Medications for adults with type 2 diabetes: a living systematic review and network meta-analysis

Kailei Nong,[1,2,#] Britta Tendal Jeppesen,[3,4,#] Qingyang Shi,[5,#] Thomas Agoritsas,[6,7,8] Gordon H Guyatt,[6,8] Heath White,[9] Yiyuan Gao,[1,2] Arnav Agarwal,[6,8,10] Helen Macdonald,[11] Xinyu Zou,[1,2] Tanya Millard,[9,12] Oliver Schnell,[13] Nikolaus Marx,[14] Frank C Brosius III,[15] Steve McDonald,[9] Matthew Quigley,[9] Xin Tian,[16] Qinlin Fan,[1,2] Barbara White,[12] Yunhe Mao,[17] Xiaohui Pan,[1] Changhai Liu,[18,19] Chunjuan Zhai,[20] Chi Yuan,[21] Qiang Li,[22] Jing An,[1,2] Yu Gan,[23] Yanyan Wang,[16] Yinghui Jin,[24] Feng Sun,[25] Zhiming Zhu,[22] Lars Rydén,[26] Eberhard Standl,[13] Tari Turner,[9] Per Olav Vandvik,[8,27,†] Sheyu Li[1,2,8,†]

For numbered affiliations see end of the article

Correspondence to: S Li, Department of Endocrinology and Metabolism, West China Hospital, Sichuan University, Chengdu, 610041, China lisheyu@gmail.com Twitter: @LisheyuSheyu; (ORCID 0000-0003-0060-0287)

Additional material is published online only. To view please visit the journal online.

Cite this as: *BMJ* 2025;390:e083039 http://dx.doi.org/10.1136/bmj-2024-083039

Accepted: 23 January 2025

ABSTRACT

OBJECTIVE

To provide up-to-date evidence on key benefits, harms, and uncertainties regarding medications for adults with type 2 diabetes.

DESIGN

Living systematic review and network meta-analysis (NMA), using frequentist random effects and GRADE (grading of recommendations, assessment, development and evaluation) approaches. Updates are planned at least two times a year.

DATA SOURCES

Medline and Embase, searched up to 31 July 2024 for the current iteration.

STUDY SELECTION

Randomised controlled trials of at least 24 weeks comparing one or more medications with standard treatment, placebo, or each other.

RESULTS

The systematic review and NMA includes 493 168 participants from 869 trials (adding 53 trials since October 2022) reporting data for 13 drug classes (63 drugs) and 26 outcomes of interest. Regarding benefits, moderate to high certainty evidence confirms the well established cardiovascular and kidney benefits of sodium-glucose cotransporter-2 (SGLT-2) inhibitors, glucagon-like peptide-1 receptor agonists (GLP-1RAs), and finerenone (the last for patients with established chronic kidney disease). The most effective drugs in reducing body weight were tirzepatide (mean difference (MD) −8.63 kg (95% confidence interval −9.34 to −7.93); moderate certainty) and orforglipron (MD −7.87 kg (−10.24 to −5.50); low certainty), followed by eight other GLP-1RAs (high to moderate certainty). Absolute benefits of medications vary substantially depending on the baseline risk of cardiovascular and kidney outcomes; risk-stratified absolute effects of medications are summarised using an interactive multiple comparisons tool (https://matchit.magicevidence.org/250709dist-diabetes/#!/). Regarding medication-specific harms, SGLT-2 inhibitors increase genital infections (odds ratio (OR) 3.29 (95% CI 2.88 to 3.77); high certainty) and ketoacidosis due to diabetes (OR 2.08 (1.45 to 2.99); high certainty), and probably increase amputations (OR 1.27 (1.01 to 1.61); moderate certainty); tirzepatide and GLP-1RAs probably increase severe gastrointestinal events (most increased risk with tirzepatide (OR 4.21 (1.87 to 9.49); moderate certainty)); finerenone increases severe hyperkalaemia (OR 5.92 (3.02 to 11.62); high certainty); and thiazolidinediones increase major osteoporotic fractures and probably increase hospitalisation for heart failure. Sulfonylureas, insulin, and dipeptidyl peptidase-4 inhibitors probably increase the risk of severe hypoglycaemia. There is low to very low certainty evidence for effects on other diabetes-related complications, including neuropathy and visual impairment. Despite interest in the issue, there is uncertainty about whether GLP-1RAs may reduce dementia (OR 0.92 (0.83 to 1.02); low certainty).

CONCLUSIONS

This living systematic review provides a comprehensive summary of the cardiovascular, kidney, and weight loss benefits, as well as medication-specific harms of medications for adults with type 2 diabetes, including effects of SGLT-2 inhibitors, GLP-1RAs, finerenone and tirzepatide.

WHAT IS ALREADY KNOWN ON THIS TOPIC

Sodium-glucose cotransporter-2 (SGLT-2) inhibitors, glucagon-like peptide-1 receptor agonists (GLP-1RAs), finerenone, and tirzepatide provide risk-dependent and differential benefits in reducing death, cardiovascular diseases, chronic kidney diseases, and body weight while demonstrating drug-specific harms

Rapidly accumulating new evidence from ongoing trials of existing and new medications, combined with changes in patents and pricing of medications, warrant living systematic reviews

WHAT THIS STUDY ADDS

This network meta-analysis provides the best current evidence on 26 outcomes identified as important to patients regarding 13 drug classes (63 medications) for adults with type 2 diabetes

Our findings demonstrate differential benefits and drug-specific harms of SGLT-2 inhibitors, GLP-1RAs, finerenone, and tirzepatide

Major uncertainties remain for complications such as neuropathy and visual impairment

This living systematic review informs two living guidelines with risk-stratified recommendations. Through a global alliance for living evidence (ALIVE), it aims to support policy and practice worldwide with timely health technology assessments and trustworthy guidelines

SYSTEMATIC REVIEW REGISTRATION

PROSPERO number: CRD42022325948. A more detailed protocol is available at https://data.aliveevidence.org/records/q02rv-km486.

READERS' NOTES

This article is the first version of a living systematic review. It is linked to a living *BMJ* Rapid Recommendation and other living clinical practice guidelines, presenting risk stratified recommendations for patients with type 2 diabetes at lower, moderate, and higher risk of cardiovascular and kidney complications. The latest evidence will be made available via the *BMJ* Rapid Recommendation and via an interactive GRADE evidence summary (MATCH-IT: https://matchit.magicevidence.org/250709dist-diabetes/#!/). Major updates will be published in *The BMJ*.

Introduction

Type 2 diabetes affects over 400 million individuals worldwide and is associated with significant cardiovascular and kidney-related morbidity and mortality.[1] The rapid emergence of novel effective medications, including sodium-glucose cotransporter-2 (SGLT-2) inhibitors and glucagon-like peptide-1 receptor agonists (GLP-1RAs), is shifting diabetes management from a glucocentric approach to one more focused on reducing these complications.[3] Evidence on the weight-lowering effects of GLP-1RAs and other emerging medications such as tirzepatide are also attracting global attention.[4] However, residual questions persist regarding long term effects and rare but serious harms.[5 6] Variability in access to medications and related costs also impact decision-making internationally.[2]

To make fully informed decisions, policymakers, payers, clinicians, and patients require reliable evidence summaries, health technology assessments (HTA), and trustworthy clinical practice guidelines that reflect emerging evidence of benefits and harms, capturing relevant alternatives across available drug classes in type 2 diabetes.[7 8] This underscores the need for systematic reviews with network meta-analyses (NMA) to evaluate the comparative effectiveness of multiple treatment options, including indirect comparisons of medications not tested head-to-head.[9] Benefits of medications on cardiovascular and kidney outcomes vary substantially in people with type 2 diabetes based on their individual risk of such complications, warranting risk-stratified decision making.[10]

Whereas traditional systematic reviews provide a static cross-sectional summary of available evidence, living systematic reviews involve dynamic updates to include the latest available evidence.[11] For type 2 diabetes, there is an "infodemic" of new publications on a plethora of medicines. Living systematic reviews and NMAs providing timely updated evidence are thus necessary. This living NMA informs living guidelines and leverages a timely and collaborative approach to evidence synthesis in an enhanced evidence ecosystem that was trialled in the covid-19 pandemic.[12] Through the Alliance for Living Evidence (ALIVE) consortium (https://www.aliveevidence.org/), this living review aims to represent the key up-to-date repository of trial evidence to inform clinical practice guidelines and HTA. Box 1 outlines related publications, including

Box 1: Linked resources

Read the linked *BMJ* Rapid Recommendations cluster:

- Agarwal A, Mustafa R, Manja V, et al. Cardiovascular, kidney-related and weight loss effects of therapeutics for type 2 diabetes: a living clinical practice guideline. *BMJ* 2025;390:e082071[13]
- Rayner DG, Shah D, Dai S-C, et al. Prognostic models for cardiovascular and kidney outcomes in people with type 2 diabetes: living systematic review and meta-analysis of observational studies. *BMJ Med* 2025;4:e001369[14]
  - A systematic review of prognostic models for estimating the likelihood of cardiovascular and kidney complications for adults with type 2 diabetes.
- Rodríguez-Gutiérrez R et al. Values, preferences, and treatment burden for initiation of GLP-1 receptor agonists, SGLT-2 inhibitors, tirzepatide and finerenone in adult patients with type 2 diabetes: a systematic review. [Pending submission to *BMJ*][15]
  - A systematic review of values and preferences of patients with type 2 diabetes

Read the linked material from MAGIC Evidence Ecosystem Foundation:

- MAGICapp (https://app.magicapp.org/#/guideline/nBkO1E)
  - Expanded version of the *BMJ* guideline with details on methods, processes, and results with multilayered recommendations, evidence summaries, and decision aids for use on all devices
- Decision support tool for visualization of the comparative absolute benefits and harms across medications to support shared decision-making (https://matchit.magicevidence.org/250709dist-diabetes/#!/). The MATCH-IT tool will provide the most up-to-date summary of evidence from this living systematic review.
- Australian Evidence-Based Clinical Practice Guideline for Diabetes[16] (https://app.magicapp.org/#/guideline/7844)
- This living systematic review and NMA supersedes two prior network meta-analyses linked with the *BMJ* Rapid Recommendation series, one addressing SGLT-2 inhibitors and GLP-1RAs[17] and one addressing all drug treatments for diabetes and incorporating evidence up to October 2022.[18] The associated living guideline supersedes the previous *BMJ* Rapid Recommendation.[19]

a living clinical practice guideline (*BMJ* Rapid Recommendation) that triggered this living NMA and is published as a linked paper.

Methods

This living systematic review and NMA is developed in collaboration with three international multidisciplinary teams: (i) a living guideline (*BMJ* Rapid Recommendation) from the MAGIC Evidence Ecosystem Foundation, incorporating input from patient partners living with type 2 diabetes, general practitioners, internists, endocrinologists, nephrologists, cardiologists, and guideline methodologists; (ii) the Australian Evidence-Based Clinical Guidelines for Diabetes developed by the Living Evidence for Diabetes Consortium, a collaboration of multiple diabetes societies in Australia supported by the Australian Living Evidence Collaboration; and (iii) the Cardiovascular Outcome Trial (CVOT) Taskforce with representation from multiple national and international professional societies across endocrinology, cardiology, and nephrology. These teams have assisted in formulating clinical questions, in identifying and rating the relative importance of outcomes to patients, and in defining risk groups to facilitate estimation of absolute effects from relative effect estimates.

