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

Hiddo L. Heerspink,1,2 Niels Jonge,1 Patrick Schloemer,3 Dustin J. Little,4 Meike Brinker,5 Christoph Tasto,5 Martin Karpefors,6 David C. Wheeler,2,7 George Bakris,8 Vlado Perkovic,2,9 Richard Nkulikiyinka,3 Jerome Rossert,4 and Samvel B. Gasparyan6

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

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
longed 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.3–5

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 impor-
tance 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 interven-
tion 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 progres-
sion 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 in-
cluded 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.6 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 accompa-
nying 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 es-
ablished 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.

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 in-
hibitor 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 trial7–13). We
selected these clinical trials because they recruited patients
with CKD and demonstrated varying effects on the pri-
mary 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 Estab-
lished Nephropathy Clinical Evaluation (CREDEENCE)
(NCT02063791), Finerenone in reducing kiDney failuRe
and disease prOgression in Diabetic Kidney Disease
(FIDELIO-DKD) (NCT02540993), Study Of diabetic Ne-
phropathy with AtRsentan (SONAR) (NCT01858532),
Reduction of Endpoints in Non-insulin-dependent diabetes
mellitus with the Angiotensin II Antagonist Losartan (RENAI

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 WOs, 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 WOs 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 WOs approach.
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 m² 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. 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, an adaptation of win ratio to include ties (a tie is considered a half loss and a half win for each group). 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. We calculated WOs and its 95% confidence interval (CI).

Maraca plots were developed to visualize HCE combining multiple time-to-event outcomes with a single continuous outcome. 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 m², 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.

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 (Supplemental 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 failure20 and coronavirus disease 201921), 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

