# Hierarchical composite endpoints: the future of randomized trials of kidney disease?

**Natalie Staplin, PhD[1] and Parminder K. Judge, PhD[1,2]**

[1] Clinical Trial Service Unit and Epidemiological Studies Unit (CTSU), Nuffield Department of Population Health, University of Oxford, Oxford, UK

[2] Oxford Kidney Unit, Churchill Hospital, Oxford University Hospitals NHS Foundation Trust, Oxford, UK

Correspondence to:

Natalie Staplin

Renal Studies Group

Clinical Trial Service Unit and Epidemiological Studies Unit (CTSU)

Nuffield Department of Population Health

Richard Doll Building, Old Road Campus,

Roosevelt Drive, Oxford OX3 7LF, UK

Email: natalie.staplin@ndph.ox.ac.uk

Phone: +44 1865 743743

Fax: +44 1865 743985

Word count

Main body: 1534 words (excluding references)

There is a substantial unmet need for new effective and safe therapies in rare kidney diseases, including focal segmental glomerulosclerosis (FSGS). Event driven trials using traditional clinical endpoints are generally not feasible when studying rare kidney diseases, so there are ongoing efforts (e.g. PARASOL) to identify suitable surrogates. Hierarchical composite endpoints (HCEs), analysed using Win statistics (i.e. win ratio/odds),[1,2] rank outcomes in order of clinical severity, and then compare patients between arms so that more important events dominate less important ones. HCEs that combine clinical outcomes with biomarkers such as eGFR slope and urine albumin:creatinine ratio (uACR) have been identified as potential candidates for primary or secondary outcomes in future randomized trials of kidney diseases, with further evaluation needed in rare diseases.[3]

In this issue of CJASN, Heerspink et al.[4] investigate the suitability of HCEs as outcomes in trials of patients with FSGS, using data from 115 such participants randomized into the DAPA-CKD trial. The authors establish that for a given sample size, the statistical power of a CKD progression HCE is higher than that for a clinical composite or eGFR slope outcome. This is particularly important for rare disease trials where numbers of patients that can be recruited is limited. The authors also presented exploratory analyses in DAPA-CKD where the CKD progression HCE suggested a benefit for dapagliflozin (win odds 1.52 [95% confidence intervals [CI] 1.01, 2.28]), but the 95%CIs for the primary composite endpoint (sustained 50% eGFR decline, kidney failure, or death due to kidney or cardiovascular causes, hazard ratio 0.45 [95%CI 0.13, 1.49]), and total eGFR slope (difference in means 0.9 mL/min/1.73m2 [95%CI -0.6, 2.3]) crossed unity (but were consistent with the overall trial results).

One of the main strengths of this study is the use of simulations to compare the sample size requirements of kidney progression HCEs with a traditional clinical composite outcome and eGFR slope. A kidney progression HCE with a regression in uACR to <0.32 g/g, in line with the findings from the PARASOL project, is considered. However, a key limitation is the post-hoc analysis of a small unbalanced subset of patients from a trial that was not focussed on FSGS, meaning that the studied population is unlikely to be representative of those that would be recruited for a FSGS trial. The simulations conducted during the study demonstrate that at least 225 FSGS patients would be needed for a reasonably well powered trial using a HCE (and larger sample sizes still for eGFR slope and composite clinical outcomes), so the subset of 115 patients from DAPA-CKD is greatly underpowered to reliably compare results from the different outcomes. Nonetheless, HCEs merit further consideration in trials of kidney diseases.

HCEs have some advantages over traditional composite outcomes that make them a potential candidate for use in trials of kidney disease. It makes intuitive sense to prioritise the most severe outcome for each participant rather than the one that occurs first chronologically, and it is appealing for trial participants to know they can contribute information to the analysis even if they don't experience a clinical event. However, there are also a number of challenges of using HCEs as outcomes in trials of kidney disease, whether in general CKD populations or for rare diseases, which are particularly noteworthy.

