Additional file 10. Risk of bias (RoB 2) by comparison, outcome and domain

RoB 2 addresses 23 outcomes from ten focused comparisons and 41 additional candidate outcomes from 32 comparisons. The primary adult set contains the 23 focused and 36 additional outcomes; five candidate assessments from these entries remain outside it. Every additional assessment is source-checked against the report that supplies the result, and each records the pages on which the result and its reporting are found.

A note on the sources used. Each comparison was assessed against the report that supplies the result used in the review: for STOP-IgAN that is the 10-year follow-up report together with the registration record of the follow-up itself (NCT03488368), whose registered primary outcome is the composite reported here, and, for the allocation procedure, the report of the original randomised trial to which the long-term report refers [27]; for DAPA-CKD and EMPA-KIDNEY it is the subgroup analysis together with the description of the parent trial's randomisation that the analysis reproduces. Where a section describes reporting from the parent trial, the parent trial itself is identified so that the statement can be checked outside this file.

Question set

#	Signalling question
1.1	Was the allocation sequence random?
1.2	Was the allocation sequence concealed until participants were enrolled and assigned to interventions?
1.3	Did baseline differences between intervention groups suggest a problem with the randomisation process?
2.1	Were participants aware of their assigned intervention during the trial?
2.2	Were carers and people delivering the interventions aware of participants' assigned intervention during the trial?
2.3	If Y/PY/NI to 2.1 or 2.2: Were there deviations from the intended intervention that arose because of the trial context?
2.4	If Y/PY/NI to 2.3: Were these deviations likely to have affected the outcome?
2.5	If Y/PY to 2.4: Were these deviations from intended intervention balanced between groups?
2.6	Was an appropriate analysis used to estimate the effect of assignment to intervention?
2.7	If N/PN/NI to 2.6: Was there potential for a substantial impact on the result of the failure to analyse participants in the group to which they were randomised?
3.1	Were data for this outcome available for all, or nearly all, participants randomised?
3.2	If N/PN/NI to 3.1: Is there evidence that the result was not biased by missing outcome data?
3.3	If N/PN to 3.2: Could missingness in the outcome depend on its true value?
3.4	If Y/PY/NI to 3.3: Is it likely that missingness in the outcome depended on its true value?

#	Signalling question
4.1	Was the method of measuring the outcome inappropriate?
4.2	Could measurement or ascertainment of the outcome have differed between intervention groups?
4.3	If N/PN/NI to 4.1 and 4.2: Were outcome assessors aware of the intervention received by study participants?
4.4	If Y/PY/NI to 4.3: Could assessment of the outcome have been influenced by knowledge of intervention received?
4.5	If Y/PY/NI to 4.4: Is it likely that assessment of the outcome was influenced by knowledge of intervention received?
5.1	Were the data that produced this result analysed in accordance with a pre-specified analysis plan that was finalised before unblinded outcome data were available for analysis?
5.2	Is the numerical result being assessed likely to have been selected, on the basis of the results, from multiple eligible outcome measurements (e.g. scales, definitions, time points) within the outcome domain?
5.3	Is the numerical result being assessed likely to have been selected, on the basis of the results, from multiple eligible analyses of the data?

Four disciplines were applied. **No overall category is filled in by default**: an unresolved question is reported as NI rather than converted into a category, and an NI answer is not treated as automatically indicating high risk — its consequence depends on what the question asks. **Design and reporting labels are not judgements**: being open-label, being a subgroup, being a secondary or exploratory endpoint, reporting a descriptive P value, or reporting a non-significant result are not, in themselves, reasons for concern in any domain. **Missing outcome data are assessed for each outcome separately**, from whether the data for that outcome were available for nearly all randomised participants and from whether the missingness could depend on the outcome's true value. The size of the analysis population, the choice of a full analysis set and the presence of a pre-specified imputation are **not** evidence that the outcome data were complete. A change in the handling of intercurrent events, and agreement between a trial's different outcomes, are **not** evidence that a result was unaffected by missing data: that requires an analysis which varies the missing-data assumption and shows the result is materially unchanged. **Stopping treatment is not missing outcome data**, and sample size belongs to imprecision, not here. Where the report does not document the completeness of an outcome, the question is answered PN or NI and the domain is decided by the tool's own path rather than by an assumption in either direction.

1. APPLAUSE-IgAN — iptacopan vs placebo (NCT04578834)

Outcomes used: 24-month annualized total eGFR slope (primary); composite kidney-failure endpoint (secondary, within the multiplicity-controlled sequence).

Domain	Questions	Answer and source-supported observation	Judgement
D1 Randomization	1.1–1.3	Y / PY / PN. Double-blind, placebo-controlled comparison reported as 1:1, with randomization described in the report and the registration record; the concealment mechanism is not described in the working record (NI on the mechanism only); no baseline imbalance in the renal variables was identified.	Low
D2 Deviations	2.1–2.7	N / N / PN / PN / NA / Y / NA. Double-blind, so neither participants nor carers were aware of the assignment. The report describes pre-specified handling of intercurrent events and continued collection of outcome data after treatment discontinuation and uses those data; analyses followed randomized assignment, so 2.6 is Y.	Low

Domain	Questions	Answer and source-supported observation	Judgement
D3 Missing outcome data — 24-month total slope	3.1–3.4	**PN / PN / PY / PN.** 3.1: the number of participants with missing eGFR values, by arm, is not reported, and the model includes values imputed under the intercurrent-event strategy, so the proportion with observed data cannot be derived. The supplementary 24-month change analysis gives 197 of 238 and 171 of 239 as having data, but the report's footnote states that this count excludes values **both** missing **and** imputed under the intercurrent-event strategy, so the two mechanisms cannot be separated and the count is not an estimate of completeness. 3.2: no analysis in the material varies the missing-data assumption. The treatment-policy supplementary analysis changes the handling of intercurrent events, and the supplementary 24-month change analysis addresses a different quantity; neither shows that the slope result is unchanged under a different missing-data assumption, so 3.2 is PN. 3.3: for an eGFR slope, deterioration is itself a trigger for rescue therapy, kidney-replacement therapy and withdrawal, so missingness could depend on the true value. 3.4: it is probably not likely to have done so to a substantial degree, because the strategy is pre-specified, identical in both arms and applied under blinding; but the extent to which the estimate depends on it cannot be bounded from the material.	**Some concerns**
D3 Missing outcome data — composite	3.1–3.4	**NI / PN / PY / PN.** 3.1: 477 of the 478 randomised participants were included in the final analysis, but the participant flow, with the numbers discontinued or censored by arm and reason, is presented in the article's supplementary figure, which is not in the material held; completeness for this outcome therefore cannot be established from the material available. 3.2: no analysis in the material tests this result against missing data. 3.3: the number of participants censored before 24 months, with reasons, is not documented in the material available, and informative censoring cannot be excluded from the material. 3.4: it is probably not likely to have been substantial — the endpoint is defined by protocol-specified events (a sustained ≥30% eGFR decline, a sustained eGFR below 15, dialysis, transplantation or kidney-failure death) and the analysis population is the randomised set — but the material does not allow the amount of censoring to be quantified.	**Some concerns**
D4 Measurement	4.1–4.5	N / PN / PY / PN / PN. eGFR and the composite components are laboratory- and event-defined; the analysis is blinded, so knowledge of allocation could not influence assessment of these quantities.	Low
D5 Selection of the result	5.1–5.3	Y / PN / PN. The primary endpoint and the statistical plan are pre-specified in the registration record; the renal composite is a secondary endpoint that the report shows to be inside the multiplicity-controlled testing sequence, and it is reported in full. Nothing indicates selection from other eligible measurements or analyses.	Low

Overall: some concerns, arising from D3 for both outcomes. The concern is that the completeness of the outcome data is not documented in the material available and no analysis in it varies the missing-data assumption; it is not a finding that the trial was conducted badly.

2. ALIGN — atrasentan vs placebo (NCT04573478)

Outcomes used: week-136 eGFR change (key secondary); total slope to week 136 (supportive); composite kidney outcome (secondary).

Domain	Questions	Answer and source-supported observation	Judgement
D1 Randomization	1.1–1.3	Y / PY / PN. Randomized, double-blind, placebo-controlled, with interactive-response randomization and masking maintained to database lock; no baseline imbalance reported.	Low

Domain	Questions	Answer and source-supported observation	Judgement
D2 Deviations	2.1–2.7	N / N / PY / PN / PN / Y / NA. Double-blind. Deviations arose from the trial context rather than from awareness of assignment: new SGLT2-inhibitor initiation during follow-up is reported and is unbalanced (25 of 170 with placebo vs 12 of 170 with atrasentan). Because the analysis followed randomized assignment, this affects the size of the estimated effect rather than introducing bias into the effect of assignment, and it is recorded as an interpretation limit rather than a domain concern.	Low
D3 Missing outcome data — eGFR outcomes (week-136 change; total slope)	3.1–3.4	**PN / PN / PY / PN.** 3.1: 154 of 170 participants in the atrasentan group and 142 of 170 in the placebo group completed the study, and the trial profile gives the numbers who left the study by arm with reasons (including loss to follow-up of 2 of 170 and 4 of 170); the number with missing **eGFR values** by arm and by outcome is not reported, and values collected after intercurrent events were censored rather than used, so completeness for these outcomes is not established. 3.2: the report states that a pre-specified supplementary analysis including all eGFR data regardless of intercurrent events was directionally supportive (appendix 1 p 32). That analysis changes the handling of intercurrent events; it does not vary the missing-data assumption, and no result of an analysis that does is in the material, so 3.2 is PN. 3.3: as for APPLAUSE-IgAN, deterioration is the trigger for rescue therapy and for leaving the study, so missingness could depend on the true value. 3.4: probably not likely to have been substantial, because the analysis is blinded and follows randomised assignment and the numbers leaving are documented; but the dependence of the eGFR estimates on the missing-data assumption cannot be bounded from the material.	**Some concerns**
D3 Missing outcome data — composite	3.1–3.4	**PN / PN / PY / PN.** 3.1: analysed in the intention-to-treat population as a time-to-event outcome, but the number censored before the end of follow-up is not reported for this outcome, so the question is not answered yes. 3.2: no analysis in the material tests this result against missing data. 3.3: participants leave the study for reasons that include progressive disease (4 of 170 and 2 of 170) and death, and informative censoring cannot be excluded from the material. 3.4: probably not likely to have been substantial — the study discontinuations are documented by arm and reason, and loss to follow-up was 2 of 170 and 4 of 170, so the extent is bounded.	**Some concerns**
D4 Measurement	4.1–4.5	N / PN / PY / PN / PN. eGFR and the composite components are laboratory- and event-defined; assessment was blinded.	Low
D5 Selection of the result	5.1–5.3	Y / PN / PN. The trial's primary endpoint (proteinuria at week 36) and the key secondary endpoint (week-136 eGFR change) are pre-specified in the registration record. The report states that the hierarchical testing sequence stopped when the key secondary endpoint was not significant and that the remaining secondary endpoints were not formally tested; all of these results, including the supportive total slope, are reported with their intervals. Stopping a pre-specified hierarchy is a stated property of the analysis plan, not evidence that the reported result was selected among multiple eligible analyses.	Low

Overall: some concerns, arising from D3 for both the eGFR outcomes and the composite. Two interpretation limits are also recorded: the halted testing hierarchy, and the unbalanced SGLT2-inhibitor co-intervention.

3. FIND-CKD, IgA subgroup — finerenone vs placebo (NCT05047263)

Outcomes used: chronic slope; total slope; composite of kidney failure or ≥40% eGFR decline.

