Check for updates

OPEN LETTER

REVISED Questioning the kidney-protective effects of tolvaptan: a testable hypothesis to revisit the current narrative

[version 2; peer review: 3 approved]

Previously titled: "Questioning the kidney-protective effect of tolvaptan: a biologically plausible hypothesis to revisit the current narrative"

Dion Groothof[1], Nicole S. Erler[2], Stephan J.L. Bakker[1]

[1]Internal Medicine, University Medical Center Groningen, Groningen, Groningen, 9713 GZ, The Netherlands
[2]Julius Center for Health Sciences and Primary Care, UMC Utrecht, Utrecht, The Netherlands

v2 **First published:** 04 Jun 2025, **5**:152
https://doi.org/10.12688/openreseurope.20428.1
Latest published: 16 Jul 2026, **5**:152
https://doi.org/10.12688/openreseurope.20428.2

Abstract

Tolvaptan is the only approved therapy for slowing disease progression in autosomal dominant polycystic kidney disease (ADPKD). Its use necessitates balancing kidney-protective effects against burdensome aquaretic side effects that limit long-term adherence. Both clinical and real-world data highlight that many eligible patients decline or discontinue therapy, complicating estimation of the true magnitude of kidney protection inferred from pivotal trials. In this perspective article, we hypothesise that two distinct sources of bias may influence estimates of treatment efficacy in opposite directions and explore the implications for the interpretation of trial results. First, we consider how treatment-related effects, particularly changes in creatinine generation, could bias creatinine-based estimated glomerular filtration rate (eGFR), potentially exaggerating the apparent benefit. Second, we evaluate how selective treatment discontinuation in the TEMPO3:4 trial may have distorted estimated treatment effects by altering which participants remain under observation over time, potentially attenuating the apparent benefit. Taken together, these considerations motivate a reassessment of placebo-adjusted treatment effects on eGFR using cystatin C measurements from TEMPO3:4, thereby enabling a within-trial, trajectory-based evaluation of potential treatment-induced changes in creatinine generation, supported by analytical approaches that better account for selective dropout. More broadly, our analysis extends beyond ADPKD, highlighting the need to re-evaluate existing clinical trials in which treatment may change creatinine generation, including those with

Open Peer Review

Approval Status ✓ ✓ ✓

	1	2	3
version 2 (revision) 16 Jul 2026	✓ view	✓ view	✓ view
version 1 04 Jun 2025	? view	? view	? view

1. **Michiel Betjes**, Erasmus Medical Center, Rotterdam, The Netherlands
2. **Amaryllis van Craenenbroeck**, KU Leuven, Louvain, Belgium
3. **Ellen Hoogeveen**, Leiden University Medical Center, Leiden, The Netherlands

Any reports and responses or comments on the article can be found at the end of the article.

sodium-glucose cotransporter-2 inhibitors and glucagon-like peptide-1 receptor agonists. Future trials should incorporate complementary filtration markers to assess treatment-related changes in creatinine generation. Across all trials, sensitivity analyses using contemporary modelling strategies are warranted to strengthen confidence in eGFR slope as a regulator-endorsed clinical endpoint. Improving the methodological rigour with which treatment effects are estimated is essential to ensure that clinical decisions accurately reflect therapeutic benefit relative to treatment burden and patient experience.

Keywords
ADPKD, Tolvaptan, Creatinine, Cystatin C, Muscle mass, Kidney function, Confounding



This article is included in the Horizon 2020 gateway.



This article is included in the Marie-Sklodowska-Curie Actions (MSCA) gateway.

Corresponding author: Stephan J.L. Bakker (s.j.l.bakker@umcg.nl)

Author roles: Groothof D: Conceptualization, Investigation, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing; **S. Erler N**: Methodology, Validation, Writing – Review & Editing; **Bakker SJL**: Conceptualization, Funding Acquisition, Investigation, Project Administration, Supervision, Writing – Original Draft Preparation, Writing – Review & Editing

Competing interests: No competing interests were disclosed.

Grant information: The senior scientist was supported by funding from the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie Actions, grant agreement 754 425 (Personalized medicine in diabetic chronic disease management)
The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Copyright: © 2026 Groothof D *et al*. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

How to cite this article: Groothof D, S. Erler N and Bakker SJL. **Questioning the kidney-protective effects of tolvaptan: a testable hypothesis to revisit the current narrative [version 2; peer review: 3 approved]** Open Research Europe 2026, **5**:152 https://doi.org/10.12688/openreseurope.20428.2

First published: 04 Jun 2025, **5**:152 https://doi.org/10.12688/openreseurope.20428.1

REVISED **Amendments from Version 1**

This revised version substantially expands and refines the original Open Letter in response to peer-review comments. The manuscript has been revised to more clearly position the hypothesis as testable rather than purely biologically plausible, and to explicitly situate it within the broader methodological context of treatment-effect estimation in longitudinal kidney disease trials.

Major additions include expanded discussion of the balance between treatment burden and efficacy, a revised and expanded Figure 1, and a new Box 1 providing a quantitative estimation of the magnitude of reduction in endogenous creatinine generation (approximated by muscle-mass loss) required to account for the observed effect of tolvaptan on creatinine-based eGFR in the TEMPO3:4 trial under stated assumptions.

Dedicated new sections have been added addressing: (i) approaches to evaluate potential confounding related to changes in muscle mass, muscle composition, and creatinine generation; (ii) detection of clinically relevant treatment-induced changes in creatinine generation in clinical trials, including quantitative sample-size considerations; and (iii) missing data, informative dropout, and model validity in longitudinal analyses of kidney function decline. The concluding section has also been substantially expanded to provide actionable recommendations for re-analysis of previous trials and design considerations for ongoing and future studies, including broader implications beyond ADPKD and tolvaptan.

Finally, a detailed Supplementary Appendix has been added describing the assumptions and calculations underlying the sample-size estimates presented in the manuscript.

The revised version also includes the addition of Dr Nicole Erler as co-author in recognition of substantial contributions to the statistical and methodological sections addressing detectability analyses, informative dropout, missing-data mechanisms, and model validity.

Any further responses from the reviewers can be found at the end of the article

Disclaimer

The views expressed in this article are those of the author(s). Publication in Open Research Europe does not imply endorsement of the European Commission.

Introduction

Autosomal dominant polycystic kidney disease (ADPKD) is an inherited systemic disorder affecting approximately 10% of patients receiving kidney replacement therapy.[1] The disease is characterised by the formation of numerous fluid-filled cysts (Figure 1), which already begin to develop in utero. Despite this early onset, most individuals remain asymptomatic for several decades. Over time, cyst growth causes increased kidney volume, irreversible disruption of the kidney parenchyma, and ultimately a relentless progression towards kidney failure, creating a constant source of uncertainty and distress for affected families.[1,2] In this context, the introduction of tolvaptan, a selective vasopressin receptor 2 antagonist, represented a significant milestone as the first—and currently the only—approved treatment shown to slow disease progression in ADPKD. Clinical trial data suggested that tolvaptan could reduce kidney growth by nearly 50% and slow the rate of decline in estimated glomerular filtration rate (eGFR) by approximately 26% over a three-year period.[3] This supported its approval for ADPKD in multiple countries, including Australia, Canada, Europe, and Japan. However, the US Food and Drug Administration (FDA) initially withheld approval due to concerns about potential liver injury and high dropout rates in the tolvaptan group, raising doubts about the reliability of the trial data and the true extent of benefit on kidney function preservation.[4] The Replicating Evidence of Preserved Renal Function: An Investigation of Tolvaptan Safety and Efficacy (REPRISE) trial addressed these concerns, showing that tolvaptan slowed the decline in eGFR by 1.27 ml/min per 1.73 m^2 compared with placebo (Figure 1). The FDA thereafter approved tolvaptan, but concerns about its long-term efficacy remain.[5,6] We share these concerns and hypothesise that the observed kidney benefits of tolvaptan were, at least in part, confounded by therapy-related muscle loss, leading to an artefactual appearance of preserved GFR with tolvaptan. At the same time, there is reason to suspect that participant dropout in the pivotal trials was informative, raising the possibility that the observed treatment effects on change in eGFR were in fact underestimations of effects on true GFR. Together, these opposing yet plausible sources of bias highlight substantial uncertainty regarding both the direction and true magnitude of tolvaptan's effect on GFR preservation. A reassessment of its impact on change in eGFR using contemporary analytical approaches capable of disentangling biological signal from measurement artefact and trial-related bias is therefore warranted.

This article substantiates the necessity of such a reassessment and is structured to provide a comprehensive and integrated evaluation of the biological and methodological factors that may have influenced the observed effects of tolvaptan on kidney disease progression in ADPKD. We first revisit the benefit-burden balance of tolvaptan therapy to establish the clinical context in which treatment decisions are made. We then examine mechanisms through which treatment-related

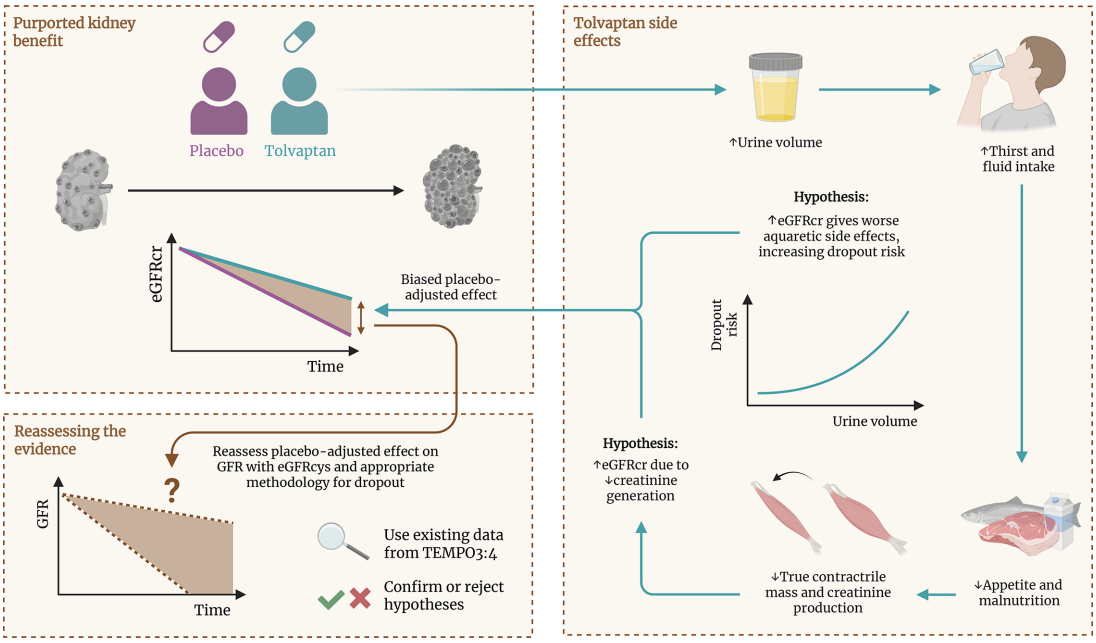


Figure 1. Hypothesized mechanism underlying the apparent kidney-protective effect of tolvaptan. We propose two distinct mechanisms through which tolvaptan therapy may have biased the estimated kidney-protective effects derived from eGFRcr in pivotal trials. First, selectively antagonising the vasopressin receptor 2 with tolvaptan leads to marked polyuria,[6] sometimes exceeding 12 litres per day,[7] necessitating substantial fluid consumption to match urine output. Increased fluid consumption may reduce appetite,[3] which, over time, can contribute to malnutrition and, consequently, gradual loss of metabolically active muscle mass. Loss of muscle mass reduces the total creatine pool, leading to lower creatinine generation and creatinine concentrations. Because creatinine concentrations are inversely related to eGFRcr, therapy-related reductions in creatinine generation may have led to an apparent preservation of eGFRcr upward, irrespective of actual changes in GFR,[8] thereby exaggerating the observed kidney-protective effect of tolvaptan. Second, higher urine volumes may also increase the risk of treatment discontinuation. Participants experiencing the most pronounced aquaretic effects, who may also have had more preserved baseline eGFRcr and higher urine output,[9] were therefore more likely to discontinue treatment and drop out prematurely. This selective attrition may have preferentially removed participants with more preserved eGFRcr from the tolvaptan arm, biasing the placebo-adjusted eGFRcr slope difference towards the null. Together, these biological and methodological pathways motivate a reassessment of the placebo-adjusted treatment effects on eGFRcr and eGFR derived from cystatin C, combined with analytical approaches that appropriately account for informative dropout. eGFRcr, creatinine-based estimated glomerular filtration rate.

effects could have influenced eGFR slope estimates in pivotal trials, and outline analytical approaches to evaluate and either confirm or exclude the presence of such effects. Next, we address a second potential source of bias by examining how missing follow-up data and treatment discontinuation arise in longitudinal studies and why standard analytical approaches may be compromised when discontinuation is related to unobserved outcomes, with specific reference to the design and analysis of past tolvaptan trials. Finally, we conclude with a structured synthesis that considers the implications for the interpretation of past, ongoing, and future trials and provide actionable directions for future research.