*Key limitation: Reduced safety information*

Most importantly, the reduced sample sizes and shorter trial follow-up durations that can be achieved using HCEs will lead to substantially reduced collection of safety outcome information, however, "there is no surrogate for safety".[5]

*Selection of clinically meaningful components*

The components included in the hierarchy should be clinically meaningful for the target population and intervention being studied. All of the HCEs considered in Heerspink et al include all-cause mortality at the top of the hierarchy. Whilst death is a highly important outcome for patients, its inclusion in the hierarchy may decrease statistical power if the treatment is not expected to affect mortality, particularly if rates of death are high.[6] This is not an issue for the analyses conducted in Heerspink et al as consistent effects on the mortality component are observed for sodium glucose co-transporter 2 inhibitors (SGLT2i). However, this is unlikely to be the case for most treatments. Removing all-cause mortality from the hierarchy (or restricting to a subset of deaths, such as CV deaths) presents other challenges as the deaths removed from the hierarchy would then be a competing risk for the remaining outcomes.[6]

The ordering of components is essential for interpretation of the win ratio as a win on the first outcome could have a substantial impact compared with a win on the final outcome.[7] Given that the majority of participants will not experience a relevant clinical event during the trial, they will contribute to the analysis with eGFR slope or reductions in uACR, neither of which are currently universally validated surrogates for trials of kidney disease. These components are ranked at the bottom of the hierarchy due to their lesser clinical importance but it is possible for these surrogate markers to exaggerate a treatment effect if most wins are decided using this element of the hierarchy.[7]

*Lack of interpretability/generalizability*

Another major limitation of HCEs is the challenges with interpretation of the win statistics used to analyze them. For example, the win ratio can only be interpreted at the population level: "what is the ratio of the probability of obtaining a better outcome to the probability of a worse outcome for a subject exposed to treatment 1 vs another subject exposed to treatment 0".[8] Given that the definition of a 'win' or better outcome is usually highly heterogeneous, ranging from a small improvement in a biomarker e.g. annual rate of change in eGFR (which might not be considered clinically meaningful) to a substantial improvement in overall survival, win statistics do not provide information on the magnitude, or importance, of the treatment effect on a particular outcome. Therefore, it is extremely challenging for clinicians to be able to understand the clinical impact of a treatment based solely on a win statistic, and makes it impossible to communicate these results to patients. Analyses using traditional composite endpoints or eGFR slope are better understood by clinicians, allowing for clear communication of trial data to guide individualised patient care. Trials using HCEs would require analyses of each separate component of the hierarchy to understand the likely size and direction of the treatment effect (which will have higher levels of uncertainty due to the smaller sample sizes generally required for trials using a HCE as the primary outcome).[7]

The interpretability and generalizability of win statistics are also limited by their dependence on the variance of the continuous components, censoring distribution and follow-up time.[1,8] For example, a treatment could be associated with a difference of 1 ml/min/1.73m$^2$/year in eGFR slope irrespective of how the individual slopes were calculated, but the estimated win ratio value would be larger for an approach that produces less variability in the slopes.[8] This is further complicated by the absence of an agreed approach for calculating individual slopes. This dependence on study design characteristics limits comparisons across multiple studies especially in meta-analyses.[1]

*Technical issues when implementing HCEs*

Issues with non-transitivity are a particular concern for HCEs (where paradoxical findings could be observed such as A wins vs B, B wins vs C, but A loses vs C), which are usually addressed by using a fixed follow-up duration.[9] In trials with staggered entry and variable follow-up, this can be achieved by establishing minimum follow-up durations to be used in the analysis. This means any information collected after the fixed follow-up time may not be used in the primary analyses, which is problematic if a larger contribution of clinical outcomes is expected in those with longer follow-up (which is often observed in trials of kidney disease e.g. with kidney failure outcomes developing later in a trial than changes in eGFR slope). This can lead to inappropriate conclusions when there is an early benefit but late harm of treatment (or vice versa).[1] However, as noted in Heerspink et al[4],

analyses using shared follow-up approaches have generally yielded similar results to fixed follow-up approaches in DAPA-CKD and other trials studied so far.