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>DAPA-CKD (N=4304)</th>
<th>CREDENCE (N=4401)</th>
<th>FIDELIO-DKD (N=5674)</th>
<th>SONAR (N=3668)</th>
<th>RENAL (N=1513)</th>
<th>IDNT (N=1715)</th>
<th>ALTITUDE (N=8561)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Median follow-up, yr</td>
<td>2.4</td>
<td>2.6</td>
<td>2.6</td>
<td>2.2</td>
<td>3.4</td>
<td>2.6</td>
<td>2.7</td>
</tr>
<tr>
<td>Age, yr (SD)</td>
<td>61.8 (12)</td>
<td>63.0 (9)</td>
<td>65.6 (9)</td>
<td>64.5 (8.8)</td>
<td>60.2 (7.4)</td>
<td>58.8 (7.7)</td>
<td>64.5 (9.8)</td>
</tr>
<tr>
<td>Female sex, N (%)</td>
<td>1425 (33.1)</td>
<td>494 (33.9)</td>
<td>1681 (29.8)</td>
<td>946 (25.8)</td>
<td>557 (36.8)</td>
<td>367 (32)</td>
<td>2735 (31.9)</td>
</tr>
<tr>
<td>Race, n (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asian</td>
<td>1467 (34.1)</td>
<td>877 (19.9)</td>
<td>1440 (25.4)</td>
<td>1198 (32.7)</td>
<td>252 (16.7)</td>
<td>51 (4.4)</td>
<td>2714 (31.7)</td>
</tr>
<tr>
<td>Black</td>
<td>191 (4.4)</td>
<td>224 (5.1)</td>
<td>264 (4.7)</td>
<td>224 (6.1)</td>
<td>230 (15.2)</td>
<td>141 (12.3)</td>
<td>277 (3.2)</td>
</tr>
<tr>
<td>Other</td>
<td>356 (8.3)</td>
<td>369 (8.4)</td>
<td>378 (6.7)</td>
<td>136 (3.7)</td>
<td>296 (19.7)</td>
<td>103 (9.0)</td>
<td>696 (8.1)</td>
</tr>
<tr>
<td>White</td>
<td>2290 (53.2)</td>
<td>2931 (66.6)</td>
<td>3592 (63.3)</td>
<td>2110 (57.5)</td>
<td>853 (74.3)</td>
<td>4850 (56.7)</td>
<td></td>
</tr>
<tr>
<td>BP, mm Hg (SD)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Systolic</td>
<td>137.1 (17)</td>
<td>140.0 (16)</td>
<td>138.0 (14)</td>
<td>133.3 (15)</td>
<td>152.5 (19.3)</td>
<td>159.4 (20.0)</td>
<td>137.3 (16)</td>
</tr>
<tr>
<td>Diastolic</td>
<td>77.5 (11)</td>
<td>78.3 (9)</td>
<td>76.0 (10)</td>
<td>71.5 (10)</td>
<td>82.4 (10.4)</td>
<td>86.9 (11.4)</td>
<td>74.2 (10)</td>
</tr>
<tr>
<td>Body weight, kg (SD)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>60</td>
<td>454 (10.5)</td>
<td>1769 (40.2)</td>
<td>656 (11.6)</td>
<td>468 (12.8)</td>
<td>79 (5.2)</td>
<td>265 (23.1)</td>
<td>2783 (32.5)</td>
</tr>
<tr>
<td>&lt;60</td>
<td>3850 (89.5)</td>
<td>2632 (59.8)</td>
<td>5018 (88.4)</td>
<td>3191 (87.0)</td>
<td>1434 (94.8)</td>
<td>873 (76.0)</td>
<td>5776 (67.5)</td>
</tr>
<tr>
<td>UACR, mg/g (IQR)</td>
<td>949 (477–1885)</td>
<td>927 (463–1833)</td>
<td>852 (446–1634)</td>
<td>828 (458–1556)</td>
<td>1245.5 (558–2544)</td>
<td>1354 (1054, 1748)</td>
<td>283 (56–889)</td>
</tr>
<tr>
<td>GFR, ml/min per 1.73 m(^2) (SD)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>≥60</td>
<td>43.1 (12)</td>
<td>56.2 (18)</td>
<td>44.3 (13)</td>
<td>42.3 (14)</td>
<td>38.6 (12.4)</td>
<td>47.2 (17.8)</td>
<td>56.9 (22.5)</td>
</tr>
<tr>
<td>&lt;60</td>
<td>3560 (83.9)</td>
<td>2931 (66.6)</td>
<td>3592 (63.3)</td>
<td>2110 (57.5)</td>
<td>853 (74.3)</td>
<td>4850 (56.7)</td>
<td></td>
</tr>
<tr>
<td>Baseline medications, n (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ACEi</td>
<td>1353 (31.4)</td>
<td>1922 (43.7)</td>
<td>1942 (34.2)</td>
<td>1319 (36.0)</td>
<td>737 (48.7)</td>
<td>507 (44.2)</td>
<td>3792 (44.3)</td>
</tr>
<tr>
<td>ARB</td>
<td>2870 (66.7)</td>
<td>2480 (56.4)</td>
<td>3725 (65.7)</td>
<td>2391 (65.2)</td>
<td>105 (6.9)</td>
<td>33 (2.9)</td>
<td>4787 (55.9)</td>
</tr>
<tr>
<td>Diuretics</td>
<td>1882 (43.7)</td>
<td>2057 (46.7)</td>
<td>3214 (56.6)</td>
<td>3157 (66.1)</td>
<td>878 (58)</td>
<td>547 (47.6)</td>
<td>5872 (68.6)</td>
</tr>
<tr>
<td>Insulin</td>
<td>1598 (37.1)</td>
<td>2884 (65.5)</td>
<td>3637 (64.1)</td>
<td>2315 (63.1)</td>
<td>910 (60.1)</td>
<td>644 (56.1)</td>
<td>4850 (56.7)</td>
</tr>
<tr>
<td>Statins</td>
<td>2794 (64.9)</td>
<td>3036 (69.0)</td>
<td>4215 (74.3)</td>
<td>2707 (73.8)</td>
<td>507 (33.5)</td>
<td>299 (26.0)</td>
<td>5576 (65.1)</td>
</tr>
</tbody>
</table>

DAPA-CKD, Dapagliflozin and Prevention of Adverse Outcomes in CKD; CREDENCE, Canagliflozin and Renal Events in Diabetes with Established Nephropathy Clinical Evaluation; FIDELIO-DKD, FInerenone in reducing kiDnEy failur and disease prgression 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 RENAL and IDNT.
Figure 1. Number (%) of individual components of the hierarchical composite kidney end point in each trial. (A) DAPA-CKD trial. (B) CREDENCE trial. (C) FIDELIO-DKD trial. (D) SONAR trial. (E) RENAAAL trial. (F) IDNT trial. (G) ALTITUDE trial. ACM, all-cause mortality; ALTITUDE, Aliskiren Trial in Type 2 Diabetes Using Cardio-Renal Endpoints; CREDENCE, 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