This review adheres to PRISMA (preferred reporting items for systematic reviews and meta-analysis) 2020 and PRISMA NMA statement reporting standards.[20 21] We have updated a pre-existing registration to reflect methods of this living systematic review and NMA (PROSPERO number: CRD42022325948). A more detailed protocol is also available in the ALIVE repository.[22]

Eligibility criteria

Eligible parallel arm randomised controlled trials (RCTs) compare medications for type 2 diabetes with each other, placebo, or standard treatment (typically representing the treatment regimens that patients received in practice before consideration of adding a new candidate medication). This NMA considers both glucose-lowering and emerging disease-modifying medications including SGLT-2 inhibitors, GLP-1RAs, dual glucose-dependent insulinotropic polypeptide and glucagon-like peptide-1 receptor agonists (tirzepatide), non-steroidal mineralocorticoid receptor antagonists (finerenone), dipeptidyl peptidase-4 (DPP-4) inhibitors, thiazolidinediones, sulfonylureas, metformin, alpha-glucosidase inhibitors, meglitinides, and insulins. Appendix 1.3 presents the detailed names of medications for each class. We limit eligibility to trials with a follow-up duration of 24 weeks or longer and exclude trials with a systematic difference of two medication classes or more between the intervention and the control. We exclude non-English language studies.

Search strategy and information sources

We conduct iterative searches in Ovid Medline and Embase, using a previously developed comprehensive search strategy and employ monthly auto alerts (appendix 1.1). The current iteration includes searches through 31 July 2024.

Study selection

Pairs of reviewers (YYG, BW, XT, MQ, TM, and KN) screen identified hits using standardised forms to identify eligible studies at title and abstract and full text levels using Covidence (https://www.covidence.org/), after filtering randomised trials with a machine learning tool (Cochrane Randomised Controlled Trial Classifier). Disagreements in judgments are resolved by discussion with senior reviewers (HW and SL).

Data collection

For newly eligible studies, paired reviewers (KN, YYG, QF, YM, XP, CL, XZ, YG, CY, JA, QL, and YYW) use standardised extraction forms to extract the following data independently: (i) general study details and setting (year, countries, setting, funding, length of follow-up, and sample sizes); (ii) baseline characteristics of included participants (age, sex, body mass index, haemoglobin A1c (HbA1c), duration of diabetes, and comorbidities including cardiovascular disease, chronic kidney disease, and obesity); (iii) interventions (medication name and dose and co-interventions administered in all eligible arms); and (iv) outcomes (trial-specific definition for each eligible outcome, number of participants analysed, number of events for binary outcomes, means or medians and variance for continuous outcomes). Where both intention-to-treat and per-protocol results are available, we prioritise the former for all outcomes.

Three linked international teams selected the outcomes of interest based on their best efforts to identify evidence regarding what patients with type 2 diabetes value, including a systematic review of patient values and preferences.[15] The *BMJ* Rapid Recommendations panel rated the perceived importance of outcomes to patients from 1 (least important) to 9 (most important) as follows:

- Critical (7-9): all-cause death, non-fatal stroke, hospitalisation for heart failure, kidney failure, health related quality of life (HRQoL), non-fatal myocardial infarction, amputation, dementia, and severe visual impairment
- Important (4-6): body weight change, severe hypoglycaemia, severe gastrointestinal events, ketoacidosis due to diabetes, severe hyperkalaemia, neuropathy, major osteoporotic fractures, falls, genital infections and urinary tract infections
- Less important (1-3): HbA1c change.

The linked Australian Guideline for Diabetes added several outcomes rated as being of critical patient importance: cardiovascular death, major adverse

cardiovascular events (MACEs), kidney disease progression (defined as a composite outcome of long term dialysis, kidney transplantation, sustained estimated glomerular filtration rate (eGFR) <15 mL/minute/1.73 m^2, sustained percent decline in eGFR of ≥40% or a doubling of serum creatinine, or kidney related death), discontinuation due to adverse events, and serious adverse events. Appendix 1.2 details the definitions used for each outcome.

Risk of bias within individual studies

Paired reviewers (TM, MQ, BW, BTJ, and CZ) assess the risk of bias for included studies independently using the modified Cochrane Risk of Bias tool produced by the Clinical Advances Through Research and Knowledge Translation (CLARITY) group at McMaster University.[23] For each included trial, reviewers evaluate the following domains: random sequence generation, allocation concealment, blinding of participants and personnel, missing outcome data, selective outcome reporting, and other bias. For each domain, rigour in trial methods is rated as either definitely yes (low risk of bias), probably yes, probably no, or definitely no (high risk of bias). Reviewers rate trials as high risk of bias overall if one or more domains were rated as such, and otherwise rate trials as being at low risk. Discrepancies are addressed via discussion and third party adjudication (HW).

Data synthesis

We use odds ratios (ORs) for dichotomous outcomes and mean differences (MDs) or standardised mean differences (SMDs) for continuous outcomes, all with their respective 95% confidence intervals (CIs). SMD is used to report HRQoL due to the heterogeneity of the measuring scale, and MD to report body weight (in kg) and HbA1c (as an absolute change of %).

Except for the body weight change and HbA1c change, the treatment nodes in the network represent medication classes rather than specific medications, assuming relative effects are similar across medications within the same medication class. Given previous NMAs demonstrated important differences on body weight change and HbA1c change across different GLP-1RAs,[4 24] we evaluate medication-specific, rather than class-related, body weight and HbA1c effects for GLP-1RAs. Placebo and standard treatment are grouped under one treatment node as the reference group across the network, representing the patient treatment regimen before adding medications.

We conduct random effects NMA using a frequentist graph theoretical approach with the weighted least-square estimator and Moore-Penrose pseudoinverse via the *netmeta* package (version 2.9) on *R* software (version 4.4.0).[25] We use the *ggraph* package (version 2.2.1) on *R* software (version 4.4.0) to draw network plots, where each node represents a separate intervention, thickness of lines between nodes represents the number of studies for each direct comparison, and the size of the nodes represents the number of participants randomised to that intervention. For trials with a single zero event, a continuity correction of 0.5 is added to all cells.[26] We use the generalised method of moments to estimate between-study variance and quantify heterogeneity in the network estimates. The statistical significance of the heterogeneity is assessed using the design-based decomposition of Cochran's Q statistic.[27]

Certainty of the evidence

We assess the certainty of the evidence using the Grading of Recommendations Assessment, Development and Evaluation (GRADE) approach for NMA. Paired reviewers (KN and QS) rate each domain for each comparison separately, with discussion and third party adjudication (SL, POV, AA, GG, and TA) to solve any discrepancies.

Certainty of each comparison is rated as high, moderate, low, or very low, accounting for risk of bias, inconsistency, indirectness, publication bias, intransitivity, incoherence (difference between direct and indirect effects), and imprecision.[28] Judgments for rating of risk of bias, inconsistency, indirectness, publication bias, and incoherence are consistent with our previous publication.[18] Briefly, we assess publication bias from two sources: global publication bias reflecting network estimates, evaluated using comparison-adjusted funnel plots[29]; and local publication bias reflecting each pairwise direct comparison, evaluated using Harbord's score test for binary outcomes and Egger's method for continuous outcomes.[30 31] To assess the robustness of local publication bias, a trim and fill analysis with both L-type and R-type estimators is conducted for meta-analysis with at least 10 trials.[32] We use node splitting and back-calculation methods to assess for local incoherence between direct and indirect estimates for each outcome-level comparison.[33] To consider potential intransitivity, we illustrate the distribution of potential effect modifiers and baseline mean outcomes for each direct comparison.

For rating of imprecision in relative estimates of effect, we apply the minimally contextualised approach.[34] This approach warrants setting thresholds for what constitutes important versus little or no effects. Our initial threshold for the target of certainty rating is, in all cases, the null, and we consequently rate our certainty in a non-zero effect. However, when the point estimate proves close to the null, we modify our target and rate certainty in little or no effect. In the absence of available research evidence to inform minimal important difference as thresholds, we take a pragmatic approach. For critical outcomes, we classify point estimates as close to the null when they prove, either as benefit or harm, less than 8%. The corresponding threshold for outcomes important but not crucial is 12%.

When rating certainty in a non-zero effect, we rate down for imprecision once if the CI crosses the null and twice if the CI also crosses either the 8% (for critical outcomes) or 12% (for important but not critical outcomes) threshold described above on the opposite side of the null. For comparisons in which the point estimate represents little or no effect, we rate down

doi: 10.1136/bmj-2024-083039 | *BMJ* 2025;390:e083039 | the bmj

one level for imprecision if the 95% CI crosses one of the close-to-the-null thresholds and twice if it crosses thresholds for both benefit and harm.

Categorisation of medication effectiveness

To facilitate the navigation of benefits and harms across 13 drug classes and 26 outcomes, we categorise medications from the most to the least effective or harmful using standard treatment as the reference intervention, following GRADE guidance.[35 36] Here, we use the relative estimates of effect, taking a minimally contextualised approach.[35] We initially select interventions with point estimates exceeding the close-to-the-null threshold compared with the reference (that is, standard treatment). We further compare these interventions to determine if they are significantly different from each other using null effects as the decision threshold. We ultimately categorise treatments into five categories: (i) among the most effective with a point estimate better than the reference and statistically no worse than any other medications; (ii) among the intermediately effective with a point estimate better than the reference and worse than at least one medication; (iii) not convincingly different from the reference; (iv) among the intermediately harmful with a point estimate more harmful than the reference but less harmful than at least one medication; and (v) among the most harmful with a point estimate more harmful than the reference and no other intervention more harmful. We then separate treatments into two groups based on the certainty of evidence for the comparison to the reference (high or moderate versus low or very low certainty).

Meta-regression, sensitivity, and subgroup analyses

For trial and aggregated patient characteristics measured as continuous variables, we perform the following four Bayesian meta-regressions:

- Proportion of patients with established cardiovascular diseases (hypothesising a larger relative effect in reducing death and cardiovascular and kidney outcomes in trials with a higher proportion of patients with cardiovascular diseases)
- Mean estimated glomerular filtration rate at baseline (hypothesising a larger relative effect in reducing death and cardiovascular and kidney outcomes in patients with lower estimated glomerular filtration rate)
- Mean body mass index at baseline (hypothesising a larger relative effect in reducing death and cardiovascular and kidney outcomes in patients with higher body mass index)
- Trial follow-up duration (hypothesising a larger relative effect in reducing death and cardiovascular and kidney outcomes in studies with longer follow-up).

The frequentist Pvalues for the subgroup effects are estimated using posterior distribution given the fact that posterior distribution converges to normal distribution.[37] The Instrument to assess the Credibility of Effect Modification Analyses (ICEMAN) [38] provides the framework for addressing the credibility of any apparent subgroup effect. In the absence of a credible subgroup effect, we infer relative effects are consistent across groups.

We perform sensitivity analyses, including a Bayesian NMA adjusted by trial duration [39]; a Mantel-Haenszel fixed effect NMA for rare events [40]; an analysis excluding trials with high risks of bias; an analysis excluding phase 2 or 3 trials; and an analysis pooling study reported hazard ratios for trials with ≥2 years' follow-up.

Absolute effect estimations

To better inform clinical decision making, we provide examples of anticipated absolute effects of medications for type 2 diabetes across benefits and harms. The absolute effects depend on the relative treatment effect and the individual's baseline risk of given events occurring without treatment. In this update, we applied three risk groups with varying baseline risks for mortality and cardiovascular and kidney related complications as defined in the linked living *BMJ* Rapid Recommendation and informed by the updated systematic review of risk prediction models.[13 14] This living guideline also informed baseline risk and absolute effect estimations for other outcomes not available for risk stratification. Here, baseline risk estimates are mainly based on pooling of control arm event rates across included trials using random effects single arm meta-analysis (appendix 5.3).