Domain	Questions	Answer and source-supported observation	Judgement
D1 Randomization	1.1–1.3	Y / PY / PN. Randomized, blinded parent trial; the IgA analysis is a pre-specified disease subgroup, so randomization is preserved within the subgroup.	Low
D2 Deviations	2.1–2.7	N / N / PN / PN / NA / Y / NA. Blinded; the subgroup analysis follows randomized assignment.	Low
D3 Missing outcome data	3.1–3.4	**NI / PN / PY / NI.** 3.1: the number of subgroup participants with missing values for the slope outcomes, and how missing data were handled within the subgroup, are **not documented in the material held** , so the proportion missing cannot be bounded. 3.2: there is no subgroup-level analysis of the effect of missing data in the material held. 3.3: for an eGFR slope in a progressive disease, missingness could depend on the true value. 3.4: **NI** — the material contains nothing that bears on the likelihood either way. Being a blinded parent trial and reporting an interval does not address whether the subgroup's missing data depended on the subgroup outcome, and that reasoning is not used. The domain is therefore decided at some concerns on the two answers that are documented, and the limit of the material is stated rather than resolved by assuming no differential missingness.	**Some concerns**
D4 Measurement	4.1–4.5	N / PN / PY / PN / PN. eGFR and the composite components are laboratory- and event-defined.	Low
D5 Selection of the result	5.1–5.3	Y / PN / PN. The IgA subgroup analysis was pre-specified in the parent trial's plan; the report labels it exploratory and gives descriptive P values. That labelling limits how the estimate should be read, but it is not evidence that a particular result was selected from multiple eligible measurements or analyses, and the subgroup results are reported in full.	Low

Overall: some concerns, arising from D3 only. The small number of subgroup events is not recorded here as a domain finding; it enters through imprecision in Additional file 11.

4. NefIgArd — targeted-release budesonide vs placebo (NCT03643965)

Outcomes used: 2-year time-weighted average eGFR (primary); composite of confirmed ≥30% eGFR decline or kidney failure.

Domain	Questions	Answer and source-supported observation	Judgement
D1	1.1–1.3	Y / Y / PN. Randomized, double-blind, placebo-controlled; allocation concealment reported; no baseline imbalance reported.	Low
D2	2.1–2.7	N / N / PN / PN / NA / Y / NA. Double-blind; the time-limited treatment course and the post-treatment follow-up are part of the pre-specified estimand, so 2.6 is Y.	Low

Domain	Questions	Answer and source-supported observation	Judgement
D3	3.1–3.4	**PN / PY / NA / NA.** 3.1: the analysis set is the full analysis set (364, 182 per group) and the report states that all patients who prematurely discontinued study treatment were followed up throughout the 2-year trial with eGFR, proteinuria and creatinine measured up to 2 years unless they withdrew consent; but neither the number who withdrew consent nor the number of participants with observed values for the 2-year time-weighted average is reported, so **being in the full analysis set and having values imputed is not evidence that nearly all participants were actually measured** , and 3.1 is PN. 3.2: the report states that missing values were imputed by multiple imputation conditional on the patient's previous eGFR measurements, and that **pre-specified sensitivity analyses were performed that varied the missing-data assumption** and that analysed the data with a mixed-model repeated-measures approach, and it reports that these **and other sensitivity analyses were consistent with the primary analysis** (Results; appendix pp 6, 9–10). That is an analysis which varies the missing-data assumption and shows the result materially unchanged, so 3.2 is PY and the conditional questions do not arise.	Low
D4	4.1–4.5	N / PN / PY / PN / PN. eGFR and the composite components are laboratory- and event-defined, with stated confirmation requirements; assessment was blinded.	Low
D5	5.1–5.3	Y / PN / PN. The 2-year renal outcomes and their confirmation requirements are pre-specified in the registration and protocol, and the analyses presented follow that plan.	Low

Overall: low risk of bias.

5. PROTECT — sparsentan vs irbesartan (NCT03762850)

Outcomes used: chronic slope (weeks 6–110); total slope (day 1 to week 110); composite kidney endpoint.

Domain	Questions	Answer and source-supported observation	Judgement
D1	1.1–1.3	Y / Y / PN. Randomized, double-blind, active-controlled (irbesartan) comparison with allocation concealment reported.	Low
D2	2.1–2.7	N / N / PY / PN / PN / Y / NA. Double-blind. The report states that an imbalance in intercurrent events was observed, particularly in the rates and reasons for premature treatment discontinuation and in the initiation of rescue immunosuppressive therapy; the analyses followed randomized assignment, so this is recorded as an interpretation limit.	Low
D3 Missing outcome data — slope outcomes	3.1–3.4	**PN / PN / PY / PN.** 3.1: the analyses used on-treatment eGFR assessments, with missing data, including data after premature treatment discontinuation, imputed by multiple imputation under a missing-at-random assumption; the number of participants with observed values for each renal outcome is not reported, so completeness is not established. 3.2: the pre-specified sensitivity analyses reported for these endpoints vary the analysis set and the handling of data after premature treatment discontinuation (including a modified intention-to-treat analysis giving a difference of 1.2 mL/min/1.73 m² per year). They do **not** vary the missing-at-random assumption on which the imputation rests, and no result of an analysis that does is in the material, so 3.2 is PN. 3.3: the report itself states that an imbalance in intercurrent events was observed, in premature treatment discontinuation and in the initiation of rescue immunosuppressive therapy, so missingness could depend on the true value. 3.4: probably not likely to have been substantial, because the analyses are blinded, follow randomised assignment and were shown to be robust to the intercurrent-event imbalance; but the dependence of the slope estimates on the imputation assumption cannot be bounded from the material.	**Some concerns**

Domain	Questions	Answer and source-supported observation	Judgement
D3 Missing outcome data — composite kidney endpoint	3.1–3.4	**PN / PN / PY / PN.** 3.1: assessed separately from the slopes. The trial profile gives the numbers who discontinued the double-blind period by arm with reasons (in one arm, 12 of 202: 9 withdrew consent, 1 death, 1 lost to follow-up, 1 investigator decision), and the trial's Figure 4 gives the number at risk over time with censored participants marked; the number censored in each arm is not tabulated as a count, so the question is not answered yes. 3.2: no analysis in the material tests this composite against missing data. 3.3: participants who discontinue are censored at that point, and the reasons include withdrawal of consent and death, so informative censoring cannot be excluded from the material. 3.4: probably not likely to have been substantial — discontinuations are documented by arm and reason and the number at risk is shown throughout follow-up, so the extent of censoring is bounded.	**Some concerns**
D4	4.1–4.5	N / PN / PY / PN / PN. eGFR and the composite components are laboratory- and event-defined, with a confirmation requirement for the ≥40% eGFR component; assessment was blinded.	Low
D5	5.1–5.3	Y / PN / PN. The primary endpoint (proteinuria) and the renal secondary endpoints are pre-specified in the registration record; a pre-specified interim analysis was published separately and both documents are read together. Reporting a pre-specified interim separately is not evidence of selection among eligible analyses.	Low

Overall: some concerns, arising from D3 for both the slope outcomes and the composite, separately assessed. The intercurrent-event imbalance is also recorded, as an interpretation limit under D2.

6. TESTING — methylprednisolone vs placebo (NCT01560052)

Outcomes used: composite of 40% eGFR decline, kidney failure or kidney death; eGFR slope.

Domain	Questions	Answer and source-supported observation	Judgement
D1	1.1–1.3	Y / Y / PN. Randomized, double-blind, placebo-controlled, with allocation concealment reported.	Low
D2	2.1–2.7	N / N / PY / PN / PN / Y / NA. Double-blind. The trial was amended during recruitment to reduce the dose and to add prophylaxis; the amendment applied to both arms and the analyses followed randomized assignment, so it does not introduce differential deviation between arms.	Low
D3 Missing outcome data — composite	3.1–3.4	**PY / NA / NA / NA.** 3.1: 493 of 503 participants (98%) completed the trial, and the report states that almost all participants were followed up to the end of the trial, that missing data were therefore negligible, and that for time-to-event analyses participants were censored when they were last known to be free of the event. Data for this outcome were available for all or nearly all randomized participants, so the conditional answers do not arise.	Low
D3 Missing outcome data — eGFR slope	3.1–3.4	**PY / NA / NA / NA.** 3.1: the slope was analysed with longitudinal linear mixed models that included all post-randomization measurements, and the report states that all analyses were conducted including all available data points and that missing data were negligible because almost all participants were followed to the end of the trial. This judgement is made on that statement and on the analysis method for the slope, not by carrying the composite's result across: the two outcomes were assessed separately, and the slope requires repeated measurements rather than a single event.	Low

Domain	Questions	Answer and source-supported observation	Judgement
D4	4.1–4.5	N / PN / PY / PN / PN. Kidney failure and the eGFR-based components are laboratory- and event-defined; assessment was blinded.	Low
D5	5.1–5.3	Y / PN / PN. The registered primary outcome is the composite reported here, and the pre-specified analysis is the one presented; the dose and prophylaxis amendment and the reporting of both phases together are described in the report, and nothing in the report contradicts the pre-specified analysis.	Low

Overall: low risk of bias. The earlier concern about missing outcome data, which had been inferred from glucocorticoid-related discontinuation of treatment, was not supported: treatment discontinuation was reported, but the report states that almost all participants were followed to the end of the trial.

7. MAIN — mycophenolate mofetil plus supportive care vs supportive care alone (NCT01854814) [26]

Outcomes assessed separately: the composite of creatinine doubling, kidney replacement therapy or kidney/cardiovascular death; and trial-phase annual eGFR decline, 1.2 (0.6) versus 3.8 (0.6) mL/min/1.73 m²/year, P<0.001. The report specifies a linear mixed-effects treatment-effect model but provides no interval for this trial-phase contrast. The arithmetic rate difference is 2.6. The 0.6 values are not assumed to be individual-slope SDs and no contrast interval is reconstructed from them. Post-trial data remain separate.

Registration record NCT01854814 was first submitted on 9 May 2013, before recruitment began: interventional, allocation randomized, parallel assignment, masking none, actual enrolment 238 (the run-in population), with the composite endpoint and its 3-year time frame as a registered primary outcome. The registration lists several participating facilities, whereas the report describes the trial as single-centre at Nanfang Hospital; that discrepancy is recorded here as a limit in the description of the setting and is not used in any domain.

Domain	Questions	Answer and source-supported observation	Judgement
D1 Randomization	1.1–1.3	**Y / NI / PN.** The report states that participants "were randomly assigned at a 1:1 ratio to receive MMF plus SC or SC alone" and that "randomization was carried out via phone call". That is the whole of what the material says about the process: it does not state how the sequence was generated or held, nor that it was concealed until assignment, so 1.2 is **NI** and the domain follows the tool's path for that answer rather than an assumption in either direction. 1.3: the published baseline table shows differences of the size expected at n = 85 per group (mean age 35.0 vs 38.2 years; 49.4% vs 61.2% male; E1 in the Oxford classification 11.8% vs 5.9%), with no pattern suggesting that assignment could be anticipated; the prespecified Cox models adjust for these same variables.	**Some concerns**
D2 Deviations	2.1–2.7	**Y / Y / PN / NA / NA / Y / NA.** Open-label, so participants and carers were aware of the assignment. 2.3 is PN rather than N on the evidence that the report gives: both groups continued comprehensive supportive care as in the run-in period, the losartan dose was comparable at randomization and during the trial phase (119 [50] vs 127 [19] mg), and mean compliance was 94.3% and reported as balanced between groups. The report does not enumerate individual protocol deviations, which is why the answer is PN. 2.6 is Y: "efficacy and safety analyses were performed using the intention-to-treat population". Being open-label is not by itself a finding in this domain, and no imbalance in the delivered treatment is documented.	Low
D3 Missing outcome data — composite	3.1–3.4	**PY / NA / NA / NA.** 3.1: the report states that 168 of the 170 randomised participants (98.8%) completed the trial and that "during the 3-year trial phase, 2 participants were lost to follow-up, 1 each from the MMF and SC groups", and the trial profile shows one loss in each arm; the analysis is a time-to-event comparison, with participants censored when last known to be free of the event. Because the completeness is documented per arm and the residual missingness is one participant per group, the question is answered probably yes rather than yes, and the conditional questions do not arise.	Low