Benefit-burden balance of tolvaptan

In clinical practice, the drug's protective effect on kidney function—delaying progression to kidney failure, on average, from 6.2 to 9.0 years[10]—must be balanced against its most burdensome side effect: polyuria of up to 12 litres per day.[6,7,11] Despite additional challenges such as disrupted sleep due to nightly urinating and a constant need of drinking water, initial reports indicated that 90% of patients accepted the treatment, often motivated by witnessing ADPKD complications in relatives.[6] However, real-world data reveal a more nuanced picture. While tolerability within the first year of tolvaptan use aligns with clinical trial findings,[12,13] discontinuation rates rise substantially after three years.[11,13] This suggests that long-term adherence to tolvaptan may be more challenging than early trials implied.

An even more striking insight from real-world data is the observation that a substantial proportion of patients at high risk of rapid disease progression identified as candidates for tolvaptan never initiate treatment. Over 57% of eligible patients in the UK and 69% in Canada choose not to initiate therapy.[11,14] The anticipated impact of tolvaptan's aquaretic effect on daily life, requiring up to 12 litres of fluid intake to match urine output,[7,11] is cited as the primary reason for this

reluctance.[14] While patients who tolerate tolvaptan beyond three months often report minimal impact on quality of life, the experience of those unable to tolerate the drug remains unexplored. Similarly, no studies have examined the potential psychological or emotional burden on patients who decline treatment.[15] For these individuals, and perhaps their families as well, the inability to adhere to a therapy that could delay kidney failure may lead to feelings of guilt or self-blame, negatively impacting their quality of life. This underscores the importance of rigorously testing hypotheses that challenge the kidney-protective effects observed in clinical trials.

Muscle loss-related confounding of creatinine-based eGFR

A critical aspect of evaluating tolvaptan's efficacy lies in understanding potential adverse effects on muscle mass and the resulting confounding of kidney function surrogates.[7] The Tolvaptan Efficacy and Safety in Management of Autosomal Dominant Polycystic Kidney Disease and Its Outcomes (TEMPO)3:4 and REPRISE trials assessed tolvaptan's effect on kidney function by measuring changes in creatinine and creatinine-based eGFR.[3,10] Creatinine is formed by the spontaneous, non-enzymatic dehydration of creatine and phosphocreatine, which are predominantly present in skeletal muscle.[16] Consequently, creatinine concentrations are directly determined by whole-body creatine pools and thus closely related to muscle mass. Once released into the circulation, creatinine is cleared almost exclusively by glomerular filtration, such that its steady-state concentration is inversely proportional to the GFR. This property makes creatinine a useful surrogate marker of GFR, provided that variation in creatinine production due to differences in muscle mass is adequately accounted for. Because muscle mass systematically varies with demographic characteristics, declining with advancing age and being, on average, lower in females than in males, creatinine concentrations also vary with age and sex independent of GFR. Estimating equations for GFR incorporate demographic variables in an attempt to account for this variability in creatinine generation, thereby providing a creatinine-based eGFR (eGFRcr) that more accurately reflects true GFR than creatinine concentrations alone.[17] Importantly, in light of the relationship between creatinine concentrations and muscle mass, both the European Medicines Agency (EMA) and FDA recommend excluding the possibility of treatment-related changes in muscle mass before evaluating treatment effects on eGFRcr in clinical trials, because such changes could alter creatinine concentrations irrespective of actual changes in GFR and hence confound eGFRcr values.[18,19] Neither the TEMPO3:4 nor the REPRISE trial explicitly reported whether this recommendation was followed or whether treatment-related muscle-mass changes were assessed and excluded.[3,10]

Against this background, it is important to consider whether tolvaptan itself could plausibly influence muscle mass or creatinine generation through its known side-effect profile. The substantially increased urinary output caused by tolvaptan induces thirst and leads to excessive fluid intake (Figure 1). To maintain fluid balance under these conditions, patients may preferentially consume palatable beverages rather than plain water, particularly when required volumes exceed several litres per day. Such choices can increase intake of sugar-sweetened or calorie-containing beverages. Regardless of how increased fluid intake is achieved, the markedly increased fluid consumption itself can suppress appetite—a possibility supported by the higher incidence of anorexia reported with tolvaptan treatment compared with placebo.[3] Appetite suppression may precede malnutrition and, over time, lead to muscle-mass loss and reduced creatinine generation, reducing blood creatinine concentrations independent of GFR (Figure 1).[7,20] Additionally, fatigue resulting from disrupted sleep and frequent urination may discourage physical activity, further contributing to the progressive muscle loss and a concomitant reduction in metabolic rate. Although body weight remained stable during the TEMPO3:4 trial,[3] this does not exclude the possibility that the observed effects of tolvaptan on changes in eGFRcr were influenced by reductions in muscle mass or, alternatively, by reductions in creatinine generation occurring despite preserved muscle mass. This distinction is crucial, because it is changes in creatinine generation, not muscle mass per se, that can bias eGFRcr. In both instances where metabolically active muscle mass and metabolic rate decline, the body enters a new steady state in which muscle mass stabilises at a lower level. Any initial muscle loss following treatment initiation would accordingly be expected to plateau over time rather than progress indefinitely. Importantly, even a transient reduction in creatinine generation early after treatment initiation would establish a new state of lower creatinine concentrations. While similar effects could in principle occur later during treatment, early changes are particularly relevant because they would persist throughout follow-up and thereby shift entire longitudinal eGFR trajectories rather than producing only transient deviations at individual time points.

To assess whether such changes could have meaningfully influenced trial outcomes, it is crucial to consider the magnitude of muscle loss that would be required to account for the treatment effect observed in the TEMPO3:4 trial.[3] A quantitative estimation of the degree of muscle-mass loss that could plausibly underlie the observed treatment effect on eGFRcr is provided in **Box S1** of the **Supplementary Appendix**.[21] Assuming that the reported between-group difference in the annual eGFRcr slope (estimated at −0.98 ml/min/1.73 m^2) was entirely attributable to reduced endogenous creatinine generation rather than true preservation of GFR, back-calculation from the trial data[3] indicates that a yearly decline in creatinine generation of approximately 1% would be sufficient. In absolute terms, this corresponds to a loss of approximately 0.33 kg of skeletal muscle per year in a typical male with 31.6 kg of muscle,[22] assuming a body weight

equal to the mean body weight in the tolvaptan arm. If body weight were to remain stable, this degree of muscle loss would necessarily be accompanied by a compensatory increase in fat mass of approximately 1 kg over the 3-year follow-up period of TEMPO3:4.[3] Such a pattern is biologically plausible in the context of reduced physical activity, a lowered metabolic rate, and increased consumption of sugar-sweetened beverages to meet the substantially increased fluid requirements associated with tolvaptan therapy.[23,24]

These considerations underscore that even modest, clinically plausible changes in muscle mass or muscle composition could meaningfully influence eGFRcr. Importantly, if such changes were sufficient to account for the apparent GFR preservation observed in the TEMPO3:4 trial,[3] they also raise the possibility that a detrimental effect of tolvaptan on true GFR may have been concealed. In particular, the drug's aquaretic effects may predispose susceptible patients to episodes of volume depletion or hypertonicity during intercurrent illness (e.g., vomiting, excessive sweating, or diarrhoea) or during periods of inadequate fluid intake if treatment is not intermittently discontinued.[25,26] Some patients may consequently experience faster disease progression due to emerging hypertonicity-related nephropathy.[26–28]

While episodes of volume depletion or hypertonicity could exacerbate kidney injury, the frequency, severity, and clinical impact of such events in tolvaptan-treated patients with ADPKD remain undefined. At present, only indirect evidence from heat-stress nephropathy is available,[26–28] which does not substitute for ADPKD-specific cohort data. Prospective monitoring of electrolytes, osmolality, and acute GFR changes during intercurrent illnesses would be necessary to substantiate this concern. Nonetheless, even this theoretical possibility of accelerated kidney injury underscores the need for analytical approaches capable of distinguishing true treatment effects on GFR from artefacts introduced by changes in creatinine generation.

Approaches to evaluate potential confounding related to changes in muscle mass and creatinine generation in clinical trials

Given the plausible dissociation between body weight, muscle mass, and endogenous creatinine generation highlighted above, potential drug effects on muscle mass, muscle quality, and creatinine generation must be reliably assessed and either confirmed or confidently excluded. When a drug effect on true GFR is approximated by evaluating longitudinal changes in eGFRcr, creatinine generation must remain stable or any co-occurring changes must be appropriately accounted for.[18,19]

When evaluating the stability of creatinine generation over time during drug treatment, muscle mass and endogenous creatinine generation are often considered interchangeable, but this assumption does not always hold. Muscle mass and creatinine generation can be differentially affected by changes in muscle composition, including variations in inter-myocellular fat (extracellular adipose tissue beneath the muscle fascia between myocytes), intramyocellular lipid content (lipid droplets within myocytes), and total muscle water content. In situations where a drug introduces clear and concordant reductions in both total body fat and muscle mass, changes in creatinine generation are likely to closely track changes in muscle mass. Notable examples are clinical trials with inhibitors of sodium-glucose cotransporter-2, as well as the glucagon-like peptide-1 (GLP-1) receptor analogue semaglutide, the glucose-dependent insulinotropic polypeptide (GIP) and GLP-1 receptor agonist tirzepatide, and the GIP, GLP-1, and glucagon receptor agonist retatrutide.[29–34] Under these circumstances, standard techniques that primarily assess muscle mass rather than total body creatine pools, including magnetic resonance imaging (MRI), computed tomography scanning, dual-energy X-ray absorptiometry, bio-electrical impedance analysis, and muscle ultrasonography, can be used to estimate the magnitude of changes in muscle mass and, by extension, creatinine generation.[35–37]

By contrast, a different scenario arises when a drug reduces whole-body creatinine generation independent of measurable changes in muscle mass or volume. In such cases, muscle mass as assessed by standard techniques may appear unchanged, while creatinine generation has in fact decreased. Drugs with such properties that have been evaluated in the renal field include the synthetic glucocorticoids budesonide and methylprednisolone.[38–40] Glucocorticoids are well known for their ability to induce muscle atrophy[41] alongside insulin resistance and a redistribution of body fat to promote fat infiltration of skeletal muscle.[42] Accordingly, glucocorticoid exposure may create a dissociation between apparent muscle mass and true contractile tissue content.

A similar situation could plausibly occur with tolvaptan. Reduced protein intake, decreased physical activity, and increased consumption of sugar-sweetened beverages to meet high fluid requirements could result in stable body weight while promoting a redistribution of fat from adipose tissue to intermuscular and intramyocellular compartments. This pattern converges on a muscle-composition phenotype that is well described in insulin-resistant states and in conditions associated with reduced physical activity. Under these conditions, muscle mass may appear preserved, while whole-body creatinine generation is nonetheless reduced. It is in this context that techniques that assess muscle mass without

specifically quantifying total body creatine pools or creatinine generation are insufficient to exclude a clinically relevant drug effect on creatinine generation. Instead, methods capable of directly assessing total body creatine pools or creatinine generation are required. Among imaging techniques, only MRI Dixon (to quantify muscle volume) combined with ^{1}H-magnetic resonance spectroscopy (for intramyocellular lipid and intramuscular creatine concentration) and ^{31}P-magnetic resonance spectroscopy (for intramuscular phosphocreatine concentration) allow comprehensive assessment of the total endogenous creatine pool.[43–46]

Detection of a clinically relevant drug effect on creatinine generation in clinical trials

Although the magnitude of change in creatinine generation required to confound eGFRcr is small and biologically plausible, confidence in the interpretation of creatinine-based trial outcomes depends on whether such changes can be empirically detected or confidently excluded using feasible study designs of relatively short duration. The yearly decline in creatinine generation of approximately 1% that would be sufficient to explain the observed treatment effect on eGFRcr in TEMPO3:4 corresponds to detectability of a 0.5% change in creatinine generation during a six-month study.

Translating this detectability threshold into a sample size estimate for a study requires assumptions regarding both the baseline size of the total endogenous creatine pool and the variability with which treatment-induced changes in that pool can be measured over time. Below, we focus on techniques that allow direct or near-direct quantification of changes in endogenous creatinine. To illustrate feasibility, we present a single condensed sample size calculation, while summarising estimates for alternative approaches. We show that required sample sizes vary substantially depending on the method used to assess creatinine generation and refer to the **Supplementary Methods**[21] for full derivations, methodological assumptions, and sample size calculations.