Given the complexity of presenting about 3000 estimates of effect from this NMA, we present relative and absolute estimates of effect, certainty ratings, and more detailed NMA results (such as number of participants and trials for each comparison) through an interactive GRADE summary of findings table available via the MATCH-IT tool (https://matchit.magicevidence.org/250709dist-diabetes/#!/). This tool has been developed for the linked *BMJ* Rapid Recommendation over several iterations and allows end users to compare any of the treatment options, including treatment effects with changeable comparators (for example, finerenone versus SGLT-2 inhibitors or GLP-1RAs for cardiovascular and kidney outcomes) and across three risk groups (from lower risk to higher risk of complications). When rating certainty of evidence for absolute treatment effect estimates, we use outcome-specific minimal important differences (MIDs) as the decision threshold informed by the living *BMJ* Rapid Recommendation guideline panel to inform judgments regarding imprecision. Treatments are similarly categorised from most to least effective using standard treatment as the reference intervention; whereas judgments for relative effect estimates in this NMA are made based on the point estimate using the null as the decision threshold, judgments for absolute effect estimates in MATCH-IT are made using outcome-specific MIDs and are based on whether 95% confidence intervals for each comparison cross the threshold.

Living model for evidence synthesis

For the ongoing review, the coordination team for the ALIVE consortium meets annually to re-visit the broad clinical question, the updating schedule, and to determine whether the living review should continue or retire based on feasibility and ongoing need for iterative updates. The review team conducts literature searches, screening of identified hits, data extraction, and risk of bias assessments for eligible studies on a bi-monthly basis. The systematic review and NMA will be updated at least two times per year, to incorporate emerging RCT evidence or to address other requests from the linked guideline panels (for example, new clinical questions such as combination therapy). NMA updates will be available at least once a year through the *BMJ* Rapid Recommendations (box 1) and—contingent on acquisition of sufficient funding—also be shared through a data repository at the ALIVE website (https://data.aliveevidence.org/). Subsequent iterations of the systematic review published in *The BMJ* may be limited to major updates. The methods team for the systematic review and NMA meets on a biweekly basis, including representation from the living *BMJ* Rapid Recommendation, and follows a consensus-based approach to decision making for the living NMA. Additional details for methods are available in the online protocol.[22]

Patient and public involvement

Patients are involved in selecting outcomes and rating their relative importance to patients, defining MIDs used to determine whether absolute effects are important from a patient perspective, and defining baseline risks for risk-stratified outcomes. Two patient partners contribute input as part of the panel for the living *BMJ* Rapid Recommendation on medications for diabetes and inform recommendations.

Results

Study selection and study characteristics

To date, the living systematic review and NMA includes 869 trials enrolling 493 168 participants (fig 1). The average age of study participants was 57.8 years, 56.9% were male, and 55.5% had a cardiovascular

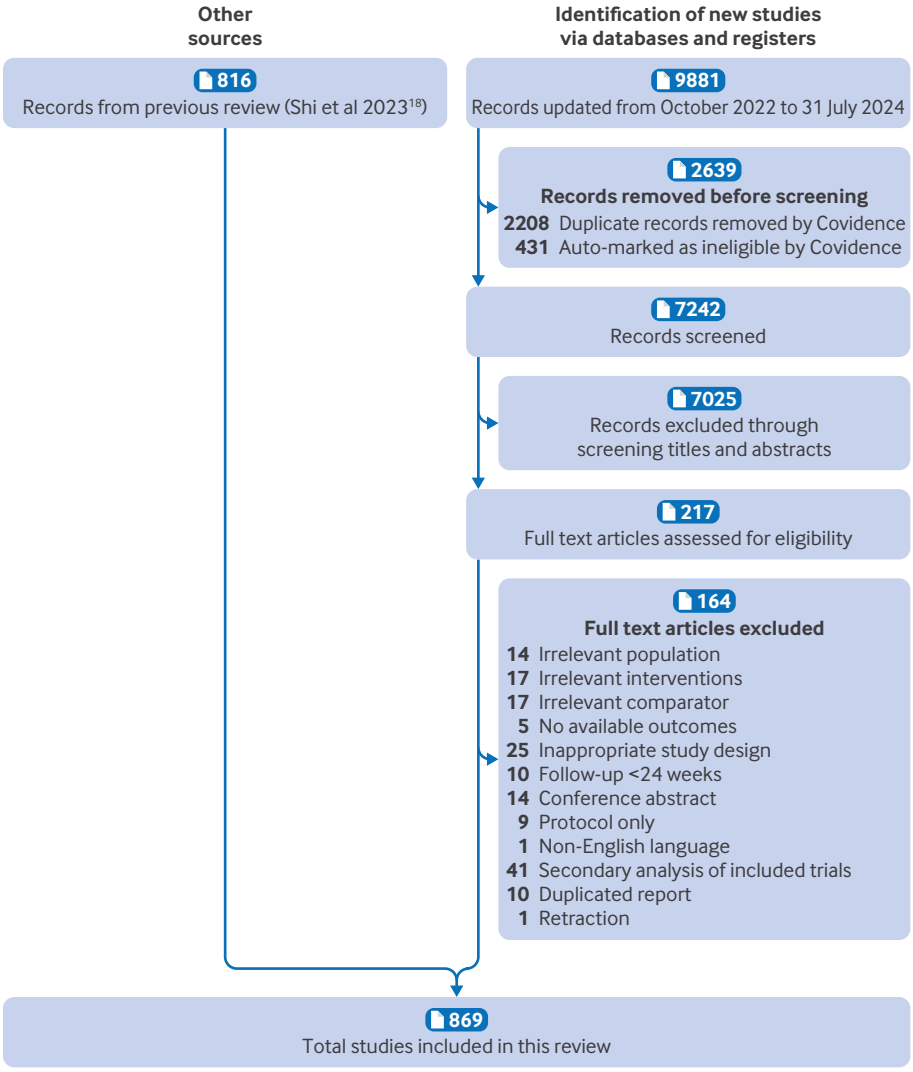

Fig 1 | Flow diagram for trial screen and selection

doi: 10.1136/bmj-2024-083039 | *BMJ* 2025;390:e083039 | the bmj

Table 1 | Baseline characteristics of included trials and their participants	
Characteristics	
Study settings (of eligible studies)	
Total No of trials	869
No of participants	493 168
Median (IQR) (range) length of follow-ups (months)	6.0 (5.5 to 12.0) (5.5 to 128)
Study characteristics (of participants)	
Pooled mean (95% CI) (95% PI) age (years)	57.8 (57.4 to 58.1) (47.4 to 68.1)
Pooled mean (95% CI) (95% PI) proportion of males (%)	56.9 (56.1 to 57.7) (34.2 to 77.0)
Pooled mean (95% CI) (95% PI) baseline BMI (kg/m^2)	29.5 (29.2 to 29.7) (22.6 to 36.3)
Pooled mean (95% CI) (95% PI) baseline HbA1c (%)	8.1 (8.1 to 8.2) (6.5 to 9.7)
Pooled mean (95% CI) (95% PI) proportion of baseline cardiovascular disease (%)	55.5 (39.3 to 70.7) (0.0 to 100.0)
Median (IQR) (range) baseline duration of diabetes (years)	7.5 (5.3 to 10.2) (0.0 to 20.7)

Notes: Pooled mean was estimated using the single-mean/proportion meta-analyses via a random-effect model.
Abbreviations: No = number, IQR = interquartile range, CI = confidence interval, PI = prediction interval, BMI = body mass index, HbA1c = haemoglobin A1C.

history (table 1). The median follow-up duration was six months (ranging from 5.5 to 128 months). In 240 of the 869 trials, there was at least one domain with high risk of bias. Of these studies at high risk of bias, 63.3% had inadequate blinding, 25.0% had serious missing outcome data, and 23.8% had inadequate allocation concealment (appendix 3). The number and characteristics of studies and participants for each outcome vary from the different included trials (appendix 2).

Comparative effectiveness of the diabetes treatment medications

Figure 2 presents the network plot summarising trials comparing treatment nodes head-to-head across outcomes. Appendix 4.1 presents network plots for each specific outcome. Figure 3 shows key benefits and harms with relative estimates of effect of all 13 interventions versus standard treatment. Appendix 4.3 shows network comparisons for all 26 outcomes, and appendix 5 provides details of the GRADE certainty ratings. Figure 4 summarises changes in body weight and HbA1c across medication classes, with effects of individual GLP1-RAs drugs reported separately given demonstrated differential effects.

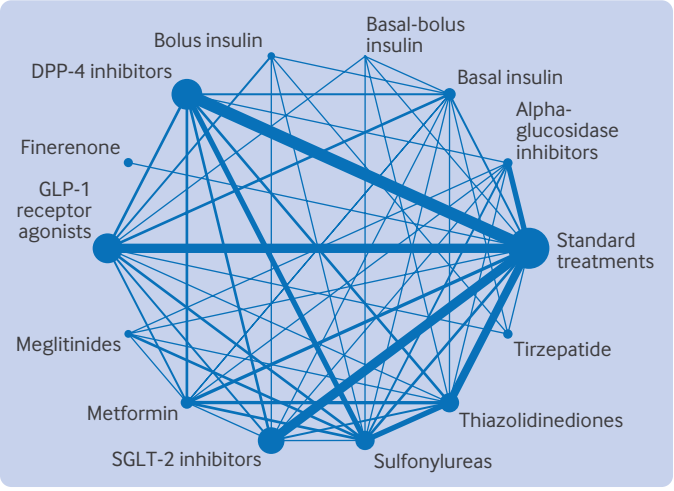

Fig 2 | Network plot for all included studies. Medications were grouped by their classes. Each node represents a medication class with node size reflecting the sample size of the treatment arm. The line between nodes represents the direct comparison between two medication classes with the thickness reflecting the number of trials. GLP-1 = glucagon-like peptide-1; SGLT-2 = sodium-glucose cotransporter-2; DPP-4 = dipeptidyl peptidase-4

Table 2 illustrates anticipated absolute effect estimates for key benefits and harms of selected medications for individuals at higher risk of death and cardiovascular and kidney complications. The MATCH-IT tool (https://matchit.magicevidence.org/250709dist-diabetes/#!/) also provides interactive evidence summaries for patients at lower and moderate risks of cardiovascular and kidney complications, demonstrating highly variable anticipated absolute benefits and harms based on variable baseline risks of outcomes occurring. Evidence summaries in MATCH-IT also allow for cross-comparisons across candidate medications, comparing effects for specific outcomes across risk groups. For example, for patients at higher risk category, GLP-1 RAs are possibly more effective in reducing stroke than SGLT-2 inhibitors (10 fewer per 1000 over 5 years, with 95% CI from 20 fewer to 2 more, low certainty evidence), while the opposite is the case for hospitalisation for heart failure (61 more per 1000 over 5 years, with 95% CI from 34 to 87 more, moderate certainty). All subsequent estimates refer to comparisons with standard treatments as the reference (comparator).

All-cause death and cardiovascular death

The analysis involves 268 trial comparisons with 357 369 participants reporting 15 703 all-cause deaths, and 153 trial comparisons with 289 167 participants reporting 9425 cardiovascular deaths. SGLT-2 inhibitors and GLP-1RAs reduce the risks of all-cause death (SGLT-2 inhibitors odds ratio (OR) 0.88 (95% CI 0.83 to 0.94), GLP-1RAs OR 0.87 (0.82 to 0.92), both high certainty) and cardiovascular death (SGLT-2 inhibitors OR 0.86 (0.80 to 0.94), GLP-1RAs OR 0.85 (0.79 to 0.92), both high certainty) (see fig 3 and appendix 5). Finerenone reduces the risk of all-cause death (OR 0.89 (0.79 to 1.00), high certainty) and probably reduces the risk of cardiovascular death (OR 0.88 (0.75 to 1.02), moderate certainty). DPP-4 inhibitors probably have little or no effect on all-cause death (OR 1.01 (0.94 to 1.08), moderate certainty) and

possibly have no effect on cardiovascular death (OR 1.00 (0.92 to 1.09), low certainty). Metformin possibly reduces the risk of all-cause death (OR 0.82 (0.64 to 1.06), low certainty), and sulfonylureas possibly increase risk (OR 1.09 (0.96 to 1.25), low certainty). Other drugs have uncertain effects on mortality outcomes (very low certainty; see fig 3 and appendix 5).