Domain	Questions	Answer and source-supported observation	Judgement
D3 Missing outcome data — annual eGFR rate	3.1–3.4	**PY / NA / NA / NA.** 3.1: the **same completion figures apply as for the composite —** 168 of the 170 randomised participants (98.8%) completed the trial **phase and the two losses are documented one per arm — so the eGFR series over three years is recorded per arm; the number contributing a value to the modelled annual rate is not tabulated, which is why the answer is probably yes rather than yes. 3.2 to 3.4 therefore do not arise. The post-trial values are not used here, so no question arises about their completeness.**	Low
D4 Measurement	4.1–4.5	**N / PN / PY / PN / PN.** The components are laboratory values (creatinine doubling) and events (kidney replacement therapy, death), and the report states that "an adjudicating committee, whose members were unaware of patient treatment assignments, adjudicated end point events". Assessors other than the adjudicators were aware of the assignment (PY), which raises the question of influence (4.4), answered PN because the components are defined by protocol criteria rather than by clinical judgement — with one component excepted: initiation of kidney replacement therapy is a clinical decision, and the clinicians making it were not blinded. That is recorded here as a limit of the outcome definition. It belongs to this domain and to this domain only, and it is not carried into the allocation judgement in D1 or into the certainty rating in Additional file 11, both of which rest on what the material says about the allocation process and about precision respectively.	Low
D5 Selection of the result	5.1–5.3	**PY / PN / PN.** The composite is a registered primary outcome with a 3-year time frame; the published definition adds "death from a kidney or cardiovascular cause" where the registration says "renal death", so 5.1 is probably yes rather than yes. The report's statistical analysis states that Cox models were fitted "with and without covariate adjustment" for this outcome; both estimates are reported, so the result used here is not selected from among several analyses that point in different directions.	Low

Overall: some concerns, arising from D1 only, and specifically from the unresolved question of whether the allocation sequence was concealed; it is not a finding that allocation was improper. **The judgement applies to both outcomes assessed here: the annual eGFR rate rests on the same allocation and on completeness that the report documents per arm, so D3 is low for it as well.** Two limits are also recorded: the composite includes a component (initiation of kidney replacement therapy) whose ascertainment could in principle be influenced by knowledge of assignment in an open-label trial, and the report describes the trial as single-centre while the registration lists several facilities.

8. STOP-IgAN — added immunosuppression vs supportive care, 10-year follow-up (NCT03488368) [29]

Outcome used: composite of all-cause death, end-stage kidney disease or a decline in eGFR exceeding 40% from baseline, analysed over a median follow-up of 7.4 years (IQR 5.7–8.3). The report calls this an "adapted primary outcome"; it is the registered primary outcome of NCT03488368, whose record was first submitted on 28 March 2018 with the endpoint defined by "available data by March 31, 2018" — three days later — and with the secondary outcomes also specified. The trial whose participants were followed is STOP-IgAN itself (starting randomization reported as 162 participants, 80 vs 82).

Domain	Questions	Answer and source-supported observation	Judgement
D1 Randomization	1.1–1.3	**Y / NI / PN.** The allocation to the two compared groups is that of the original randomized trial. The long-term report refers for the details of "eligibility criteria and treatment allocation" to that trial's report [27], which was consulted rather than relying on the long-term report alone. 1.1 is **Y** on its statement that "Randomization codes that were used to assign patients in a 1:1 ratio were generated by means of covariate adaptive randomization with respect to factors that had the potential to modify the treatment effect (i.e., eGFR and proteinuria)". 1.2 is **NI** : neither that report nor the 10-year follow-up report states how the codes were held or kept out of the reach of those enrolling participants, so the concealment question is unresolved in the material and the domain follows the tool's path for that answer. What is unresolved is concealment alone, not whether the sequence was random. 1.3: the report reports no imbalance between the randomized groups and states that its comparison of the baseline characteristics of the 13 participants without long-term data against those of the 149 with data "failed to reveal major differences" (Supplementary Table S1).	**Some concerns**
D2 Deviations	2.1–2.7	**Y / Y / PN / NA / NA / Y / NA.** Open-label trial. 2.3 is PN on the evidence available: the report states that 156 of the 162 randomized patients (96.3%) finished the original 3-year trial, including the two who died, and it does not describe deviations in the delivery of the assigned trial treatment. 2.6 is Y: the long-term analysis compares the groups as randomized, 72 assigned to supportive care and 77 to added immunosuppression. A limit is recorded separately: the follow-up after the third year was observational and treatment during it was at the discretion of physicians who knew the original allocation, so the 10-year estimate describes the consequences of assignment rather than of continued therapy; comparability of post-trial care between groups is not documented, and this is returned to in Additional file 11.	Low
D3 Missing outcome data — composite	3.1–3.4	**PN / PY / NA / NA.** 3.1: data on mortality, end-stage kidney disease and eGFR were captured for 149 of the 162 randomized participants (92.0%; 72 vs 77), and long-term data were unavailable for 13, so the question is not answered yes. 3.2: the statistical methods state that "sensitivity analyses for the primary endpoint were performed using best- and worst-case strategies", replacing missing data by the assumption that the primary endpoint was not achieved by 31 March 2018 in one analysis and that it was achieved immediately after the last available value in the other — two opposite assumptions applied to the same missing values — and the results state that "worst- and best-case analyses confirmed the lack of a significant difference between the 2 groups (Supplementary Table S2)". That is an analysis which varies the missing-data assumption and shows the result unchanged, so 3.2 is PY and the conditional questions do not arise. This is the second comparison in this file, after NefIgArd, for which such an analysis exists in the material.	Low
D4 Measurement	4.1–4.5	**N / PN / PY / PN / PN.** Death, end-stage kidney disease and a >40% eGFR decline are event- and laboratory-defined, with eGFR calculated by the CKD-EPI creatinine equation; the sites provided the values from routine records and the report states that independent double data entry was performed. Those providing the data knew the original allocation (4.3 PY), but nothing in the material suggests this influenced whether an event met the definition (4.4 PN); the endpoint components do not depend on adjudicated judgement.	Low
D5 Selection of the result	5.1–5.3	**Y / PN / PN.** The composite analysed is the outcome registered as primary in NCT03488368, with the same components, the same thresholds and the same time frame, registered before the data assembled for this analysis existed. The report presents the composite and each component separately (all-cause death, end-stage kidney disease, >40% eGFR loss, >30% eGFR loss) and adds a multivariate analysis in which "the treatment regimen in the trial phase was not predictive"; no result is presented in isolation. The report's word "adapted" describes the endpoint's relation to the original trial's primary endpoint, not a departure from the follow-up study's own registered plan.	Low

Overall: some concerns, arising from D1 only, on the unresolved question of how the original allocation was concealed — the generation of the sequence is described in the original trial report [27], and what the material does not state is how the codes were held. One limit is also recorded: treatment during the observational follow-up was neither standardized nor shown to be comparable between the groups.

9. DAPA-CKD, IgA subgroup — dapagliflozin vs placebo (NCT03036150) [30]

Outcomes used: (i) the parent trial's primary composite — a sustained decline in eGFR of ≥50% confirmed over at least 28 days, end-stage kidney disease (initiation of dialysis or transplantation, or sustained eGFR below 15 mL/min/1.73 m²), or death from a kidney disease-related or cardiovascular cause — within the pre-specified IgA nephropathy subgroup: 6/137 vs 20/133, hazard ratio 0.29 (0.12–0.73), median follow-up 2.1 years. Two further outcomes are used from the same report and are assessed separately below: the total slope difference, 1.2 mL/min/1.73 m² per year (−0.12 to 2.51), and the chronic slope difference, 2.4 (1.08 to 3.71), whose window begins at week 2. All three results appear in Tables 2 and 3 of the main text.

Registration record NCT03036150 records allocation randomized, parallel assignment and quadruple masking (participant, care provider, investigator, outcomes assessor), with 4304 participants enrolled.

Domain	Questions	Answer and source-supported observation	Judgement
D1 Randomization	1.1–1.3	**Y / Y / PN.** Participants were randomly assigned "in accordance with the sequestered, fixed randomization schedule, using balanced blocks", and randomization "was conducted using an interactive voice- or web-based system and stratified on the diagnosis of type 2 diabetes and UACR"; the sequence was therefore concealed until assignment, so 1.2 is Y. For the subgroup itself the report states that "participants assigned to dapagliflozin or placebo had similar baseline characteristics (Table 1)". The subgroup is defined by a characteristic recorded at enrolment, so randomization is preserved within it.	Low
D2 Deviations	2.1–2.7	**N / N / PN / PN / NA / Y / NA.** Blinded, with drug and placebo identically packaged; the report states that study personnel other than the independent data monitoring committee and the participants were blinded to allocation. 2.6 is Y: "we included data from all randomized patients according to the intention-to-treat principle". The parent trial was stopped early on the recommendation of that committee; that affects how many events accumulated and is recorded as an interpretation limit, not as a deviation domain finding.	Low
D3 Missing outcome data — composite	3.1–3.4	**PN / PN / PY / NI.** 3.1: the material does not report, for the IgA subgroup, how many participants lacked the data needed to ascertain this composite, so completeness within the subgroup cannot be established; analysing all randomized participants is not taken as evidence that their outcome data were complete. 3.2: the sensitivity analysis reported for this endpoint was restricted to participants with biopsy-proven IgA nephropathy (254 of 270) and gave a hazard ratio of 0.28 (0.11–0.72). That analysis varies the population, not the missing-data assumption, and no analysis in the material varies that assumption, so 3.2 is PN. 3.3: participants who progress may stop attending visits or stop the study drug, so missingness could depend on the true value. 3.4: **NI** — the material contains nothing that bears on how likely that dependence was; being masked and having an interval reported is not used as an answer to this question.	**Some concerns**
D3 Missing outcome data — total slope	3.1–3.4	**PN / PN / PY / NI.** 3.1: the **number of subgroup participants with missing eGFR values, by arm, is not reported, so the completeness of the slope data cannot be derived, and participants who started kidney replacement therapy contribute no further values. 3.2: the biopsy-proven sensitivity analysis reported for the composite** varies the population, not the missing-data assumption, and no analysis in the material varies **how the slope model handles missing values. 3.3: missingness could depend on the true value, since participants whose kidney function is failing attend fewer visits. 3.4: NI — nothing in the material bears on how likely that dependence was.**	**Some concerns**
D3 Missing outcome data — chronic slope	3.1–3.4	**PN / PN / PY / NI, for the same reasons as the total slope and with one addition: the chronic window starts at week 2, so participants who discontinued the study drug immediately after randomization contribute less to this estimate than to the total slope, and the report does not give that number by arm.**	**Some concerns**

Domain	Questions	Answer and source-supported observation	Judgement
D4 Measurement	4.1–4.5	**N / PN / PY / PN / PN.** The ≥50% decline requires a sustained fall confirmed over at least 28 days, and end-stage kidney disease and death are protocol-defined; the report states that "an independent event adjudication committee assessed all clinical endpoints using these prespecified endpoint definitions", and the registration records masked outcome assessment.	Low
D5 Selection of the result	5.1–5.3	**Y / PN / PN.** The statistical analysis states that "we prespecified analyses of the effects of dapagliflozin on the primary and secondary efficacy endpoints in participants according to the etiology of kidney disease, with the glomerulonephritis category further subcategorized by underlying cause, including IgA nephropathy", and that "IgA nephropathy was included as a prespecified category among participants with glomerulonephritis". The report presents the primary composite, the kidney-specific secondary composite and the other components for the subgroup, with forest plots across pre-specified baseline strata, so the result used here was not selected from among reported alternatives that pointed the other way.	Low

Overall: some concerns, arising from D3 **only, and the judgement applies to each of the three outcomes assessed here.** The small number of events in the subgroup is not a finding here; it enters through imprecision in Additional file 11.