Reproducibility studies combining Dixon MRI with ^{1}H- and ^{31}P-magnetic resonance spectroscopy suggest that total-body creatine pool size, used as a proxy for creatinine generation, can be measured with relatively low within-subject variability across repeated measurements.[43–46] Assuming a baseline pool of 135 g in a 79 kg male[16] and a standard deviation of 1.0 g for individual baseline-to-6-month changes, a 0.5% difference between treatment groups over six months corresponds to an absolute difference of 0.675 g. Under these assumptions, a total sample size of 70 participants (35 per group) would be required to detect such a difference with 80% power at a two-sided α of 0.05 in an independent-samples comparison of the mean change in creatine pool between placebo and intervention arms.

Direct isotope dilution techniques, while providing near-direct quantification of total body creatine pool size, are subject to greater variability and would require larger samples (approximately 250 participants in total). An indirect, regulator-endorsed approach to assessing treatment-related changes in creatinine generation is the use of alternative filtration markers that are less dependent on muscle mass, such as cystatin C.[18,19] Because cystatin C is produced at a relatively constant rate by all nucleated cells,[47] independent of creatinine generation, regulatory agencies recommend its use as a confirmatory marker when drug-related effects on creatinine generation cannot be excluded. Comparison of creatinine-based eGFR and cystatin C-based eGFR (eGFRcys) therefore provides a practical means to detect divergence attributable to non-GFR determinants, including creatinine generation. However, approaches based on longitudinal discordance between eGFRcr and eGFRcys, although operationally simple and readily implementable in clinical trials, would require considerably larger sample sizes (on the order of several thousand participants) to detect similarly small changes over short time intervals. Lastly, estimates based on 24-hour urinary creatinine excretion fall between imaging-based and isotope dilution-based approaches (approximately 150 participants under ideal conditions), but are highly sensitive to collection error in unsupervised, free-living populations. This would likely necessitate substantially larger sample sizes in contemporary clinical trials.

Although evaluation of discordance between eGFRcr and eGFRcys trajectories requires substantially larger sample sizes than approaches based on 24-hour urinary creatinine excretion under ideal conditions, it offers a unique practical advantage. Because creatinine is already measured repeatedly to derive eGFRcr trajectories in clinical trials, cystatin C can be measured in the same blood samples, allowing longitudinal discordance between eGFRcr and eGFRcys to be evaluated at relatively low cost, without additional participant burden or the need for a separate confirmatory study. In this framework, treatment-induced changes in creatinine generation can be evaluated by testing whether the placebo-adjusted trajectory of eGFRcr diverges over time from the corresponding trajectory of eGFRcys.

In patients with ADPKD, eGFRcr and eGFRcys are moderately correlated, sharing approximately 30% of their variance,[48] consistent with their common dependence on true GFR. The remaining variance attributable to non-GFR determinants of each marker, including creatinine generation for eGFRcr, provides the statistical basis for detecting treatment-induced divergence between the two markers over time. In TEMPO3:4, creatinine as a secondary efficacy endpoint and cystatin C as a prespecified pharmacodynamic variable were both defined in the trial protocol

(Supplementary Material of the original report),[3] allowing derivation of repeated measurements of eGFRcr and eGFRcys and, in principle, assessment of the divergence between these estimates. Given the longitudinal structure and sample size of the trial, such a within-trial, trajectory-based approach would have allowed detection of a minimum cumulative treatment-induced change in endogenous creatinine generation of approximately 1.2% over three years (**Supplementary Methods S6**[21]). This estimate assumes a conservative correlation of 0.3 between longitudinal changes in eGFRcr and eGFRcys, allowing for progressive discordance due to time-varying non-GFR determinants of each marker, and is anchored to the number of participants remaining under observation at three years in TEMPO3:4.[3] Because detection of such a cumulative change corresponds to resolving an average change of approximately 0.2% per six months, this approach offers substantially greater statistical power and practical feasibility than conducting a separate short-term study with two repeated measurements.

Taken together, these considerations indicate that therapy-induced changes in creatinine generation of clinically relevant magnitude are not only biologically plausible but also empirically detectable within existing trial datasets. However, whether such signals can be validly interpreted depends critically on the assumptions underlying models for longitudinal data when follow-up is incomplete.

Missing data, informative dropout, and model validity

The primary analytical approach used in the TEMPO3:4 trial relied on linear mixed-effects models.[3] While such models are widely used in clinical trials with repeatedly measured continuous outcomes, their validity depends on several assumptions, including that missing outcome data are missing at random (MAR).[49–51] Here, 'missing data' refers to the absence of measurements originally scheduled as part of a study, a situation that commonly occurs in clinical research due to factors such as protocol-defined measurement discontinuation, treatment cessation, withdrawal of consent, intercurrent illness, loss to follow-up including death, and logistical or administrative constraints. Under the MAR assumption, the probability that a measurement is missing depends only on information that has already been observed and is included in the model (i.e., covariates or previously observed longitudinal values). Examples of MAR in nephrology include situations in which follow-up eGFR measurements are discontinued after a patient's eGFR exceeds a predefined safety threshold, or when biomarker testing is purposefully omitted because levels of another marker fall within expected ranges. A notable example is the omission of cystatin C testing when a patient's muscle mass, serving as a proxy of creatinine generation, is considered typical. In such situations, the missing-data mechanism can be considered MAR only if the variable driving the decision to omit testing (e.g., muscle mass or its proxy) is explicitly included in the model. More generally, missingness may arise from protocol-defined criteria or from clinical decision-making based on prior information. In practice, this means that relevant observed measurements that inform the missingness (e.g., prior eGFR or related biomarkers) should be included in the modelling approach. When this assumption holds, linear mixed-effects models can provide valid inference for longitudinal treatment effects.[49]

A distinct situation arises when the probability that a measurement is missing depends on information that has not been observed, in which case missing outcome data are missing not at random (MNAR).[49] In clinical trials, this may occur when patients discontinue treatment because of worsening symptoms or declining eGFR that is not fully captured in prior measurements. In such cases, dropout is informative, with missing data depending directly on unobserved outcomes that would have been measured had follow-up continued. As a result, patients who leave the study prematurely differ systematically from those who remain, increasing the risk that analyses assuming MAR yield biased estimates of treatment effects.[49]

In practice, dropout mechanisms may reflect a combination of MAR and MNAR processes, with different factors operating simultaneously or at different stages of follow-up. Although such mixed mechanisms have been examined in methodological work,[52–54] their detailed consideration is beyond the scope of the present discussion. Regardless of the exact situation, the magnitude and even the direction of the ensuing bias in the estimation of eGFR slope depend critically on how and why follow-up data become missing. For this reason, contemporary methodological guidance, aligned with current FDA recommendations, emphasises that sensitivity analyses are an indispensable component of longitudinal trial analysis when inference relies on missing-data assumptions.[55,56] This emphasis is grounded in the recognition that, as clearly noted in the work of Rizopoulos, the observed data alone cannot distinguish between MAR and MNAR dropout mechanisms, making sensitivity analyses the only pragmatic way to evaluate the impact of potential violations of these assumptions.[51]

The extent to which such analyses can be effectively performed varies across the pivotal ADPKD trials. Although the REPRISE trial permitted partial evaluation of missing-data assumptions within an intention-to-treat framework, as follow-up measurements of eGFRcr continued after treatment discontinuation, it does likely not allow assessment of potential divergence of trajectories of eGFRcr and eGFRcys. This is because, unlike the TEMPO3:4 trial, measurement of

cystatin C was not specified in the trial protocol (Supplementary Material of the original report).[10] As a result, sensitivity analyses that jointly address missing-data assumptions and distinguish effects on true GFR from artifacts related to treatment-induced changes in endogenous creatinine generation would require post-hoc cystatin C assessment in stored samples.

In contrast to REPRISE, the design of the TEMPO3:4 trial imposes more fundamental constraints on analogous sensitivity analyses. In TEMPO3:4, follow-up measurements of eGFRcr were, by design, discontinued when treatment was stopped,[3,57] precluding observation of eGFRcr trajectories beyond the point of discontinuation. This discrepancy with the intention-to-treat framework employed in REPRISE is important, because it places greater reliance on assumptions about unobserved eGFRcr values after discontinuation. In models fit under a MAR assumption, including standard linear mixed-effects models, missing post-discontinuation eGFRcr values are implicitly extrapolated from each participant's model-estimated trajectory based on their observed data. In effect, this amounts to assuming that, conditional on the data observed up to the point of discontinuation, participants who stop treatment would, on average, have continued to follow the same eGFRcr trajectory as those who remain under observation. Admittedly, the plausibility of this assumption becomes questionable when the reasons for treatment discontinuation are related to a participant's underlying kidney function trajectory and are not fully captured by the observed data included in the model. In such situations, the assumption that participants who discontinue would, on average, have continued to follow the same trajectory as those remaining under observation may no longer be tenable. Frequent discontinuation and systematic between-group differences may further amplify the impact of such mechanisms, although these features alone do not establish informative (MNAR-type) dropout. This motivates closer examination of the extent and determinants of treatment discontinuation in TEMPO3:4.

In that trial, treatment discontinuation occurred in 221 of 961 participants (23.0%) assigned to tolvaptan compared with 67 of 483 participants (13.8%) assigned to placebo ($\chi^2 = 16.8$; $P < 0.0001$).[3] This pronounced imbalance between the two groups suggests that leaving the study may have been linked to effects or side-effects of the treatment, even if the underlying reasons for discontinuation were similar across groups but occurred more frequently in the tolvaptan arm. Combined with the possibility that reasons for leaving the study may not be fully captured by the observed data, this raises the possibility of an MNAR-type mechanism in the context of missing-data analysis.

A clinically plausible explanation for the observed differential discontinuation relates to the side-effect profile of tolvaptan. The drug's aquaretic effects, particularly severe polyuria, are a leading cause of rapid treatment discontinuation, as consistently documented in real-world cohorts.[11,14] Similar patterns were observed in TEMPO3:4, where aquaresis-associated adverse effects were reported by more than half of tolvaptan-treated participants and were the commonest reason for treatment discontinuation.[3] Importantly, urine volume during tolvaptan treatment is not randomly distributed across patients. Kramers et al. demonstrated that higher baseline measured GFR (determined using continuous infusion of ^{125}I-iothalamate and ^{131}I-hippuran) and higher eGFRcr were independently associated with higher urine volume during tolvaptan treatment.[9] Given that greater urine volume directly translates into more pronounced aquaretic symptoms, these observations suggest that participants with more preserved kidney function were more likely to experience intolerable aquaretic symptoms during tolvaptan treatment, thereby increasing the likelihood of treatment discontinuation and early study dropout (Figure 1). Importantly, consistent with observations of increasing discontinuation rates with longer-term tolvaptan use,[11,13] the burden of aquaretic symptoms may evolve over time and is only partially reflected in observed eGFR values, which themselves are subject to biological variability and measurement fluctuations.[58–60] As a result, the probability of treatment discontinuation may depend on factors not fully captured by prior observed measurements, providing a plausible basis for MNAR-type dropout mechanisms.

In the context of longitudinal analyses based on linear mixed-effects models fitted under a MAR assumption, selective discontinuation of this kind has important implications for the estimation of eGFRcr slopes. Preferential discontinuation of tolvaptan among participants with higher baseline eGFRcr would shift the composition of the remaining cohort in the tolvaptan arm towards participants with lower eGFRcr and more advanced disease, thereby leading to a steeper estimated population-level decline in eGFRcr than would have been observed in the absence of selective dropout. In contrast, discontinuation in the placebo arm is more plausibly driven by disease progression, such that participants experiencing more rapid declines in kidney function are more likely to drop out. This mechanism would similarly shift the composition of the remaining cohort in the placebo arm towards participants with higher eGFRcr and slower progression, yielding an artificially flatter estimated eGFRcr slope. Together, these opposing selection processes could, if not fully captured by the previously observed eGFR measurements and covariates included in the model, bias the placebo-adjusted treatment effect towards the null (Figure 1), resulting in underestimation of the true effect of tolvaptan on GFR decline.

A similar selection process may also help explain why the reported effect of tolvaptan on annual rate of total kidney volume growth, the primary endpoint of the TEMPO3:4 trial, appears largest early after treatment initiation and attenuates thereafter.[3] This temporal pattern has previously been discussed, particularly with respect to whether early reductions in total kidney volume growth necessarily translate into sustained effects on kidney function decline, while it has also been emphasised that such early structural changes may still reflect a meaningful disease-modifying mechanism with longer-term implications.[5,61] Importantly, while the latter interpretation supports the biological plausibility of a beneficial effect of tolvaptan, it does not preclude the possibility that concurrent mechanisms unrelated to GFR, as well as selective treatment discontinuation, may bias estimates of treatment effect on eGFRcr. From this perspective, if informative dropout is present, conventional analyses may attenuate the true effect of tolvaptan on GFR decline.