Non-fatal myocardial infarction and non-fatal stroke

This analysis involves 214 trial comparisons with 303 654 participants reporting 8889 non-fatal myocardial infarctions, and 182 trial comparisons with 294 387 participants reporting 4948 non-fatal strokes. SGLT-2 inhibitors (OR 0.90 (0.82 to 0.98), high certainty) and GLP-1RAs (OR 0.91 (0.85 to 0.98), high certainty) reduce the risk of non-fatal myocardial infarction; finerenone may also reduce risk (OR 0.91 (0.74 to 1.12), low certainty). GLP-1RAs reduce the risk of non-fatal stroke (OR 0.87 (0.79 to 0.96), high certainty), as may DPP-4 inhibitors (OR 0.90 (0.79 to 1.03), low certainty). Other medications may have little impact on or have very uncertain effects on non-fatal myocardial infarction and stroke compared with standard treatment (low to very low certainty; see fig 3 and appendix 5).

Hospitalisation for heart failure

The analysis involves 144 trial comparisons with 257 718 participants reporting 6702 events. SGLT-2 inhibitors (OR 0.66 (0.60 to 0.72), high certainty) and finerenone (OR 0.78 (0.66 to 0.92), high certainty) are among the best in reducing the risk of hospitalisation for heart failure, followed by GLP-1RAs (OR 0.91 (0.83 to 0.99), high certainty). Thiazolidinediones probably increase the risk of hospitalisation for heart failure (OR 1.54 (1.27 to 1.88), moderate certainty). Other drugs may have little or no effect or uncertain effects on hospitalisation for heart failure (low to very low certainty; see fig 3 and appendix 5).

3-point major adverse cardiovascular events (3-P MACE)

This analysis involves 29 trial comparisons with 167 880 participants and 15 761 events for 3-P MACE. SGLT-2 inhibitors (OR 0.89 (0.83 to 0.95), high certainty) and GLP-1RAs (OR 0.85 (0.80 to 0.91), high certainty) reduce the risk of 3-P MACE; tirzepatide (OR 0.88 (0.26 to 3.05), low certainty) and thiazolidinediones (OR 0.91 (0.73 to 1.14), low certainty) may also reduce risk of the same. DPP-4 inhibitors probably have little or no effect on 3-P MACE (OR 1.01 (0.92 to 1.12), moderate certainty). Other drugs may have little or no or uncertain effect on 3P-MACE (low to very low certainty; see fig 3 and appendix 5).

Kidney failure and kidney disease progression

The analysis involves 19 trial comparisons with 123 636 participants reporting 1593 kidney failure events, and 56 trial comparisons with 223 586 participants reporting 8167 events for kidney disease progression. SGLT-2 inhibitors reduce the risk of kidney failure (OR 0.68 (0.56 to 0.83), high certainty). GLP-1RAs (OR 0.86 (0.70 to 1.06), moderate certainty) and finerenone (OR 0.85 (0.71 to 1.01), moderate certainty) probably reduce the risk of kidney failure. SGLT-2 inhibitors (OR 0.61 (0.55 to 0.69), high certainty) and finerenone (OR 0.84 (0.73 to 0.96), high certainty) reduce likelihood of kidney disease progression, as probably do GLP-1RAs (OR 0.84 (0.76 to 0.93), moderate certainty). SGLT-2 inhibitors are among the most effective medications and are probably superior to GLP-1RAs and finerenone in reducing the risk of kidney disease progression (both moderate certainty). Other drugs may have little or no effect or uncertain effects on kidney failure and kidney disease progression relative to standard treatment (low to very low certainty; see fig 3 and appendix 5).

Health-related quality of life (HRQoL)

The analysis involves 35 trial comparisons with 24 446 participants using 13 questionnaires (see details in appendix 1.2). GLP-1RAs (standardised mean difference (SMD) 0.20 (95% CI 0.10 to 0.29), high certainty), tirzepatide (SMD 0.46 (0.22 to 0.69), moderate certainty), and SGLT-2 inhibitors (SMD 0.32 (0.13 to 0.51), moderate certainty) probably improve HRQoL. Other drugs may have little or no impact on HRQoL (all low certainty; see fig 3 and appendix 5).

Body weight change

The analysis involves 485 trial comparisons with 263 620 participants. Figure 4 gives the results. Tirzepatide (mean reduction 8.63 kg (7.93 to 9.34), moderate certainty) and orforglipron (mean reduction 7.87 kg (5.5 to 10.24), low certainty) are the most effective drugs in reducing body weight. Other individual GLP-1RA medications, SGLT-2 inhibitors, metformin, and alpha-glucosidase inhibitors variably reduce body weight (mean reduction ranging from 4.44 to 0.52 kg, and from high to low certainty). Thiazolidinediones probably (mean increase 2.79 kg (2.54 to 3.04 kg), moderate certainty) and basal-bolus insulin possibly (mean increase 3.11 kg (1.93 to 4.29 kg), low certainty) increase body weight the most. Four other medication classes probably increase body weight (all moderate certainty): 1.94 kg for basal insulin, 1.87 kg for sulfonylurea, 1.29 kg for meglitinides, and 0.27 kg for DPP-4 inhibitors. Bolus insulin may increase body weight by 2.04 kg (low certainty evidence).

Severe visual impairment

The analysis involves 116 trial comparisons with 247 964 participants, of whom 7570 developed severe visual impairment, including blindness, diabetic retinopathy, cataract, and macular oedema. We did not identify moderate to high certainty evidence supporting any medication in increasing or decreasing the risk of severe visual impairment (low to very low certainty; see appendix 5).

Dementia

The analysis involves 23 trial comparisons with 161 179 participants, 1933 of whom developed

Table 2 | Anticipated absolute effects (95% confidence intervals) for patients with type 2 diabetes and established cardiovascular or chronic kidney disease at higher risk of complications

Interventions	All-cause death	Non-fatal MI	Non-fatal stroke	Hospitalised for heart failure (CKD)*	Hospitalised for heart failure (CVD)†	3-P MACE	Kidney disease progression	HRQoL score	Severe hypoglycaemia	Discontinued due to adverse events	Drug-specific adverse events
Baseline risks (per 1000 patients)	240	110	90	60	300	440	100	50 points	60	132	—
SGLT-2 inhibitors	**23 fewer (32 fewer to 11 fewer)**	**10 fewer (18 fewer to 2 fewer)**	1 fewer (10 fewer to 9 more)	**20 fewer (23 fewer to 16 fewer)**	**80 fewer (95 fewer to 64 fewer)**	**28 fewer (45 fewer to 13 fewer)**	**37 fewer (42 fewer to 29 fewer)**	**3.2 higher (1.3 higher to 5.1 higher)**	6 fewer (13 fewer to 1 more)	**14 more (0 to 28 more)**	**Genital infection: 128 more (108 more to 151 more)** **Amputation: 8 more (0 to 17 more)** **Ketoacidosis due to diabetes: 11 more (4 more to 19 more)**
GLP-1 receptor agonists	**24 fewer (34 fewer to 15 fewer)**	**9 fewer (15 fewer to 2 fewer)**	**11 fewer (18 fewer to 3 fewer)**	**5 fewer (10 fewer to 1 fewer)**	**19 fewer (38 fewer to 2 fewer)**	**40 fewer (54 fewer to 23 fewer)**	**15 fewer (22 fewer to 6 fewer)**	**2.0 higher (1.0 higher to 2.9 higher)**	1 fewer (6 fewer to 4 more)	**72 more (57 more to 88 more)**	**Severe GI events: 40 more (19 more to 67 more)**
Finerenone	**21 fewer (40 fewer to 0 fewer)**	9 fewer (26 fewer to 12 more)	0 fewer (15 fewer to 18 more)	**13 fewer (20 fewer to 5 fewer)**	—	—	**15 fewer (25 fewer to 4 fewer)**	—	**21 fewer (33 fewer to 2 fewer)**	21 more (24 fewer to 80 more)	**Severe hyperkalaemia: 297 more (151 more to 464 more)**
Tirzepatide	55 fewer (118 fewer to 35 more)	31 fewer (100 fewer to 319 more)	—	20 fewer (49 fewer to 75 more)	82 fewer (232 fewer to 211 more)	31 fewer (270 fewer to 266 more)	39 fewer (91 fewer to 209 more)	**4.6 higher (2.2 higher to 6.9 higher)**	20 fewer (42 fewer to 25 more)	**93 more (46 more to 150 more)**	**Severe GI events: 131 more (40 more to 283 more)**
Metformin	34 fewer (72 fewer to 11 more)	5 fewer (29 fewer to 26 more)	12 fewer (36 fewer to 20 more)	24 more (42 fewer to 259 more)	82 more (193 fewer to 459 more)	—	52 more (62 fewer to 348 more)	0.8 higher (2.1 lower to 3.6 higher)	39 more (6 fewer to 116 more)	**39 more (14 more to 68 more)**	—
Alpha-glucosidase inhibitors	81 fewer 178 fewer to 119 more)	71 fewer (103 fewer to 83 more)	388 more (21 fewer to 829 more)	112 more (52 fewer to 780 more)	282 more (247 fewer to 672 more)	—	—	0.7 higher (2.9 lower to 4.2 higher)	16 more (41 fewer to 196 more)	**87 more (42 more to 139 more)**	—
Thiazolidinediones	9 fewer (32 fewer to 16 more)	2 fewer (18 fewer to 15 more)	13 fewer (26 fewer to 2 more)	**30 more (15 more to 47 more)**	**98 more (52 more to 146 more)**	23 fewer (75 fewer to 32 more)	7 fewer (32 fewer to 24 more)	2.3 higher (0.9 lower to 5.5 higher)	22 more (3 fewer to 56 more)	**30 more (11 more to 50 more)**	**Major osteoporotic fractures: 22 more (1 more to 54 more)**
DPP-4 inhibitors	2 more (11 fewer to 14 more)	1 more (7 fewer to 11 more)	8 fewer (18 fewer to 2 more)	3 more (3 fewer to 9 more)	10 more (11 fewer to 34 more)	2 more (20 fewer to 28 more)	4 fewer (14 fewer to 8 more)	0.7 higher (0.7 lower to 2.0 higher)	**7 more (1 more to 13 more)**	6 fewer (16 fewer to 7 more)	—
Sulfonylureas	16 more (7 fewer to 43 more)	1 more (17 fewer to 22 more)	2 more (16 fewer to 23 more)	0 fewer (11 fewer to 14 more)	0 fewer (45 fewer to 49 more)	14 more (34 fewer to 65 more)	7 fewer (26 fewer to 14 more)	1.1 higher (1.1 lower to 3.2 higher)	**186 more (134 more to 245 more)**	7 more (9 fewer to 23 more)	—
Meglitinides	83 more (106 fewer to 358 more)	77 fewer (104 fewer to 56 more)	52 more (66 fewer to 436 more)	—	—	—	—	2.0 higher (2.5 lower to 6.5 higher)	109 more (3 fewer to 345 more)	1 more (30 fewer to 38 more)	—
Basal insulin	13 more (40 fewer to 73 more)	2 fewer (55 fewer to 92 more)	12 fewer (55 fewer to 70 more)	2 fewer (20 fewer to 24 more)	6 fewer (82 fewer to 82 more)	—	5 fewer (24 fewer to 18 more)	1.1 higher (0.6 lower to 2.7 higher)	**68 more (42 more to 102 more)**	**52 fewer (68 fewer to 32 fewer)**	—
Basal-bolus insulin	42 fewer (183 fewer to 268 more)	71 fewer (106 fewer to 177 more)	33 fewer (79 fewer to 169 more)	—	—	—	—	—	**176 more (2 more to 529 more)**	68 fewer (108 fewer to 30 more)	—
Bolus insulin	34 fewer (138 fewer to 128 more)	17 more (63 fewer to 191 more)	10 fewer (73 fewer to 222 more)	21 fewer (56 fewer to 224 more)	85 fewer (271 fewer to 427 more)	—	121 more (89 fewer to 776 more)	0.7 lower (2.4 lower to 1.1 higher)	**104 more (38 more to 201 more)**	**52 fewer (77 fewer to 19 fewer)**	—
Standard treatments	Reference	Reference	Reference	Reference	Reference	Reference	Reference	Reference	Reference	Reference	Reference

The table shows absolute benefits and harms of the drugs for patients with type 2 diabetes and established cardiovascular or chronic kidney disease at higher risk of complications. The bold estimates are statistically significant.
*This risk group includes patients with type 2 diabetes and established chronic kidney disease at higher risk of complications.
†This risk group includes patients with type 2 diabetes and established cardiovascular disease at higher risk of complications.
Abbreviations: MI = myocardial infarction; HRQoL = health-related quality of life; SGLT-2 = sodium-glucose cotransporter-2; GLP-1 = glucagon-like peptide-1; CKD = chronic kidney disease; CVD = cardiovascular disease; 3-P MACE = 3-point major adverse cardiovascular events; GI = gastrointestinal.

