis for at least 28 days or kidney transplantation; (3) sustained GFR <15 ml/min per 1.73 m<sup>2</sup> for at least 28 days; (4) sustained GFR decline from baseline for at least 28 days of 57%; (5) 50%; (6) 40%; or (7) GFR slope. In a [Supplemental Analysis](#), we assessed treatment effects on a kidney-specific HCE which was defined in the same way as the primary HCE without inclusion of all-cause mortality. In this [Supplemental Analysis](#), patients who died contributed to the analysis with their event of highest priority before dying.

### Statistical Analyses

The patients were analyzed using the intention-to-treat principle: All patients were followed and analyzed irrespective of their compliance to the planned course of treatment and included in the analysis as randomized. Hence, the dichotomous outcomes occurring during the 36 months of follow-up were included in the analysis irrespective of treatment discontinuation. The follow-up duration varied among trials. Therefore, not every patient had 36 months of follow-up ([Table 1](#)). To account for the variable follow-up when constructing the HCE, we extended the follow-up for patients with a shorter follow-up by using the clinically most important outcome from the observed follow-up for the analysis on month 36.

We used proportional hazards (Cox) regression models to assess the effect of the active intervention compared with placebo on the risk for first relevant composite kidney end point. We stratified the Cox models for factors used at randomization and adjusted for covariates as originally defined in each clinical trial.

To estimate treatment effect on total GFR slope, we used a two-slope mixed-effects model accounting for acute and chronic phase of each trial, where the acute phase was the period up to the first postrandomization visit when the acute treatment effect on GFR was considered fully present.<sup>14</sup> The model adjusts for baseline GFR and accounts for different sources of variation in GFR between and within participants and treatment arms. Only on-treatment observations were selected for analysis of GFR slope to avoid potential bias in GFR slope which may result because of early discontinuation of treatments with acute reversible effects in GFR. Patients not experiencing any of the dichotomous events defined in the

hierarchy contributed to the analysis with their individual GFR slope obtained from this two-slope model.

The HCE is analyzed using WOs,<sup>15</sup> an adaptation of win ratio<sup>16</sup> to include ties (a tie is considered a half loss and a half win for each group).<sup>17</sup> Every patient in the active group is compared with every patient in the control group, and the patient with an event of a higher priority (more severe) loses against the other patient. The hierarchical comparison of the components of the kidney HCE is shown in [Supplemental Table 2](#). After all possible comparisons are completed, the total number of the wins of the active treatment, the total number of losses, and the total number of ties are used to derive win statistics. The WOs is defined as total number of wins plus half of the ties divided by the total number of losses and half of the ties. It should be noted that for the proposed kidney HCE, the proportion of ties is negligible because of the use of timing of events and continuous GFR slope, and hence, WOs is essentially equal to win ratio. To account for the differential follow-up times between patients, we performed a [Supplemental Analysis](#) where the shared follow-up of two patients was used to select the outcome with the highest priority in a pairwise comparison of patients.<sup>16</sup> We calculated WOs and its 95% confidence interval (CI).<sup>18</sup>

Maraca plots were developed to visualize HCE combining multiple time-to-event outcomes with a single continuous outcome.<sup>19</sup> A maraca plot is formed by end-to-end adjoining, from left to right by declining severity of uniformly scaled Kaplan–Meier plots of times to each dichotomous outcome among those without more severe outcomes, with superimposed boxplot of the continuous outcome. The maraca plot visualizes the contribution of components of an HCE over time.

## RESULTS

Patient characteristics of the participants in each clinical trial are shown in [Table 1](#). Mean age ranged between 59 and 66 years, mean GFR between 39 and 57 ml/min per 1.73 m<sup>2</sup>, and median urinary albumin/creatinine ratio between 283 and 1354 mg/g. In all clinical trials, baseline characteristics were well balanced across randomized patient groups.<sup>7–13</sup>

### Contributions of Individual Components to the Composite End Point

The components of the original primary kidney end point in each trial are shown in [Supplemental Table 3](#). Declines in GFR of 40%, 50%, or 57% were the most common components of the primary composite kidney end point in each trial. For the HCE, all-cause mortality and 40% eGFR decline were the most common components ([Figure 1](#)). In comparing the original primary kidney end point from each trial with the HCE, the latter included more kidney failure events because all such events were included in patients who did not die during the observation period ([Table 2](#) and [Supplemental Table 3](#)).

In the DAPA-CKD, CREDENCE, FIDELIO-DKD, and SONAR trials, 15%–20% of participants experienced one of the time-to-event outcomes (death, kidney failure, or a 57%, 50%, or 40% GFR decline) during the follow-up. GFR slope contributed to the end point in the remaining 80%–85% of participants in each trial. By contrast, in the RENAAL and IDNT trials, approximately 50% of all participants died, experienced kidney failure, or an end point based on the different GFR thresholds.

### Comparison of HR and WOs

The effects of the interventions on each component of the HCE, analyzed with Cox proportional hazards regression, were broadly consistent except that in the SONAR, RENAAL, and IDNT trials; the interventions reduced the risks of the kidney-related components but did not reduce the risk of all-cause mortality (Table 2). The effects on GFR slope in all trials were directionally similar when compared with the effects on kidney end points (Table 2).

The WOs in the six clinical trials that had shown a reduction in the risk of the primary composite kidney end point ranged from 1.13 to 1.41 indicating a more favorable outcome for a patient assigned to active compared with placebo treatment (Table 2). In the ALTITUDE trial, which showed no risk reduction of the kidney end point, the WOs was <1 (0.84), also indicating no benefit.

To visualize treatment effects for hierarchical composite kidney end points, we developed maraca plots (Figure 2). These plots can be used to visualize an HCE combining multiple time-to-event outcomes and a single continuous outcome. The maraca plot shows the cumulative percentages of patients experiencing each dichotomous outcome during the fixed follow-up period, among those who avoid worse outcomes during that period combined with a box-whisker plot showing the median and 25th–75th percentiles of the continuous GFR slope distribution for patients without a dichotomous event in each treatment group. The maraca plot demonstrates fewer dichotomous outcomes in the active compared to placebo group in all trials except for the ALTITUDE trial which did not demonstrate efficacy of active treatment. The box-whisker component demonstrates that the median rate of GFR decline for patient not experiencing dichotomous outcomes was slower (shift to the right in the maraca plot) in the active compared with the placebo group in all trials except for the ALTITUDE trial.

In comparing the HCE with the original primary kidney trial end points, we observed similar directions and magnitudes of the treatment effect estimates (Figures 3 and 4 and Table 2). For example, in DAPA-CKD, the HR for the primary outcome of sustained 50% GFR decline, kidney failure, or renal death was 0.61 (95% CI, 0.51 to 0.73). The WOs for the HCE was 1.41 (95% CI, 1.32 to 1.52; Figures 3 and 4 and Table 2). Similarly, in FIDELIO-DKD, the HR for the primary outcome of sustained 40% GFR decline, kidney failure, or renal death was 0.82 (95% CI, 0.73 to 0.93), and the WO was 1.26 (95%

CI, 1.19 to 1.34). Removing all-cause mortality from the kidney HCE did not substantially alter the results, but in some trials led to numerically higher WOs (Supplement Figures 1 and 2). The WOs from a shared follow-up approach demonstrated similar results compared with our main analyses (Supplemental Table 4), which supports the robustness of our findings. For example, the WOs using a shared follow-up time in the DAPA-CKD trial was 1.42 (95% CI, 1.32 to 1.52) versus 1.41 (95% CI, 1.32 to 1.52) in our main analysis (Supplemental Table 5).