At the same time, given that creatinine concentrations (and eGFRcr values derived from those) were measured at regular 4-month intervals in TEMPO3:4,[3] it remains uncertain to what extent these evolving symptom-driven mechanisms are fully captured by prior observed measurements, and hence whether the resulting dropout process materially violates a MAR assumption. As emphasized by Rizopoulos,[51] the observed data cannot distinguish between MAR and MNAR mechanisms, underscoring the need for sensitivity analyses to evaluate the impact of potential departures from this assumption on estimates of treatment effect. Such analyses are increasingly applied in contemporary renal trials,[30,34,39] including the FLOW trial evaluating the effect of semaglutide on kidney outcomes.[30] Notably, in that study, sensitivity analyses were reported for eGFRcr but not for eGFRcys, despite the latter being less frequently measured. This discrepancy is relevant, as less frequent assessment of cystatin C increases the likelihood that prior observed values do not fully capture the processes underlying dropout. As a result, the MAR assumption may be more readily violated for eGFRcys-based analyses, increasing the likelihood of MNAR-type mechanisms and, consequently, biased estimation of eGFR slope.

Importantly, the limitations of the pivotal trial design, arising from selective treatment discontinuation and reliance on unverifiable MAR assumptions inherent to linear mixed-effects modelling in the presence of missing outcome data, were not addressed in the TEMPO4:4 open-label extension trial.[62] Because enrolment into TEMPO4:4 was restricted to participants who completed TEMPO3:4, continued follow-up was conditioned on treatment persistence during the randomised phase. As a result, inclusion of the TEMPO4:4 extension data in the analyses[62] cannot be expected to mitigate informative dropout arising during the pivotal trial[3] and may instead reinforce attenuation of the estimated treatment effects.

Taken together, these considerations indicate that selective treatment discontinuation related to unobserved (parts of) outcome trajectories may introduce bias into estimates derived from standard mixed-effects models fitted under a MAR assumption. In the setting of TEMPO3:4, the opposing selection mechanisms described above could, if not fully captured by the observed data, bias the placebo-adjusted treatment effect towards the null, leading to potential underestimation of the true treatment effect of tolvaptan on GFR decline. At the same time, because the observed data alone cannot distinguish between MAR and MNAR dropout mechanisms, it remains uncertain whether such bias materially affected the original analyses. This uncertainty underscores the importance of sensitivity analyses that both exploit creatinine and cystatin C to explicitly evaluate the impact of alternative assumptions regarding missing-data mechanisms.

Concluding remarks and future directions

In this article, we examined the benefit-burden balance of tolvaptan therapy and reviewed mechanisms through which treatment-induced effects, most notably changes in endogenous creatinine generation, could have influenced eGFRcr slope estimates in pivotal trials. We highlighted that such changes may not only exaggerate apparent kidney-protective effects when GFR is assessed with eGFRcr, but also obscure a detrimental effect on true GFR. We discussed analytical approaches to evaluate and either confirm or exclude the presence of such effects, emphasising that a within-trial, trajectory-based approach leveraging existing data from TEMPO3:4 offers the greatest statistical efficiency and practical feasibility for this purpose. We further illustrated how missing follow-up data due to treatment discontinuation can compromise standard analyses when missingness is related to unobserved (parts of) outcome trajectories. This concern was motivated by the observed imbalance in treatment discontinuation in TEMPO3:4[3] and supported by plausible biological mechanisms.

Taken together, these considerations indicate that estimates of the placebo-adjusted eGFRcr slope difference may have been influenced by two distinct sources of bias, each acting through a different mechanism. First, treatment-related changes in endogenous creatinine generation may have reduced the apparent decline in eGFRcr in the tolvaptan arm, thereby exaggerating the observed kidney-protective effect. Second, selective, arm-specific treatment discontinuation may have preferentially removed participants with better kidney function and slower progression from the tolvaptan arm, while preferentially retaining slower progressors in the placebo arm, thereby potentially attenuating the observed kidney-protective effect.

Addressing these intertwined biological and methodological challenges requires analytical approaches that can disentangle potential treatment-induced changes in creatinine generation from true kidney-protective effects, while simultaneously accounting for treatment discontinuation and loss to follow-up. Re-analyses of trials on tolvaptan and other drugs that may influence creatinine generation, along with future clinical studies, should all consider the following analytical strategies:

- Joint models for longitudinal and time-to-event data, in which biomarker trajectories and time-to-treatment discontinuation are modelled simultaneously, allowing the dependence between biomarker evolution and discontinuation to be explicitly represented rather than assumed away.[51] When relevant, competing-risk structures, preferably using cause-specific hazards,[63] should be specified to distinguish different dropout or intercurrent event processes;

- Flexible longitudinal sub-models, including spline-based representations of time, to accommodate potential non-linearity in biomarker change over time and reduce the risk of mis-specifying population-level slopes[49];

- Explicit modelling of informative dropout and intercurrent events, including competing and mixed dropout mechanisms, thereby considering treatment discontinuation and related intercurrent events as outcomes of interest rather than as nuisance processes[52–54];

- Parallel modelling of eGFRcr and eGFRcys, enabling direct evaluation of whether treatment effects differ by filtration marker and facilitating evaluation of potential treatment-induced changes in endogenous creatinine generation;

- Sensitivity analyses under alternative missing-data assumptions, to examine how estimated treatment effects change under plausible departures from MAR-type mechanisms and to characterise the range of effects compatible with the observed data.

These strategies can be implemented within either frequentist or Bayesian inferential frameworks. Fully Bayesian approaches are particularly attractive, as they allow longitudinal outcomes, dropout processes, and incomplete covariate information to be modelled jointly under a single coherent probability model. In this setting, imputation of missing covariate data is performed simultaneously with model estimation, ensuring compatibility between analysis and imputation models and appropriately propagating uncertainty due to missing data into all inferential quantities. Such approaches are now readily implementable using contemporary statistical software.[64]

The foregoing considerations have immediate implications for the design and interpretation of ongoing and future clinical trials. The ongoing HYDRO-PROTECT trial (ClinicalTrials.gov number, NCT05373264), which evaluates whether hydrochlorothiazide can be added to tolvaptan to mitigate aquaretic side effects without compromising efficacy, is directly addressing a key determinant of treatment discontinuation.[65] However, the primary analyses described in the trial protocol rely on linear mixed-effects models and cystatin C measurements have not been specified, thereby limiting the ability to evaluate treatment effects on creatinine generation or to apply joint modelling strategies of the type discussed here. Considering treatment discontinuation, the possibility of guilt and self-blame among patients who refuse to start or discontinue using tolvaptan currently remains largely conjectural, as qualitative or quantitative evidence on patient-reported emotional outcomes in this context is lacking. We recommend performance of studies (e.g., structured interviews, validated quality-of-life instruments) capable of capturing these dimensions.

More broadly, the issues highlighted here extend beyond tolvaptan and ADPKD. Trials across nephrology increasingly rely on eGFR slope as a regulator-endorsed surrogate endpoint, supported by extensive methodological evaluation and by its practical advantages for trial feasibility.[18,19,66,67] At the same time, may contemporary interventions introduce precisely the complexities discussed in this article, including acute treatment effects on eGFR, substantial treatment discontinuation, informative missingness, and treatment-related changes in body composition that may alter creatinine generation independently of true GFR. These concerns are particularly relevant for trials of therapies associated with weight loss or altered muscle composition, including SGLT2 inhibitors and incretin-based therapies such as semaglutide, tirzepatide, retatrutide, and the oral non-peptide GLP-1 receptor agonist orfoglipron.[29–34,68] In these settings, kidney outcomes and kidney-related parameters are increasingly being evaluated using creatinine-based eGFR slopes. Careful consideration of missing-data mechanisms, incorporation of complementary filtration markers such as cystatin C, and application of sensitivity analyses under alternative missing-data assumptions should therefore be integral to trial design and analysis.

Ultimately, improving the methodological rigour with which treatment effects are estimated is not merely a statistical exercise. It directly informs the balance between therapeutic benefit and treatment burden experienced by patients. Ensuring that estimates of efficacy are as unbiased and interpretable as possible is therefore essential to guide clinical decision-making and to ensure that treatment decisions are grounded in robust evidence.

Ethics and consent

Ethics and consent is not required.

Data availability

Supplementary material supporting this perspective article has been deposited in the Zenodo repository.[21]

Repository name: Zenodo.

Title of project: Supplementary Appendix: Quantitative Estimation of the Muscle-Mass Decline Required to Bias Creatinine-Based eGFR and Sample Size Considerations for Detecting Treatment-Induced Changes in Endogenous Creatinine Generation in Trials Using Creatinine-Based eGFR Outcomes.

DOI: https://doi.org/10.5281/zenodo.21158706

The repository contains **Box S1**, which presents the quantitative estimation of the muscle-mass decline required to meaningfully bias creatinine-based eGFR. It also contains the **Supplementary Methods**, which describe the methodological assumptions, derivations, and sample size calculations related to detecting treatment-induced changes in endogenous creatinine generation in creatinine-based eGFR analyses.[21]

The material is available under the terms of the Creative Commons Attribution 4.0 International license (CC BY 4.0).

References

1. Logeman C, Cho Y, Sautenet B, *et al.*: **"A sword of Damocles": patient and caregiver beliefs, attitudes and perspectives on presymptomatic testing for autosomal dominant polycystic kidney disease: a focus group study.** *BMJ Open.* 2020; **10**(10): e038005.
PubMed Abstract | Publisher Full Text | Free Full Text
2. Grantham J: **Clinical practice: Autosomal dominant polycystic kidney disease.** *N Engl J Med.* 2008; **359**(25): 1477–1485.
PubMed Abstract | Publisher Full Text
3. Torres VE, Chapman AB, Devuyst O, *et al.*: **Tolvaptan in Patients with Autosomal Dominant Polycystic Kidney Disease.** *N Engl J Med.* 2012; **367**(25): 2407–2418.
PubMed Abstract | Publisher Full Text | Free Full Text
4. Mustafa RA, Yu ASL: **Burden of Proof for Tolvaptan in ADPKD: Did REPRISE Provide the Answer?** *Clin J Am Soc Nephrol.* 2018; **13**(7): 1107–1109.
PubMed Abstract | Publisher Full Text | Free Full Text
5. Anderson CL: **Doubts about the efficacy of tolvaptan for polycystic kidney disease.** *Clin Nephrol.* 2020; **93**(6): 307–309.
PubMed Abstract | Publisher Full Text
6. Capuano I, Buonanno P, Riccio E, *et al.*: **Tolvaptan vs. somatostatin in the treatment of ADPKD: A review of the literature.** *Clin Nephrol.* 2022; **97**(3): 131–140.
PubMed Abstract | Publisher Full Text
7. Groothof D, Post A, Bakker SJL: **Tolvaptan, Kidney Function Decline, and Potential Confounding by Muscle Wasting.** *Kidney Med.* 2023; **5**(10): 100711.
PubMed Abstract | Publisher Full Text | Free Full Text
8. Groothof D, Shehab NBN, Erler NS, *et al.*: **Creatinine, cystatin C, muscle mass, and mortality: Findings from a primary and replication population-based cohort.** *J Cachexia Sarcopenia Muscle.* 2024; **15**(4): 1528–1538.
PubMed Abstract | Publisher Full Text | Free Full Text
9. Kramers BJ, van Gastel MDA, Boertien WE, *et al.*: **Determinants of Urine Volume in ADPKD Patients Using the Vasopressin V2 Receptor Antagonist Tolvaptan.** *Am J Kidney Dis.* 2019; **73**(3): 354–362.
PubMed Abstract | Publisher Full Text
10. Torres VE, Chapman AB, Devuyst O, *et al.*: **Tolvaptan in Later-Stage Autosomal Dominant Polycystic Kidney Disease.** *N Engl J Med.* 2017; **377**(20): 1930–1942.
PubMed Abstract | Publisher Full Text
11. Chong J, Harris T, Ong ACM: **Regional variation in tolvaptan prescribing across England: national data and retrospective evaluation from an expert centre.** *Clin Kidney J.* 2023; **16**(1): 61–68.
PubMed Abstract | Publisher Full Text | Free Full Text
12. Thomas M, Gois PHF, Butcher BE, *et al.*: **Treatment persistence to tolvaptan in patients with autosomal dominant polycystic kidney disease: a secondary use of data analysis of patients in the IMADJIN® dataset.** *BMC Nephrol.* 2021; **22**(1): 400.
PubMed Abstract | Publisher Full Text | Free Full Text
13. Saito R, Yamamoto H, Ichihara N, *et al.*: **Persistence of tolvaptan medication for autosomal dominant polycystic kidney disease: A retrospective cohort study using Shizuoka Kokuho Database.** *Medicine (Baltimore).* 2022; **101**(40): e30923.
PubMed Abstract | Publisher Full Text | Free Full Text
14. Calvaruso L, Yau K, Akbari P, *et al.*: **Real-life use of tolvaptan in ADPKD: a retrospective analysis of a large Canadian cohort.** *Sci Rep.* 2023; **13**(1): 22257.
PubMed Abstract | Publisher Full Text | Free Full Text
15. Todorova P, Arjune S, Hendrix C, *et al.*: **Interaction Between Determinants Governing Urine Volume in Patients With ADPKD on Tolvaptan and its Impact on Quality of Life.** *Kidney Int Rep.* 2023; **8**(8): 1616–1626.
PubMed Abstract | Publisher Full Text | Free Full Text
16. Wyss M, Kaddurah-Daouk R: **Creatine and Creatinine Metabolism.** *Physiol Rev.* 2000; **80**(3): 1107–1213.
Publisher Full Text
17. Disease K: **Improving Global Outcomes (KDIGO) CKD Work Group. KDIGO 2024 Clinical Practice Guideline for the Evaluation and Management of Chronic Kidney Disease.** *Kidney Int.* 2024; **105**(4-Suppl): S117–S314.
Publisher Full Text
18. Levey AS, Inker LA, Matsushita K, *et al.*: **GFR decline as an end point for clinical trials in CKD: A scientific workshop sponsored by the national kidney foundation and the US food and drug**