Interventions	All-cause death	Non-fatal MI	Non-fatal stroke	Hospitalisation for heart failure	3-P MACE	Kidney disease progression	HRQoL score (SMD (95%CI))	Severe hypoglycaemia	Discontinuation due to adverse events	Drug-specific adverse events
SGLT-2 inhibitors	0.88 (0.83 to 0.94)	0.90 (0.82 to 0.98)	0.99 (0.88 to 1.11)	0.66 (0.60 to 0.72)	0.89 (0.83 to 0.95)	0.61 (0.55 to 0.69)	0.32 (0.13 to 0.51)	0.89 (0.78 to 1.01)	1.12 (1.00 to 1.25)	Genital infection 3.29 (2.88 to 3.77) Amputation 1.27 (1.01 to 1.61) Ketoacidosis due to diabetes 2.08 (1.45 to 2.99)
GLP-1 receptor agonists	0.87 (0.82 to 0.92)	0.91 (0.85 to 0.98)	0.87 (0.79 to 0.96)	0.91 (0.83 to 0.99)	0.85 (0.80 to 0.91)	0.84 (0.76 to 0.93)	0.20 (0.10 to 0.29)	0.98 (0.90 to 1.07)	1.69 (1.53 to 1.86)	Severe gastrointestinal event 1.88 (1.41 to 2.52)
Finerenone	0.89 (0.79 to 1.00)	0.91 (0.74 to 1.12)	1.00 (0.82 to 1.22)	0.78 (0.66 to 0.92)	—	0.84 (0.73 to 0.96)	—	0.64 (0.43 to 0.96)	1.19 (0.80 to 1.77)	Severe hyperkalaemia 5.92 (3.02 to 11.62)
Tirzepatide	0.72 (0.44 to 1.20)	0.69 (0.08 to 6.08)	—	0.65 (0.17 to 2.44)	0.88 (0.26 to 3.05)	0.58 (0.08 to 4.02)	0.46 (0.22 to 0.69)	0.65 (0.29 to 1.45)	1.91 (1.42 to 2.58)	Severe gastrointestinal event 4.21 (1.87 to 9.49)
Metformin	0.82 (0.64 to 1.06)	0.95 (0.71 to 1.27)	0.85 (0.58 to 1.25)	1.44 (0.28 to 7.35)	—	1.61 (0.36 to 7.29)	0.08 (-0.21 to 0.36)	1.72 (0.89 to 3.35)	1.36 (1.12 to 1.64)	—
Alpha-glucosidase inhibitors	0.60 (0.21 to 1.77)	0.33 (0.06 to 1.93)	9.25 (0.75 to 114.27)	3.25 (0.13 to 82.49)	—	—	0.07 (-0.29 to 0.42)	1.29 (0.31 to 5.40)	1.84 (1.39 to 2.44)	—
Thiazolidinediones	0.95 (0.83 to 1.09)	0.98 (0.82 to 1.16)	0.84 (0.69 to 1.02)	1.54 (1.27 to 1.88)	0.91 (0.73 to 1.14)	0.92 (0.66 to 1.27)	0.23 (-0.09 to 0.55)	1.40 (0.95 to 2.06)	1.27 (1.10 to 1.46)	Major osteoporotic fractures 1.60 (1.03 to 2.48)
DPP-4 inhibitors	1.01 (0.94 to 1.08)	1.01 (0.93 to 1.11)	0.90 (0.79 to 1.03)	1.05 (0.95 to 1.17)	1.01 (0.92 to 1.12)	0.96 (0.85 to 1.09)	0.07 (-0.07 to 0.20)	1.12 (1.01 to 1.24)	0.95 (0.86 to 1.06)	—
Sulfonylureas	1.09 (0.96 to 1.25)	1.01 (0.83 to 1.23)	1.03 (0.81 to 1.29)	1.00 (0.80 to 1.25)	1.06 (0.87 to 1.30)	0.92 (0.72 to 1.16)	0.11 (-0.11 to 0.32)	5.10 (3.77 to 6.88)	1.06 (0.92 to 1.21)	—
Meglitinides	1.51 (0.49 to 4.71)	0.28 (0.05 to 1.61)	1.68 (0.25 to 11.22)	—	—	—	0.20 (-0.25 to 0.65)	3.18 (0.95 to 10.65)	1.01 (0.75 to 1.35)	—
Basal insulin	1.07 (0.79 to 1.44)	0.98 (0.47 to 2.05)	0.85 (0.37 to 1.92)	0.97 (0.65 to 1.44)	—	0.94 (0.74 to 1.20)	0.11 (-0.06 to 0.27)	2.31 (1.77 to 3.02)	0.57 (0.45 to 0.73)	—
Basal-bolus insulin	0.78 (0.19 to 3.27)	0.33 (0.03 to 3.26)	0.61 (0.11 to 3.53)	—	—	—	—	4.84 (1.04 to 22.48)	0.45 (0.16 to 1.27)	—
Bolus insulin	0.82 (0.36 to 1.84)	1.18 (0.40 to 3.49)	0.88 (0.17 to 4.59)	0.64 (0.07 to 6.22)	—	2.56 (0.10 to 63.53)	-0.07 (-0.24 to 0.11)	3.08 (1.71 to 5.54)	0.57 (0.38 to 0.84)	—
Standard treatments	Reference	Reference	Reference	Reference	Reference	Reference	Reference	Reference	Reference	Reference

High to moderate certainty evidence	Low to very low certainty evidence
Among the most effective	Possibly among the most effective
Among the intermediate effective	Possibly among the intermediate effective
Not convincingly different from standard treatment	Possibly not convincingly different from standard treatment
Among the intermediate harmful	Possibly among the intermediate harmful
Among the most harmful	Possibly among the most harmful

Notes: The study adopts minimally contextualised framework with decision thresholds and null effect to rate and categorise the drugs from among the most effective to among the most harmful. The decision thresholds for relative effects are 8% for critical binary outcomes, and 12% for moderate important binary outcomes. We first categorise the drugs that are superior to (or inferior to) standard treatments (that is, point estimate exceeds (or falls below) decision threshold) into among the most effective group (or the most harmful group). We then categorise the drugs that are among the most effective (or the most harmful) but inferior to (that is, point estimate falls below and 95% CI does not cross null effect) at least one medication in that group into among the intermediate effective group (or the intermediate harmful group).
Abbreviations: MI = myocardial infarction; 3-P MACE = 3-point major adverse cardiovascular events; HRQoL = Health-related quality of life; SMD = standardised mean difference; CI = confidence interval; SGLT-2 = sodium-glucose cotransporter-2; GLP-1 = glucagon-like peptide-1

Fig 3 | Benefits and harms of pharmacotherapy for type 2 diabetes, showing comparative effect estimates versus standard treatments, which typically represent the treatment regimens the patient received before the clinician considered adding a new medication. Values are odds ratios (95% confidence intervals) unless stated otherwise.

Interventions	Bodyweight change (kg)	Haemoglobin A1C (%)
Tirzepatide	-8.63 (-9.34 to -7.93)	-1.78 (-1.96 to -1.59)
Orforglipron	-7.87 (-10.24 to -5.50)	-1.82 (-2.37 to -1.26)
Semaglutide (subcutaneous)	-4.44 (-5.02 to -3.86)	-1.36 (-1.51 to -1.21)
Beinaglutide	-4.20 (-9.79 to 1.39)	-0.41 (-1.53 to 0.71)
Semaglutide (oral)	-2.89 (-3.51 to -2.26)	-1.15 (-1.31 to -0.99)
Efpeglenatide	-2.59 (-4.46 to -0.72)	-1.20 (-1.60 to -0.79)
Liraglutide	-2.34 (-2.70 to -1.97)	-0.88 (-0.97 to -0.78)
SGLT-2 inhibitors	-1.94 (-2.14 to -1.74)	-0.59 (-0.64 to -0.55)
Exenatide immediate-release	-1.86 (-2.56 to -1.16)	-0.73 (-0.90 to -0.57)
Dulaglutide	-1.64 (-2.14 to -1.15)	-0.97 (-1.10 to -0.83)
Exenatide extended-release	-1.07 (-1.74 to -0.40)	-0.83 (-1.01 to -0.65)
Metformin	-0.79 (-1.11 to -0.46)	-0.78 (-0.91 to -0.65)
Lixisenatide	-0.75 (-1.26 to -0.25)	-0.54 (-0.68 to -0.39)
Albiglutide	-0.63 (-1.28 to 0.02)	-0.69 (-0.87 to -0.52)
Alpha-glucosidase inhibitors	-0.52 (-0.94 to -0.10)	-0.61 (-0.72 to -0.50)
Loxenatide	0.16 (-1.78 to 2.10)	-0.92 (-1.23 to -0.61)
DPP-4 inhibitors	0.27 (0.10 to 0.44)	-0.56 (-0.60 to -0.52)
Finerenone	0.39 (-1.07 to 1.84)	0.09 (-0.29 to 0.47)
Meglitinides	1.29 (0.60 to 1.98)	-0.68 (-0.91 to -0.45)
Sulfonylureas	1.87 (1.59 to 2.15)	-0.97 (-1.18 to -0.77)
Basal insulin	1.94 (1.56 to 2.32)	-0.73 (-0.94 to -0.52)
Bolus insulin	2.04 (1.39 to 2.69)	-0.74 (-0.93 to -0.56)
Thiazolidinediones	2.79 (2.54 to 3.04)	-0.67 (-0.73 to -0.62)
Basal-bolus insulin	3.11 (1.93 to 4.29)	-0.52 (-0.80 to -0.23)
Visepegenatide	—	-0.74 (-1.33 to -0.15)
Standard treatments	Reference	Reference

High to moderate certainty evidence
Among the most effective
Among the intermediate effective
Not convincingly different from standard treatment
Among the intermediate harmful
Among the most harmful

Low to very low certainty evidence
Possibly among the most effective
Possibly among the intermediate effective
Possibly not convincingly different from standard treatment
Possibly among the intermediate harmful
Possibly among the most harmful

Notes: The figure shows comparative effect estimates versus standard treatments, which typically represented the treatment regimen the patient received before the clinician considered adding a new medication. The study adopts minimally contextualised framework with a null effect threshold to rate and categorise the drugs from among the most effective to among the most harmful. We first categorise the drugs that are superior to (or inferior to) standard treatments (that is, the point estimate exceeds (or falls below) the null effect and the 95% confidence interval does not cross it) into among the most effective group (or the most harmful group). We then categorise the drugs that are among the most effective (or the most harmful) but inferior to at least one drug in that group (that is, the point estimate falls below and the 95% CI does not cross) into among the intermediate effective group (or the intermediate harmful group).
Abbreviations: SGLT-2 = sodium glucose cotransporter-2; DPP-4 = dipeptidyl peptidase-4

Fig 4 | Effects of pharmacotherapy for type 2 diabetes on body weight and haemoglobin A1c. Values are mean differences (95% confidence intervals)

dementia. GLP-1RAs may reduce the risk of dementia (OR 0.92 (0.83 to 1.02), low certainty). Other drugs have uncertain effects on dementia compared with standard treatments (all very low certainty; see appendix 5).