10. EMPA-KIDNEY, IgA subgroup — empagliflozin vs placebo (NCT03594110) [31]

Outcomes used: kidney disease progression within the IgA nephropathy group — sustained ≥40% eGFR decline from randomization, end-stage kidney disease, sustained eGFR below 10 mL/min/1.73 m², or death from kidney failure — 51/413 vs 67/404, hazard ratio 0.67 (0.46–0.97), median follow-up 2.0 years; and the chronic eGFR slope (from 2 months to the final follow-up visit), reported as a relative reduction of 43% (95% CI 27–59). The IgA group is one of three categories into which the pre-specified glomerular disease subgroup was split; the report states explicitly that this split was made post hoc.

Registration record NCT03594110 records allocation randomized, parallel assignment and double masking (participant, investigator), with 6609 participants enrolled.

Domain	Questions	Answer and source-supported observation	Judgement
D1 Randomization	1.1–1.3	**Y / PY / PN.** The parent trial randomized 6609 participants (1:1) "using a minimisation algorithm" with adjustment for categorised baseline variables, and the registration records masking of participants and investigators. 1.2 is **PY** rather than Y: a minimisation algorithm requires the allocation to be determined at the moment of assignment, but the report does not describe how the sequence was held or concealed, and that limit is stated rather than resolved. 1.3: no imbalance is reported; the subgrouping is by primary kidney disease recorded at enrolment, so randomization is preserved within the split, and the per-arm baseline values within the IgA group are not tabulated in the material held.	Low
D2 Deviations	2.1–2.7	**N / N / PN / PN / NA / Y / NA.** Double-blind parent trial. 2.6 is Y: the analyses were by randomized assignment, with time at risk running from randomization; participants were censored at the earliest of death, loss to follow-up or withdrawal of consent. The trial was recommended to stop early by its data monitoring committee; this affects the number of events and belongs to interpretation, not to this domain.	Low

Domain	Questions	Answer and source-supported observation	Judgement
D3 Missing outcome data — composite progression outcome	3.1–3.4	**PN / PN / PY / NI.** 3.1: within the IgA group the numbers lost to follow-up or withdrawing consent are not reported, so completeness for this outcome in this split is not established. 3.2: no analysis in the material varies the missing-data assumption for this comparison. 3.3: participants whose disease progresses are the most likely to stop attending or to withdraw consent, so missingness could depend on the true value. 3.4: **NI** — the material held says nothing about how likely that dependence was.	**Some concerns**
D3 Missing outcome data — chronic slope	3.1–3.4	**PN / PN / PY / NI.** 3.1: the completeness of eGFR measurements within the IgA group is not documented; the models used all available central-laboratory values, which is not the same as their being complete. 3.2: the slope was estimated with shared parameter models, which jointly model the rate of change in eGFR and the time to end-stage kidney disease or death. That addresses one specific mechanism — dropout produced by the event itself — but it is a modelling assumption rather than a sensitivity analysis, and no analysis in the material varies the missing-data assumption for this comparison, so 3.2 is PN. 3.3 and 3.4: as above, PY then NI.	**Some concerns**
D4 Measurement	4.1–4.5	**N / PN / PY / PN / PN.** The ≥40% decline component is defined with a persistence requirement over consecutive scheduled visits; eGFR values are from central-laboratory measurement wherever available; masking of participants and investigators is recorded in the registration.	Low
D5 Selection of the result	5.1–5.3	**N / PN / PN.** 5.1 is answered no on the report's own statement: the prespecified subgroups were diabetic, glomerular, hypertensive or renovascular, and other-or-unknown causes, while "the glomerular disease subgroup was further split post hoc into IgA nephropathy, focal segmental glomerulosclerosis, and other glomerulonephritides", a split made "to test hypotheses raised by subsidiary reports from the DAPA-CKD trial". The IgA comparison therefore did not have a pre-specified analysis plan finalised before outcome data were available. 5.2 and 5.3 are PN: the outcome is the parent trial's registered primary outcome rather than one of several measurements of it, one prespecified Cox model adjusted for the minimisation variables with an interaction term is used, and all three glomerular categories are reported side by side with their intervals and a heterogeneity test ($p = 0.58$), as are the total and chronic slopes, so no result is presented in isolation. The domain is nevertheless not judged low: the material establishes that the split was post hoc but not that it was defined independently of the results — an external reason is given, yet whether a different split would have been chosen cannot be checked from the report. That is recorded as an unresolved uncertainty about this comparison and is not converted into a finding that the result was selected.	**Some concerns**

Overall: some concerns, arising from D3 for both outcomes, assessed separately, and from D5 for the post hoc definition of the comparison. The post hoc status of the split is a property of how this comparison came to be defined; it does not by itself show that the reported estimate was chosen from among alternatives, and the review presents all three glomerular categories together.

11. Part B. The other results that entered the synthesis

Part B reports additional outcome-specific appraisals. Previously completed assessments are distinguished from new or corrected source-checked assessments. Each result retains its own outcome, analysis population and five domain judgements; trial membership does not transfer an appraisal automatically.

GRADE remains restricted to the 23 focused outcomes. The result-level coverage map identifies assessed, unassessed and unresolved-eligibility outcomes and the provenance of each assessment.

The abstract-only blisibimod result remains unassessed and is excluded from appraisal counts. The full Masutani report supports its stated appraisal. Five results from these entries retain candidate appraisals without entering the primary adult synthesis.

The rows below give the domain answers with the source-supported observation for each; "NI" means the material held for this review does not answer the question, and the domain then follows the tool's own path for that answer.

Comparison and result assessed	**D1**	**D2**	**D3**	**D4**	**D5**	**Overall**	**Source and appraisal provenance**
Manno 2009 — Creatinine doubling or end-stage kidney disease [adult eligibility unresolved; candidate only]	Low	Some concerns	Some concerns	Low	Some concerns	Some concerns	D1 low on its own statement: the sequence "was generated at the coordinating centre by random number tables; a list divided into blocks of 10 was adequately concealed", with "central telephone randomization for every eligible patient". D2: open-label, and the report enumerates no individual deviations, so 2.3 to 2.5 are NI. D3: completeness for these two outcomes is not documented; the report states that data from participants who dropped out "were collected and analysed for the primary outcome", which covers those participants rather than establishing that nearly all participants had values for these outcomes, and no analysis varies the missing-data assumption. D5: no registration record or protocol was located in the material, so whether the analysis plan was finalised before unblinded data were available cannot be checked. Provenance: previously completed assessment retained. Not established by available files.
Manno 2009 — End-stage kidney disease alone [adult eligibility unresolved; candidate only]	Low	Some concerns	Some concerns	Low	Some concerns	Some concerns	D1 low on its own statement: the sequence "was generated at the coordinating centre by random number tables; a list divided into blocks of 10 was adequately concealed", with "central telephone randomization for every eligible patient". D2: open-label, and the report enumerates no individual deviations, so 2.3 to 2.5 are NI. D3: completeness for these two outcomes is not documented; the report states that data from participants who dropped out "were collected and analysed for the primary outcome", which covers those participants rather than establishing that nearly all participants had values for these outcomes, and no analysis varies the missing-data assumption. D5: no registration record or protocol was located in the material, so whether the analysis plan was finalised before unblinded data were available cannot be checked. Provenance: previously completed assessment retained. Not established by available files.
Pozzi 1999 — ≥50% rise in plasma creatinine [adult eligibility unresolved; candidate only]	Some concerns	Some concerns	Low	Low	Some concerns	Some concerns	D1: 1.2 is NI; the report states that patients "were randomly assigned to the steroid or the control group by means of a centralised table of random numbers" but not how it was held. D2: open-label with a supportive-care comparator; no deviations are enumerated, so those questions are NI; 2.6 is Y on "Analyses were by intention to treat". D3 low because the outcome counts are reported with denominators equal to the randomised numbers (9 of 43 versus 14 of 43 for the 50% threshold; 1 of 43 versus 9 of 43 for the 100% threshold), so the data were available for nearly all randomised participants. D5: no registration record was located for this trial Provenance: Previously completed assessment retained. Not established by available files.

Comparison and result assessed	D1	D2	D3	D4	D5	Overall	Source and appraisal provenance
Pozzi 1999 — 100% plasma creatinine rise (doubling) [adult eligibility unresolved; candidate only]	Some concerns	Some concerns	Low	Low	Some concerns	Some concerns	D1: 1.2 is NI; the report states that patients "were randomly assigned to the steroid or the control group by means of a centralised table of random numbers" but not how it was held. D2: open-label with a supportive-care comparator; no deviations are enumerated, so those questions are NI; 2.6 is Y on "Analyses were by intention to treat". D3 low because the outcome counts are reported with denominators equal to the randomised numbers (9 of 43 versus 14 of 43 for the 50% threshold; 1 of 43 versus 9 of 43 for the 100% threshold), so the data were available for nearly all randomised participants. D5: no registration record was located for this trial Provenance: Previously completed assessment retained. Not established by available files.
Ballardie 2002 — renal survival at five years	Some concerns	Some concerns	Some concerns	Low	Some concerns	Some concerns	D1: the report states that patients were randomized to two arms but gives neither the method nor concealment, so 1.2 is NI. D2: open-label across two arms with a provision that "patients could exit the trial at 2 yr"; neither the number exercising it nor individual deviations are reported. D3: the result is a Kaplan–Meier estimate with the proportion surviving given but without the number at risk or the numbers withdrawn, so completeness for this outcome is not established Provenance: Previously completed assessment retained. Not established by available files.
Dipyridamole plus low-dose warfarin (Lee 1997) — end-stage kidney disease during the trial	Some concerns	Some concerns	Low	Low	Some concerns	Some concerns	D1: allocation is described only as randomization, with no method and no concealment, and the report documents baseline differences favouring the control group: patients randomized to treatment "were younger, and had worse scores on histology for tubular atrophy and arteriolar hyperplasia", so 1.3 is answered PN only because the trial randomized 21 participants in total. D2: the comparator received no study treatment, so participants were aware of assignment, and no deviations are enumerated. D3 low: the event counts are given with denominators equal to the randomized numbers (1 of 10 versus 4 of 11) and the report describes how end-of-trial values were recorded for participants withdrawn for reaching a creatinine threshold Provenance: Previously completed assessment retained. Not established by available files.
Lv 2009 — kidney survival, defined by a ≥50% rise in serum creatinine	Some concerns	High	Some concerns	Low	Some concerns	High	D1: randomization "by means of a computer-generated randomization code" is stated, concealment is not, and the report documents an imbalance in the direction of worse prognosis for the combination arm: those randomized to combination therapy "had a significantly greater degree of proteinuria and tended to have more glomerulosclerosis". D2 is high : the only between-group estimate the report gives for this outcome comes from a model adjusted for on-study time-averaged proteinuria, which is a variable measured after randomization and affected by the treatment itself, so the analysis does not estimate the effect of assignment (2.6 PN), and the imbalance it adjusts for is itself unfavourable to the experimental arm, so the impact could be substantial (2.7 PY). This review reaches no treatment conclusion from this row precisely because of that. D3: completeness is not documented. D5: no registration record was located Provenance: Previously completed assessment retained. Not established by available files.