### Sample Size

Figure 5 compares the sample size requirements and statistical power of the novel HCE using bootstrap resampling of clinical trials with the original primary kidney end point and GFR slope to detect the observed treatment effect for each end point. The resampling procedure used 1000 iterations at each sample size ( $n=200, 500$ , increments of 500 until 3000). In four of the six trials that reported a benefit of the examined intervention (except the RENAAL and IDNT trials), the sample size requirements using the HCE were smaller compared with the original kidney end point as defined in each trial. The ALTITUDE trial was not included in Figure 5 because we did not consider that future trials will be powered to detect a nonbeneficial treatment effect. When all-cause mortality was excluded from the HCE, the sample size requirements using the HCE were smaller than those for the original kidney end point in all trials (Supplemental Figure 3).

## DISCUSSION

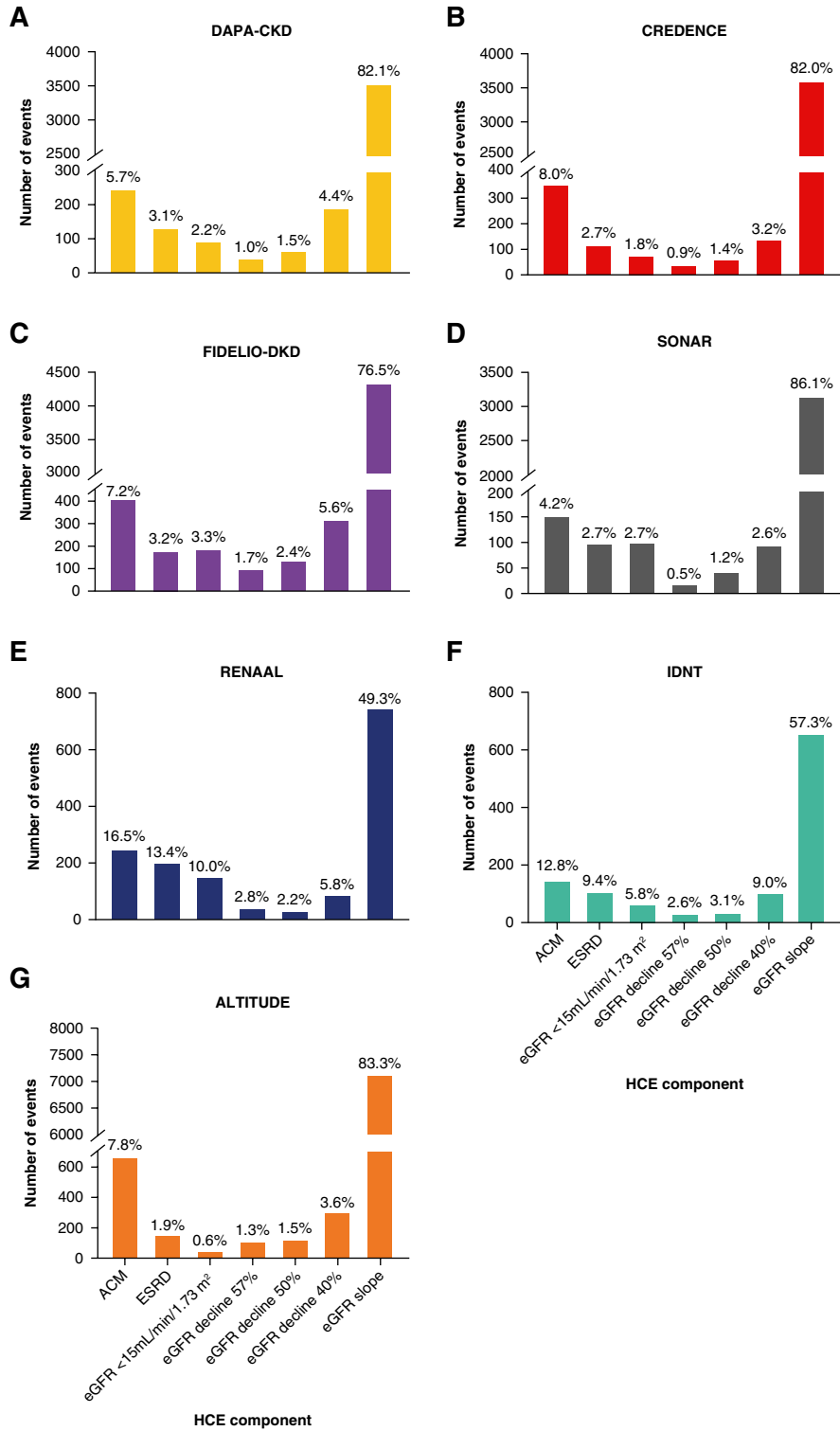
HCEs are flexible end points that provide a clinically meaningful measure of a patient's condition throughout the follow-up of a clinical trial as opposed to the traditional time-to-event end point that prioritizes the first event during the follow-up period, without taking into account its clinical importance relative to those of other and potentially later occurring components of the composite outcome. HCEs have been used in various therapeutic domains (*e.g.*, heart failure<sup>20</sup> and coronavirus disease 2019<sup>21</sup>), but their use in nephrology clinical trials is uncommon. We developed a novel hierarchical composite kidney end point and demonstrated that in various kidney outcome trials with different interventions, the WOs analysis provided estimates that were directionally consistent with and of similar magnitudes to estimates derived from time-to-first event Cox analysis. Use of the HCE increased statistical power compared with conventional time-to-first event analysis. The newly defined hierarchical composite kidney end point prioritizes clinically impactful outcomes and seems to be more sensitive to detect treatment effects than the traditional time-to-event CKD end points as would be expected because it more effectively incorporates information on disease progression at all levels.