Neuropathy

The analysis involves 69 trial comparisons with 180 420 participants reporting 4880 events for neuropathy. GLP-1RAs (OR 1.01 (0.92 to 1.11), moderate certainty) probably have little or no effect

on risk of neuropathy. Sulfonylureas (OR 1.22 (1.04 to 1.44), low certainty) and thiazolidinediones (OR 1.21 (0.97 to 1.51), low certainty) may increase the risk of neuropathy. Other medications have uncertain effects on neuropathy events (low to very low certainty; see appendix 5).

Severe hypoglycaemia

The analysis involves 210 trial comparisons with 308 839 participants reporting 5699 events of severe hypoglycaemia. Sulfonylureas (OR 5.10 (3.77 to 6.88)), basal-bolus insulin (OR 4.84 (1.04 to 22.48)), and bolus insulin (OR 3.08 (1.71 to 5.54)) probably increase the risk of severe hypoglycaemic events, with likely smaller increases in risk with basal insulin (OR 2.31 (1.77 to 3.02)) and DPP-4 inhibitors (OR 1.12 (1.01 to 1.24)) (all moderate certainty). Meglitinides (OR 3.18 (0.95 to 10.65)), metformin (OR 1.72 (0.89 to 3.35)), and thiazolidinediones (OR 1.40 (0.95 to 2.06)) possibly increase the risk of severe hypoglycaemic events (all low certainty). SGLT-2 inhibitors (OR 0.89 (0.78 to 1.01), moderate certainty) and GLP-1RAs (OR 0.98 (0.90 to 1.07), moderate certainty) probably have little or no effect on the risk of severe hypoglycaemia. Finerenone reduces severe hypoglycaemia compared with standard treatment (OR 0.64 (0.43 to 0.96), high certainty). Tirzepatide and alpha-glucosidase inhibitors have uncertain effects on severe hypoglycaemia compared with standard treatments (both very low certainty; see fig 3 and appendix 5).

Discontinuation due to adverse events and serious adverse events

The analysis involves 530 trial comparisons with 399 079 participants, 27 729 of whom stopped medication due to adverse events and 484 trial comparisons with 400 006 participants reporting 71 997 serious adverse events. Figure 3 shows relative estimates of effect for discontinuation due to adverse events, with moderate certainty evidence demonstrating highest risks for tirzepatide (OR 1.91 (1.42 to 2.58)), alpha-glucosidase inhibitors (OR 1.84 (1.39 to 2.44)), and GLP-1RAs (OR 1.69 (1.53 to 1.86)). Other medications may have little or no effect or uncertain effects on discontinuation due to adverse events and serious adverse events (low to very low certainty; see fig 3 and appendix 5).

Severe gastrointestinal events

The analysis involves 46 trial comparisons with 73 270 participants reporting 1952 events. Tirzepatide (OR 4.21 (1.87 to 9.49), moderate certainty) and GLP-1RAs (OR 1.88 (1.41 to 2.52), moderate certainty) probably increase severe gastrointestinal adverse events. Of the two, tirzepatide probably increases gastrointestinal adverse events to a greater extent compared with GLP-1RAs. Basal insulin probably decreases the risk of severe gastrointestinal adverse events (OR 0.27 (0.10 to 0.71), moderate certainty). Other drugs have uncertain effects on severe gastrointestinal events compared with standard treatments (low to very low certainty; see fig 3 and appendix 5).

Genital infections and urinary tract infections

The analysis involves 104 trial comparisons with 105 129 participants reporting 2423 genital infections, and 285 trial comparisons with 284 602 participants reporting 10 172 urinary tract infections. SGLT-2 inhibitors increase genital infections (OR 3.29 (2.88 to 3.77), high certainty) but have no effects on urinary tract infections (OR 1.04 (0.99 to 1.11), high certainty). Sulfonylureas probably reduce the risks of genital infections (OR 0.52 (0.37 to 0.75), moderate certainty) and urinary tract infections (OR 0.87 (0.75 to 1.00), moderate certainty). Other drugs may have little or no or uncertain effects on genital infections and urinary tract infections compared with standard treatments (low to very low certainty; see fig 3 and appendix 5).

Amputation

The analysis involves 19 trial comparisons with 108 256 participants reporting 1165 amputations. SGLT-2 inhibitors probably increase the risk of amputation (OR 1.27 (1.01 to 1.61), moderate certainty; fig 3). GLP-1RAs (OR 0.71 (0.07 to 6.95)) and DPP-4 inhibitors (OR 0.91 (0.51 to 1.60)) may reduce the risk of amputation (both low certainty). Other drugs may have little or no or uncertain effects on amputation (low to very low certainty; see appendix 5).

Ketoacidosis due to diabetes

The analysis involves 39 trial comparisons with 142 555 participants reporting 284 ketoacidosis events. SGLT-2 inhibitors increase the risk of ketoacidosis due to diabetes (OR 2.08 (1.45 to 2.99), high certainty; fig 3), as possibly do GLP-1RAs (OR 1.12 (0.68 to 1.86), low certainty). Finerenone possibly reduces the risk of ketoacidosis due to diabetes (OR 0.68 (0.30 to 1.56), low certainty). Other drugs have uncertain effects on ketoacidosis due to diabetes (all very low certainty; see appendix 5).

Severe hyperkalaemia

The analysis involves two trials with 12 999 participants reporting 71 events demonstrating that finerenone increases severe hyperkalaemia (OR 5.92 (3.02 to 11.62), high certainty).

Major osteoporotic fractures and falls

The analysis involves 112 trial comparisons with 257 188 participants reporting 990 major osteoporotic fractures and 84 trial comparisons with 202 303 participants and 813 events for falls. Thiazolidinediones increase major osteoporotic fractures (OR 1.60 (1.03 to 2.48), high certainty; fig 3) and possibly increase the risk of falls (OR 1.74 (0.56 to 5.45), low certainty). DPP-4 inhibitors probably decrease the risks of major osteoporotic fractures (OR 0.82 (0.63 to 1.05), moderate certainty) and falls (OR 0.74 (0.52 to 1.06), moderate certainty). Finerenone

possibly decreases the risks of major osteoporotic fractures (OR 0.85 (0.49 to 1.49), low certainty) and falls (OR 0.56 (0.20 to 1.55), low certainty). SGLT-2 inhibitors possibly increase the risk of falls (OR 1.14 (0.86 to 1.50), low certainty) but may have little effect on major osteoporotic fractures (OR 0.97 (0.77 to 1.22), low certainty). Other drugs may have little or no or uncertain effects compared with standard treatments on fractures or falls (low to very low certainty; see appendix 5).

Haemoglobin A1c change

The analysis involves 774 trial comparisons with 419 698 participants. Figure 4 gives the results. All medication classes result in HbA1c reduction except for finerenone and beinaglutide. Tirzepatide (mean difference (MD) −1.78% (95% CI −1.96% to −1.59%), moderate certainty) and orforglipron (MD −1.82% (−2.37% to −1.26%), low certainty) prove the most effective, followed by other drugs with reductions in HbA1c ranging from 1.36% to 0.52% (moderate to very low certainty; fig 4)

Subgroup analyses and sensitivity analyses

Our study did not identify any credible subgroup effects (appendix 6), and all sensitivity analyses confirmed the robustness of our findings (appendix 7).

Discussion

Principal findings

The current iteration of this living systematic review includes 869 randomised controlled trials enrolling nearly half a million people with type 2 diabetes, adding 53 trials since October 2022. Key findings include the confirmation of cardiovascular and kidney benefits of SGLT-2 inhibitors, GLP-1Ras, and finerenone with moderate to high certainty evidence; some notable harms of established medications (including increased risks of genital infection, amputation, and ketoacidosis due to diabetes for SGLT-2 inhibitors, severe gastrointestinal events for GLP-1RAs and tirzepatide, severe hyperkalaemia for finerenone, heart failure exacerbation and fractures for thiazolidinediones, and severe hypoglycaemia for insulin, sulfonylureas, and DPP-4 inhibitors); and the remaining uncertainty for numerous outcomes of importance to patients. With absolute benefits determined by baseline risks for cardiovascular and kidney outcomes balanced against relevant harms, our findings underscore the need to provide personalised diabetes care for medications and directly inform the linked guidelines to facilitate shared decision making based on values and preferences.[13 17-19]

For weight loss and HbA1c, tirzepatide and orforglipron represent the most effective among many emerging molecules.[4] Tirzepatide also carries the highest risk of severe gastrointestinal events, and its cardiovascular and kidney benefits continue to be evaluated in a large ongoing trial.[41] Orforglipron, an emerging non-peptide oral GLP-1RA, demonstrates potential for significant weight and HbA1c reductions, but evidence is limited to one phase 2 trial providing low certainty evidence. Although it is fair to assume similar cardiovascular and kidney benefits as observed for GLP-1 RAs as a medication class, orforglipron is now undergoing phase 3 randomised trials and has not yet received marketing approval in any countries.

Major uncertainties for all medications include benefits on complications such as severe visual impairment and neuropathy, which take many years to develop and have not been sufficiently captured in the existing trials. Similarly, we found low certainty evidence to investigate the hypothesis that GLP-1RAs may reduce dementia, and more trials are needed to investigate this potential benefit.[42]

Strengths and limitations

A key strength of our living systematic review and NMA is the comprehensive summary of best current evidence across benefits and harms for available medications for type 2 diabetes. Our wide inclusion of 13 medication classes and 26 outcomes was informed by three international teams representing guideline panels in the global perspective, reflecting a harmonised effort to better inform a global target audience and reduce duplication of efforts. Our study incorporated current and rigorous approaches to NMA and GRADE assessment, wide expert input, as well as risk-stratified evidence summaries in digestible formats for absolute effects across key cardiovascular and kidney outcomes.