Comparison and result assessed	D1	D2	D3	D4	D5	Overall	Source and appraisal provenance
Frisch 2005 — ≥50% rise in serum creatinine	Some concerns	Low	Some concerns	Low	Some concerns	Some concerns	D1: double-blind and placebo-controlled, but how the sequence was generated and concealed is not stated, so 1.2 is NI; the report notes baseline differences "that arose despite randomization", which it adjusted for, and they are answered PN rather than as indicating a problem. D2: blinded, and 2.6 is Y on "All analyses were done as intention to treat"; the trial was terminated early at the second scheduled interim analysis, which is recorded as an interpretation limit because it bounds the number of events rather than as a deviation finding. D3: the numbers not completing treatment are given by arm with reasons, but the number with values for each outcome is not stated and the reporting is interrupted by the early termination, so completeness is not established Provenance: Previously completed assessment retained. Not established by available files.
Frisch 2005 — End-stage renal disease	Some concerns	Low	Some concerns	Low	Some concerns	Some concerns	D1: double-blind and placebo-controlled, but how the sequence was generated and concealed is not stated, so 1.2 is NI; the report notes baseline differences "that arose despite randomization", which it adjusted for, and they are answered PN rather than as indicating a problem. D2: blinded, and 2.6 is Y on "All analyses were done as intention to treat"; the trial was terminated early at the second scheduled interim analysis, which is recorded as an interpretation limit because it bounds the number of events rather than as a deviation finding. D3: the numbers not completing treatment are given by arm with reasons, but the number with values for each outcome is not stated and the reporting is interrupted by the early termination, so completeness is not established Provenance: Previously completed assessment retained. Not established by available files.
Tang 2005 — ≥50% rise in serum creatinine	Some concerns	Some concerns	Some concerns	Low	Some concerns	Some concerns	D1: "randomly assigned to receive MMF ... or to continue contemporaneous medications without addition of MMF" is the whole statement; no method, no concealment. D2: the comparator was continuation of ongoing treatment rather than placebo and blinding is not described; 2.6 is Y on "The main efficacy analysis was performed on an intention-to-treat basis and included patients who underwent randomization". D3: 40 participants were randomized and the report gives no numbers withdrawn or missing; it notes that the design assumed a 10% dropout rate, which is not a statement about what occurred. D5: the outcome is a pre-specified secondary endpoint of a trial whose primary endpoint was proteinuria, and no registration record was located Provenance: Previously completed assessment retained. Not established by available files.
Mayo fish-oil trial, long-term report — ≥50% rise in serum creatinine over a mean of 6.4 years	Some concerns	Some concerns	Low	Low	Some concerns	Some concerns	D1: 1.2 is NI, the long-term report not restating the procedure of the trial it follows. D2 is recorded at some concerns rather than high, and the reason matters: 17 of the 30 placebo-assigned participants at risk at the end of the randomized trial switched to fish oil, so the intervention actually received diverged from the assignment and was unbalanced, but the report analyses the groups as randomized, so what is estimated is still the effect of assignment; the consequence is that the long-term contrast does not isolate the effect of receiving fish oil throughout, which is already recorded in Additional file 12. D3 low: one participant was lost to follow-up after completing two years, so nearly all randomized participants contributed. D5: no registration record was located for a trial that ran before registration was customary Provenance: Previously completed assessment retained. Not established by available files.

Comparison and result assessed	D1	D2	D3	D4	D5	Overall	Source and appraisal provenance
Alexopoulos 2004 — ≥50% rise in serum creatinine	Some concerns	High	Some concerns	Low	Some concerns	High	D1: 1.2 is NI; patients were "randomly assigned to receive either fish-oil or supportive treatment" with no method stated. D2 is high : six of the 34 randomized patients, four in the fish-oil arm and two in the control arm, "either did not comply with the treatment or were lost to follow-up and excluded from the analysis", so participants were excluded from the analysis for reasons that include non-compliance, the exclusions were unequal, and in a trial of this size the exclusion of a fifth participant's data from one arm could have affected the direction of the result (2.6 PN, 2.7 PY). D3: 28 of 34 randomized participants were analysed, so the question of completeness is answered no rather than probably, and no analysis varies the missing-data assumption Provenance: Previously completed assessment retained. Not established by available files.
Praga 2003 — ≥50% rise in baseline serum creatinine (renal survival, the trial's primary endpoint)	Low	Some concerns	Low	Some concerns	Some concerns	Some concerns	D1 low: simple randomization with a table of random numbers, and allocation concealment by enclosing assignments in sequentially numbered, opaque closed envelopes; the groups were similar at baseline for age, sex, blood pressure, serum creatinine, creatinine clearance and proteinuria (Table 1). Physical pp2–3. D2 some concerns: the trial is unblinded, so assignment was known to participants and investigators; blood-pressure targets (≤140/90 mmHg) and the permitted add-on agents were specified for both arms, the report states that there were no differences in blood-pressure control during follow-up, and the primary and secondary endpoints were analysed by intention-to-treat on the randomized denominators of 23 and 21. Differential co-intervention cannot be excluded from the report, and the concern is recorded rather than dismissed. Physical pp1–3. D3 low: five of the 44 randomized participants withdrew — enalapril was interrupted in one woman after 36 months because of an intention to become pregnant, and four (two per arm) were lost to follow-up between 29 and 76 months — while the primary-endpoint counts are reported against the full randomized denominators under a stated intention-to-treat analysis. Physical pp1,3–4. D4 some concerns: the outcome is an objective laboratory threshold, a 50% rise in serum creatinine against the first post-randomization value, sampled at every protocol visit, three-monthly in year one and six-monthly thereafter in both arms; no adjudication is described and measurement was unblinded. Physical pp2–3. D5 some concerns: renal survival defined by a 50% rise in baseline serum creatinine is the stated primary endpoint and is the endpoint reported, with one primary analysis by intention-to-treat; the secondary serum-creatinine endpoint and proteinuria are reported separately rather than substituted for it. No dated protocol or statistical analysis plan was held, so its pre-specification could not be verified, which is not evidence that it was chosen for its result. Physical pp2–4. Source: J Am Soc Nephrol 2003;14:1578–1583, physical PDF pages 1, 2, 3, 4. New source-checked assessment.

Comparison and result assessed	D1	D2	D3	D4	D5	Overall	Source and appraisal provenance
Liu 2014 — ≥25% fall in eGFR within 12 months	Some concerns	Some concerns	Some concerns	Some concerns	Some concerns	Some concerns	D1: the report states that eligible patients "were assigned to the MP group or MP plus CsA group successively and randomly"; neither the generation of the sequence nor any concealment mechanism is described, and the wording does not distinguish random from sequential allocation. Physical p2. D2: open-label treatment with no placebo; two combination-arm patients discontinued CsA after three months because of severe pulmonary infection and one withdrew, and no assignment-preserving analysis is stated, the comparisons being made on the 48 analysed. Physical pp2,4. D3: 51 patients were treated (26 MP, 25 combination) and 48 analysed (25/23); three were lost to follow-up, one in the MP and two in the combination arm, so missing outcome data differ between arms. Physical p4. D4: eGFR is calculated from serum creatinine, an objective measurement, but assessors were not blinded and no adjudication of the 25% threshold is described. Physical pp2,4. D5: the 25% eGFR fall is a pre-specified secondary endpoint reported with counts, but several secondary endpoints are analysed and no dated protocol was held. Physical pp2,4. Source: Intern Med 2014;53:675–681, physical PDF pages 1, 2, 3, 4. New source-checked assessment.
Li 2013 — eGFR at 60 months	Low	Some concerns	Some concerns	Low	Some concerns	Some concerns	D1 low: patients were randomized by a computer-generated list generated and maintained by a third party not involved in conducting the study. Physical p2. D2 some concerns: open-label with no placebo; analysis was stated to be on an intention-to-treat basis and the two patients who stopped ramipril for cough or dizziness were retained, but blood pressure was marginally lower in the ramipril group throughout, so a difference in co-intervention cannot be separated from the assigned treatment. Physical pp2,4. D3 some concerns: one treatment-arm and two control-arm participants were lost to follow-up, while the CONSORT figure reports 30 of 30 analysed in each arm; the tabulated 60-month eGFR uses those denominators. Physical pp4,5. D4 low: eGFR is derived from serum creatinine by a validated equation and is an objective measurement, and blood pressure was recorded by paramedical staff; no adjudication is described. Physical pp2,5. D5 some concerns: the 60-month eGFR and the rate of decline are secondary analyses of a trial whose registered primary endpoint was a composite of hypertension, proteinuria of 1 g/day or more, or a 20% eGFR decline; both are reported in full, and no dated protocol was held. Physical pp2,4. Source: Am J Med 2013;126:162–168, physical PDF pages 1, 2, 3, 4, 5. New source-checked assessment.
Li 2013 — rate of decline of eGFR over 60 months	Low	Some concerns	Some concerns	Low	Some concerns	Some concerns	D1 low: patients were randomized by a computer-generated list generated and maintained by a third party not involved in conducting the study. Physical p2. D2 some concerns: open-label with no placebo; analysis was stated to be on an intention-to-treat basis and the two patients who stopped ramipril for cough or dizziness were retained, but blood pressure was marginally lower in the ramipril group throughout, so a difference in co-intervention cannot be separated from the assigned treatment. Physical pp2,4. D3 some concerns: one treatment-arm and two control-arm participants were lost to follow-up, while the CONSORT figure reports 30 of 30 analysed in each arm; the tabulated 60-month eGFR uses those denominators. Physical pp4,5. D4 low: eGFR is derived from serum creatinine by a validated equation and is an objective measurement, and blood pressure was recorded by paramedical staff; no adjudication is described. Physical pp2,5. D5 some concerns: the 60-month eGFR and the rate of decline are secondary analyses of a trial whose registered primary endpoint was a composite of hypertension, proteinuria of 1 g/day or more, or a 20% eGFR decline; both are reported in full, and no dated protocol was held. Physical pp2,4. Source: Am J Med 2013;126:162–168, physical PDF pages 1, 2, 3, 4, 5. New source-checked assessment.

Comparison and result assessed	D1	D2	D3	D4	D5	Overall	Source and appraisal provenance
NEFIGAN 2b — mean percentage change from baseline in eGFR at 12 months	Low	Low	Some concerns	Low	Some concerns	Some concerns	D1 low: permuted blocks generated by computer algorithm, stratified by baseline UPCR, with allocation performed by an independent group (Pharma Consulting Group AB) and masking maintained for patients, all trial staff, the sponsor and the DSMB. Physical pp3,4. D2 low: double-blind with identical placebo capsules; the full analysis set corresponds to a modified intention-to-treat set; treatment was tapered and stopped at month 9 with follow-up to month 12, so discontinuation did not truncate outcome ascertainment. Physical pp3,4. D3 some concerns: 51 patients were randomized to 16 mg/day, 49 received the intervention and 48 formed the full analysis set; the 12-month value is a model-based mean percentage change with a reported interval and is not accompanied by absolute eGFR levels in the material held, so the handling of missing visits cannot be reconstructed here. Physical pp4,5. D4 low: double-blind outcome measurement of an objective CKD-EPI eGFR; no adjudication required. Physical pp3,5. D5 some concerns: the primary outcome was change in UPCR; the 12-month eGFR change is a secondary outcome reported in the text and appendix of a registered trial, and no dated analysis plan was held. Physical pp4,5. Source: Lancet 2017;389:2117–27, physical PDF pages 1, 2, 3, 4, 5, 6. New source-checked assessment.
Abelmoschus manihot — creatinine doubling, the eligible component of the trial's composite	Low	Low	Some concerns	Low	Some concerns	Some concerns	D1 low: the sequence was generated by a centre-stratified random order (Proc Plan, SAS 9.4), held at the coordinating institute, and an interactive voice-response system assigned trial units, "in order to ensure the allocation concealment". D2 low: double-blind and double-dummy. D3: the component counts are reported without per-arm denominators for this outcome and the completeness of the component is not documented. D5: whether the components of the composite were analysed separately in the pre-specified plan is not stated, and no registration record was located in the material held Provenance: Previously completed assessment retained. Not established by available files.
Calcitriol (Joo 2026) — annual decline in eGFR over 12 months	Some concerns	Low	Some concerns	Low	Low	Some concerns	D1: allocation used "randomly permuted blocks and stratification", but concealment of the sequence is not described, so 1.2 is NI. D2 low: double-blind with matching placebo. D3: 115 of 138 randomized participants (83.3%) completed the study and the report lists last-observation-carried-forward among its methods, which is an imputation rather than an analysis varying the missing-data assumption, so completeness for the eGFR outcome is not established. D5 low: the report names the change in eGFR among the trial's primary outcomes and presents the analyses that follow from that plan Provenance: Previously completed assessment retained. Not established by available files.
ENVISION — change in eGFR at month 12	Low	Low	Some concerns	Low	Low	Some concerns	D1 low: randomization "with the use of an interactive voice-response or Web-response system". D2 low: double-blind, placebo-controlled. D3: 155 participants were randomized and the analyses are described for the modified intention-to-treat population, but the number with observed eGFR values at month 12, by arm, is not reported, so completeness is not established. D5 low: the eGFR change is a pre-specified secondary endpoint and the dose arms and placebo were analysed and reported together Provenance: Previously completed assessment retained. Not established by available files.
IGNAZ — change in eGFR at month 24	Low	Low	Some concerns	Low	Low	Some concerns	D1 low: randomization "1:1:1:1 in permuted blocks by interactive response technology". D2 low: double-blind, placebo-controlled. D3: 48 participants were randomized in part 1 and 35 were evaluable at month 24, so about a quarter contributed no value at that time point, and the handling of those participants is not described. D5 low: the trial is registered (NCT05065970) and the eGFR endpoint is pre-specified among the secondary endpoints Provenance: Previously completed assessment retained. Not established by available files.