<table>
<thead>
<tr>
<th>Treatment Comparisons</th>
<th>DAPA-CKD</th>
<th>CREDENCE</th>
<th>FIDELIO-DKD</th>
<th>SONAR</th>
<th>RENAAAL</th>
<th>IDNT</th>
<th>ALTITUDE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Dapagliflozin versus Placebo</td>
<td>Canagliflozin versus Placebo</td>
<td>Finerenone versus Placebo</td>
<td>Atrasentan versus Placebo</td>
<td>Losartan versus Placebo</td>
<td>Irbesartan versus Placebo</td>
<td>Aliskiren versus Placebo</td>
</tr>
<tr>
<td>Event</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>All-cause mortality</td>
<td>247</td>
<td>369</td>
<td>463</td>
<td>313</td>
<td>180</td>
<td>734</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>0.69 (0.53 to 0.88)</td>
<td>0.83 (0.68 to 1.02)</td>
<td>0.90 (0.75 to 1.07)</td>
<td>1.02 (0.81 to 1.27)</td>
<td>0.92 (0.69 to 1.23)</td>
<td>1.07 (0.92 to 1.23)</td>
<td></td>
</tr>
<tr>
<td>Kidney replacement</td>
<td>174</td>
<td>176</td>
<td>258</td>
<td>287</td>
<td>341</td>
<td>183</td>
<td>229</td>
</tr>
<tr>
<td></td>
<td>0.66 (0.49 to 0.90)</td>
<td>0.74 (0.55 to 1.00)</td>
<td>0.86 (0.67 to 1.10)</td>
<td>0.70 (0.55 to 0.88)</td>
<td>0.71 (0.58 to 0.88)</td>
<td>0.77 (0.57 to 1.03)</td>
<td>1.09 (0.84 to 1.41)</td>
</tr>
<tr>
<td>GFR &lt;15 ml/min per 1.73 m²</td>
<td>204</td>
<td>203</td>
<td>366</td>
<td>114</td>
<td>409</td>
<td>196</td>
<td>175</td>
</tr>
<tr>
<td></td>
<td>0.67 (0.51 to 0.88)</td>
<td>0.60 (0.45 to 0.80)</td>
<td>0.82 (0.67 to 1.01)</td>
<td>0.76 (0.52 to 1.10)</td>
<td>0.76 (0.62 to 0.91)</td>
<td>0.61 (0.46 to 0.81)</td>
<td>0.83 (0.83 to 1.51)</td>
</tr>
<tr>
<td>57% GFR decline</td>
<td>201</td>
<td>156</td>
<td>412</td>
<td>103</td>
<td>359</td>
<td>166</td>
<td>304</td>
</tr>
<tr>
<td></td>
<td>0.61 (0.46 to 0.82)</td>
<td>0.41 (0.29 to 0.57)</td>
<td>0.68 (0.55 to 0.82)</td>
<td>0.62 (0.42 to 0.92)</td>
<td>0.74 (0.60 to 0.92)</td>
<td>0.65 (0.48 to 0.89)</td>
<td>0.82 (0.88 to 1.37)</td>
</tr>
<tr>
<td>50% GFR decline</td>
<td>313</td>
<td>262</td>
<td>638</td>
<td>193</td>
<td>443</td>
<td>248</td>
<td>468</td>
</tr>
<tr>
<td></td>
<td>0.53 (0.42 to 0.67)</td>
<td>0.53 (0.41 to 0.69)</td>
<td>0.73 (0.62 to 0.85)</td>
<td>0.58 (0.44 to 0.78)</td>
<td>0.80 (0.67 to 0.97)</td>
<td>0.61 (0.47 to 0.79)</td>
<td>1.08 (0.90 to 1.30)</td>
</tr>
<tr>
<td>40% GFR decline</td>
<td>538</td>
<td>454</td>
<td>1056</td>
<td>329</td>
<td>598</td>
<td>400</td>
<td>382</td>
</tr>
<tr>
<td></td>
<td>0.63 (0.53 to 0.74)</td>
<td>0.59 (0.48 to 0.71)</td>
<td>0.81 (0.72 to 0.92)</td>
<td>0.81 (0.65 to 1.01)</td>
<td>0.88 (0.75 to 1.04)</td>
<td>0.83 (0.68 to 1.01)</td>
<td>1.12 (0.98 to 1.28)</td>
</tr>
<tr>
<td>GFR slope&lt;sup&gt;a&lt;/sup&gt;</td>
<td>1.12 (0.80, 1.43)</td>
<td>1.66 (1.30, 2.00)</td>
<td>0.64 (0.40 to 0.89)</td>
<td>0.60 (0.23 to 0.97)</td>
<td>1.00 (0.40 to 1.76)</td>
<td>1.10 (0.47 to 1.74)</td>
<td>-0.30 (-0.6 to 0.01)</td>
</tr>
<tr>
<td>Treatment effect composite end point</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HR (Cox)</td>
<td>0.61 (0.51 to 0.73)</td>
<td>0.70 (0.59 to 0.82)</td>
<td>0.82 (0.73 to 0.93)</td>
<td>0.71 (0.58 to 0.88)</td>
<td>0.79 (0.66 to 0.94)</td>
<td>0.74 (0.59 to 0.94)</td>
<td>1.08 (0.95 to 1.23)</td>
</tr>
<tr>
<td>WO&lt;sup&gt;b&lt;/sup&gt;</td>
<td>1.41 (1.32 to 1.52)</td>
<td>1.48 (1.38 to 1.58)</td>
<td>1.26 (1.19 to 1.34)</td>
<td>1.16 (1.07 to 1.25)</td>
<td>1.13 (1.00 to 1.27)</td>
<td>1.17 (1.02 to 1.34)</td>
<td>0.84 (0.80 to 0.88)</td>
</tr>
</tbody>
</table>