**Table 1. Baseline characteristics**

Characteristic	DAPA-CKD (N=4304)	CREDESCENCE (N=4401)	FIDELIO-DKD (N=5674)	SONAR (N=3668)	RENAAL (N=1513)	IDNT (N=1715)	ALTITUDE (N=8561)
Enrollment period	2016–2018	2014–2017	2015–2018	2013–2018	1996–1999	1996–1999	2007–2010
Median follow-up, yr	2.4	2.6	2.6	2.2	3.4	2.6	2.7
Characteristic							
Age, yr (SD)	61.8 (12)	63.0 (9)	65.6 (9)	64.5 (8.8)	60.2 (7.4)	58.8 (7.7)	64.5 (9.8)
Female sex, N (%)	1425 (33.1)	494 (33.9)	1681 (29.8)	946 (25.8)	557 (36.8)	367 (32)	2735 (31.9)
Race, n (%)							
Asian	1467 (34.1)	877 (19.9)	1440 (25.4)	1198 (32.7)	252 (16.7)	51 (4.4)	2714 (31.7)
Black	191 (4.4)	224 (5.1)	264 (4.7)	224 (6.1)	230 (15.2)	141 (12.3)	277 (3.2)
Other	356 (8.3)	369 (8.4)	378 (6.7)	136 (3.7)	296 (19.7)	103 (9.0)	696 (8.1)
White	2290 (53.2)	2931 (66.6)	3592 (63.3)	2110 (57.5)	735 (48.6)	853 (74.3)	4850 (56.7)
BP, mm Hg (SD)							
Systolic	137.1 (17)	140.0 (16)	138.0 (14)	133.3 (15)	152.5 (19.3)	159.4 (20.0)	137.3 (16)
Diastolic	77.5 (11)	78.3 (9)	76.0 (10)	71.5 (10)	82.4 (10.4)	86.9 (11.4)	74.2 (10)
Body weight, kg (SD)	81.7 (21)	87.0 (20.7)	87.2 (20)	85.7 (20)	82.2 (20.7)	87.3 (18.8)	82.7 (19.4)
Hba1c, % (SD)	7.06 (1.7)	8.3 (1.3)	7.7 (1.3)	7.8 (1.5)	8.48 (1.6)	8.1 (1.7)	7.79 (1.6)
GFR, ml/min per 1.73 m <sup>2</sup> (SD)	43.1 (12)	56.2 (18)	44.3 (13)	42.3 (14)	38.6 (12.4)	47.2 (17.8)	56.9 (22.5)
GFR, ml/min per 1.73 m <sup>2</sup> , n (%)							
≥60	454 (10.5)	1769 (40.2)	656 (11.6)	468 (12.8)	79 (5.2)	265 (23.1)	2783 (32.5)
<60	3850 (89.5)	2632 (59.8)	5018 (88.4)	3191 (87.0)	1434 (94.8)	873 (76.0)	5776 (67.5)
UACR, mg/g (IQR)	949 (477–1885)	927 (463–1833)	852 (446–1634)	828 (458–1556)	1245.5 (558–2544)	1354 (1054, 1748)	283 (56–889)
UACR, mg/g, n (%)							
>1000	2079 (48.3)	2053 (46.7)	2480 (43.7)	892 (24.5)	857 (56.6)	912 (79.4)	1904 (22.2)
≤1000	2225 (51.7)	2348 (53.3)	3191 (56.2)	2771 (75.5)	656 (43.4)	224 (19.5)	6527 (76.2)
Baseline medications, n (%)							
ACEi	1353 (31.4)	1922 (43.7)	1942 (34.2)	1319 (36.0)	737 (48.7)	507 (44.2)	3792 (44.3)
ARB	2870 (66.7)	2480 (56.4)	3725 (65.7)	2391 (65.2)	105 (6.9)	33 (2.9)	4787 (55.9)
Diuretics	1882 (43.7)	2057 (46.7)	3214 (56.6)	3157 (86.1)	878 (58)	547 (47.6)	5872 (68.6)
Insulin	1598 (37.1)	2884 (65.5)	3637 (64.1)	2315 (63.1)	910 (60.1)	644 (56.1)	4850 (56.7)
Statins	2794 (64.9)	3036 (69.0)	4215 (74.3)	2707 (73.8)	507 (33.5)	299 (26.0)	5576 (65.1)

DAPA-CKD, Dapagliflozin and Prevention of Adverse Outcomes in CKD; CREDESCENCE, Canagliflozin and Renal Events in Diabetes with Established Nephropathy Clinical Evaluation; FIDELIO-DKD, Flnerenone in reducing kiDnEy faiLure and dlsease prOgression in Diabetic Kidney Disease; SONAR, Study Of diabetic Nephropathy with AtRasentan; RENAL, Reduction of Endpoints in Non-insulin-dependent diabetes mellitus with the Angiotensin II Antagonist Losartan; IDNT, Irbesartan Diabetic Nephropathy Trial; ALTITUDE, Aliskiren Trial in Type 2 Diabetes Using Cardio-Renal Endpoints; SD, standard deviation; Hba1c, hemoglobin A1c; UACR, urinary albumin/creatinine ratio; IQR, interquartile range; ACEi, angiotensin-converting-enzyme inhibitor; ARB, angiotensin receptor blocker.

\*Mean follow-up duration provided for RENAAL and IDNT.



**Figure 1. Number (%) of individual components of the hierarchical composite kidney end point in each trial.** (A) DAPA-CKD trial. (B) CRENDENCE trial. (C) FIDELIO-DKD trial. (D) SONAR trial. (E) RENAAL trial. (F) IDNT trial. (G) ALTITUDE trial. ACM, all-cause mortality; ALTITUDE, Aliskiren Trial in Type 2 Diabetes Using Cardio-Renal Endpoints; CRENDENCE, Canagliflozin and Renal Events in Diabetes with Established Nephropathy Clinical Evaluation; DAPA-CKD, Dapagliflozin and Prevention of Adverse Outcomes in CKD; FIDELIO-DKD, FInerenone in reducing kiDnEy faiLure and disease prOgression in Diabetic Kidney Disease; IDNT, Irbesartan Diabetic Nephropathy Trial; RENAL, Reduction of Endpoints in Non-insulin-dependent diabetes mellitus with the Angiotensin II Antagonist Losartan; SONAR, Study Of diabetic Nephropathy with AtRasentan.

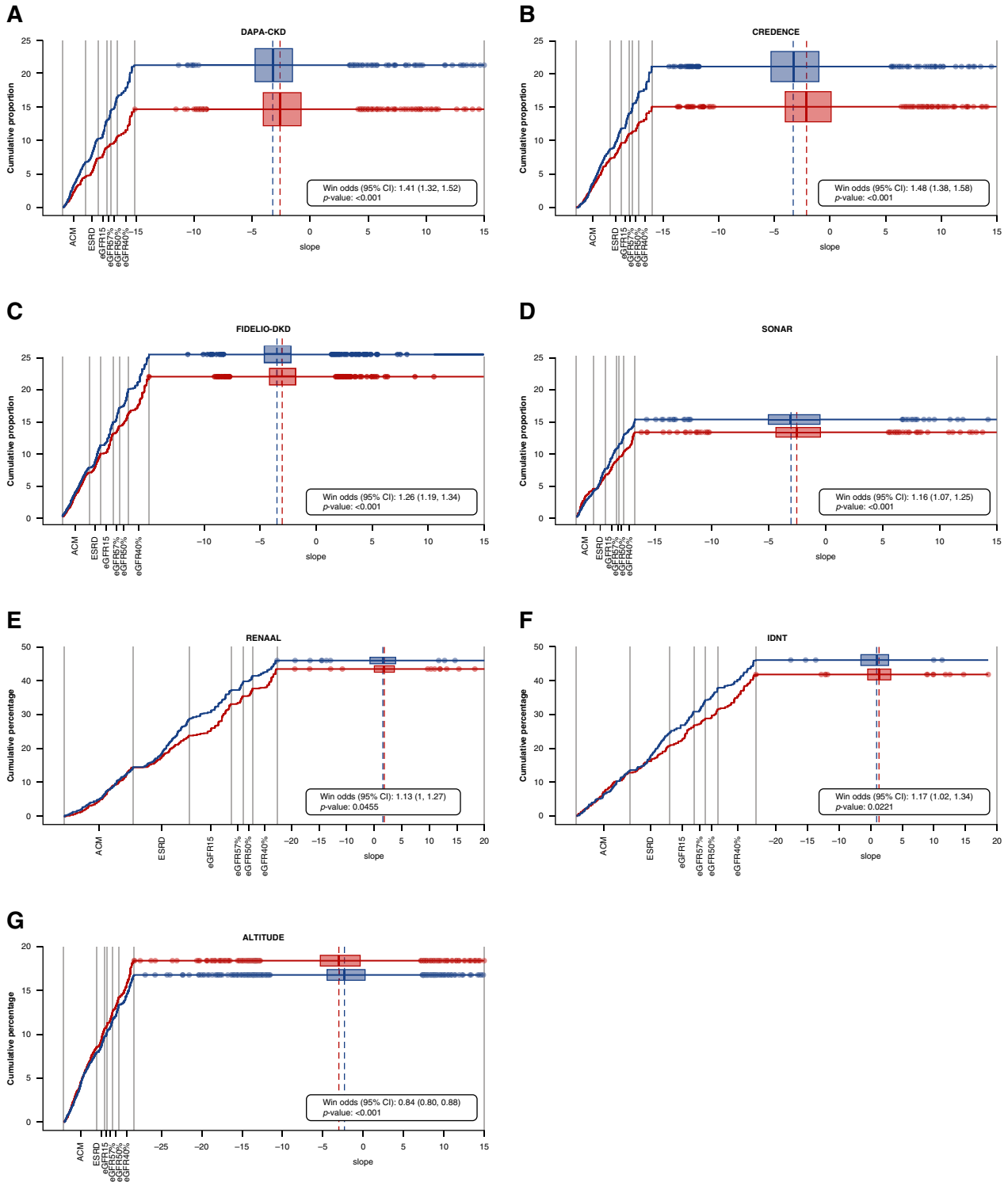
**Table 2.** Comparison of time to first event analysis and win odds in the seven selected trials