administration. *Am J Kidney Dis.* 2014; **64**(6): 821–835.
PubMed Abstract | Publisher Full Text

19. 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.
PubMed Abstract | Publisher Full Text

20. Vandewoude MFJ, Alish CJ, Sauer AC, *et al.*: **Malnutrition-sarcopenia syndrome: Is this the future of nutrition screening and assessment for older adults?** *J Aging Res.* 2012; **2012**: 651570.
PubMed Abstract | Publisher Full Text | Free Full Text

21. Groothof D, Erler NS, SJL B: *Supplementary Appendix: Quantitative Estimation of the Muscle-Mass Decline Required to Meaningfully Bias Creatinine-Based eGFR and Supplementary Methods for Detecting Treatment-Induced Changes in Endogenous Creatinine Generation.* 2026.
Publisher Full Text

22. Heymsfield SB, Arteaga C, Mcmanus C, *et al.*: **Perspective in Nutrition Measurement of muscle mass in humans: validity of the 24-hour urinary creatinine method.** *Am J Clin Nutr.* 1983; **37**(3): 478–494.
PubMed Abstract | Publisher Full Text

23. Maersk M, Belza A, Stødkilde-Jørgensen H, *et al.*: **Sucrose-sweetened beverages increase fat storage in the liver, muscle, and visceral fat depot: a 6-mo randomized intervention study.** *Am J Clin Nutr.* 2012; **95**(2): 283–289.
PubMed Abstract | Publisher Full Text

24. Cioffi CE, Welsh JA, Alvarez JA, *et al.*: **Associations of Added Sugar from All Sources and Sugar-Sweetened Beverages with Regional Fat Deposition in US Adolescents: NHANES 1999–2006.** *Curr Dev Nutr.* 2019; **3**(12): nzz130.
PubMed Abstract | Publisher Full Text | Free Full Text

25. Müller RU, Messchendorp AL, Birn H, *et al.*: **An update on the use of tolvaptan for autosomal dominant polycystic kidney disease: consensus statement on behalf of the ERA Working Group on Inherited Kidney Disorders, the European Rare Kidney Disease Reference Network and Polycystic Kidney Disease International.** *Nephrol Dial Transplant.* 2022; **37**(5): 825–839.
PubMed Abstract | Publisher Full Text | Free Full Text

26. Groothof D, Post A, Shehab NBN, *et al.*: **Adapting Clinical and Research Practice to Global Warming: Physiology-Based Interim Recommendations for Managing Hypertonicity Risk during Extreme Heat.** *Kidney360.* 2024; **5**(12): 1917–1921.
PubMed Abstract | Publisher Full Text | Free Full Text

27. Johnson R, Sánchez-Lozada L: **Chronic kidney disease: Mesoamerican nephropathy—new clues to the cause.** *Nat Rev Nephrol.* 2013; **9**(10): 560–561.
PubMed Abstract | Publisher Full Text

28. Yang X, Wu H, Li H: **Dehydration-associated chronic kidney disease: a novel case of kidney failure in China.** *BMC Nephrol.* 2020; **21**(1): 159.
PubMed Abstract | Publisher Full Text | Free Full Text

29. Groothof D, Shehab NBN, Post A, *et al.*: **SGLT2 Inhibitors, Muscle Loss, and Creatinine-Based Estimated GFR: An Integrative Conceptual Review of Renoprotection.** *Kidney Med.* 2026; **8**(4): 101297.
PubMed Abstract | Publisher Full Text | Free Full Text

30. Perkovic V, Tuttle KR, Rossing P, *et al.*: **Effects of Semaglutide on Chronic Kidney Disease in Patients with Type 2 Diabetes.** *N Engl J Med.* 2024; **391**(2): 109–121.
Publisher Full Text

31. Heerspink HJL, Sattar N, Pavo I, *et al.*: **Effects of Tirzepatide Versus Insulin Glargine on Cystatin C–Based Kidney Function: A SURPASS-4 Post Hoc Analysis.** *Diabetes Care.* 2023; **46**(8): 1501–1506.
PubMed Abstract | Publisher Full Text | Free Full Text

32. Packer M, Zile MR, Kramer CM, *et al.*: **Interplay of Chronic Kidney Disease and the Effects of Tirzepatide in Patients With Heart Failure, Preserved Ejection Fraction, and Obesity: The SUMMIT Trial.** *J Am Coll Cardiol.* 2025; **85**(18): 1721–1735.
PubMed Abstract | Publisher Full Text

33. Locatelli JC, Costa JG, Haynes A, *et al.*: **Incretin-Based Weight Loss Pharmacotherapy: Can Resistance Exercise Optimize Changes in Body Composition?** *Diabetes Care.* 2024; **47**(10): 1718–1730.
PubMed Abstract | Publisher Full Text

34. Jastreboff AM, Kaplan LM, Frías JP, *et al.*: **Triple–Hormone-Receptor Agonist Retatrutide for Obesity — A Phase 2 Trial.** *N Engl J Med.* 2023; **389**(6): 514–526.
PubMed Abstract | Publisher Full Text

35. Borga M, West J, Bell JD, *et al.*: **Advanced body composition assessment: From body mass index to body composition profiling.** *J Investig Med.* 2018; **66**(5): 1–9.
PubMed Abstract | Publisher Full Text | Free Full Text

36. Li GHY, Lee GKY, Au PCM, *et al.*: **The effect of different measurement modalities in the association of lean mass with mortality: A systematic review and meta-analysis.** *Osteoporos Sarcopenia.* 2021; **7**(Suppl 1): S13–S18.
PubMed Abstract | Publisher Full Text | Free Full Text

37. Li L, Xia Z, Zeng X, *et al.*: **The agreement of different techniques for muscle measurement in diagnosing sarcopenia: a systematic review and meta-analysis.** *Quant Imaging Med Surg.* 2024; **14**(3): 2177–2192.
PubMed Abstract | Publisher Full Text | Free Full Text

38. Smerud HK, Bárány P, Lindström K, *et al.*: **New treatment for IgA nephropathy: enteric budesonide targeted to the ileocecal region ameliorates proteinuria.** *Nephrology, Dialysis, Transplantation: Official Publication of the European Dialysis and Transplant Association - European Renal Association.* 2011; **26**(10): 3237–3242.
Publisher Full Text

39. Lafayette R, Kristensen J, Stone A, *et al.*: **Efficacy and safety of a targeted-release formulation of budesonide in patients with primary IgA nephropathy (NefIgArd): 2-year results from a randomised phase 3 trial.** *Lancet.* 2023; **402**(10405): 859–870.
PubMed Abstract | Publisher Full Text

40. Lv J, Wong MG, Hladunewich MA, *et al.*: **Effect of Oral Methylprednisolone on Decline in Kidney Function or Kidney Failure in Patients With IgA Nephropathy: The TESTING Randomized Clinical Trial.** *JAMA.* 2022; **327**(19): 1888–1898.
PubMed Abstract | Publisher Full Text | Free Full Text

41. Goldberg AL: **Protein turnover in skeletal muscle. II. Effects of denervation and cortisone on protein catabolism in skeletal muscle.** *J Biol Chem.* 1969; **244**(12): 3223–3229.
Publisher Full Text

42. Geer EB, Islam J, Buettner C: **Mechanisms of glucocorticoid-induced insulin resistance: Focus on adipose tissue function and lipid metabolism.** *Endocrinol Metab Clin North Am.* 2014; **43**(1): 75–102.
PubMed Abstract | Publisher Full Text | Free Full Text

43. Layec G, Bringard A, Le Fur Y, *et al.*: **Reproducibility assessment of metabolic variables characterizing muscle energetics in vivo: A 31P-MRS study.** *Magn Reson Med.* 2009; **62**(4): 840–854.
PubMed Abstract | Publisher Full Text

44. Edwards LM, Tyler DJ, Kemp GJ, *et al.*: **The reproducibility of 31-phosphorus MRS measures of muscle energetics at 3 tesla in trained men.** *PLoS One.* 2012; **7**(6): e37237.
Publisher Full Text

45. Lanza IR, Bhagra S, Nair KS, *et al.*: **Measurement of human skeletal muscle oxidative capacity by 31P-MR spectroscopy: A cross-validation with in vitro measurements.** *J Magn Reson Imaging.* 2011; **34**(5): 1143–1150.
PubMed Abstract | Publisher Full Text | Free Full Text

46. Grimm A, Meyer H, Nickel MD, *et al.*: **Repeatability of Dixon magnetic resonance imaging and magnetic resonance spectroscopy for quantitative muscle fat assessments in the thigh.** *J Cachexia Sarcopenia Muscle.* 2018; **9**(6): 1093–1100.
PubMed Abstract | Publisher Full Text | Free Full Text

47. Abrahamson M, Olafsson I, Palsdottir A, *et al.*: **Structure and expression of the human cystatin C gene.** *Biochem J.* 1990; **268**(2): 287–294.
PubMed Abstract | Publisher Full Text | Free Full Text

48. Sans L, Radosevic A, Quintian C, *et al.*: **Cystatin C estimated glomerular filtration rate to assess renal function in early stages of autosomal dominant polycystic kidney disease.** *PLoS One.* 2017; **12**(3): e0174583.
PubMed Abstract | Publisher Full Text | Free Full Text

49. Fitzmaurice G, Davidian M, Verbeke G, Molenberghs G. *Longitudinal Data Analysis: Handbooks of Modern Statistical Methods.* Fitzmaurice G, Davidian M, Verbeke G, *et al.*, editors. 1st ed. Chapman & Hall/CRC Press; 2008.
Publisher Full Text

50. Ibrahim JG, Molenberghs G: **Missing data methods in longitudinal studies: A review.** *Test.* 2009; **18**(1): 1–43.
PubMed Abstract | Publisher Full Text | Free Full Text

51. Rizopoulos D: *Joint Models for Longitudinal and Time-to-Event Data: With Applications in R.* 1st ed. Chapman & Hall/CRC Press; 2012.
Publisher Full Text

52. Thomadakis C, Meligkotsidou L, Pantazis N, *et al.*: **Longitudinal and time-to-drop-out joint models can lead to seriously biased estimates when the drop-out mechanism is at random.** *Biometrics.* 2019; **75**(1): 58–68.
PubMed Abstract | Publisher Full Text

53. Vonesh EF, Greene T: **Biased estimation with shared parameter models in the presence of competing dropout mechanisms.**

Biometrics. 2022; **78**(1): 399–406.
PubMed Abstract | **Publisher Full Text**

54. Thomadakis C, Meligkotsidou L, Pantazis N, *et al.*: **Rejoinder to "Biased Estimation With Shared Parameter Models in the Presence of Competing Dropout Mechanisms".** *Biometrics.* 2022; **78**(1): 407–408.
Publisher Full Text

55. Kunselman AR: **A brief overview of sensitivity analyses.** *Fertil Steril.* 2023; **119**(6): 904–906.
PubMed Abstract | **Publisher Full Text**