This living systematic review also has limitations, some of which reflect limited trial evidence as noted above.[18] Heterogeneity of trial design and baseline characteristics of the participants may amplify the beneficial effects of some therapeutics for specific outcomes. Our sensitivity and subgroup analyses demonstrate consistent results, supporting the robustness of our findings.[43]

We also recognise limited nuance in our analysis and questions left unanswered in this first iteration of the living review. A recent NMA focusing on 15 different GLP-1RAs demonstrated molecule-specific and dose-dependent benefits and harms for weight loss and gastrointestinal side effects.[4] The extent to which these are credible subgroup effects is uncertain, and we aim to include such extended analyses in future iterations. A pertinent clinical question concerns the effectiveness of two or more therapeutics used in combination; SGLT-2 inhibitors and GLP-1 receptor agonists demonstrate improvements in cardiovascular outcomes among individuals with type 2 diabetes, and some evidence suggests that there may be additional benefit when they are used in combination.[44]

Finally, we acknowledge methodological limitations for our choices in rating certainty of evidence for imprecision and categorisation of effectiveness of medications (fig 3) using relative estimates of effect in this NMA. To consistently interpret and rate imprecision of relative effects, we adopted a minimally contextualised approach according to GRADE, and used arbitrary relative effect thresholds

for what constitutes important effects versus little or no effect.[35] To help readers navigate across a perhaps overwhelming number of drug classes and outcomes, we categorised effectiveness of medications using the same thresholds. These limitations serve as a reminder that health technology assessments (HTA) and guidelines are more suitable tools for decision support for policy and practice. The MATCH-IT tool exemplifies application of a fully contextualised approach, making use of risk-stratified absolute estimates of effect and outcome-specific minimal important differences as decision thresholds; the tool and linked practice guidelines therefore demonstrate how the findings of this NMA may be translated to real-world clinical decision-making.[45 46]

Implications for policy and practice

As reflected in largely unchanged recommendations in the linked living guidelines (see box 1), the addition of 53 new trials in this updated NMA does not necessarily call for major changes in policy or practice, if already based on best current trial evidence. However, the only way to know is through dynamic production and use of timely comparative evidence to inform HTA and guidelines, and ideally also trialists.[47] As such, this NMA exemplifies the need for global collaboration on living evidence, following initial successes through covid-19.[48] Representing the first pilot from the ALIVE consortium, this systematic review will—if funding permits—be iteratively updated to reflect latest evidence in a "living" model.[49] We hope this review can serve as a central data source to accelerate translation of evidence into policy and practice. Representatives for leading professional societies have confirmed the need for this NMA to inform their guidelines.[48] We believe it is equally relevant for HTA agencies to move to global collaboration on living evidence and comparative effectiveness, fully looped into an enhanced evidence ecosystem.[50]

Maintaining the living NMA will warrant direct bilateral collaboration with organisations making use of the updated evidence. It is also resource-demanding to dynamically update such a large NMA, and ALIVE needs funding for this purpose. Despite organisations showing considerable interest in access to the NMA, we have not yet identified organisations willing to pay enough for further development and maintenance of this living NMA. A recent call from Wellcome Trust on funding living evidence synthesis infrastructure at the global level holds promise.[51]

Conclusions

This living systematic review and NMA provides a comprehensive and up-to-date summary of the comparative effectiveness of available medications for patients with type 2 diabetes. It is designed to allow global collaboration on living evidence synthesis. As such, it holds the potential to facilitate a global living evidence ecosystem, informed decisions by policy-makers, clinicians, and patients, and reduced waste in research.[50]

AUTHOR AFFILIATIONS

[1]Department of Endocrinology and Metabolism, Laboratory of Diabetes and Metabolism Research, West China Hospital, Sichuan University, Chengdu, 610041, China

[2]MAGIC China Centre, Cochrane China Centre, Chinese Evidence-based Medicine, West China Hospital, Sichuan University, Chengdu, 610041, China

[3]Future Evidence Foundation, Melbourne, Australia

[4]School of Public Health and Preventive Medicine, Monash University, Melbourne, Australia

[5]Groningen Research Institute of Pharmacy, Faculty of Science and Engineering, University of Groningen, Groningen, The Netherlands

[6]Department of Health Research Methods, Evidence and Impact, McMaster University, ON, Canada

[7]Division of General Internal Medicine, Division of Clinical Epidemiology, University Hospitals of Geneva, Geneva, Switzerland

[8]MAGIC Evidence Ecosystem Foundation, Oslo, Norway

[9]Australian Living Evidence Collaboration (ALEC), Cochrane Australia, School of Public Health and Preventive Medicine, Monash University, Melbourne, Australia

[10]Division of General Internal Medicine, University of Alberta, Edmonton, Alberta, Canada

[11]BMJ Editorial Tavistock Square, London, UK

[12]Australian Diabetes Society, Sydney, Australia

[13]Forschergruppe Diabetes eV at the Helmholtz Centre, Munich-Neuherberg, Germany

[14]Clinic for Cardiology, Angiology, and Intensive Care Medicine, RWTH Aachen University, University Hospital Aachen, Aachen, Germany

[15]Division of Nephrology, University of Arizona College of Medicine Tucson, Tucson, AZ, USA

[16]Healthcare Innovation Research Laboratory, Nursing Key Laboratory of Sichuan Province, National Clinical Research Centre for Geriatrics, West China Hospital, Sichuan University/West China School of Nursing, Sichuan University, Chengdu, China

[17]Department of Orthopaedics, Orthopaedic Research Institute, West China Hospital, Sichuan University, Chengdu, China

[18]Center of Infectious Diseases, West China Hospital, Sichuan University, Chengdu, China

[19]Division of Infectious Diseases, State Key Laboratory of Biotherapy and Centre of Infectious Disease, West China Hospital, Sichuan University, Chengdu, China

[20]Department of Cardiology, Shandong Provincial Hospital affiliated to Shandong First Medical University, Jinan, China

[21]Department of Paediatric Surgery, West China Hospital, Sichuan University, Chengdu, China

[22]Department of Hypertension and Endocrinology, Daping Hospital, Centre for Hypertension and Metabolic Diseases, Army Medical University of PLA, Chongqing Institute of Hypertension, Chongqing, China

[23]Department of Anaesthesiology, Laboratory of Anaesthesia and Critical Care Medicine, National-Local Joint Engineering Research Centre of Translational Medicine of Anaesthesiology, Frontiers Science Centre for Disease-related Molecular Network, West China Hospital, Sichuan University, Chengdu, China

[24]Center for Evidence-Based and Translational Medicine, Zhongnan Hospital of Wuhan University, Wuhan, China

[25]Department of Epidemiology and Biostatistics, School of Public Health, Peking University Health Science Centre, Beijing, China

[26]Department of Medicine K2, Karolinska Institutet, Stockholm, Sweden

[27]Department of Medicine, Lovisenberg Diaconal Hospital, Oslo, Norway

[#]co-first authors who contribute equally to this study

[†]co-senior authors of this study

We thank Frankie Achille for developing the MATCH-IT tool for this NMA.

Contributors: KNong, BTJeppesen, and QShi are co-first authors who contribute equally to this study. SLi and POVandvik are co-senior authors of this study. SLi, POVandvik, GGuyatt, KNong, BTJeppesen, AAgarwal, TMillard, HWhite, MQuigley, SMcdonald, BWhite, TTurner, and ZZhu conceived and designed the study.

BTJeppesen, KNong, TMillard, HWhite, MQuigley, SMcdonald, BWhite, AAgarwal, POVandvik, TTurner, and SLi discussed and drafted the study protocol. TMillard, MQuigley, BWhite, YGao, XTian, and KNong screened and selected the articles. KNong, YYGao, QFan, YMao, XPan, CLi, XZou, YGan, CYuan, JAn, QLi, and YYWang extracted the data. TMillard, MQuigley, BWhite, BTJeppesen, and CZhai assessed the risk of bias of newly included trials. KNong analysed the data. YJin and FSun provided methodological consultation. KNong, QShi, SLi, GGuyatt, POVandvik, AAgarwal, and TAgoritsas rated and revised the GRADE certainty of evidence. KNong, SLi, POVandvik, GGuyatt, AAgarwal, HMacdonald, OSchnell, NMarx, FBrosiusIII, LRydén, and EStandl interpreted the results. KNong, BTJeppesen, SLi, GGuyatt, POVandvik, and AAgarwal drafted the manuscript. KNong, SLi, GGuyatt, POVandvik, AAgarwal, HMacdonald, BTJeppesen, QShi, HWhite, TTurner, OSchnell, NMarx, FBrosiusIII, LRydén, and EStandl critically revised the manuscript. All authors contributed to revising the manuscript. All authors had fully assessed to all the data in the study and had final responsibility for the decision to submit for publication. SLi, POVandvik, GGuyatt, OSchnell, NMarx, FBrosiusIII, LRydén, EStandl, and TAgoritsas supervised the study. SLi is the guarantor. The corresponding author attests that all listed authors meet authorship criteria and that no others meeting the criteria have been omitted.

Funding: This study is supported by 1.3.5 Projects for Disciplines of Excellence West China Hospital, Sichuan University (Grant No ZYYC24001). The funders had no role in considering the study design or in the collection, analysis, interpretation of data, writing of the report, or decision to submit the article for publication.

Competing interests: All authors have completed the ICMJE uniform disclosure form at https://www.icmje.org/disclosure-of-interest/ and declare: support from the West China Hospital, Sichuan University for the submitted work; KNong, BTJeppesen, QShi, TAgoritsas, GGuyatt, HWhite, YGao, AAgarwal, HMacdonald, XZou, TMillard, SMcdonald, MQuigley, XTian, QFan, BWhite, YMao, XPan, CLiu, CZhai, CYuan, QLi, JAn, YGan, YWang, YJin, FSun, ZZhu, TTurner, and POVandvik: no support from any organisation for the submitted work; no financial relationships with any organisations that might have an interest in the submitted work in the previous three years; no other relationships or activities that could appear to have influenced the submitted work. EStandl reported personal fees from Oxford Diabetes Trials Unit, Bayer, Berlin Chemie, Boehringer Ingelheim, Menarini, Merck Serono, EXCEMED, Novartis, Novo Nordisk, and Sanofi. LRydén reported grants or contracts from Swedish Heart Lung Foundation, Stockholm County Council, Erling Perssons Foundation and Boehringer-Ingelheim, and payment or honoraria for lectures, presentations, speakers bureaus, manuscript writing or educational events from Bayer AG, Boehringer Ingelheim, Novo Nordisk. FBrosiusIII reported grants or contracts from National Institutes of Health, and consulting fees from Gilead Sciences. OSchnell reported payment or honoraria for lectures, presentations, speakers bureaus, manuscript writing or educational events from Abbott Diagnostics, Lilly Deutschland, Boehringer Ingelheim, Bayer, Mannkind, and Lifescan and is a founder and CEO of Sciarc GmbH. NMarx reported grants or contracts from Boehringer Ingelheim, Merck, Novo Nordisk, Deutsche Forschungsgesellschaft (German Research Foundation; TRR 219), and consulting fees from Boehringer Ingelheim, Merck, Novo Nordisk, AstraZeneca, BMS, and payment or honoraria for lectures, presentations, speakers bureaus, manuscript writing or educational events from Boehringer Ingelheim, Merck, Novo Nordisk, Lilly, BMS, AstraZeneca. SLi received the funds from West China Hospital of Sichuan University, National Natural Science Foundation of China, and Sichuan Science and Technology Program.

Ethical approval: Not required.

Data sharing: No additional data available.

Patient consent: Not required.

Transparency: The manuscript's guarantors (SLi) affirm that the manuscript is an honest, accurate and transparent account of the study being reported; that no important aspects of the study have been omitted; and that any discrepancies from the study as planned and registered have been explained.

Dissemination to participants and related patient and public communities: Two patient partners contribute input as part of the panel for the living *BMJ* Rapid Recommendation on medications for type 2 diabetes and inform clinical questions, minimal important differences, recommendations as part of each guideline programmes. This study will support the update of a *BMJ* Rapid Recommendation (https://www.bmj.com/rapid-recommendations).

Provenance and peer review: Not commissioned; externally peer reviewed.