Treatment Comparisons	DAPA-CKD		CREDESCENCE		FIDELIO-DKD		SONAR		RENAAL		IDNT		ALTITUDE	
	Dapagliflozin versus Placebo		Canagliflozin versus Placebo		Finerenone versus Placebo		Atrasentan versus Placebo		Losartan versus Placebo		Irbesartan versus Placebo		Aliskiren versus Placebo	
	n	HR (95% CI)	n	HR (95% CI)	n	HR (95% CI)	n	HR (95% CI)	n	HR (95% CI)	n	HR (95% CI)	n	HR (95% CI)
<b>Event</b>														
All-cause mortality	247	0.69 (0.53 to 0.88)	369	0.83 (0.68 to 1.02)	463	0.90 (0.75 to 1.07)	162	0.80 (0.67 to 0.96)	313	1.02 (0.81 to 1.27)	180	0.92 (0.69 to 1.23)	734	1.07 (0.92 to 1.23)
Kidney replacement	174	0.66 (0.49 to 0.90)	176	0.74 (0.55 to 1.00)	258	0.86 (0.67 to 1.10)	287	0.70 (0.55 to 0.88)	341	0.71 (0.58 to 0.88)	183	0.77 (0.57 to 1.03)	229	1.09 (0.84 to 1.41)
GFR <15 ml/min per 1.73 m <sup>2</sup>	204	0.67 (0.51 to 0.88)	203	0.60 (0.45 to 0.80)	366	0.82 (0.67 to 1.01)	114	0.76 (0.52 to 1.10)	409	0.76 (0.62 to 0.91)	196	0.61 (0.46 to 0.81)	175	1.12 (0.83 to 1.51)
57% GFR decline	201	0.61 (0.46 to 0.82)	156	0.41 (0.29 to 0.57)	412	0.68 (0.55 to 0.82)	103	0.62 (0.42 to 0.92)	359	0.74 (0.60 to 0.92)	166	0.65 (0.48 to 0.89)	304	1.10 (0.88 to 1.37)
50% GFR decline	313	0.53 (0.42 to 0.67)	262	0.53 (0.41 to 0.69)	638	0.73 (0.62 to 0.85)	193	0.58 (0.44 to 0.78)	443	0.80 (0.67 to 0.97)	248	0.61 (0.47 to 0.79)	468	1.08 (0.90 to 1.30)
40% GFR decline	538	0.63 (0.53 to 0.74)	454	0.59 (0.48 to 0.71)	1056	0.81 (0.72 to 0.92)	329	0.81 (0.65 to 1.01)	598	0.88 (0.75 to 1.04)	400	0.83 (0.68 to 1.01)	832	1.12 (0.98 to 1.28)
GFR slope <sup>a</sup>		1.12 (0.80,1.43)		1.66 (1.30,2.00)		0.64 (0.40 to 0.89)		0.60 (0.23 to 0.97)		1.08 (0.40 to 1.76)		1.10 (0.47 to 1.74)		-0.30 (-0.6 to 0.01)
<b>Treatment effect composite end point</b>														
HR (Cox)		0.61 (0.51 to 0.73)		0.70 (0.59 to 0.82)		0.82 (0.73 to 0.93)		0.71 (0.58 to 0.88)		0.79 (0.66 to 0.94)		0.74 (0.59 to 0.94)		1.08 (0.95 to 1.23)
WOs <sup>b</sup>		1.41 (1.32 to 1.52)		1.48 (1.38 to 1.58)		1.26 (1.19 to 1.34)		1.16 (1.07 to 1.25)		1.13 (1.00 to 1.27)		1.17 (1.02 to 1.34)		0.84 (0.80 to 0.88)

Hazard ratios for the composite end point and components were calculated using Cox proportional hazards regression models. The win odds for the kidney hierarchical composite end point are shown in the bottom row. Values are n (%). Hazard ratios were calculated using Cox proportional hazards regression models and were adjusted for covariates as described in the primary publication of each trial. DAPA-CKD, Dapagliflozin and Prevention of Adverse Outcomes in CKD; CREDESCENCE, Canagliflozin and Renal Events in Diabetes with Established Nephropathy Clinical Evaluation; FIDELIO-DKD, Finerenone in reducing kidney failure and disease progression in Diabetic Kidney Disease; SONAR, Study Of diabetic Nephropathy with AtRasentan; RENAL, Reduction of Endpoints in Non-insulin-dependent diabetes mellitus with the Angiotensin II Antagonist Losartan; IDNT, Irbesartan Diabetic Nephropathy Trial; ALTITUDE, Aliskiren Trial in Type 2 Diabetes Using Cardio-Renal Endpoints; HR, hazard ratio; CI, confidence interval; WO, win odd.

<sup>a</sup>In each trial, the total GFR slope is defined as the annual decline in GFR from randomization until 36 months of follow-up time. Inclusion criteria differ between trials and do not allow direct comparison of results.

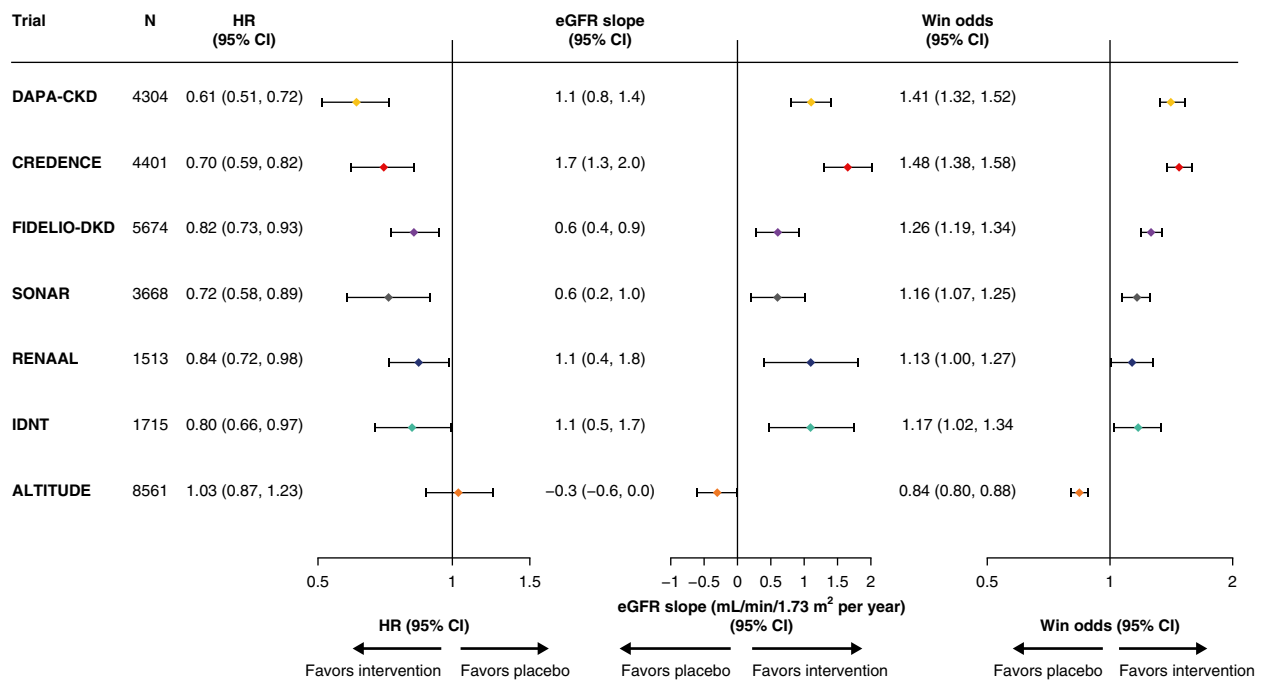
<sup>b</sup>Win odds were computed in a hierarchy: all-cause mortality, kidney replacement, GFR <15 ml/min per 1.73 m<sup>2</sup>, 57%, 50%, 40% GFR decline, and GFR slope.



