Additional file 12. Pooling decisions for each candidate combination of eligible results

How the decision was made

A set of eligible results was carried into a pooled estimate where they answer **one treatment question** — that is, where the intervention and comparator, the outcome being measured and the estimand are comparable enough for a single underlying effect to be a quantity worth estimating. Differences in dose, in duration of follow-up or in the control condition do not by themselves exclude a synthesis. What they require is that they be stated, and that their bearing on the interpretation of a common effect be given. Where the differences are large enough that no single treatment question remains, so that a pooled number would average several different effects rather than one, the results are reported side by side instead.

Three positions are therefore available for each candidate combination, and each row below states which was taken: an exploratory synthesis, side-by-side reporting, or insufficient material to form the comparison at all. **The number of studies available, the width of the resulting interval and the absence of a published weight structure are not grounds for refusing a synthesis.** A pooled interval may be wide and still be the right summary, and its width is then reported rather than avoided. A random-effects model does not by itself make unlike trials comparable; it describes between-study variance, it does not create a shared question.

One exploratory synthesis was carried out, of the two older mycophenolate mofetil trials, and is reported in Additional file 4. It is post hoc, it is not part of the clinical synthesis, and no treatment conclusion in this review rests on it.

1. Kidney-progression event outcomes

Candidate combination	Treatment and comparator	Endpoint composition	Observation window	Effect estimand	Decision, and what is reported instead
All ten principal comparisons as a group — APPLAUSE-IgAN, ALIGN, FIND-CKD (IgA), NefIgArd, PROTECT, TESTING, MAIN, STOP-IgAN (10-year), DAPA-CKD (IgA) and EMPA-KIDNEY (IgA)	Different drug classes against placebo, an active comparator or optimized supportive care; the review does not rank classes, so cross-class pooling was not attempted. MAIN and STOP-IgAN both compare added immunosuppression with supportive care, and DAPA-CKD and EMPA-KIDNEY both compare SGLT2 inhibition with placebo, but neither pair is pooled with the rest for the reasons in the endpoint and window columns	The composites are not the same endpoint: the eGFR-decline component is ≥30% (APPLAUSE, NefIgArd, ALIGN) or ≥40% (PROTECT, TESTING, FIND-CKD, EMPA-KIDNEY, STOP-IgAN), or ≥50% sustained with 28-day confirmation (DAPA-CKD); one composite carries no eGFR-decline component at all, MAIN, which counts creatinine doubling instead; some composites include cardiovascular death, some kidney death and one all-cause death; and the confirmation and persistence requirements differ between trials	24 months; 136 weeks; to month 32 (FIND-CKD); 2 years; 110 weeks; a median of 3.5 years (IQR 2.4–6.2) across two phases (TESTING); 3 years (MAIN); a median of 7.4 years (STOP-IgAN); a median of 2.1 and 2.0 years for the two SGLT2 subgroups — the windows do not coincide, so the events counted are not the same events	Hazard ratios for nine of the ten and a relative risk for PROTECT; four of the ten are subgroup estimates and two are analyses of a subset of randomized participants rather than of all of them, and one comparison (EMPA-KIDNEY's IgA group) was defined after the subgroup had been split	Reported side by side, each labelled with its own measure, window and composition (Table 2, Figures 2–3). No single treatment question spans ten drug classes, and even the two pairs that share a mechanism do not describe the same comparison
TESTING vs NefIgArd — the two placebo-controlled glucocorticoid or gut-targeted budesonide comparisons	Both versus placebo, but targeted-release budesonide is not systemic glucocorticoid therapy; its systemic exposure and safety profile differ	TESTING's composite is 40% eGFR decline, kidney failure or kidney death; NefIgArd's is confirmed ≥30% eGFR decline or kidney failure. A 10-percentage-point difference in the threshold changes which patients are counted, and the confirmation rules differ	A median of 3.5 years (IQR 2.4–6.2) across two phases (TESTING) versus 2 years (NefIgArd)	Both hazard ratios, but computed over different event definitions and windows	Study-level estimates, with the compositional difference stated in the row