Comparison and result assessed	D1	D2	D3	D4	D5	Overall	Source and appraisal provenance
HR19042 — difference in eGFR change at month 12	Low	Low	Some concerns	Some concerns	Some concerns	Some concerns	Assessed against the conference abstract only, which is the source of the row. D1 low for what the abstract states: randomization "via an interactive response technology system, stratified by baseline UPCR". D2 low: double-blind, placebo-controlled. D3 and D4: the abstract reports neither the numbers contributing to this comparison nor how measurement was carried out, so both domains rest on NI answers. D5: the abstract describes the 12-month eGFR change as one of two co-primary endpoints whereas the registration entry records the month-9 urine protein-to-creatinine ratio as the primary endpoint; the discrepancy is unresolved in the material held Provenance: Previously completed assessment retained. Not established by available files.
Mizoribine (Hirai 2017) — eGFR over 36 months	Some concerns	Some concerns	High	Low	Some concerns	High	D1: 1.2 is NI; patients were "randomly assigned (1:1) to receiving standard treatment plus MZR or standard treatment" with no method stated. D2: open-label; nine of 21 in the mizoribine group and 15 of 21 in the control group completed the study, so around half the intervention arm did not, and while some of this occurs after randomization it leaves the delivered intervention very different from the assigned one. D3 is high : data were available for fewer than half of the randomized participants in one arm, the difference between arms is large and in one direction, no analysis in the report varies the missing-data assumption, and progression is among the reasons a participant stops participating, so missingness could well have depended on the outcome's true value Provenance: Previously completed assessment retained. Not established by available files.
Tacrolimus (Yu 2017) — eGFR over the 5-year follow-up after a 16-week randomized course	Some concerns	Some concerns	Some concerns	Low	Some concerns	Some concerns	D1: the report states randomization stratified by renin–angiotensin-system inhibitor use but neither how the sequence was generated nor whether it was concealed, so 1.2 is NI. D2: the comparison is 16 weeks of randomized treatment followed by observation, so the interventions compared over the long term are no longer those assigned, and neither deviations nor the analysis population for the five-year values is described. D3: completeness of eGFR ascertainment over five years is not documented. D5: the long-term eGFR analysis is not described as pre-specified, and no registration record or protocol was located in the material held Provenance: Previously completed assessment retained. Not established by available files.
Masutani 2016 — creatinine doubling or ≥50% eGFR decline (no event in either group)	Some concerns	Some concerns	Some concerns	Low	Low	Some concerns	Obtained in full for this version. D1: the report states registration (UMIN000000593), a 1:1 randomization and allocation "using a minimiz" [minimisation algorithm], without describing how the allocation was concealed, so 1.2 is NI. D2: open-label; no deviations are enumerated, though 2.6 is Y on "Intention-to-treat (ITT) analyses were performed". D3: neither the number of participants contributing values to this endpoint nor withdrawals for it are reported, and no analysis varies the missing-data assumption. D5 low: the trial is registered in UMIN with a protocol cited in the report, and this endpoint is presented as a secondary one alongside the others rather than in place of them Provenance: Previously completed assessment retained. Not established by available files.
JANUS — renal function over 72 weeks	Some concerns	Low	Some concerns	Low	Some concerns	Some concerns	D1: the report states that participants were randomized 1:1:1 and refers for "details on randomization and blinding" to supplementary material not held for this review, so both the sequence and its concealment are NI. D2 low: double-blind and placebo-controlled. D3: 16 participants were randomized and the values used here are shown graphically only, with no completeness stated for either arm. D5: no registration record was located for this trial Provenance: Previously completed assessment retained. Not established by available files.

Comparison and result assessed	D1	D2	D3	D4	D5	Overall	Source and appraisal provenance
Rituximab trial (Lafayette 2017) — change in eGFR over 1 year	Some concerns	Some concerns	Some concerns	Low	Some concerns	Some concerns	D1: open-label, randomized 1:1, with no sequence or concealment described, and the report notes that "randomization resulted in a higher but not statistically significant baseline serum creatinine concentration in the rituximab group", which in a 34-participant trial is answered PN rather than as indicating a problem. D2: participants were aware of assignment and the report describes no deviations, but 2.6 is Y on the intention-to-treat analysis it reports. D3: the report gives dropouts as two versus three with 15 versus 14 completing, and separately states that all rituximab-treated subjects had post-treatment data for the intention-to-treat analysis; those numbers do not reconcile as stated, so completeness for this outcome is not established rather than assumed Provenance: Previously completed assessment retained. Not established by available files.
ALIGN SGLT2-inhibitor stratum — Change in eGFR at week 136	Low	Low	Some concerns	Low	Some concerns	Some concerns	D1: Interactive-response randomization within a defined SGLT2-inhibitor stratum, with patients, investigators and sponsor masked through week 136. Chance baseline eGFR imbalance alone is not evidence of failed randomization. D2: Double-blind treatment; eGFR collection continued after drug discontinuation. Post-intercurrent-event censoring belongs to the reported hypothetical estimand; stopping medication is not automatically a missing renal outcome. D3: 32 per arm randomized and 29/28 completed. The small stratum-specific estimate is subject to censoring after intercurrent events; robustness to missingness assumptions was not independently established in retrieved pages. D4: Central clinical chemistry, a common creatinine-based modified CKD-EPI method and masked measurement. D5: The stratum is explicitly exploratory; independently dated stratum-specific analysis plans were not verified. A large favourable effect is not evidence of selective reporting. Source: ALIGN SGLT2-inhibitor stratum, physical PDF pages 3, 4, 5. New source-checked assessment.
Lv 2009 — CKDstage5 during follow-up	Some concerns	Low	Some concerns	Some concerns	Some concerns	Some concerns	D1: Computer-generated randomization code reported; concealment method not reported. Nine pre-randomization exclusions do not create randomized attrition. Baseline proteinuria was higher in combination group, but that alone does not prove allocation bias. PDF pp2–4 (journal27–29). D2: Open-label treatment with a defined prednisone course and cilazapril titration in both groups; no important trial-context deviations are documented. No inappropriate exclusion of the lost participants is established from the narrative; the outcome-availability uncertainty is assessed in D3. PDF pp2–4. D3: Three patients were lost to follow-up. Completeness of CKDstage5 ascertainment for those three is not documented and missing status could matter with only two events. This assessment is specific to CKDstage5 and is not transferred mechanically from SCr50%. PDF p4. D4: SCr was measured at scheduled visits using the same study schedule, but CKDstage5 is reported without an explicit operational case definition or independent adjudication. Do not invent the eGFR threshold or require dialysis when neither is stated for these cases. PDF pp2,4. D5: The declared primary endpoint is confirmed SCr50% increase; declared secondaries are eGFR25% decrease and proteinuria reduction. CKDstage5 appears in Results without a retrievable dated prespecified analysis plan. An independent monitor prespecified a2-year interim look and the trial stopped for better outcomes, but this does not by itself prove selective reporting of CKDstage5. PDF pp3–4. Source: Lv 2009, physical PDF pages 2, 3, 4. New source-checked assessment.

Comparison and result assessed	D1	D2	D3	D4	D5	Overall	Source and appraisal provenance
Alexopoulos 2004 — ESRD over4 years	Some concerns	High	Some concerns	Some concerns	Some concerns	High	D1: Prospective randomized assignment is stated but sequence generation and allocation concealment are not described; baseline comparisons refer to the28 retained, not all34 assigned. PDF pp3–4 (journal454–455). D2: For an effect-of-assignment target, the reported analysis excludes6 of34 assigned patients because of noncompliance or loss (4/18 fish-oil vs2/16 control). This is not an assignment-preserving analysis and the six exclusions are large relative to the seven ESRD events. The precise split between nonadherence exclusions and missing outcomes is unavailable. This judgement is based on the demonstrably inappropriate restricted analysis, not merely the absence of blinding. PDF p4. D3: ESRD ascertainment for the six excluded patients is not reported; their missing status could materially change the1vs6 event comparison. The paper does not separate loss from noncompliance or give reasons linking missingness to renal outcome. Do not assert a proven direction of missingness bias; do not add another high judgement solely for the same exclusions already considered under D2. PDF p4. D4: ESRD is defined as dialysis initiation or SCr>10mg/dL (p3), whereas results describe dialysis/transplantation (p6); whether the threshold component contributed is unclear. Clinicians were apparently aware of assignment and no blinded renal-endpoint adjudication or standardized dialysis-start criteria are given. The blinded biopsy pathologist is not a blinded ESRD assessor. PDF pp3,6. D5: ESRD is identified as a secondary endpoint and is reported separately (pp4–5), but no dated analysis plan can establish the final analysis population/endpoint specification before results. No selective-reporting high-risk conclusion is established. PDF pp4–5. Source: Alexopoulos 2004, physical PDF pages 3, 4, 5, 6. New source-checked assessment.
Donadio 1999 / Mayo extension — ESRD through extended follow-up	Some concerns	Low	Low	Some concerns	Some concerns	Some concerns	D1: The extension describes the original randomized placebo-controlled double-blind trial but does not reproduce sequence generation or concealment; the primary1994 report was not available in the indexed sources reviewed for this assessment. Do not claim these methods were verified from the extension. PDF p1. D2: For the stated effect of original assignment to2-year fish oil followed by subsequent care, ESRD analyses retain original groups, including noncompliant patients. After planned blinding ended,44/45 at-risk fish-oil patients continued and17/30 at-risk placebo patients switched to fish oil; these were explicitly permitted post-trial care choices. Such crossover limits a sustained-treatment interpretation, but is not automatically a high-risk deviation for the original-assignment estimand. PDF pp1–2. D3: The extension explicitly ascertained vital and ESRD status of all noncompliant participants and retained original groups, unlike censoring used for the separate SCr50% endpoint. One fish-oil patient was lost after2 years; this isolated uncertainty is small relative to106 randomized and27ESRD events. Do not transfer the SCr-endpoint censoring judgement to ESRD. PDF p2. D4: Repeated dialysis or transplant is clearly defined. However, after year2 treating physicians knew original assignment and no independent blinded endpoint adjudication or standardized dialysis-initiation rule is described. There is plausible clinical-timing uncertainty without evidence supporting high risk. PDF pp1–2. D5: ESRD is explicitly the secondary endpoint; all106 original assignments and all27events are presented, as are2-year and post-trial components. A dated plan for this long-term observational extension is unavailable. The additional as-treated continuation comparisons must not replace the original-assignment8vs19 result. PDF pp1–2,4. Source: Donadio 1999 / Mayo extension, physical PDF pages 1, 2, 4. New source-checked assessment.