56. Cheng DM, Hogan JW: **The Sense and Sensibility of Sensitivity Analyses.** *N Engl J Med.* 2024; **391**(11): 972–974.
PubMed Abstract | **Publisher Full Text**

57. Torres VE, Meijer E, Bae KT, *et al.*: **Rationale and Design of the TEMPO (Tolvaptan Efficacy and Safety in Management of Autosomal Dominant Polycystic Kidney Disease and its Outcomes) 3-4 Study.** *Am J Kidney Dis.* 2011; **57**(5): 692–699.
PubMed Abstract | **Publisher Full Text**

58. Hilderink JM, Van Der Linden N, Kimenai DM, *et al.*: **Biological variation of creatinine, cystatin C, and eGFR over 24 hours.** *Clin Chem.* 2018; **64**(5): 851–860.
PubMed Abstract | **Publisher Full Text**

59. Rowe C, Sitch AJ, Barratt J, *et al.*: **Biological variation of measured and estimated glomerular filtration rate in patients with chronic kidney disease.** *Kidney Int.* 2019; **96**(2): 429–435.
PubMed Abstract | **Publisher Full Text**

60. Thöni S, Keller F, Denicolò S, *et al.*: **Biological variation and reference change value of the estimated glomerular filtration rate in humans: A systematic review and meta-analysis.** *Front Med (Lausanne).* 2022; **9**: 1009358.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

61. Gansevoort RT: **Reply to Dr. Anderson's letter.** *Clin Nephrol.* 2020; **93**(6): 310–311.
PubMed Abstract | **Publisher Full Text**

62. Torres VE, Chapman AB, Devuyst O, *et al.*: **Multicenter, open-label, extension trial to evaluate the long-term efficacy and safety of early versus delayed treatment with tolvaptan in autosomal dominant polycystic kidney disease: The TEMPO 4:4 Trial.** *Nephrol Dial Transplant.* 2018; **33**(3): 477–489.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

63. Bonneville EF, de Wreede LC, Putter H: **Why you should avoid using multiple Fine–Gray models: insights from (attempts at) simulating proportional subdistribution hazards data.** *J R Stat Soc Ser A Stat Soc.* 2024; **187**(3): 580–593.
Publisher Full Text

64. Erler NS, Rizopoulos D, Lesaffre EMEH: **JointAI: Joint Analysis and Imputation of Incomplete Data in R.** *J Stat Softw.* 2021; **100**(20): 1–56.
Publisher Full Text

65. Bais T, Meijer E, Kramers BJ, *et al.*: **HYDROchlorothiazide versus placebo to PROTECT polycystic kidney disease patients and improve their quality of life: study protocol and rationale for the HYDRO-PROTECT randomized controlled trial.** *Trials.* 2024; **25**(1): 120.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

66. 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.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

67. Inker LA, Collier W, Greene T, *et al.*: **A meta-analysis of GFR slope as a surrogate endpoint for kidney failure.** *Nat Med.* 2023; **29**(7): 1867–1876.
Publisher Full Text

68. Aronne LJ, Horn DB, le Roux CW, *et al.*: **Orforglipron for maintenance of body weight reduction: the double-blind, randomized phase 3b ATTAIN-MAINTAIN trial.** *Nat Med.* Published online May 2026; **13**.
PubMed Abstract | **Publisher Full Text**

Open Peer Review

Current Peer Review Status: ✔ ✔ ✔

Version 2

Reviewer Report 18 August 2026

https://doi.org/10.21956/openreseurope.26036.r77464

© **2026 Betjes M.** This is an open access peer review report distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

✔ **Michiel Betjes**
Erasmus Medical Center, Rotterdam, The Netherlands

The authors have made substantial and valuable changes to the revised manuscript.

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: general nephrology and kidney transplantation

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Reviewer Report 25 July 2026

https://doi.org/10.21956/openreseurope.26036.r77466

© **2026 van Craenenbroeck A.** This is an open access peer review report distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

✔ **Amaryllis van Craenenbroeck**
KU Leuven, Louvain, Belgium

Thank you for this revised version of the manuscript. The authors addressed all raised concerns.

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: nephrology, vascular dysfunction and its reversibility by exercise training

I confirm that I have read this submission and believe that I have an appropriate level of

expertise to confirm that it is of an acceptable scientific standard.

Reviewer Report 21 July 2026

https://doi.org/10.21956/openreseurope.26036.r77465

© 2026 Hoogeveen E. This is an open access peer review report distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Ellen Hoogeveen
Clinical Epidemiology, Leiden University Medical Center, Leiden, The Netherlands

The Open Letter has been substantially improved and is now well considered, thoroughly thought out, and clearly presented. I have no further comments.

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: nephrology and clinical epidemiology

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Version 1

Reviewer Report 15 July 2025

https://doi.org/10.21956/openreseurope.22106.r55603

© 2025 Hoogeveen E. This is an open access peer review report distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Ellen Hoogeveen
Clinical Epidemiology, Leiden University Medical Center, Leiden, The Netherlands

Summary
Groothof and Bakker present a thought-provoking hypothesis that the apparent kidney-protective effect of tolvaptan in ADPKD trials may, at least partially, be attributed to therapy related muscle wasting - leading to lower serum creatinine and a spurious rise in creatinine-based eGFR (eGFRcr) - rather than true preservation of glomerular filtration. They propose that tolvaptan's aquaretic effects (polyuria up to 10 L/day) drive excessive fluid intake, appetite suppression, malnutrition, and gradual sarcopenia, which in turn raises eGFRcr. The authors note that cystatin C - a filtration marker insensitive to muscle mass - was measured in TEMPO3:4 but never reported, and call for

reanalysis of those data. They further speculate that intermittent hypertonic dehydration events may actually worsen kidney injury, and highlight the psychological burden on patients who decline or discontinue tolvaptan due to its side effects.

Major Comments

1. Speculative muscle-mass loss and eGFR confounding
The central premise - that tolvaptan induces sufficient muscle wasting to alter creatinine pools and confound eGFRcr - is plausible but unsupported by direct longitudinal body-composition or nutritional data from trial participants. The extrapolation of a 1.3% annual muscle-mass decline derives from general sarcopenia literature rather than ADPKD cohorts on tolvaptan, and the TEMPO3:4 trial did not report serial lean-mass measurements or nutritional assessments to substantiate clinically meaningful sarcopenia.

2. Absence of published cystatin C results
Although cystatin C was prespecified as a pharmacodynamic endpoint in TEMPO3:4, it is not clarified whether these data exist in full, nor discussed why they were omitted.[1] The authors of the TEMPO3:4 trial could investigate and report whether cystatin C trajectories differed between tolvaptan and placebo arms, and if not, whether non-significance or logistical constraints explain the absence of publication.

3. Existing cystatin C literature in ADPKD
Several studies (e.g., Sans et al., 2017) have evaluated cystatin C-based eGFR in early ADPKD stages and found concordance with eGFRcr, suggesting limited creatinine bias in this setting.[2] The manuscript would be strengthened by situating the hypothesis within this existing body of work.

4. Psychological burden claims need empirical backing
The discussion of guilt and self-blame among patients who forgo or stop tolvaptan is an important humanistic angle but remains conjectural. There is currently no published qualitative or quantitative research exploring patient-reported emotional outcomes in this context. The authors should recommend or initiate studies (e.g., structured interviews, validated quality-of-life instruments) to capture these dimensions.

5. Hypertonicity-related nephropathy hypothesis
While episodes of volume depletion or hypertonicity could theoretically exacerbate kidney injury, the frequency, severity, and clinical impact of such events in tolvaptan-treated ADPKD patients remain undefined. Citing indirect evidence from heat-stress nephropathy does not substitute for ADPKD-specific cohort data. Prospective monitoring of electrolytes, osmolality, and acute GFR changes during intercurrent illnesses would be necessary to substantiate this concern.

Minor Comments

- **Body weight vs. composition**: The manuscript notes stable body weight during TEMPO3:4, and suggests that fat-mass increases could mask lean-mass loss; recommending DXA or bioimpedance analyses in future trials would be helpful.
- **Statistical framework**: Propose power calculations and minimal detectable difference for a cystatin C - based reanalysis to guide feasibility.
- **Figure clarity**: In Figure 1, consider annotating with estimated effect sizes (e.g., expected

decline in creatinine per kg muscle lost) to aid interpretability.

Conclusion
This perspective raises a biologically plausible challenge to the current narrative on tolvaptan's kidney-protective effect and underscores important recommendations to account for muscle-mass changes. However, it remains speculative without direct empirical support. To advance the field, the authors could (1) propose to seek and publish the TEMPO3:4 cystatin C data; (2) propose prospective body-composition and patient-reported outcome measures, including a qualitative study; and (3) rigorously evaluate the hypertonicity hypothesis in clinical practice. Such efforts will ensure that decisions about tolvaptan use are grounded in robust, multifaceted evidence.

References
1.Torres VE, Meijer E, Bae KT, *et al:* Rationale and design of the TEMPO(Tolvaptan Efficacy and Safety in Management of autosomal dominant
polycystic kidney disease and its outcomes) 3-4 Study. *Am J Kidney Dis.* 2011;
57(5): 692–9. PMID: 21333426

2. Sans *et al:* Cystatin C estimated glomerular filtration rate to assess renal function in early stages of autosomal dominant polycystic kidney disease*PLoS One.* 2017;12(3):e0174583. PMID: 28346513

References
1. Torres V, Meijer E, Bae K, Chapman A, et al.: Rationale and Design of the TEMPO (Tolvaptan Efficacy and Safety in Management of Autosomal Dominant Polycystic Kidney Disease and Its Outcomes) 3-4 Study. *American Journal of Kidney Diseases*. 2011; **57** (5): 692-699 Publisher Full Text
2. Sans L, Radosevic A, Quintian C, Montañés R, et al.: Cystatin C estimated glomerular filtration rate to assess renal function in early stages of autosomal dominant polycystic kidney disease. *PLOS ONE*. 2017; **12** (3). Publisher Full Text

Is the rationale for the Open Letter provided in sufficient detail? (Please consider whether existing challenges in the field are outlined clearly and whether the purpose of the letter is explained)
Yes

Does the article adequately reference differing views and opinions?
Yes

Are all factual statements correct, and are statements and arguments made adequately supported by citations?
Yes

Is the Open Letter written in accessible language? (Please consider whether all subject-specific terms, concepts and abbreviations are explained)
Yes

Where applicable, are recommendations and next steps explained clearly for others to follow? (Please consider whether others in the research community would be able to

implement guidelines or recommendations and/or constructively engage in the debate)
Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: nephrology and clinical epidemiology

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Author Response 17 May 2026

Dion Groothof

Detailed, itemised response to the comments by Reviewer Dr. Ellen Hoogeveen
Summary Groothof and Bakker present a thought-provoking hypothesis that the apparent kidney-protective effect of tolvaptan in ADPKD trials may, at least partially, be attributed to therapy related muscle wasting - leading to lower serum creatinine and a spurious rise in creatinine-based eGFR (eGFRcr) - rather than true preservation of glomerular filtration. They propose that tolvaptan's aquaretic effects (polyuria up to 10 L/day) drive excessive fluid intake, appetite suppression, malnutrition, and gradual sarcopenia, which in turn raises eGFRcr. The authors note that cystatin C - a filtration marker insensitive to muscle mass - was measured in TEMPO3:4 but never reported, and call for reanalysis of those data. They further speculate that intermittent hypertonic dehydration events may actually worsen kidney injury, and highlight the psychological burden on patients who decline or discontinue tolvaptan due to its side effects. **Major Comments Comment 1:** *"1. Speculative muscle-mass loss and eGFR confounding The central premise - that tolvaptan induces sufficient muscle wasting to alter creatinine pools and confound eGFRcr - is plausible but unsupported by direct longitudinal body-composition or nutritional data from trial participants. The extrapolation of a 1.3% annual muscle-mass decline derives from general sarcopenia literature rather than ADPKD cohorts on tolvaptan, and the TEMPO3:4 trial did not report serial lean-mass measurements or nutritional assessments to substantiate clinically meaningful sarcopenia."*
Response to Comment 1: We agree with the reviewer that the TEMPO3:4 trial did not include serial assessments of muscle mass, body composition, or nutritional status and therefore does not provide direct evidence to substantiate clinically meaningful sarcopenia in this setting. Our hypothesis is therefore not intended to assert the presence of such effects, but to highlight a biologically plausible mechanism that could, in principle, influence creatinine-based eGFR (eGFRcr). To address this important limitation and to strengthen the empirical basis of our argument, we have substantially revised the manuscript by adding a dedicated section entitled "Detection of a clinically relevant drug effect on creatinine generation in clinical trials". In this section, we move beyond conceptual reasoning and explicitly examine whether treatment-induced changes in endogenous creatinine generation of a magnitude sufficient to affect eGFRcr could be empirically detected or excluded within feasible study designs. Specifically, we:

1. quantify the magnitude of change in creatinine generation required to meaningfully influence eGFRcr trajectories based on calculations that use representative data from TEMPO3:4;

2. evaluate the detectability of such changes using direct and indirect measurement approaches (including imaging-based techniques, isotope dilution, and 24-hour urinary creatinine excretion), and;
3. (iii) highlight that a within-trial, longitudinal evaluation of discordance between trajectories of creatinine-based eGFR and cystatin C-based eGFR (using cystatin C data that were prespecified and collected in TEMPO3:4), provides a practical and sufficiently powered approach to detect or refute such effects without the need for additional data collection.