1 GBD 2021 Diabetes Collaborators. Global, regional, and national burden of diabetes from 1990 to 2021, with projections of prevalence to 2050: a systematic analysis for the Global Burden of Disease Study 2021. *Lancet* 2023;402:203-34. doi:10.1016/S0140-6736(23)01301-6.
2 Moucheraud C, Lenz C, Latkovic M, Wirtz VJ. The costs of diabetes treatment in low- and middle-income countries: a systematic review. *BMJ Glob Health* 2019;4:e001258. doi:10.1136/bmjgh-2018-001258.
3 Galindo RJ, Trujillo JM, Low Wang CC, McCoy RG. Advances in the management of type 2 diabetes in adults. *BMJ Med* 2023;2:e000372. doi:10.1136/bmjmed-2022-000372.
4 Yao H, Zhang A, Li D, et al. Comparative effectiveness of GLP-1 receptor agonists on glycaemic control, body weight, and lipid profile for type 2 diabetes: systematic review and network meta-analysis. *BMJ* 2024;384:e076410. doi:10.1136/bmj-2023-076410.
5 Hathaway JT, Shah MP, Hathaway DB, et al. Risk of Nonarteritic Anterior Ischemic Optic Neuropathy in Patients Prescribed Semaglutide. *JAMA Ophthalmol* 2024;142:732-9. doi:10.1001/jamaophthalmol.2024.2296.
6 Barthold D, Li J, Basu A. Patient Out-of-Pocket Costs for Type 2 Diabetes Medications When Aging Into Medicare. *JAMA Netw Open* 2024;7:e2420724. doi:10.1001/jamanetworkopen.2024.20724.
7 Djulbegovic B, Guyatt GH. Progress in evidence-based medicine: a quarter century on. *Lancet* 2017;390:415-23. doi:10.1016/S0140-6736(16)31592-6.
8 Zhou YL, Shi QY, Chen XY, Li SY, Shen BR. [Ontologies Applied in Clinical Decision Support Systems for Diabetes]. *Sichuan Da Xue Xue Bao Yi Xue Ban* 2023;54:208-16. doi:10.12182/20220860201.
9 Veroniki AA, Florez I, Hutton B, Straus SE, Tricco AC. Two decades of network meta-analysis: Roadmap to their applications and challenges. *Res Synth Methods* 2024;15:741-6. doi:10.1002/jrsm.1744.
10 Rodriguez-Valadez JM, Tahsin M, Fleischmann KE, et al. Cardiovascular and Renal Benefits of Novel Diabetes Drugs by Baseline Cardiovascular Risk: A Systematic Review, Meta-analysis, and Meta-regression. *Diabetes Care* 2023;46:1300-10. doi:10.2337/dc22-0772.
11 Elliott JH, Synnot A, Turner T, et al, Living Systematic Review Network. Living systematic review: 1. Introduction-the why, what, when, and how. *J Clin Epidemiol* 2017;91:23-30. doi:10.1016/j.jclinepi.2017.08.010.
12 Siemieniuk RA, Bartoszko JJ, Zeraatkar D, et al. Drug treatments for covid-19: living systematic review and network meta-analysis. *BMJ* 2020;370:m2980. doi:10.1136/bmj.m2980.
13 Agarwal A, Mustafa R, Manja V, et al. Cardiovascular, kidney-related, and weight loss effects of therapeutics for type 2 diabetes: a living clinical practice guideline. *BMJ* 2025;390:e082071. doi:10.1136/bmj-2024-082071.
14 Rayner D, Shah D, Dai S, et al. Prognostic models for cardiovascular and renal outcomes in patients with type 2 diabetes: a living systematic review and meta-analysis of observational studies. *BMJ Med* 2025;4:e001369.
15 Rodríguez-Gutiérrez R, et al. Values, preferences, and treatment burden for initiation of GLP-1 receptor agonists, SGLT-2 inhibitors, tirzepatide and finerenone in adult patients with type 2 diabetes: a systematic review. [Pending submission to *BMJ*].
16 Australian Evidence-Based Clinical Guidelines for Diabetes. https://app.magicapp.org/#/guideline/7844
17 Palmer SC, Tendal B, Mustafa RA, et al. Sodium-glucose cotransporter protein-2 (SGLT-2) inhibitors and glucagon-like peptide-1 (GLP-1) receptor agonists for type 2 diabetes: systematic review and network meta-analysis of randomised controlled trials. *BMJ* 2021;372:m4573. doi:10.1136/bmj.m4573.
18 Shi Q, Nong K, Vandvik PO, et al. Benefits and harms of drug treatment for type 2 diabetes: systematic review and network meta-analysis of randomised controlled trials. *BMJ* 2023;381:e074068. doi:10.1136/bmj-2022-074068.
19 Li S, Vandvik PO, Lytvyn L, et al. SGLT-2 inhibitors or GLP-1 receptor agonists for adults with type 2 diabetes: a clinical practice guideline. *BMJ* 2021;373:n1091. doi:10.1136/bmj.n1091.
20 Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ* 2021;372:n71. doi:10.1136/bmj.n71.
21 Hutton B, Salanti G, Caldwell DM, et al. The PRISMA extension statement for reporting of systematic reviews incorporating network meta-analyses of health care interventions: checklist and explanations. *Ann Intern Med* 2015;162:777-84. doi:10.7326/M14-2385.

22 Jeppesen BT, Nong K, Millard T, et al. Protocol for a living evidence dataset and systematic review of medications for people with type 2 diabetes. Alliance for Living Evidence 2024. https://data.aliveevidence.org/records/q02rv-km486
23 Guyatt G, Busse J. Methods Commentary: Risk of Bias in Randomized Trials. DistillerSR, 2021. https://www.distillersr.com/resources/methodological-resources/risk-of-bias-commentary
24 Shi Q, Wang Y, Hao Q, et al. Pharmacotherapy for adults with overweight and obesity: a systematic review and network meta-analysis of randomised controlled trials. *Lancet* 2024;403:e21-31. doi:10.1016/S0140-6736(24)00351-9.
25 Rücker G. Network meta-analysis, electrical networks and graph theory. *Res Synth Methods* 2012;3:312-24. doi:10.1002/jrsm.1058.
26 Bhaumik DK, Amatya A, Normand S-L, et al. Meta-Analysis of Rare Binary Adverse Event Data. *J Am Stat Assoc* 2012;107:555-67. doi:10.1080/01621459.2012.664484.
27 Higgins JPT, Jackson D, Barrett JK, Lu G, Ades AE, White IR. Consistency and inconsistency in network meta-analysis: concepts and models for multi-arm studies. *Res Synth Methods* 2012;3:98-110. doi:10.1002/jrsm.1044.
28 Balshem H, Helfand M, Schünemann HJ, et al. GRADE guidelines: 3. Rating the quality of evidence. *J Clin Epidemiol* 2011;64:401-6. doi:10.1016/j.jclinepi.2010.07.015.
29 Chaimani A, Salanti G. Using network meta-analysis to evaluate the existence of small-study effects in a network of interventions. *Res Synth Methods* 2012;3:161-76. doi:10.1002/jrsm.57.
30 Harbord RM, Egger M, Sterne JAC. A modified test for small-study effects in meta-analyses of controlled trials with binary endpoints. *Stat Med* 2006;25:3443-57. doi:10.1002/sim.2380.
31 Egger M, Davey Smith G, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. *BMJ* 1997;315:629-34. doi:10.1136/bmj.315.7109.629.
32 Duval S, Tweedie R. Trim and fill: A simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics* 2000;56:455-63. doi:10.1111/j.0006-341X.2000.00455.x.
33 König J, Krahn U, Binder H. Visualizing the flow of evidence in network meta-analysis and characterizing mixed treatment comparisons. *Stat Med* 2013;32:5414-29. doi:10.1002/sim.6001.
34 Zeng L, Brignardello-Petersen R, Hultcrantz M, et al. GRADE Guidance 34: update on rating imprecision using a minimally contextualized approach. *J Clin Epidemiol* 2022;150:216-24. doi:10.1016/j.jclinepi.2022.07.014.
35 Brignardello-Petersen R, Florez ID, Izcovich A, et al, GRADE working group. GRADE approach to drawing conclusions from a network meta-analysis using a minimally contextualised framework. *BMJ* 2020;371:m3900. doi:10.1136/bmj.m3900.
36 Phillips MR, Sadeghirad B, Busse JW, et al. Development and design validation of a novel network meta-analysis presentation tool for multiple outcomes: a qualitative descriptive study. *BMJ Open* 2022;12:e056400. doi:10.1136/bmjopen-2021-056400.
37 Van der Vaart AW. *Asymptotic statistics*. Vol 3. Cambridge university press, 2000.
38 Schandelmaier S, Briel M, Varadhan R, et al. Development of the Instrument to assess the Credibility of Effect Modification Analyses (ICEMAN) in randomized controlled trials and meta-analyses. *CMAJ* 2020;192:E901-6. doi:10.1503/cmaj.200077.
39 Wiksten A, Hawkins N, Piepho H-P, Gsteiger S. Nonproportional Hazards in Network Meta-Analysis: Efficient Strategies for Model Building and Analysis. *Value Health* 2020;23:918-27. doi:10.1016/j.jval.2020.03.010.
40 Efthimiou O, Rücker G, Schwarzer G, Higgins JPT, Egger M, Salanti G. Network meta-analysis of rare events using the Mantel-Haenszel method. *Stat Med* 2019;38:2992-3012. doi:10.1002/sim.8158.
41 Nicholls SJ, Bhatt DL, Buse JB, et al, SURPASS-CVOT investigators. Comparison of tirzepatide and dulaglutide on major adverse cardiovascular events in participants with type 2 diabetes and atherosclerotic cardiovascular disease: SURPASS-CVOT design and baseline characteristics. *Am Heart J* 2024;267:1-11. doi:10.1016/j.ahj.2023.09.007.
42 Tang B, Sjölander A, Wastesson JW, et al. Comparative effectiveness of glucagon-like peptide-1 agonists, dipeptidyl peptidase-4 inhibitors, and sulfonylureas on the risk of dementia in older individuals with type 2 diabetes in Sweden: an emulated trial study. *EClinicalMedicine* 2024;73:102689. doi:10.1016/j.eclinm.2024.102689.
43 Guyatt G, Zhao Y, Mayer M, et al. GRADE guidance 36: updates to GRADE's approach to addressing inconsistency. *J Clin Epidemiol* 2023;158:70-83. doi:10.1016/j.jclinepi.2023.03.003.
44 Packer M. Should We Be Combining GLP-1 Receptor Agonists and SGLT2 Inhibitors in Treating Diabetes?*Am J Med* 2018;131:461-3. doi:10.1016/j.amjmed.2017.11.052.
45 Løvsletten PO, Hunskaar BS, Heen AF, et al. Physicians found an interactive tool displaying structured evidence summaries for multiple comparisons understandable and useful: a qualitative user testing study. *J Clin Epidemiol* 2024;172:111399. doi:10.1016/j.jclinepi.2024.111399.
46 Hunskaar B, Løvsletten P, Achille F, et al. MATCH-IT: A decision support tool for multiple comparisons presenting data from network meta-analysis to facilitate guideline development. *Clin Public Health Guide* 2024;1. doi:10.1002/gin2.70003.
47 Naci H, Kesselheim AS, Røttingen J-A, Salanti G, Vandvik PO, Cipriani A. Producing and using timely comparative evidence on drugs: lessons from clinical trials for covid-19. *BMJ* 2020;371:m3869. doi:10.1136/bmj.m3869.
48 Ceriello A, Rodbard HW, Battelino T, et al, Taskforce of the Guideline Workshop. Data from network meta-analyses can inform clinical practice guidelines and decision-making in diabetes management: perspectives of the taskforce of the guideline workshop. *Cardiovasc Diabetol* 2023;22:277. doi:10.1186/s12933-023-01993-3.
49 Vandvik PO, Brignardello-Petersen R, Guyatt GH. Living cumulative network meta-analysis to reduce waste in research: A paradigmatic shift for systematic reviews?*BMC Med* 2016;14:59. doi:10.1186/s12916-016-0596-4.
50 Vandvik PO, Brandt L. Future of Evidence Ecosystem Series: Evidence ecosystems and learning health systems: why bother?*J Clin Epidemiol* 2020;123:166-70. doi:10.1016/j.jclinepi.2020.02.008.
51 Wellcome. Evidence Synthesis Infrastructure Collaborative. 2024. https://wellcome.org/news/evidence-synthesis-infrastructure-collaborative

Web appendix: Supplementary appendix