**Figure 2. Maraca plots in each trial.** The Maraca plots show the contribution of the different time-to-event end point components; the treatment effect on the different time-to-event components of the composite; and the treatment effect on the continuous GFR slope component, for patients not experiencing any of the dichotomous outcomes. (A) DAPA-CKD trial. (B) CREDENCE trial. (C) FIDELIO-DKD trial. (D) SONAR trial. (E) RENAAL trial. (F) IDNT trial. (G) ALTITUDE trial. CI, confidence interval.

Downloaded from <http://journals.lww.com/jasn> by BMDM5ePHKav1ZEurnt1QIN4akJLHeZgbsIH04XMM10N0yWCX1AVV nYQp/1QH3D3D00dRy7ITV5F14C3VC1y0abg9QZXdG5j2mW1ZlE1= on 11/06/2023





**Figure 3.** Forest plot of the treatment effects on primary kidney end point as defined in each trial, GFR slope, and HCE. The left forest plot shows the HR of the treatment effect (time-to-first event) on the primary kidney end point; the forest plot in the middle, the treatment effect on GFR slope; and the right forest plot, the WOs of the treatment effect on the HCE. HCE, hierarchical composite end point; HR, hazard ratio; WO, win odd.

One of the advantages of the HCE approach is that it prioritizes the clinically most relevant component of the end point. In clinical trials in nephrology, this is of particular relevance because most of the end point components in a time-to-first event approach comprise percentage reductions in GFR rather than kidney failure end point which is the ultimate outcome patients care about. Indeed, in the trials included in our analysis, more than half of all primary end points were based on GFR changes rather than kidney failure. By contrast, the HCE approach prioritized kidney failure over GFR decline, leading to a higher number of kidney failure outcomes that contributed to the analysis. Despite the increase in kidney failure outcomes included in our HCE, GFR slope remained the most common contributor to the kidney HCE. GFR slope is considered a valid surrogate end point by regulatory agencies, a notion supported by a recent study showing excellent agreement between treatment effects on GFR slope and clinical kidney end points across a diverse set of populations with varying disease severity and treated with different interventions.<sup>22</sup> It should be noted that the HCE captures treatment effects on GFR slope only in patients who did not experience a dichotomous outcome, indicating that the competing events of death and kidney failure are handled by including them in the HCE as components of higher priority, which highlights the importance of inclusion of all-cause mortality in the HCE.

The prioritization of components in an HCE has generally been a matter of debate largely because it is subjective and

depends on patients' and physicians' preferences and beliefs. However, in contrast to other disease areas, the kidney HCE we developed follows the progression of CKD; the clinical end point of kidney failure is achieved through a reduction in GFR. Hence, by definition, declines in GFR are connected to kidney failure, and the prioritization of the end point components is therefore less subjective.

The direction and magnitude of the treatment effects expressed as WOs or HR were similar across all trials as summarized by the correlation between HRs and WOs. The HRs of all trials that reported a clinical benefit of the therapeutic intervention ranged between 0.61 and 0.86 which corresponded to WOs between 1.13 and 1.41. Accordingly, in sufficiently powered clinical trials, WOs ranging between these estimates represent a clinically relevant benefit. We included the ALTITUDE trial which reported neutral effects of aliskiren on the secondary kidney end point as a negative control. The results of the HCE analysis also demonstrated no benefit for aliskiren, with a WO of 0.84 (95% CI, 0.80 to 0.88). Although we do not consider that this *post hoc* analysis should be used to generate new conclusions about these previously conducted trials, this result shows that the increased statistical power associated with our kidney HCE may improve detection of both beneficial and detrimental treatment effects.

The effect estimates of the WOs in each trial were graphically represented in the maraca plot which is a new method to visualize a HCE that incorporates multiple severity-ordered time-to-event end points, represented by connected

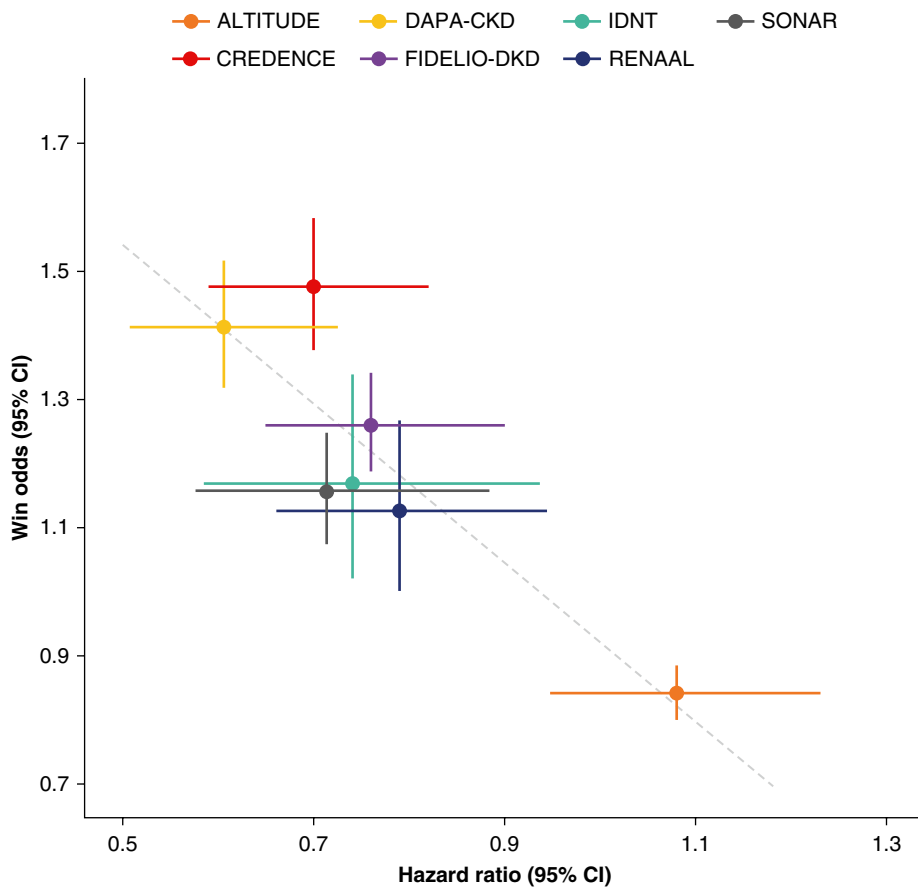


Figure 4. Scatter plot of treatment estimates on the primary kidney end point and HCE.