Candidate combination	Treatment and comparator	Endpoint composition	Observation window	Effect estimand	Decision, and what is reported instead
STOP-IgAN (10-year follow-up) vs MAIN — added immunosuppression versus supportive care in both cases	Both compare added immunosuppression with optimized supportive care, so these two were the closest candidate pair in the immunosuppression group	Not the same composite: STOP-IgAN counts all-cause death, end-stage kidney disease and a >40% eGFR decline; MAIN counts creatinine doubling, kidney failure and death from a kidney or cardiovascular cause. Two components differ in kind: all-cause death is not kidney or cardiovascular death, and a >40% eGFR decline is not creatinine doubling	Median 7.4 years of follow-up versus a 3-year trial phase; the windows barely overlap	Both hazard ratios. MAIN's is adjusted for baseline covariates and STOP-IgAN's is not, but an adjusted estimate from a randomized trial targets the same treatment contrast when the adjustment set contains only baseline covariates, so that difference is not what separates them; the composite and the window are	Both estimates reported separately, with the compositional and window differences stated
MAIN vs Frisch 2005 and Tang 2005 — mycophenolate across three trials	MAIN added mycophenolate to optimized supportive care; the older trials compared it with placebo (Frisch, double-blind) or with continuation of ongoing treatment (Tang, not placebo-controlled). The control conditions are not one comparison	MAIN's composite is creatinine doubling, kidney failure or death from a kidney or cardiovascular cause; the older trials report a ≥50% rise in serum creatinine, with different confirmation requirements. Creatinine doubling and a 50% rise are different thresholds, and MAIN's composite additionally counts events the older trials' outcome does not	3 years (MAIN) versus an achieved mean total follow-up of 59 and 75 weeks (Frisch arms) and 72 weeks (Tang)	Adjusted hazard ratio (MAIN) versus odds ratios calculated from the reported counts (the older trials). As in the row above, adjusted and unadjusted estimates are not incommensurable in general; here the outcome itself differs	MAIN is discussed on its own and is not combined with the older trials. The two older trials are treated as a separate candidate pair and assessed on their own in section 2.1
Older corticosteroid and immunosuppressive trials using a ≥50% creatinine rise or renal survival (Pozzi 1999 [adult eligibility unresolved] and 2010, Manno 2009 [adult eligibility unresolved], Ballardie 2002, Katafuchi 2003, Stangou 2011, Min 2017 and comparable entries)	Corticosteroid regimens with or without cytotoxic agents against control or against each other; the regimens differ between trials, so even the treatments compared are not one intervention	Renal survival is defined by different thresholds (50%, 50% or 100%) and different confirmation rules, and several trials report events only within survival curves, from which no count was extracted	4 to 10 years, with different censoring rules	Most report log-rank P values or curve positions only; several give no between-group estimate at all	Descriptive reporting, with "data not extractable" kept distinct from "outcome not eligible"; no estimate was digitized or inferred. Insufficient material, not a comparability judgement

2. Two candidate combinations examined on their own

These two pairs were the only ones in which a single treatment question could plausibly be stated, so they were examined item by item rather than by category.

2.1 Mycophenolate mofetil: Frisch 2005 and Tang 2005

Both trials randomized mycophenolate mofetil against no mycophenolate, and both report a ≥50% rise in serum creatinine, so a common question can be stated: whether mycophenolate mofetil reduces the risk of that event. Four things were checked against the trial reports.