It should also be noted that while there is a role for dedicated trials of novel therapies targeted at specific subtypes of FSGS, for treatments targeting final common kidney disease pathways, inclusion of patients with FSGS into large scale CKD trials will allow assessments of effects on clinical outcomes whilst also providing more reliable safety data. Relative effect sizes are often broadly similar irrespective of the underlying cause of kidney disease, as was demonstrated in the trials of SGLT2i.[10,11]

In summary, HCEs offer some advantages to traditional composite outcomes and other surrogate markers that make them appealing alternatives for trials of kidney disease, including rare conditions such as FSGS, but are substantially limited by a number of major interpretational and technical challenges. While they may provide more power for a given sample size in some trials, the lack of a clear interpretable measure of the treatment effect for both clinicians and patients mean that they are not a panacea for future trials of kidney disease and traditional composites endpoints will remain the standard in clinical trials for the foreseeable future.

# REFERENCES


1. Cunningham JM, Fuest S, Burke H. Win Ratio Method for Hierarchical Composite Outcomes in Randomized Clinical Trials. *JAMA*. 2026;335(8):715-716. doi:10.1001/jama.2025.25686

2. 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. *European Heart Journal*. 2011;33(2):176-182. doi:10.1093/eurheartj/ehr352

3. International Society of Nephrology's International Consensus Meeting on Changing Paradigms of Studies in Kidney D, Levin A, Jaure A, et al. Changing paradigms of studies in kidney diseases. *Kidney International*. 2025;108(6):1002-1017. doi:10.1016/j.kint.2025.09.016

4. Heerspink HJL, Little D, Gasparyan SB, et al. A Hierarchical Composite Endpoint for Chronic Kidney Disease Progression in Patients with Focal Segmental Glomerulosclerosis. *CJASN*.

5. Temple R. Are Surrogate Markers Adequate to Assess Cardiovascular Disease Drugs? *JAMA*. 1999;282(8):790-795. doi:10.1001/jama.282.8.790

6. Redfors B, Gregson J, Crowley A, et al. The win ratio approach for composite endpoints: practical guidance based on previous experience. *European Heart Journal*. 2020;41(46):4391-4399. doi:10.1093/eurheartj/ehaa665

7. Butler J, Stockbridge N, Packer M. Win Ratio: A Seductive But Potentially Misleading Method for Evaluating Evidence from Clinical Trials. *Circulation*. 2024;149(20):1546-1548. doi:doi:10.1161/CIRCULATIONAHA.123.067786

8. Thomsen HF, Gasparyan SB, Furberg JF, et al. Interpretational challenges of the Win Ratio in analyzing Hierarchical Composite Endpoints in Chronic Kidney Disease. 2504.05909. 2025. Accessed 24/02/2026. https://arxiv.org/abs/2504.05909

9. Little DJ, Gasparyan SB, Schloemer P, et al. Validity and Utility of a Hierarchical Composite End Point for Clinical Trials of Kidney Disease Progression: A Review. *Journal of the American Society of Nephrology*. 2023;34(12):1928-1935. doi:10.1681/asn.0000000000000244

10. Judge PK, Staplin N, Mayne KJ, et al. Impact of primary kidney disease on the effects of empagliflozin in patients with chronic kidney disease: secondary analyses of the EMPA-KIDNEY trial. *The Lancet Diabetes & Endocrinology*. 2024/01/01/ 2024;12(1):51-60. doi:https://doi.org/10.1016/S2213-8587(23)00322-4

11. The Nuffield Department of Population Health Renal Studies Group, The SGLT2 inhibitor Meta-Analysis Cardio-Renal Trialists' Consortium. Impact of diabetes on the effects of sodium glucose co-transporter-2 inhibitors on kidney outcomes: collaborative meta-analysis of large placebo-controlled trials. *The Lancet*. 2022;400(10365):1788-1801. doi:10.1016/S0140-6736(22)02074-8