In this way, the revised manuscript clarifies that, although direct measurements of muscle mass are lacking, the hypothesis is testable, as it can be evaluated using existing trial data and contemporary analytical approaches. **Comment 2:** *"2. Absence of published cystatin C results Although cystatin C was prespecified as a pharmacodynamic endpoint in TEMPO3:4, it is not clarified whether these data exist in full, nor discussed why they were omitted.[1] The authors of the TEMPO3:4 trial could investigate and report whether cystatin C trajectories differed between tolvaptan and placebo arms, and if not, whether non-significance or logistical constraints explain the absence of publication."* **Response to Comment 2:** We agree that clarification regarding the availability and analysis of cystatin C data from the TEMPO3:4 trial is highly relevant, particularly in light of its prespecification as a pharmacodynamic endpoint, as well as its potential to inform the interpretation of eGFRcr in the presence of treatment-related changes in creatinine generation. As noted in the revised manuscript, cystatin C measurements were indeed included in the trial protocol, indicating that such data were collected. However, to our knowledge, analyses of cystatin C trajectories or derived cystatin C-based eGFR (eGFRcys) outcomes have not been reported. The reasons for this are not described in the original publications, and we have therefore refrained from speculating on this point. In the revised manuscript, we now explicitly highlight that the availability of these data provides a unique opportunity to evaluate discordance between eGFRcr and eGFRcys trajectories within the trial. Such analyses would allow assessment of whether treatment effects differ by filtration marker and could help clarify the extent to which creatinine-based estimates may be influenced by non-GFR determinants. This point is now more explicitly emphasised in the revised manuscript. **Comment 3:** *"3. Existing cystatin C literature in ADPKD Several studies (e.g., Sans et al., 2017) have evaluated cystatin C-based eGFR in early ADPKD stages and found concordance with eGFRcr, suggesting limited creatinine bias in this setting.[2] The manuscript would be strengthened by situating the hypothesis within this existing body of work."* **Response to Comment 3:** We thank the reviewer for highlighting this important evidence. We agree that studies such as Sans et al. (2017) demonstrate a moderate correlation between eGFRcr and eGFRcys in patients with ADPKD, reflecting their shared dependence on underlying glomerular filtration rate. In that study, a correlation coefficient of 0.557 was observed at preserved values of eGFRcr and eGFRcys, indicating that approximately one-third of the variance is shared between the two measures ($0.557 \times 0.557 \times 100\% \approx 31\%$). At the same time, this correlation is incomplete, with a substantial proportion of variability remaining unexplained, consistent with contributions from non-GFR determinants such as creatinine generation for eGFRcr. Rather than contradicting our hypothesis, this observation provides the statistical basis for detecting treatment-induced divergence between eGFRcr and eGFRcys trajectories over time. To address the reviewer's comment, we have now explicitly incorporated this perspective into the revised manuscript. The added text reads as follows: "In patients with ADPKD, eGFRcr and eGFRcys are moderately correlated, sharing approximately 30% of their variance (Ref:

Sans L et al., PLoS One 2017;12(3):e0174583), consistent with their common dependence on true GFR. The remaining variance attributable to non-GFR determinants of each marker, including creatinine generation for eGFRcr, provides the statistical basis for detecting treatment-induced divergence between the two markers over time." **Comment 4.** *"4. Psychological burden claims need empirical backing The discussion of guilt and self-blame among patients who forgo or stop tolvaptan is an important humanistic angle but remains conjectural. There is currently no published qualitative or quantitative research exploring patient-reported emotional outcomes in this context. The authors should recommend or initiate studies (e.g., structured interviews, validated quality-of-life instruments) to capture these dimensions."* **Response to Comment 4:** We thank the reviewer for this suggestion and agree that the discussion of potential psychological burden, including feelings of guilt of self-blame among patients who decline of discontinue tolvaptan, remains speculative in the absence of empirical evidence. To address this point, we have revised the manuscript to explicitly acknowledge this limitation and to recommend further research in this area. Specifically, we now suggest that future studies should incorporate qualitative and quantitative approaches, such as the ones suggested by this reviewer, to systematically evaluate patient-reported emotional outcomes related to treatment decisions. The added text in the revised manuscript reads as follows: "Considering treatment discontinuation, the possibility of guilt and self-blame among patients who refuse to start or discontinue using tolvaptan currently remains largely conjectural, as qualitative or quantitative evidence on patient-reported emotional outcomes in this context is lacking. We recommend performance of studies (e.g., structured interviews, validated quality-of-life instruments) capable of capturing these dimensions." **Comment 5.** *"5. Hypertonicity-related nephropathy hypothesis While episodes of volume depletion or hypertonicity could theoretically exacerbate kidney injury, the frequency, severity, and clinical impact of such events in tolvaptan-treated ADPKD patients remain undefined. Citing indirect evidence from heat-stress nephropathy does not substitute for ADPKD-specific cohort data. Prospective monitoring of electrolytes, osmolality, and acute GFR changes during intercurrent illnesses would be necessary to substantiate this concern."* **Response to Comment 5:** We thank the reviewer for this thoughtful comment. We agree that, while episodes of volume depletion or hypertonicity could theoretically contribute to kidney injury, the frequency, severity, and clinical relevance of such events in tolvaptan-treated patients with ADPKD remain insufficiently characterised. We also concur that indirect evidence from heat-stress nephropathy cannot substitute for ADPKD-specific cohort data. To address this point, we have revised the manuscript to explicitly acknowledge these limitations and to emphasise the need for prospective evaluation. Specifically, we now highlight that studies incorporating longitudinal monitoring of electrolytes, osmolality, and acute changes in kidney function during intercurrent illness would be required to substantiate this hypothesis. The added text in the revised manuscript reads as follows: "While episodes of volume depletion or hypertonicity could theoretically exacerbate kidney injury, the frequency, severity, and clinical impact of such events in tolvaptan-treated patients with ADPKD remain undefined. At present, only indirect evidence from heat-stress nephropathy is available (Groothof D et al., Kidney360;5(12):1917-1921; Johnson R & Sánchez-Lozada L, Nat Rev Nephrol 2013;9(10):560-561; Yang X et al., BMC Nephrol 2020;21(1):159), which does not substitute for ADPKD-specific cohort data. Prospective monitoring of electrolytes, osmolality, and acute GFR changes during intercurrent illnesses would be necessary to substantiate this concern. Nonetheless, even this theoretical possibility of accelerated kidney injury underscores the need for analytical approaches

capable of distinguishing true treatment effects on GFR from artefacts introduced by changes in creatinine generation." **Minor comments Comment 6.** *"Body weight vs. composition: The manuscript notes stable body weight during TEMPO3:4, and suggests that fat-mass increases could mask lean-mass loss; recommending DXA or bioimpedance analyses in future trials would be helpful."* **Response to Comment 6:** We thank the reviewer for highlighting this important point. The conceptual framework underlying our manuscript explicitly recognises that stable body weight does not preclude meaningful changes in body composition, and that changes in fat mass may indeed mask concurrent changes in lean tissue. We further acknowledge that the original version of the manuscript did not explicitly address how such discrepancies could be assessed in practice. We concur that discussing modalities capable of detecting these changes adds important methodological clarity. To address this, we have expanded the manuscript with a dedicated section entitled "Approaches to evaluate potential confounding related to changes muscle mass and creatinine generation in clinical trials". In this section, we go beyond the assessment of muscle mass alone and emphasise that it is ultimately endogenous creatinine generation, rather than muscle mass per se, that is most relevant for the interpretation of eGFRcr. Specifically, we distinguish between scenarios in which changes in muscle mass and fat mass occur concordantly and scenarios in which these components change discordantly, for example due to alterations in muscle composition or fat redistribution. In the former cases, standard techniques such as dual-energy X-ray absorptiometry (DXA) or bioelectrical impedance analysis (BIA) can provide informative estimates, whereas in the case of discordant changes more advanced approaches capable of assessing total body creatine pools are required. These additions aim to clarify how potential discrepancies between body weight, body composition, and creatinine generation can be appropriately evaluated in future clinical and observational studies. **Comment 7:** *"Statistical framework: Propose power calculations and minimal detectable difference for a cystatin C - based reanalysis to guide feasibility."* **Response to Comment 7:** We thank the reviewer for this helpful suggestion. In response to this comment, and in alignment with **Comment 1** of this reviewer, we have added a dedicated section entitled "Detection of a clinically relevant drug effect on creatinine generation in clinical trials". As detailed in our **Response to Comment 1**, this section explicitly addresses the feasibility of a cystatin C-based re-analysis by providing quantitative estimates of the magnitude of treatment-induced changes in creatinine generation required to meaningfully influence eGFRcr, along with corresponding minimal detectable differences and sample size requirements under realistic assumptions. We further outline how a within-trial, longitudinal comparison of eGFRcr and eGFRcys trajectories in TEMPO3:4 can be used to detect such effects with adequate statistical power. We believe these additions directly address the reviewer's request by demonstrating that a cystatin C-based re-analysis is not only conceptually appropriate, but also practically feasible within the existing trial dataset of TEMPO3:4. **Comment 8:** *"Figure clarity: In Figure 1, consider annotating with estimated effect sizes (e.g., expected decline in creatinine per kg muscle lost) to aid interpretability."* **Response to Comment 8:** We thank the reviewer for this suggestion and agree that providing quantitative context is important for interpreting the plausibility of the proposed mechanism. Rather than incorporating numerical estimates directly in **Figure 1**, which is intended as a conceptual schematic illustration, we have chosen to present these calculations in a dedicated Box (**Box 1**) in the manuscript. In this Box, we explicitly quantify the magnitude of change in endogenous creatinine generation (and corresponding muscle mass) that would be required to account for the observed

difference in eGFRcr between treatment arms in the TEMPO3:4 trial, under clearly stated assumptions. In addition, the detectability and sample size implications of such changes are further elaborated in the section “Detection of a clinically relevant drug effect on creatinine generation in clinical trials” (see also our **Responses to Comments 1 and 7**). We believe that this approach maintains the clarity of Figure 1 while providing a transparent and appropriately contextualised quantitative framework elsewhere in the manuscript.

Competing Interests: No competing interests were disclosed.

Reviewer Report 09 July 2025

https://doi.org/10.21956/openreseurope.22106.r55137

© 2025 van Craenenbroeck A. This is an open access peer review report distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

?

Amaryllis van Craenenbroeck
KU Leuven, Louvain, Belgium

I have read with great interest the letter by Groothof and Bakker questioning the kidney-protective effect of tolvaptan. Some comments:

- I am not sure whether the title fully covers the message (especially the adjective ‘biologically’). The authors hypothesize that use of tolvaptan can induce loss of muscle mass. This is an interesting, novel and challenging viewpoint and merits further investigation.
However, in their further explanation, no discussion is present yet on including direct measurements of muscle mass/strength in future trials. In contrast with cystatin C based eGFR estimations, this would provide direct info on the ‘biological’ ground of the hypothesis. The same would be true for markers of malnutrition (part of the key hypothesis in the figure). I would suggest broadening this perspective a bit more.

- A discussion on the effect of tolvaptan on change in kidney volume, the primary endpoint of the TEMPO3:4 trial, is lacking. The authors might want to discuss why their hypothesis would be more plausible than the underlying hypothesis of the TEMPO3:4 trial in a separate paragraph.