Comparison and result assessed	D1	D2	D3	D4	D5	Overall	Source and appraisal provenance
Tang 2010 — Composite doubling of SCr or ESRD over6 years	Some concerns	Some concerns	Low	Some concerns	Some concerns	Some concerns	D1: Open-label random assignment by drawing envelopes; sequence-generation method and envelope opacity/sealing/sequential numbering are not stated. Original cohort40 with20 per group. PDF pp5–6 (journal547–548). D2: All MMF patients completed the planned6-month course and all randomized participants were included in assignment analyses. However, during extended follow-up more MMF patients received combined ACEi/ARB treatment; authors note this potential cointervention contribution. In an open-label extension, the cause of that imbalance and whether it reflects trial-context deviations are uncertain, so high risk cannot be inferred. PDF pp2,5–6. D3: The paper explicitly reports all40 patients followed for at least6years and all alive. The composite adds one SCr-doubling event per group to ESRD2vs9. No outcome missingness is described. This is a new outcome-specific judgement, not copied from the earlier18-month report. PDF p2. D4: SCr assays were performed at a central laboratory and doubling is objective. Most composite events are ESRD, defined by dialysis/transplant; in an open-label trial no standardized dialysis-start threshold or blinded adjudication is described. Hence some uncertainty for the clinical component, without evidence of high measurement bias. PDF pp2,6. D5: The paper and Figure2 report the composite, but endpoint hierarchy is inconsistent (Figure2 calls it primary; Methods places doubling/ESRD/death among secondaries), and a dated extension analysis plan is unavailable. All patients were alive, so inclusion of death does not change3vs10. This supports uncertainty, not a proven selected-result claim. PDF pp2,6. Source: Tang 2010, physical PDF pages 2, 5, 6. New source-checked assessment.
Pozzi 2010 — Initiation of dialysis during follow-up	Some concerns	Low	Some concerns	Some concerns	Some concerns	Some concerns	D1: Two centralized computer-generated randomization lists stratified by timing of biopsy are stated; the actual method preventing recruiters from anticipating or accessing the next assignment is not described. Random sequence generation is adequate but concealment cannot be independently verified. The chance age imbalance alone is not evidence of bias. PDF pp2,6–7 (journal1784,1788–1789). D2: Open-label but identical steroid schedules, comparable RAS-blocker exposure, one rescue-immunosuppressant protocol violation in each group, and main analyses retaining all207 randomized participants. Treatment cessation from azathioprine toxicity is an effect of treatment, not by itself a trial-context deviation. Do not replace the reported6/101vs4/106 by the per-protocol subset. PDF pp2,4,7. D3: Table3 lists loss to follow-up2vs3 and transfer to another center1vs1. The paper does not establish that dialysis status was known after each loss/transfer. With only10 dialysis events, even these few unavailable statuses may matter. The18 incomplete treatment courses are not automatically18 missing dialysis outcomes. PDF pp4–5. D4: Dialysis initiation is an objective procedure but a clinician-timed decision. The open-label report gives neither a standardized dialysis-initiation threshold nor blinded endpoint adjudication. No evidence shows actual differential initiation; uncertainty supports Some concerns, not an automatic high-risk label. PDF pp2,6–7. D5: The formal primary endpoint is confirmed SCr50% increase; secondary endpoints are proteinuria and adverse events. Dialysis is separately reported as a clinical event. Although a protocol publication is cited, the dated dialysis-specific analysis specification was not available in the reviewed source. No selective-reporting high-risk claim is supported. PDF pp5–7. Source: Pozzi 2010, physical PDF pages 2, 4, 5, 6, 7. New source-checked assessment.

Comparison and result assessed	D1	D2	D3	D4	D5	Overall	Source and appraisal provenance
Pozzi 2013 — Initiation of dialysis during follow-up	Some concerns	Low	Some concerns	Some concerns	Some concerns	Some concerns	D1: The advanced-CKD stratum used a separate randomization list within the253-patient programme. The linked2010 report describes centralized computer-generated lists, but concealment mechanics for this separate list are not independently documented in the2013 paper. The20:26 imbalance is not sufficient proof of randomization bias. PDF pp3–5; linked2010Methodspp6–7. D2: Assigned-group intention-to-treat analysis is reported for all46 randomized participants; the per-protocol35 are an additional analysis. Both arms followed the same12-month steroid schedule and RAS-blocker use was similar. Toxicity-driven treatment withdrawal and one protocol violation do not demonstrate important trial-context deviations or an inappropriate main analysis. PDF pp4–6,8(Fig1). D3: Figure1 shows1lost to follow-up in the azathioprine group;6vs4 stopped treatment for side effects. It does not establish complete long-term dialysis ascertainment after withdrawal, despite the stated ITT analyses. Only35 complete12-month treatment; do not claim all11 discontinuers had missing renal outcomes or all46 had verified complete follow-up. PDFp8 image; pp5–6. D4: Dialysis is clearly reported separately in narrative andTableIII, but the unmasked comparison provides no uniform dialysis-start rule or blinded clinical adjudication. The timing of a clinical decision may be influenced by assignment; actual differential bias is not demonstrated. PDFpp4–5,10. D5: SCr50% change is declared primary; proteinuria and safety are stated secondary outcomes. Dialysis9vs10 is consistently reported as a clinical event, but a dated prespecified dialysis analysis was unavailable. Do not borrow the adjusted SCr-endpoint Cox estimate for dialysis. PDFpp3–5,10. Source: Pozzi 2013, physical PDF pages 3, 4, 5, 6, 8, 10. New source-checked assessment.
Liang 2022 — eGFR decline>25% at6months [adult eligibility unresolved; candidate only]	Some concerns	High	Some concerns	Low	Some concerns	High	SAS sequence; separate sealed envelopes prepared by statistical assistant and used by research assistant. Opacity/sequential opening unreported, so concealment remains uncertain. PDFp3. Effectiveness used per-protocol analysis after6post-assignment withdrawals for toxicity/nonremission. Excluding2/36versus4/38 can materially affect an outcome with only2events. No crossover reported. PDFpp3–5. Only retained68 completed6months; withdrawn participants' renal outcomes are unavailable. Reasons were treatment-related illness/nonremission; outcome-dependence remains uncertain. Avoid duplicating the D2high rationale. PDFp4. Objective creatinine assay and specified MDRD equation; common follow-up. Unmasked treatment alone does not imply biased laboratory threshold ascertainment. PDFp3. eGFR25% deterioration is explicitly secondary and separately tabulated; dated outcome-specific plan unavailable. Table4 says>25%,Methods25%; retain reporting distinction. PDFpp3,8. Provenance: new source-checked assessment. Not established by available files.
Hou 2017 — End-stage renal disease by 12 months	Low	Low	Some concerns	Low	Some concerns	Some concerns	D1: Independent organization generated block randomization; sequentially numbered concealed envelopes opened in sequence after consent. D2: Open-label treatment alone does not establish deviations affecting ESRD. Exclusions from the primary remission analysis cannot automatically be imposed on separately tabulated ESRD events. D3: 0 versus 2 ESRD events but no explicit ESRD denominator or complete outcome-ascertainment account; primary efficacy and safety populations differ. Do not invent an ESRD denominator. D4: ESRD reported as an objective clinical safety event; no differential measurement demonstrated in the retrieved report. D5: ESRD appears among adverse events rather than the primary efficacy endpoint. A dated ESRD-specific analysis plan was not independently verified. Source: Hou 2017, physical PDF pages 2, 3, 6. New source-checked assessment.

Comparison and result assessed	D1	D2	D3	D4	D5	Overall	Source and appraisal provenance
Cruzado 2011 — Dialysis at 12 months	Some concerns	Low	Low	Some concerns	Some concerns	Some concerns	D1: Permuted blocks of six and a 1:1 allocation are described, but sequence generation details and allocation concealment are not. The observed 14:9 imbalance alone does not demonstrate manipulation. PDF p2 (journal p3597). D2: Open-label trial, but no important trial-context deviations are described; enalapril/atorvastatin doses and additional antihypertensive use were comparable, and no participant withdrew treatment. Open label alone is not a high-risk judgement. PDF pp2–4. D3: The safety section explicitly reports all patients alive and dialysis-free at the end of the study and no treatment withdrawals; the event denominator is all23 randomized participants. PDF p4. D4: Dialysis occurrence is a clinical procedure, but its initiation involves clinician judgement. The trial was open label and provides neither a standardized dialysis-initiation threshold nor blinded endpoint adjudication. This permits some concern about timing/ascertainment, without evidence that assignment actually altered it. PDF pp2,4. D5: Therapeutic failure, including dialysis or transplantation, is defined as a secondary endpoint in the paper; a dated analysis plan is unavailable. Both randomized groups' zero events are reported. No selective-reporting high-risk finding is supported. PDF pp2,4. Source: Cruzado 2011, physical PDF pages 2, 3, 4. New source-checked assessment.
Cheng 2015 — Haemodialysis therapy during 24 months	Low	Low	Some concerns	Low	Some concerns	Some concerns	D1: Variable-block computer randomization; independent pharmacist alone held the allocation; opaque sealed numbered boxes concealed treatment before assignment. PDF p2 (journal p78). D2: Participants, healthcare providers and outcome assessors were blinded. Same valsartan run-in/background care and prohibited additional immunosuppressants/antiplatelet agents. No evidence of important trial-context deviations. The 32 run-in exclusions occurred before randomization, not post-randomization attrition. PDF pp2–3. D3: The paper says all patients were followed for 24 months, and Table 4 uses randomized denominators, but the flow chart has no post-randomization completion counts and does not explain ascertainment after treatment discontinuation or dialysis. Complete event ascertainment cannot be independently confirmed. This is an information gap, not established missingness bias. PDF pp3,6. D4: Haemodialysis is an objective treatment event; outcome staff were blinded. No differential ascertainment is described. However, duration or maintenance status is not specified: retain the authors' label haemodialysis, rather than relabeling all five events as adjudicated ESKD. PDF pp2,6. D5: The main paper identifies proteinuria reduction as primary and reports renal measures at several visits; no dated prespecified outcome-specific analysis plan was available for checking haemodialysis analyses. No evidence proves results-based selection. PDF pp3,6. Source: Cheng 2015, physical PDF pages 2, 3, 6. New source-checked assessment.