What was checked	Frisch 2005 [45]	Tang 2005 [46]	Consequence for a common effect
Confirmation requirement	The primary outcome was a **sustained** 50% increase in serum creatinine from baseline, observed for **at least two consecutive months** of follow-up	No persistence requirement is stated; the report gives the number of patients who "experienced a 50% or more increase in serum creatinine levels" during follow-up	The definitions differ in strictness. The direction of the resulting bias cannot be determined from the reports
Observation period	One year of treatment with follow-up to two years; the achieved mean total follow-up was 59 weeks in the MMF arm and 75 weeks in the placebo arm	24 weeks of treatment, 72 weeks of follow-up	Windows of the same order. Neither is long enough for the event counts to be treated as complete
Risk population	Moderately advanced disease: creatinine clearance 20–80 mL/min, mean baseline serum creatinine 2.4 mg/dL, at least 1 g of proteinuria per day and two further risk factors; all patients received angiotensin inhibition	Persistent proteinuria above 1 g per 24 hours despite renin–angiotensin blockade; patients with serum creatinine above 300 µmol/L (3.4 mg/dL) were excluded	Different entry criteria, but the observed control-group event rate was similar (2/15 and 3/20), so the difference in entry criteria did not translate into a large difference in observed risk over these windows
Treatment and comparator	MMF titrated up to 1000 mg twice daily, double-blind against identical placebo	MMF 1.5–2 g/day for 24 weeks added to ongoing treatment, against continuation of that treatment; not placebo-controlled, and blinding is not described	Both compare mycophenolate with no mycophenolate, but the control conditions differ, and one trial was blinded while the other was not

Decision: exploratory synthesis carried out. The shared question holds, so the combination was taken into a synthesis, reported in Additional file 4 with fixed-effect, DerSimonian–Laird and Hartung–Knapp–Sidik–Jonkman intervals. The differences above are stated there and here, and they are why the result is exploratory rather than clinical: the two trial estimates point in opposite directions (odds ratio 2.71, 0.44 to 16.68 in Frisch and 0.30, 0.03 to 3.15 in Tang; $I^2 = 52.6\%$), so the pooled interval is compatible with appreciable benefit and with appreciable harm. Carrying the pair into a synthesis is a statement that the question is shared, not a claim that the trials are exchangeable, and the pooled interval is not used to support any treatment conclusion.

2.2 Omega-3 fatty acids: the two fish-oil trials

Both trials randomized omega-3 fatty acids, and both report a ≥50% rise in serum creatinine with usable event counts in both arms — 17/55 versus 29/51 and 1/14 versus 6/14 — so the same checks were applied.

What was checked	Mayo/Dobos trial [47,48]	Alexopoulos 2004 [49]	Consequence for a common effect
Dose	12 g/day of fish oil, providing 1.9 g of EPA and 1.4 g of DHA	A "very low dose" of purified omega-3 fatty acids: 0.85 g of EPA and 0.57 g of DHA	The preparations differ and the daily EPA dose differs by more than a factor of two. The second trial was designed specifically to test whether a much lower dose would work, so the dose is part of the question rather than a nuisance variable around it
Comparator	Olive oil placebo, double-blind, multicentre	Symptomatic care; no placebo, and blinding is not described	One comparison is blinded against an identical placebo, the other is open against no added treatment. The two do not estimate the same contrast
Event definition	Primary endpoint an increase of ≥50% in serum creatinine, ascertained over follow-up	A ≥50% rise in serum creatinine at 4 years, reported separately from a ≥50% decline in GFR.	The creatinine threshold is shared, but confirmation requirements and assessment windows differ; the Alexopoulos result is not a combined creatinine/GFR endpoint.
Observation period	A mean of 6.4 years in the long-term report, which continues past the 2-year randomized trial into open-label observation; that report documents that some originally placebo-assigned patients switched to fish oil	4 years of treatment	Both windows are long. In the first trial part of the long-term follow-up accrued outside the randomized allocation, which limits how far the long-term result can be attributed to the randomized comparison; it does not by itself take away the value of reporting the comparison by originally allocated group, which remains readable with the switching stated alongside it

Decision: reported side by side without pooling. The trials share the ≥50% serum-creatinine threshold but differ in preparation, dose, placebo versus symptomatic-care comparator, confirmation requirements and follow-up windows. Individual-study odds ratios preserve the treatment contrast and observation period of each report; their concordant direction is described without a pooled treatment estimate.