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; CREDENCE, 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 m², 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.
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 that corresponded to WOs between 1.13 and 1.41. Accordingly, in therapeutic intervention ranged between 0.61 and 0.86 which expressed as WOs or HR were similar across all trials as summarized by the correlation between HRs and WOs. 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 were 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 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 were 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
Kaplan–Meier plots in hierarchical subgroups, and a continuous outcome, represented by the box–whisker plot. 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 RENAAAL 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. 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.
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) CREDENCE 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 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 Ailynlam, AstraZeneca, Bayer, Glaxo Smith Kline, InREGEN, Ionis, Janssen, KBP Biosciences, and Novo Nordisk. M. Brinker reports ownership interest: Bayer AG. M. Brinker, R. Nkulikiyinka, P. Schloemer, and C. Tasto are Bayer employees. S.B. Gasparian, M. Karpefors, D.J. Little, and J. Rossert are AstraZeneca employees. H.L. Heerspink is consultant for AstraZeneca, Bayer, Boehringer Ingelheim, Chino, CSL Behring, Di- merix, Eli-Lilly, Gilead, Janssen, Merck, Novo Nordisk, ProKidney, Traver- Therapeutics, and Vifor Fresenius. He has received research support from AstraZeneca, Boehringer Ingelheim, Janssen, and Novo Nordisk; honor- raria: 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, Chino, Durect, Eli Lilly, Gilead, GSK, Janssen, Merck, Mitsubishi Tanabe, Mundipharma, Novartis, Novo Nordisk, Otsuka, Pfizer, Pharmalink, Reata, Relypsa, Roche, Sanofi, Servier, Traver, 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 consulting fees from Amgen, Astellas, AstraZeneca, Boehringer Ingelheim, Bayer, Elelon, Gelderma, George Clinical, Gilead, GlaxoSmithKline, Jans- sen, 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, Ailynlam, Medscape; Honoraria: Merck, Novo Nordisk, Astra Zeneca, Ionis, Ailynlam, 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: AstraZe- neca, Bayer, Boehringer Ingelheim, Chino, Eli Lilly, Gilead, GlaxoSmithKline, Janssen, Mitsubishi Tanabe, Mundipharma, Novartis, Novo Nordisk, Otsuka, Traver, UdoptDate; Ownership Interest: George Clinical; Research Funding: AstraZeneca, Bayer, Chino, Gilead, GlaxoSmithKline, Janssen, Novartis, Novo Nordisk, Otsuka, Traver, Tricida; Honoraria: AstraZeneca, Bayer, Boehringer Ingelheim, Chino, Eli Lilly, Gilead, GlaxoSmithKline, Janssen, Mitsubishi Tanabe, Mundipharma, Novartis, Novo Nordisk, Otsuka, Traver, Tricida, UdoptDate; and Advisory or Leadership Role: Steering committees for Bayer, Chino, GlaxoSmithKline, Janssen, Novartis, Novo Nordisk, Otsuka, Pfizer, Traver; 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 Jorg Powlitschko 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


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

Writing – original draft: Hiddo L. Heerspink.


DATA SHARING STATEMENT

Previously published data were used for this study.

SUPPLEMENTAL MATERIAL


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.

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) DAPO-CKD trial. (B) CREDEnCE trial. (C) FIDELIO-DKD trial. (D) SONAR trial. (E) RENAAL trial. (F) IDNT trial.


Supplemental Excel File 1. ADSI—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


**AFFILIATIONS**

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