Is the rationale for the Open Letter provided in sufficient detail? (Please consider whether existing challenges in the field are outlined clearly and whether the purpose of the letter is explained)
Yes

Does the article adequately reference differing views and opinions?
Yes

Are all factual statements correct, and are statements and arguments made adequately

supported by citations?
Yes

Is the Open Letter written in accessible language? (Please consider whether all subject-specific terms, concepts and abbreviations are explained)
Yes

Where applicable, are recommendations and next steps explained clearly for others to follow? (Please consider whether others in the research community would be able to implement guidelines or recommendations and/or constructively engage in the debate)
Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: nephrology, vascular dysfunction and its reversibility by exercise training

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Author Response 17 May 2026

Dion Groothof

Detailed, itemised response to the comments by Reviewer Dr. Amaryllis van Craenenbroeck I have read with great interest the letter by Groothof and Bakker questioning the kidney-protective effect of tolvaptan. Some comments: **Comment 1:** *"I am not sure whether the title fully covers the message (especially the adjective 'biologically'). The authors hypothesize that use of tolvaptan can induce loss of muscle mass. This is an interesting, novel and challenging viewpoint and merits further investigation. However, in their further explanation, no discussion is present yet on including direct measurements of muscle mass/strength in future trials. In contrast with cystatin C based eGFR estimations, this would provide direct info on the 'biological' ground of the hypothesis. The same would be true for markers of malnutrition (part of the key hypothesis in the figure). I would suggest broadening this perspective a bit more."* **Response to Comment 1:** We thank the reviewer for the thoughtful and constructive comment. We agree that, in the original version of the manuscript, the adjective 'biologically' in the title was insufficiently substantiated and that a broader discussion of direct measurements of muscle mass, muscle composition, and muscle strength would strengthen the biological basis of our hypothesis. To address this, we have made three substantive revisions. First, we have modified the title by replacing the terms 'biologically plausible' with 'testable', to more accurately reflect the nature of the hypothesis. Second, we have substantially expanded the manuscript to explicitly discuss how direct measurements of muscle mass, muscle composition, and muscle strength could be incorporated in future clinical trials, thereby broadening the biological perspective as suggested. Third, and importantly, we have added a new section demonstrating that treatment-induced changes in endogenous creatinine generation of clinically relevant magnitude are not only biologically plausible but also empirically detectable/testable using

feasible study designs. In this context, we outline both direct approaches (such as imaging- and isotope-based quantification of creatine pools) and indirect approaches based on longitudinal discordance between trajectories of creatinine-based eGFR and cystatin C-based eGFR within existing trial datasets. We further emphasise in the concluding section how these measurement strategies can be employed into re-analyses of existing trials and integrated into future trial designs. **Comment 2:** "A discussion on the effect of tolvaptan on change in kidney volume, the primary endpoint of the TEMPO3:4 trial, is lacking. The authors might want to discuss why their hypothesis would be more plausible than the underlying hypothesis of the TEMPO3:4 trial in a separate paragraph." **Response to Comment 2:** We thank the reviewer for this important comment and agree that the effects of tolvaptan on total kidney volume (TKV), the primary endpoint of the TEMPO3:4 trial, warrant explicit consideration in the context of our hypothesis. To address this, we have added a dedicated paragraph to the revised manuscript in which we discuss the observed temporal pattern of TKV growth rate reduction and its interpretation in the literature. In particular, we now acknowledge that early reductions in TKV growth may reflect a biologically plausible disease-modifying effect of tolvaptan, as previously emphasised. At the same time, we clarify that this biological interpretation does not preclude the possibility that concurrent mechanisms unrelated to GFR, as well as selective treatment discontinuation, may influence estimates of treatment effect on creatinine-based eGFR. Accordingly, we have revised the manuscript to explicitly position our hypothesis as complementary to, rather than in competition with, the established biological framework derived from TKV outcomes. The relevant text can be found in the revised manuscript in the section running from "A similar selection process may also help to explain..." to "...by highlighting potential sources of bias in the quantification of treatment effects on kidney function".

Competing Interests: No competing interests were disclosed.

Reviewer Report 26 June 2025

https://doi.org/10.21956/openreseurope.22106.r55601

© 2025 Betjes M. This is an open access peer review report distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

?

Michiel Betjes
Erasmus Medical Center, Rotterdam, The Netherlands

The open letter bij Groothof and Bakker questions the kidney protective effect of tolvaptan in patients with polycystic disease. There main argument is that tolvaptan could potentially lower muscle mass thereby lowering the serum creatinine leading to positive effect on eGFR. This potential bias could be assessed by an alternative measurement of eGFR assessment by using cystatine C.
There are no data to support this hypothesis as cystatine C measurement and measurements on

muscle mass from tolvaptan trials are not available/done.

The letter has value as a hypothesis is generated that questions the true effect of tolvaptan. The authors think this is an important issue as it is a matter of truth seeking and tolvaptan may be poorly tolerated.

Concerning the latter, the addition of a thiazide may reduce urinary output during tolvaptan while maintaining efficacy. This should be mentioned.
Also, as stated by the authors, tolvaptan does reduce the growth of kidney cysts which is one of the most important predictors of the rate of eGFR loss. Thus, a biological plausible explanation for the action of tolvaptan is documented. Therefore, the sentence "We hypothesize that a reevaluation of the TEMPO3:4 trial data using the available cystatin C measurements would reveal no significant effect of tolvaptan on slowing kidney function decline compared with placebo" is to bold and should be rephrased to ".. could reveal a lesser efficacy of tolvaptan than based on creatine measurements..".
The need for cystatin C measurement if muscle mass loss could be expected is in itself an important message. Given the side effects of tolvaptan, balancing these against the true efficacy in slowing down eGFR loss is another important message.
In conclusion, the hypothesis has no measured data to build on but in itself is of interest as it highlights that serum creatine may sometimes be a doggy measure of establishing kidney function. I would recommend to rewrite the letter such that this is the main message and take tolvaptan trials as an example of study design that would have benefited from addition of cystatin measurement. In addition, this could still be done in the recently started HYDROPROTECT trial.

Is the rationale for the Open Letter provided in sufficient detail? (Please consider whether existing challenges in the field are outlined clearly and whether the purpose of the letter is explained)
Yes

Does the article adequately reference differing views and opinions?
Partly

Are all factual statements correct, and are statements and arguments made adequately supported by citations?
Yes

Is the Open Letter written in accessible language? (Please consider whether all subject-specific terms, concepts and abbreviations are explained)
Yes

Where applicable, are recommendations and next steps explained clearly for others to follow? (Please consider whether others in the research community would be able to implement guidelines or recommendations and/or constructively engage in the debate)
Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: general nephrology and kidney transplantation

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Author Response 17 May 2026

Dion Groothof

Detailed, itemised response to the comments of Reviewer Dr. Michiel Betjes The open letter bij Groothof and Bakker questions the kidney protective effect of tolvaptan in patients with polycystic disease. There main argument is that tolvaptan could potentially lower muscle mass thereby lowering the serum creatinine leading to positive effect on eGFR. This potential bias could be assessed by an alternative measurement of eGFR assessment by using cystatine C. **Comment 1:** *"There are no data to support this hypothesis as cystatine C measurement and measurements on muscle mass from tolvaptan trials are not available/done."* **Response to Comment 1:** We thank the Reviewer for this important comment. We agree that there are no direct data on muscle mass from the pivotal tolvaptan trials to support our hypothesis. However, we would like to clarify that cystatin C measurements are in fact available in the pivotal TEMPO3:4 trial (Torres VE et al., N Engl J Med 2012;367(25):2407-2418). As specified in the trial protocol, published as Supplementary Material of the original report, cystatin C was included as a prespecified pharmacodynamic variable, alongside repeated measurement of creatinine as a secondary efficacy endpoint. The availability of these longitudinal data enables, in principle, a within-trial, trajectory-based assessment of divergence between creatinine-based eGFR (eGFRcr) and cystatin C-based eGFR (eGFRcys). As outlined in the revised manuscript, such an approach would allow detection of small but clinically relevant treatment-induced changes in endogenous creatinine generation with adequate statistical power, leveraging the existing sample size and repeated measurements collected during the trial. We have clarified and expanded this point in the revised manuscript, which can be found in between lines "Although evaluation of discordance between eGFRcr and eGFRcys trajectories requires substantially larger sample sizes than approaches based on 24-hour urinary creatinine excretion under ideal conditions, it offers a unique practical advantage" through "..., this approach offers substantially greater statistical power and practical feasibility than conducting a separate short-term study with two repeated measurements." We believe that this addition addresses the Reviewer's concern by demonstrating that, while direct measurements of muscle mass are indeed lacking, the available data nonetheless permit an indirect but empirically testable evaluation of the proposed hypothesis. **Comment 2:** *"The letter has value as a hypothesis is generated that questions the true effect of tolvaptan. The authors think this is an important issue as it is a matter of truth seeking and tolvaptan may be poorly tolerated."* **Response to Comment 2:** We thank the reviewer for these supportive comments and for recognising the relevance of the hypothesis presented. **Comment 3:** *"Concerning the latter, the addition of a thiazide may reduce urinary output during tolvaptan while maintaining efficacy. This should be mentioned."* **Response to Comment 3:** We thank the Reviewer for this valuable suggestion and agree that the potential for thiazide diuretics to mitigate aquaretic side effects of tolvaptan is clinically relevant and warrants mention. Accordingly, we have revised the manuscript to include discussion of this approach in the context of the ongoing HYDRO-PROTECT trial,

which evaluates whether hydrochlorothiazide can be added to tolvaptan to reduce urinary output without compromising efficacy. As noted in the revised text, this strategy directly addresses a key determinant of and treatment discontinuation. At the same time, we clarify that whether efficacy is fully preserved remains under investigation, and we highlight that the current trial design does not include cystatin C measurements, which limits the ability to assess potential effects on creatinine generation using the approaches discussed in our manuscript. **Comment 4:** *"Also, as stated by the authors, tolvaptan does reduce the growth of kidney cysts which is one of the most important predictors of the rate of eGFR loss. Thus, a biological plausible explanation for the action of tolvaptan is documented. Therefore, the sentence "We hypothesize that a reevaluation of the TEMPO3:4 trial data using the available cystatin C measurements would reveal no significant effect of tolvaptan on slowing kidney function decline compared with placebo" is to bold and should be rephrased to ".. could reveal a lesser efficacy of tolvaptan than based on creatine measurements.."."* **Response to Comment 4:** We agree that tolvaptan has a biologically plausible mechanism of action, as reflected by its demonstrated effect on reducing the growth of kidney cysts, a key predictor of subsequent decline in eGFR. We also agree that our original statement was too strong in light of this evidence. To address this, we have removed the statement from the abstract and have used more nuanced stance in the Introduction to adopt a more balanced interpretation, placing emphasis on the uncertainty regarding both the direction and true magnitude of tolvaptan's effect on GFR preservation. In addition, this comment, together with aspects of other Reviewers' comments, prompted us to further expand the manuscript with a dedicated section on "Missing data, informative dropout, and model validity". In this section, we outline how alternative analytical approaches could either attenuate or strengthen the estimated treatment effect, depending on the underlying assumptions regarding missing data. This revision clarifies that our objective is not to refute the biological plausibility of tolvaptan's effect, but to highlight the importance of fully evaluating potential sources of bias in its quantification. **Comment 5:** *"The need for cystatin C measurement if muscle mass loss could be expected is in itself an important message. Given the side effects of tolvaptan, balancing these against the true efficacy in slowing down eGFR loss is another important message."* **Response to Comment 5:** We thank the Reviewer for highlighting these important points and agree that the need for cystatin C measurement in settings where changes in muscle mass or creatinine generation could be expected is a key message of our work. We also agree that, given the side-effect profile of tolvaptan, careful balancing of tolerability against true efficacy in slowing eGFR decline is clinically highly relevant. To address this, we revised the manuscript to place greater emphasis on the role of cystatin C as a complementary filtration marker. In particular, we expanded the discussion to more clearly describe how combined analyses of eGFRcr and eGFRcys can be used to detect or exclude treatment-related changes in muscle mass and creatinine generation, thereby improving interpretation of treatment effects. **Comment 6:** *"In conclusion, the hypothesis has no measured data to build on but in itself is of interest as it highlights that serum creatine may sometimes be a doggy measure of establishing kidney function. I would recommend to rewrite the letter such that this is the main message and take tolvaptan trials as an example of study design that would have benefited from addition of cystatin measurement. In addition, this could still be done in the recently started HYDROPROTECT trial."* **Response to Comment 6:** We thank the Reviewer for this thoughtful recommendation. In line with this suggestion, we have revised the manuscript to more clearly position the central message that reliance on creatinine alone may be insufficient to fully characterise treatment effects on kidney

function, particularly in settings where non-GFR determinants such as creatinine generation may be altered. We now present the tolvaptan trials as a representative example of a study context in which complementary assessment of cystatin C could have strengthened interpretation of treatment effects. We also refer to our **Response to Comment 5** of this reviewer. In addition, we have incorporated discussion of the ongoing HYDRO-PROTECT trial, noting that inclusion of cystatin C measurements in such studies would further enhance the ability to evaluate potential effects on creatinine generation and improve robustness of outcome assessment.

Competing Interests: No competing interests were disclosed.