3. Longitudinal eGFR outcomes

Candidate combination	Treatment and comparator	Endpoint composition	Observation window	Effect estimand	Decision, and what is reported instead
Total slope vs chronic slope within the same trial (APPLAUSE-IgAN, ALIGN, FIND-CKD, PROTECT) and across trials	Same treatment comparison, so pooling is not blocked by the treatments	A total slope includes the initial treatment period, in which a haemodynamic effect is expected; a chronic slope excludes it. The two answer different questions about the same trial	Both are annualized, but over different spans (24 months; to week 136; about 1 year; weeks 6–110 versus day 1 to week 110)	Both are annualized differences in mL/min/1.73 m² per year, but they are different functions of the same data and were reported side by side, never averaged	Both slopes are shown, labelled, in Figure 4 panels a and b. Not one question: a slope that includes the initial period and one that excludes it are different estimands
Slope differences vs fixed-time eGFR changes vs time-weighted averages (ALIGN week-136 change; NefIgArd 2-year time-weighted average; several older trials' fixed-time values)	Not comparable at the treatment level either, but the primary issue is the estimand	A difference in eGFR **at** a time point, a **rate over** an interval and a **time-weighted average over** a period are three different quantities; no anchor exists for converting one into another	Fixed time versus rate versus average	Differences in mL/min/1.73 m² in all three cases, which is why the shared unit cannot be read as a shared quantity	Figure 4 panel c separates time-weighted averages from fixed-time changes by label, and no cross-panel comparison is made
DAPA-CKD IgA slope vs EMPA-KIDNEY IgA chronic slope	Both are SGLT2 inhibition versus placebo, the closest candidate pair in this family	The DAPA-CKD IgA subgroup reports a total slope difference of 1.2 mL/min/1.73 m² per year (−0.12 to 2.51) and a chronic slope difference of 2.4 (1.08 to 3.71); the EMPA-KIDNEY IgA value is a relative reduction of 43% (27 to 59), for which the material gives no absolute group values	Not applicable — the combination was never formed	An absolute difference and a relative reduction are different quantities, and the chronic windows differ: the DAPA-CKD chronic slope starts at week 2	Reported side by side in Table 3, each with its own measure, window and interval. No pooled slope difference was calculated, and the EMPA-KIDNEY value cannot be converted to an absolute rate from the material held
The remaining longitudinal estimates (sibeprenlimab, felzartamab, atacicept, calcitriol, HR19042, mizoribine, tacrolimus, rituximab and others)	Different drugs, several with dose arms sharing one placebo group, which makes the arm-level values non-independent	Change-from-baseline values, least-squares mean changes and graphical-only curves	6 months to 5 years	Some rows have no between-group estimate at all; where an estimate exists it is a change, not a slope	Study-level values in Table 3 and Figure 4, each labelled with its design limit

4. What these decisions rest on, and what they do not

The decisions above rest on whether a treatment question is shared, not on a rule that every condition must be satisfied identically before any two results may be combined. Trials that differ in dose, in follow-up or in control condition can still be synthesized, and the differences then belong to the interpretation of the common effect.

One point of method deserves a precise statement. In the mycophenolate synthesis the pooled interval is highly sensitive to the method of inference: with two studies, where the DerSimonian–Laird estimate of τ^2 is greater than zero, the Hartung–Knapp–Sidik–Jonkman standard error equals the DerSimonian–Laird standard error and only the critical value differs ($t = 12.706$ versus 1.96). **With very few studies an interval is sensitive to the method of inference**, which is why all three intervals are reported side by side in Additional file 4. This is not a statement that the interval reflects the method rather than the data: the data determine the estimate and the between-study variance, and the method determines how that variance is propagated into an interval. Nor is it a general statement that every two-study synthesis is inadmissible.

Two clarifications on scope. First, the absence of a pooled estimate for the combinations reported side by side is a property of how these trials report their endpoints and define their comparisons, not a judgement that the treatments can never be compared; where a future analysis has individual participant data, a common endpoint and a common window, the obstacles listed here do not apply. Second, a small number of studies, a wide interval and the absence of a published weight structure are recorded above as things that do not by themselves exclude a synthesis, and none of the decisions in this file was taken on those grounds.

Population eligibility update. Pozzi 1999 and Manno 2009 are candidate reports of unresolved adult eligibility; combinations naming them are outside the primary adult synthesis. The two-trial mycophenolate exploratory calculation is unaffected by this age query but is excluded under the original outcome standard.