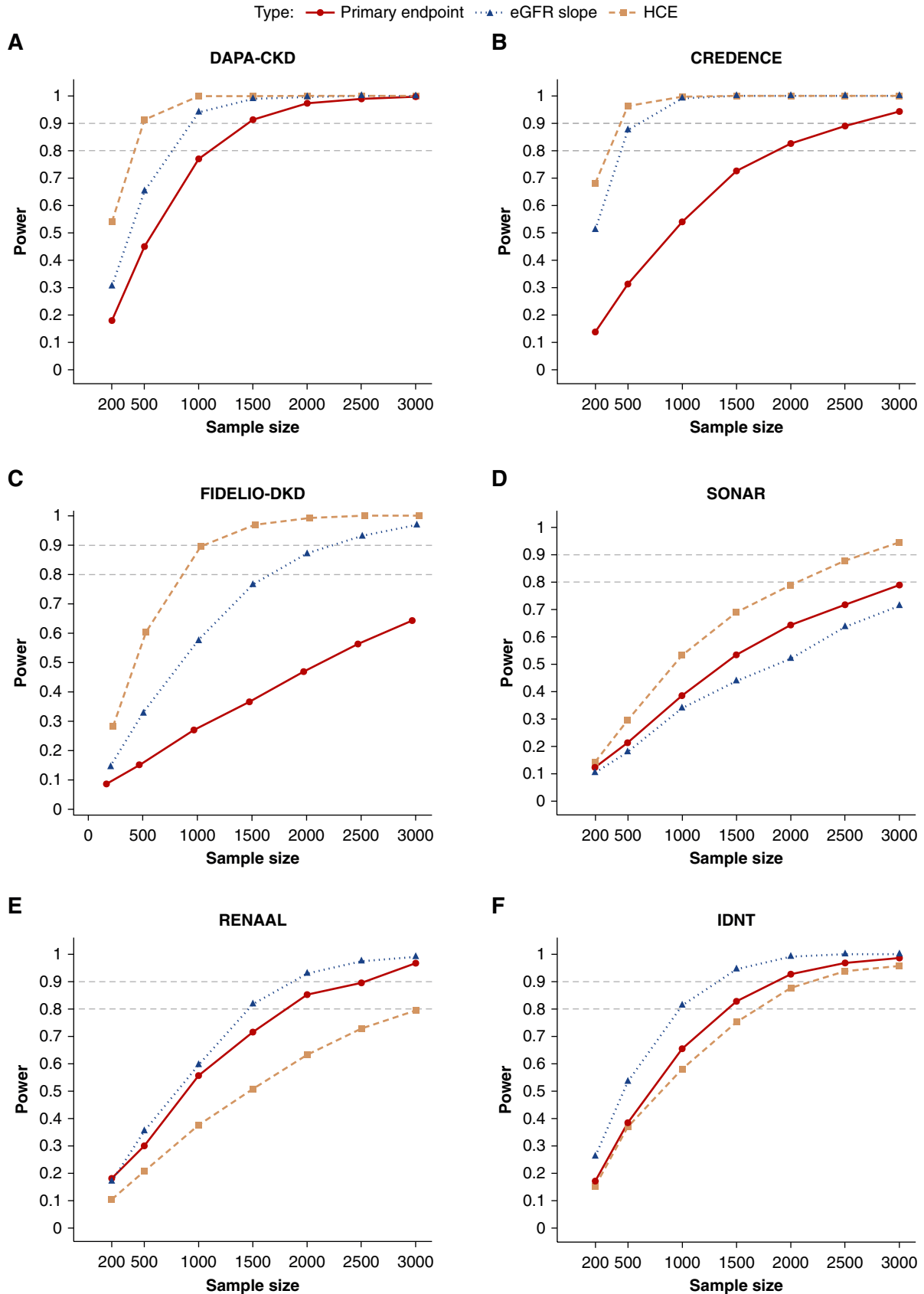
Kaplan–Meier plots in hierarchical subgroups, and a continuous outcome, represented by the box–whisker plot.<sup>19</sup> The maraca plot captures the relevant aspects of the HCE treatment effect estimates: the contributions of the different time-to-event end point components; the treatment effects on the different time-to-event components of the composite; and the treatment effect on the continuous GFR slope component, for patients not experiencing any of the dichotomous outcomes. Thus, the maraca plot visualizes the contributions of both the dichotomous and continuous outcomes to the overall treatment effect.

The end points we tested included and excluded death. One could argue that death should generally be included as the top of the hierarchy because it is the outcome of ultimate clinical importance. However, for treatments not expected to affect death, inclusion of death in the composite end point poses a dilemma because it may dilute the treatment effect and decrease statistical power. This was observed in the RENAAL and IDNT trials where the incidence of all-cause mortality was higher compared with the other trials and the interventions, losartan and irbesartan, respectively, did not reduce mortality decreasing statistical power for the HCE.<sup>9,10</sup> Indeed, excluding all-cause mortality from the HCE in these trials increased statistical power. However, if death is excluded from the HCE, it will be a competing risk for the remaining

outcomes. We propose an alternative definition of the kidney HCE which includes all kidney outcomes in all patients but excludes mortality for any patient who dies during follow-up. Eventually, inclusion of all-cause mortality (or the cause-specific mortality that can be attributable to the underlying disease, in this case cardiovascular and kidney disease) is relevant to any disease and any composite end point. Therefore, considerations pertaining to the objective of the clinical study and the potential effect of the intervention on cause-specific or all-cause death should be considered for deciding on this issue.

The effect of an intervention on a clinical kidney end point is determined by the number of patients reaching the outcomes included in the end point, that is, in clinical trials of CKD progression, the patients with the fastest rate of progression or those at advanced stages of CKD who are likely to experience kidney failure. By contrast, the effect of an intervention on GFR slope provides an estimate in all patients including both fast and slow progressors. Because the HCE comprises both the clinical end point and the GFR slope, all patients contribute to the end point, thereby increasing its ability to detect clinically relevant effect sizes in all patients. Indeed, the results of our study demonstrated that in all trials the HCE increased statistical power compared with time-to-event end points, highlighting the potential efficiency gains

Downloaded from http://journals.lww.com/jasn by BMDMfsePHKavIzEoumTtQIN4a+kJLHEZgbsHh04XM10hOwCX1AW on 11/06/2023



**Figure 5.** Sample size curves for all trials showing the sample size and statistical power of the original kidney end point in each trial, GFR slope, and HCE. (A) DAPA-CKD trial. (B) CRENDENCE trial. (C) FIDELIO-DKD trial. (D) SONAR trial. (E) RENAAL trial. (F) IDNT trial.

that may be anticipated when using an HCE. However, use of HCEs pertain to specific study questions and their potential advantages should be considered in the context of the drug's mechanism of action, study population, clinical trial setting, and the extent to which the continuous component is established as a genuine surrogate for clinical events in the given population.

The results of this study should be interpreted in the context of its limitations. First, the clinical trials used in this *post hoc* analysis were not designed to estimate the effect of interventions using an HCE. In particular, in all trials, the duration of follow-up varied among participants because of the event-driven clinical trial designs which present analytical challenges for pair-wise comparisons. We adopted a fixed 36-month follow-up approach and included only outcomes occurring during this period. We extended the follow-up for patients with a shorter follow-up by using the clinically most important outcome from the observed follow-up for the analysis on month 36. This is essentially carrying forward the patient's most severe outcome for the inclusion in the analysis, which is not an optimal imputation method. The results were however unchanged in an additional analysis using the shared follow-up approach. Nevertheless, clinical trials using the HCE in the future should ideally be planned with a requirement of a minimum follow-up time for all participants, and the analysis time point should be selected such that most participants complete the required follow-up. In this design, the statistical analysis plan could consider alternate approaches for missing follow-up time, such as multiple imputation. Second, GFR measurements were not available for all patients at all follow-up visits. This may have influenced the precision of effect estimates on GFR slope.

In conclusion, using data from seven landmark clinical kidney outcome trials, we developed and validated a hierarchical composite kidney end point that both conceptually and empirically seems to provide increased statistical power compared with the established kidney end point. The prioritization of clinical outcomes and ability to include GFR slope make the HCE an attractive alternative end point to the established kidney end point.