Comparison and result assessed	D1	D2	D3	D4	D5	Overall	Source and appraisal provenance
Cheng 2015 — Fixed-time eGFR at 24 months	Low	Low	Some concerns	Some concerns	Some concerns	Some concerns	D1: Computer variable-block sequence and numbered opaque sealed boxes held by an independent pharmacist. PDF p2. D2: Double-blinded participants/providers/assessors, shared background care and no important reported trial-context deviations. PDF pp2–3. D3: Visit-specific analysis numbers and handling of five participants who began haemodialysis are not given. It is unclear whether 24-month eGFR includes all randomized participants or how post-dialysis data were handled. PDF pp3,5–6. D4: Assessors were blinded, but two different eGFR equations are printed: Methods p3 specifies Chinese-adjusted MDRD (175, exponents −1.234/−0.179, female factor0.79); Table 1 p4 gives the conventional MDRD (186, −1.154/−0.203, female factor0.742). The equation actually used for Figure3 is unclear. This is measurement-method uncertainty, not bias inferred from lack of blinding. PDF pp3–4,6. D5: Multiple follow-up times and four treatment groups are reported; a dated plan for the specific fixed-time result and comparison is unavailable. No evidence establishes selective omission. PDF pp3,5–6. Source: Cheng 2015, physical PDF pages 2, 3, 4, 5, 6. New source-checked assessment.
Donadio 2001 — ESRD: chronic repetitive dialysis or renal transplant	Some concerns	Low	Some concerns	Some concerns	Some concerns	Some concerns	D1: Random assignment stratified by baseline creatinine, prior fish-oil use and prior corticosteroid use is stated. How the sequence was generated or concealed is not reported. Most baseline variables were similar; baseline platelet difference alone does not establish allocation bias. PDF pp2,4. D2: The trial was open label, but treatment adherence was assessed objectively with fatty-acid measurements and was described as excellent. Shared blood-pressure management was specified. Investigators attempted to follow renal status after noncompliance and used assigned groups for secondary survival analyses. Important trial-context deviations are not demonstrated; incomplete outcome follow-up is considered under D3. PDF pp2–4,7. D3: One high-dose participant who stopped before treatment was censored at time zero for ESRD. Thirteen of55 participants without ESRD left early (refusal8, moved2, lost1, adverse events2). Investigators attempted renal follow-up, but completeness after withdrawal and sensitivity to censoring are not established. With18 ESRD events this uncertainty could matter. PDF pp2–3. D4: ESRD is explicitly defined as chronic repetitive dialysis or transplantation. However, the study was open label and standardized dialysis-initiation criteria or blinded endpoint adjudication are not described; clinical timing could therefore be influenced by knowledge of assignment. Objective event definition alone does not resolve that gap, and there is no evidence supporting high risk. PDF pp2,7. D5: ESRD and the separate ESRD/adverse-event/noncompliance composite are defined and reported, including survival methods; a dated prespecified analysis plan is unavailable. Use ESRD8vs10, not the composite12vs16, and exclude the historical placebo comparison. PDF pp2–3,6–8. Source: Donadio 2001, physical PDF pages 2, 3, 4, 6, 7, 8. New source-checked assessment.

Comparison and result assessed	D1	D2	D3	D4	D5	Overall	Source and appraisal provenance
TESTING 2022 — Kidney failure requiring maintenance dialysis or kidney transplantation [B34_1]	Low	Low	Low	Low	Low	Low	D1: Central randomization used a minimization algorithm with site, proteinuria, eGFR and biopsy strata; participants, investigators, site staff and sponsors were masked. Baseline treatment groups were similar. Physical pp2–3,5. D2: The Methods specify analysis by randomized group regardless of adherence or protocol violations; the standalone kidney-failure Table 2 uses the full 257/246 randomized denominators and an adjusted Cox model. Recruitment pause, dose reduction and prophylaxis applied within corresponding randomized phases; scheduled follow-up continued after treatment was stopped. The assignment effect is assessed here, not a per-protocol effect. Physical pp2–4,6. D3: For the kidney-failure time-to-event result, the report states almost all participants were followed until trial end, missing data were negligible, and censoring occurred when last known event-free. The trial profile records three versus seven lost/withdrawn, and Figure 2B gives kidney-failure-specific risk sets. The judgement uses this outcome-ascertainment statement and its own event analysis, not vital-status completeness alone or the composite judgement. Signalling 3.1 PY; conditional questions NA. Physical pp3–4,6–7. D4: Kidney failure is defined as maintenance dialysis or kidney transplantation and adjudicated by a blinded independent committee. Visits after year one included three-monthly telephone or face-to-face endpoint ascertainment. The same event definition and masked adjudication apply to the separately reported component. Physical p3. D5: The article lists individual primary/secondary composite components as secondary outcomes, specifies adjusted Cox analysis for secondary time-to-event outcomes, and reports the standalone component in Table 2 and Figure 2B. The authors checked the dated protocol/SAP and confirmed that the standalone kidney-failure analysis was prespecified (5.1 Y;5.2 PN;5.3 PN). The article also states that kidney failure was one of 11 prespecified secondary outcomes (physical p1). Physical pp1,3,6–7. Source: JAMA 2022, physical PDF pages 2–7. New source-checked assessment.

Part B contains 41 outcome assessments from 32 candidate comparisons: one low risk, 35 some concerns and five high risk. Of these, 36 assessments from 29 adult-eligible comparisons comprise one low risk, 31 some concerns and four high risk. Five assessments from three comparisons of unresolved adult eligibility comprise four some concerns and one high risk. Comparisons and assessments are not counts of independent trials.

The additional results are not GRADE-rated. Their source values and outcome-specific appraisals are retained in the outcome inventory; treatment-level conclusions are tied to the 23 explicitly graded outcomes. Results of unresolved adult eligibility are not used in the primary adult synthesis.

12. Summary of all assessed results

Comparison	Outcomes assessed	D1	D2	D3	D4	D5	Overall
APPLAUSE-IgAN	24-month total slope; composite kidney-failure endpoint	Low	Low	**Some concerns** (both outcomes, separately assessed)	Low	Low	**Some concerns**

Comparison	Outcomes assessed	D1	D2	D3	D4	D5	Overall
ALIGN	Week-136 eGFR change; total slope; composite kidney outcome	Low	Low	**Some concerns** (eGFR outcomes and composite assessed separately)	Low	Low	**Some concerns**
FIND-CKD (IgA subgroup)	Chronic slope; total slope; composite	Low	Low	**Some concerns**	Low	Low	**Some concerns**
NefIgArd	2-year time-weighted average eGFR; composite	Low	Low	Low	Low	Low	**Low**
PROTECT	Chronic slope; total slope; composite kidney endpoint	Low	Low	**Some concerns** (slopes and composite assessed separately)	Low	Low	**Some concerns**
TESTING	Composite of 40% eGFR decline, kidney failure or kidney death; eGFR slope	Low	Low	Low (composite and slope assessed separately)	Low	Low	**Low**
MAIN	Composite: creatinine doubling, kidney failure or death from a kidney or cardiovascular cause; annual rate of eGFR decline	**Some concerns** (1.2 NI)	Low	Low	Low (see the recorded limit on kidney replacement therapy)	Low	**Some concerns**
STOP-IgAN (10-year follow-up)	Composite: all-cause death, end-stage kidney disease or >40% eGFR decline	**Some concerns** (1.2 NI)	Low	Low (3.2 PY: best- and worst-case analyses)	Low	Low	**Some concerns**
DAPA-CKD (IgA subgroup)	Composite: sustained ≥50% eGFR decline, end-stage kidney disease or renal/cardiovascular death; total and chronic eGFR slope	Low	Low	**Some concerns**	Low	Low	**Some concerns**
EMPA-KIDNEY (IgA subgroup)	Kidney disease progression; chronic eGFR slope	Low	Low	**Some concerns** (both outcomes, separately assessed)	Low	**Some concerns** (post hoc definition of the split)	**Some concerns**

Among the ten principal comparisons, none was judged at high risk of bias overall. Of the ten, two are at low risk and eight carry a concern: four of those concerns arise only from missing outcome data (APPLAUSE-IgAN, ALIGN, PROTECT, FIND-CKD), two only from an unresolved question about concealment of the allocation sequence (MAIN, STOP-IgAN), one only from missing outcome data within the subgroup (DAPA-CKD), and one from missing outcome data together with the post hoc definition of the IgA comparison (EMPA-KIDNEY). A missing-outcome-data concern means that the completeness of the outcome data is not documented in the material available and no

analysis in it varies the missing-data assumption; for two comparisons, NefIgArd and the STOP-IgAN 10-year follow-up, such an analysis does exist and is reported as showing the result unchanged, and those comparisons' missing-data-related rows are answered through question 3.2 rather than by assuming completeness.

The candidate appraisal inventory contains 64 results: five low risk, 54 some concerns and five high risk. The primary adult subset contains 59 assessed results: five low risk, 50 some concerns and four high risk; the five assessments from these entries are separate. Figure 5 continues to show only the 23 focused outcomes. GRADE covers 23 outcomes (three high, eight moderate, twelve low) and measures a different property.

13. Outcome-level assessment coverage

Coverage is recorded by result. The standalone TESTING kidney-failure component has its own source-checked five-domain assessment (B34_1); the composite and eGFR assessments were not transferred to it. Blisibimod remains unassessed because the held material is insufficient. Narrative-only older outcomes retain their stated extraction and coverage limits.

Result	**Coverage status**
TESTING — standalone kidney failure	Separately assessed in Part B, B34_1 (source-checked assessment). Not formally GRADE-rated in the retained 23-outcome set.
Blisibimod — abstract-only ESKD count	Five NI domains do not constitute a completed assessment. More primary material is required; excluded from appraisal totals.
Other narrative-only older outcomes	No separate numeric result was inferred from curves or non-significant statements. Coverage remains outcome-specific.

New and corrected outcome-specific appraisals

Nineteen new or corrected assessments are recorded at result level, each checked against its source. These include the ALIGN SGLT2-inhibitor stratum and Donadio 2001 randomized high- versus low-dose ESRD result. The CKD stage 5 result in Lv 2009 is assessed from its own counts, not from the post-treatment-adjusted creatinine model. Cheng’s allocation and masking descriptions, Tang’s non-placebo comparator and the outcome-specific handling of withdrawals are corrected from the original reports. Previously completed assessments are retained unchanged.

The nineteen are: the ALIGN SGLT2-inhibitor stratum (week-136 eGFR change); Lv 2009 CKD stage 5; Alexopoulos 2004 ESRD; Donadio 1999/Mayo extension ESRD; Tang 2010 composite; Pozzi 2010 dialysis; Pozzi 2013 dialysis; Liang 2022 >25% eGFR decline; Hou 2017 ESRD; Cruzado 2011 dialysis; Cheng 2015 haemodialysis and 24-month eGFR; Donadio 2001 ESRD; TESTING 2022 kidney failure (B34_1); Praga 2003 ≥50% serum-creatinine rise (B35_1); Liu 2014 ≥25% eGFR fall (B36_1); Li 2013 60-month eGFR and rate of decline (B37_1, B37_2); and NEFIGAN 2b 12-month eGFR change (B38_1). Each retains its five domain judgements, the source observation supporting each, and its overall judgement in the table above. The nineteen new or corrected outcome-specific assessments were reviewed by the authors against the source reports, and the domain and overall judgements were confirmed.

Source checks retain Cheng’s haemodialysis counts of 4, 1, 0 and 0 across four 42-person arms and its graphical eGFR trajectory as separate results. Donadio 2001 contributes only its randomized dose comparison (ESRD 8/36 versus 10/37); the historical-placebo contrast is excluded. Liang’s >25% eGFR decline is 1/34 versus 1/34 among 68 analysed

after 74 assigned, but adult eligibility remains unresolved. Eligible alternative outcomes enter the original-standard adult restriction only when both outcome and population conditions are met.

14. Limits of this assessment

The original principal-comparison assessments are retained unchanged. New or corrected assessments record their source pages and domain reasoning at result level. Missing information is recorded explicitly rather than resolved by assumptions about blinding, objective measurement or treatment-effect direction.

Two assessment limits are recorded about allocation concealment, and they differ in what is missing. For MAIN the material states neither how the sequence was generated nor how it was held. For STOP-IgAN the original trial report [27] was consulted and describes generation — codes assigned in a 1:1 ratio by covariate adaptive randomization with respect to eGFR and proteinuria — while neither that report nor the 10-year report states how the codes were held; the unanswered question is therefore 1.2 alone. In both, the answer was recorded as NI on the question that the material does not answer, and the consequence was expressed as a domain judgement about the allocation process, not as a finding that the allocation was improper. Whether knowledge of an open-label assignment could have influenced how a component of a composite was ascertained is a different question, asked under domain 4, and it is not used to support either of these two judgements.

An unassessed result has no overall risk category; it is not classified as low risk. The new source and eligibility checks are reflected in Additional files 13 and 3. Assessment counts, eligible-entry counts and displayed result rows use distinct units.

Older reports often provide limited allocation or analysis-plan details. Existing appraisals preserve those reporting limitations. New assessments cite the outcome-specific evidence actually examined. Additional RoB 2 coverage does not extend GRADE beyond the 23 focused outcomes.