The R software (R Core Team [2023], version 4.3.1) implementation of the derivation and analysis of kidney HCE is provided in the [Supplemental Appendix](#), which includes also synthetic datasets as an example ([Supplemental Excel Files 1–3](#)). The synthetic datasets were created and kindly provided by the Analytics Data Preparation Team, Data Office, Data Science and AI, AstraZeneca.

## DISCLOSURES

G. Bakris is supported by T32 NIH grant DK07011 and is a consultant to Alnylam, AstraZeneca, Bayer, Glaxo Smith Kline, InREGEN, Ionis, Janssen, KBP Biosciences, and Novo Nordisk. M. Brinker reports ownership in-

terest: Bayer AG. M. Brinker, R. Nkulikiyinka, P. Schloemer, and C. Tasto are Bayer employees. S.B. Gasparyan, M. Karpfors, D.J. Little, and J. Rossert are AstraZeneca employees. H.L. Heerspink is consultant for AstraZeneca, Bayer, Boehringer Ingelheim, Chinook, CSL Behring, Dimerix, Eli-Lilly, Gilead, Janssen, Merck, Novo Nordisk, ProKidney, Travere Therapeutics, and Vifor Fresenius. He has received research support from AstraZeneca, Boehringer Ingelheim, Janssen, and Novo Nordisk; honoraria: lecture fees from AstraZeneca and NovoNordisk; and speakers bureau: AstraZeneca. N. Jongs reports travel grants from AstraZeneca. D.J. Little reports ownership interest: AstraZeneca and other interests or relationships: volunteer as a nephrologist at Walter Reed National Military Medical Center. V. Perkovic serves as a Board Director for St. Vincents Health Australia, George Clinical, and several Medical Research Institutes. He has received honoraria for Steering Committee roles and scientific presentations and/or advisory board attendance from Abbvie, Amgen, AstraZeneca, Baxter, Bayer, Boehringer Ingelheim, Chinook, Durect, Eli Lilly, Gilead, GSK, Janssen, Merck, Mitsubishi Tanabe, Mundipharma, Novartis, Novo Nordisk, Otsuka, Pfizer, Pharmalink, Reata, Relypsa, Roche, Sanofi, Servier, Travere, and Tricida. J. Rossert reports ownership interest: Amgen, AstraZeneca, and Vertex. C. Tasto reports ownership interest: Bayer AG. D.C. Wheeler has received honoraria and/or consultancy fees from Amgen, Astellas, AstraZeneca, Boehringer Ingelheim, Bayer, Eledon, Galderma, George Clinical, Gilead, GlaxoSmithKline, Janssen, Medscape, Merck Sharp and Dohme, Mundipharma, Napp, Pfizer, Pharmacosmos, ProKidney, Reata, Takeda, Tricida, Vifor Fresenius, and Zydus and advisory or leadership role: AstraZeneca; and speakers bureau: Amgen, Astellas, AstraZeneca, Janssen, Merck Sharp and Dohme, Mundipharma, Napp, and Vifor Fresenius. R. Nkulikiyinka reports Employer: Bayer AG; and Ownership Interest: Bayer AG. G. Bakris reports Consultancy: Janssen, Bayer, KBP Biosciences, Novo Nordisk, Astra-Zeneca, Ionis, Alnylam, Medscape; Honoraria: Merck, Novo Nordisk, Astra Zeneca, Ionis, Alnylam, KBP Biosciences, and Bayer; Advisory or Leadership Role: KBP Biosciences, American J Nephrology, Editor, Diabetes Care, Assoc. Ed.; American Heart Assoc.; UpToDate-Nephrology; and Other Interests or Relationships: American Diabetes Association, American Heart Association, Blood Pressure Council. V. Perkovic reports Consultancy: AstraZeneca, Bayer, Boehringer Ingelheim, Chinook, Eli Lilly, Gilead, GlaxoSmithKline, Janssen, Mitsubishi Tanabe, Mundipharma, Novartis, Novo Nordisk, Otsuka, Travere, Tricida, UptoDate; Ownership Interest: George Clinical; Research Funding: AstraZeneca, Bayer, Chinook, Gilead, GlaxoSmithKline, Janssen, Novartis, Novo Nordisk, Otsuka, Travere, Tricida; Honoraria: AstraZeneca, Bayer, Boehringer Ingelheim, Chinook, Eli Lilly, Gilead, GlaxoSmithKline, Janssen, Mitsubishi Tanabe, Mundipharma, Novartis, Novo Nordisk, Otsuka, Travere, Tricida, UptoDate; and Advisory or Leadership Role: Steering committees for Bayer, Chinook, GlaxoSmithKline, Janssen, Novartis, Novo Nordisk, Otsuka, Pfizer, Travere; Board Director for St Vincents Health Australia, George Clinical.

## FUNDING

None.

## ACKNOWLEDGMENTS

The authors thank all investigators, patients, and research teams for their contribution to the reported clinical trials. We thank Jörg Pawlitschko for support with statistical analyses. We thank Joan Buenconsejo for the initial discussion on this topic. Editorial support in creating figures was provided by Josh Abbott, BSc, and Moamen Hammad, PhD, both of Scion, London, United Kingdom, supported by Bayer AG according to Good Publication Practice guidelines (<https://www.acpjournals.org/doi/10.7326/M22-1460>).

## AUTHOR CONTRIBUTIONS

**Conceptualization:** George Bakris, Meike Brinker, Samvel B. Gasparyan, Hiddo L. Heerspink, Niels Jongs, Martin Karpefors, Dustin J. Little, Richard Nkulikiyinka, Vlado Perkovic, Jerome Rossert, Patrick Schloemer, Christoph Tasto, David C. Wheeler.

**Data curation:** George Bakris, Meike Brinker, Samvel B. Gasparyan, Hiddo L. Heerspink, Niels Jongs, Martin Karpefors, Dustin J. Little, Richard Nkulikiyinka, Vlado Perkovic, Jerome Rossert, Patrick Schloemer, Christoph Tasto, David C. Wheeler.

**Formal analysis:** Niels Jongs, Patrick Schloemer, Christoph Tasto.

**Writing – original draft:** Hiddo L. Heerspink.

**Writing – review & editing:** George Bakris, Meike Brinker, Samvel B. Gasparyan, Hiddo L. Heerspink, Niels Jongs, Martin Karpefors, Dustin J. Little, Richard Nkulikiyinka, Vlado Perkovic, Jerome Rossert, Patrick Schloemer, Christoph Tasto, David C. Wheeler.

## DATA SHARING STATEMENT

Previously published data were used for this study.

## SUPPLEMENTAL MATERIAL

This article contains the following supplemental material online at <http://links.lww.com/JSN/E542>, <http://links.lww.com/JSN/E543>, <http://links.lww.com/JSN/E544>, and <http://links.lww.com/JSN/E545>.

**Supplemental Table 1.** Definition of established primary kidney end points in each trial.

**Supplemental Table 2.** Hierarchical comparison of the components of the kidney HCE.

**Supplemental Table 3.** Contribution of individual end point components to the primary composite kidney end point as defined in each trial.

**Supplemental Table 4.** Number of wins on active and placebo treatment, number of ties, and win odds for each component of the HCE using a shared follow-up time for each pairwise comparison.

**Supplemental Table 5.** Comparison of the win odds using a nonshared and shared follow-up time.

**Supplemental Figure 1.** Number (%) of individual components of the hierarchical composite kidney end point excluding all-cause mortality.

**Supplemental Figure 2.** Maraca plot of the hierarchical composite kidney end point excluding mortality.

**Supplemental Figure 3.** Sample size curves for all trials showing the sample size and statistical power of the original kidney end point in each trial, GFR slope, and HCE excluding all-cause mortality. (A) DAPA-CKD trial. (B) CREDENCE trial. (C) FIDELIO-DKD trial. (D) SONAR trial. (E) RENAAL trial. (F) IDNT trial.

**Supplemental Document: Technical Appendix.**

**Supplemental Excel File 1.** ADSL—patient-level demographic data.

**Supplemental Excel File 2.** 2ADLB—eGFR measurements of patients over time.

**Supplemental Excel File 3.** ADET—time-to-event outcomes of patients.

## REFERENCES

- Levin A, Agarwal R, Herrington WG, et al. International consensus definitions of clinical trial outcomes for kidney failure: 2020. *Kidney Int.* 2020;98(4):849–859. doi:10.1016/j.kint.2020.07.013
- Levey AS, Gansevoort RT, Coresh J, et al. Change in albuminuria and GFR as end points for clinical trials in early stages of CKD: a scientific workshop sponsored by the National Kidney Foundation in collaboration with the US Food and Drug Administration and European Medicines Agency. *Am J Kidney Dis.* 2020;75(1):84–104. doi:10.1053/j.ajkd.2019.06.009
- Lambers Heerspink HJ, Weldegiorgis M, Inker LA, et al. Estimated GFR decline as a surrogate end point for kidney failure: a post hoc analysis from the reduction of end points in non-insulin-dependent diabetes with the angiotensin II antagonist losartan (RENAAL) study and irbesartan diabetic nephropathy trial (IDNT). *Am J Kidney Dis.* 2014;63(2):244–250. doi:10.1053/j.ajkd.2013.09.016
- Inker LA, Lambers Heerspink HJ, Mondal H, et al. GFR decline as an alternative end point to kidney failure in clinical trials: a meta-analysis of treatment effects from 37 randomized trials. *Am J Kidney Dis.* 2014;64(6):848–859. doi:10.1053/j.ajkd.2014.08.017
- Inker LA, Heerspink HJL, Tighiouart H, et al. GFR slope as a surrogate end point for kidney disease progression in clinical trials: a meta-analysis of treatment effects of randomized controlled trials. *J Am Soc Nephrol.* 2019;30(9):1735–1745. doi:10.1681/ASN.2019010007
- Redfors B, Gregson J, Crowley A, et al. The win ratio approach for composite endpoints: practical guidance based on previous experience. *Eur Heart J.* 2020;41(46):4391–4399. doi:10.1093/eurheartj/ehaa665
- Perkovic V, Jardine MJ, Neal B, et al. Canagliflozin and renal outcomes in type 2 diabetes and nephropathy. *New Engl J Med.* 2019;380(24):2295–2306. doi:10.1056/NEJMoa1811744
- Parving H-H, Brenner BM, McMurray JJV, et al. Cardiorenal end points in a trial of aliskiren for type 2 diabetes. *New Engl J Med.* 2012;367(23):2204–2213. doi:10.1056/NEJMoa1208799
- 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. *New Engl J Med.* 2001;345(12):851–860. doi:10.1056/NEJMoa011303
- 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. *New Engl J Med.* 2001;345(12):861–869. doi:10.1056/NEJMoa011161
- Bakris GL, Agarwal R, Anker SD, et al. Effect of finerenone on chronic kidney disease outcomes in type 2 diabetes. *New Engl J Med.* 2020;383(23):2219–2229. doi:10.1056/NEJMoa2025845
- Heerspink HJL, Parving HH, Andress DL, et al. Atrasentan and renal events in patients with type 2 diabetes and chronic kidney disease (SONAR): a double-blind, randomised, placebo-controlled trial. *Lancet.* 2019;393(10184):1937–1947. doi:10.1016/S0140-6736(19)30772-X
- Heerspink HJL, Stefánsson BV, Correa-Rotter R, et al. Dapagliflozin in patients with chronic kidney disease. *New Engl J Med.* 2020;383(15):1436–1446. doi:10.1056/NEJMoa2024816
- Vonesh E, Tighiouart H, Ying J, et al. Mixed-effects models for slope-based endpoints in clinical trials of chronic kidney disease. *Stat Med.* 2019;38(22):4218–4239. doi:10.1002/sim.8282
- Gasparyan SB, Kowalewski EK, Folkvaljon F, et al. Power and sample size calculation for the win odds test: application to an ordinal endpoint in COVID-19 trials. *J Biopharm Stat.* 2021;31(6):765–787. doi:10.1080/10543406.2021.1968893
- Pocock SJ, Ariti CA, Collier TJ, Wang D. The win ratio: a new approach to the analysis of composite endpoints in clinical trials based on clinical priorities. *Eur Heart J.* 2012;33(2):176–182. doi:10.1093/eurheartj/ehr352
- Gasparyan SB, Folkvaljon F, Bengtsson O, Buenconsejo J, Koch GG. Adjusted win ratio with stratification: calculation methods and interpretation. *Stat Methods Med Res.* 2021;30(2):580–611. doi:10.1177/0962280220942558
- Yu RX, Ganju J. Sample size formula for a win ratio endpoint. *Stat Med.* 2022;41(14):2691–2692. doi:10.1002/sim.9388
- Karpefors M, Lindholm D, Gasparyan SB. The maraca plot: a novel visualization of hierarchical composite endpoints. *Clin Trials.* 2023;20(1):84–88. doi:10.1177/17407745221134949

20. Packer M. Development and evolution of a hierarchical clinical composite end point for the evaluation of drugs and devices for acute and chronic heart failure: a 20-year perspective. *Circulation*. 2016;134(21):1664–1678. doi:[10.1161/CIRCULATIONAHA.116.023538](https://doi.org/10.1161/CIRCULATIONAHA.116.023538)
21. Gasparyan SB, Buenconsejo J, Kowalewski EK, et al. Design and analysis of studies based on hierarchical composite endpoints: insights from the DARE-19 trial. *Ther Innov Regul Sci*. 2022;56(5):785–794. doi:[10.1007/s43441-022-00420-1](https://doi.org/10.1007/s43441-022-00420-1)
22. Collier W, Inker LA, Haaland B, et al. Evaluation of variation in the performance of GFR slope as a surrogate end point for kidney failure in clinical trials that differ by severity of CKD. *Clin J Am Soc Nephrol*. 2023;18(2):183–192. doi:[10.2215/CJN.0000000000000050](https://doi.org/10.2215/CJN.0000000000000050)

## AFFILIATIONS

- <sup>1</sup>Department of Clinical Pharmacy and Pharmacology, University Medical Center Groningen, University of Groningen, Groningen, The Netherlands
- <sup>2</sup>The George Institute for Global Health, Sydney, New South Wales, Australia
- <sup>3</sup>Pharmaceuticals, Research and Development, Bayer AG, Berlin, Germany
- <sup>4</sup>Late Stage Development, Cardiovascular, Renal and Metabolism (CVRM), Biopharmaceuticals R&D, AstraZeneca, Gaithersburg, Maryland
- <sup>5</sup>Pharmaceuticals, Research and Development, Bayer AG, Wuppertal, Germany
- <sup>6</sup>Late Stage Development, Cardiovascular, Renal and Metabolism (CVRM), BioPharmaceuticals R&D, AstraZeneca, Gothenburg, Sweden
- <sup>7</sup>Department of Renal Medicine, University College London, London, United Kingdom
- <sup>8</sup>Department of Medicine, University of Chicago Medicine, Chicago, Illinois
- <sup>9</sup>Faculty of Medicine & Health, University New South Wales, Sydney, New South Wales, Australia