Comparative effects of drugs for adults with overweight or obesity: systematic review and network meta-analysis

Kailei Nong,[1,2] Qingyang Shi,[3] Xinran Xie,[1,2] Yang Wang,[4] Arnav Agarwal,[5,6,7] Gordon H Guyatt,[6,7] Haojie Zhang,[1,2] Yiyuan Gao,[1,2] Kamlesh Khunti,[8] Carel W Le Roux,[9] Indah S Widyahening,[10] Qinlin Fan,[1,2] Tianyou Liu,[11,12] Yunhe Mao,[13] Ivan D Florez,[14,15,16] Heyue Du,[17] Xiaohui Pan,[1] Xinyu Zou,[1,2] Chaoyang Wang,[18] Xiaodong Sun,[19] Jing Li,[1] Qiukui Hao,[7,20] Qingyi Jia,[1] Feng Sun,[21] Zhiming Zhu,[22] Thomas Agoritsas,[6,7,23] Haoming Tian,[1] Per Olav Vandvik,[6,24] Sheyu Li[1,2,6]

For numbered affiliations see end of the article

Correspondence to: S Li
lisheyu@scu.edu.cn
(or @LisheyuSheyu on X;
ORCID 0000-0003-0060-0287)

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

Cite this as: *BMJ* 2026;394:e372161
http://dx.doi.org/10.1136/bmj-2026-372161

Accepted: 28 May 2026

ABSTRACT

OBJECTIVE

To provide an up-to-date evidence summary about the comparative benefits and harms of drugs for adults with overweight or obesity to inform decision making for policymakers, payers, clinicians, and patients.

DESIGN

Systematic review and network meta-analysis of 24 outcomes using frequentist random effects models and bayesian dose-response models, the GRADE (Grading of Recommendations Assessment, Development, and Evaluation) approach, and the Cochrane Risk of Bias 2 tool.

DATA SOURCES

Medline, Embase, and Cochrane Library, searched up to 12 November 2025.

STUDY SELECTION

Randomised controlled trials of 12 weeks' duration or longer comparing one or more drugs with lifestyle modification, placebo, or another drug.

RESULTS

This network meta-analysis comprised 262 trials (99 791 participants) evaluating 19 drugs with follow-up from 12 to 172 weeks. Compared with lifestyle modification alone, at one year, moderate to high certainty evidence shows substantial weight loss with tirzepatide (mean difference -14.9%, 95% confidence interval -16.0% to -13.9%), cagrilintide-semaglutide (CagriSema, -14.8%, -16.9% to -12.7%), oral semaglutide (-10.9%, -12.7% to -9.1%), orforglipron (-9.9%, -12.4% to -7.5%), subcutaneous semaglutide (-9.8%, -10.6% to -9.1%), and phentermine-topiramate (-8.1%, -9.7% to -6.5%). Emerging agents (ecnoglutide, mazdutide, retatrutide) may produce similar or greater reductions (13.1-14.6%; very low to low certainty). Moderate to high certainty evidence supports discontinuation because of adverse events to be highest with orforglipron, naltrexone-bupropion, liraglutide, phentermine-topiramate, CagriSema, and oral semaglutide (risk ratios from 1.9 to 4.2); gastrointestinal events were most increased with naltrexone-bupropion, oral semaglutide, orforglipron, and tirzepatide (risk ratios from 3.1 to 4.2). Fatigue risk increased, particularly with naltrexone-bupropion (risk ratio 8.9; absolute increase 331 per 1000 people over one year), orforglipron (3.4; 100 more per 1000), and CagriSema (3.2; 92 more per 1000). Tirzepatide reduced fat mass the most (by 25.7%) but also lean mass the most (by 8.3%). Subcutaneous semaglutide was the only drug associated with reduced all cause mortality (risk ratio 0.81, 95% confidence interval 0.72 to 0.93) and myocardial infarction (0.72, 0.61 to 0.85), with these estimates largely informed by cardiovascular outcome trials in high risk populations. Subcutaneous semaglutide (0.43, 0.21 to 0.84) and tirzepatide (0.49, 0.27 to 0.88) reduced heart failure risk. No drugs convincingly reduced kidney failure or improved quality of life (43 trials with 45 663 participants) beyond established minimally important differences (all mean differences ‹5 points; minimally important difference 10). Except for larger weight reductions in trials with longer duration (shown for subcutaneous semaglutide), subgroup analyses for drug dosages and key patient characteristics did not identify credible differences in relative effects of treatment.

CONCLUSIONS

Obesity drugs produce variable weight loss at one year, with larger benefits generally accompanied by greater harms and discontinuation. Most agents do not improve quality of life meaningfully and few show cardiovascular benefits. Decisions in clinical practice should consider trade-offs between benefits and harms within the context of shared decision making.

SYSTEMATIC REVIEW REGISTRATION

PROSPERO CRD42024507993.

WHAT IS ALREADY KNOWN ON THIS TOPIC

Several drugs for adults with overweight or obesity produce substantial weight loss, but most have not been compared directly in head-to-head trials

Previous network meta-analyses have focused mainly on approved drugs and weight loss outcomes, leaving uncertainty about the broader balance of benefits and harms

WHAT THIS STUDY ADDS

Tirzepatide and cagrilintide-semaglutide (CagriSema) achieved (with moderate to high certainty) the greatest weight loss, but larger benefits are generally accompanied by more discontinuations, gastrointestinal symptoms, fatigue, and lean mass loss

Subcutaneous semaglutide is the only drug with evidence of reduced all cause mortality, myocardial infarction, and heart failure (also shown for tirzepatide) with no drug convincingly reducing kidney failure

Despite weight reduction, none of the drugs produced improvements in quality of life at one year beyond established minimally important difference thresholds

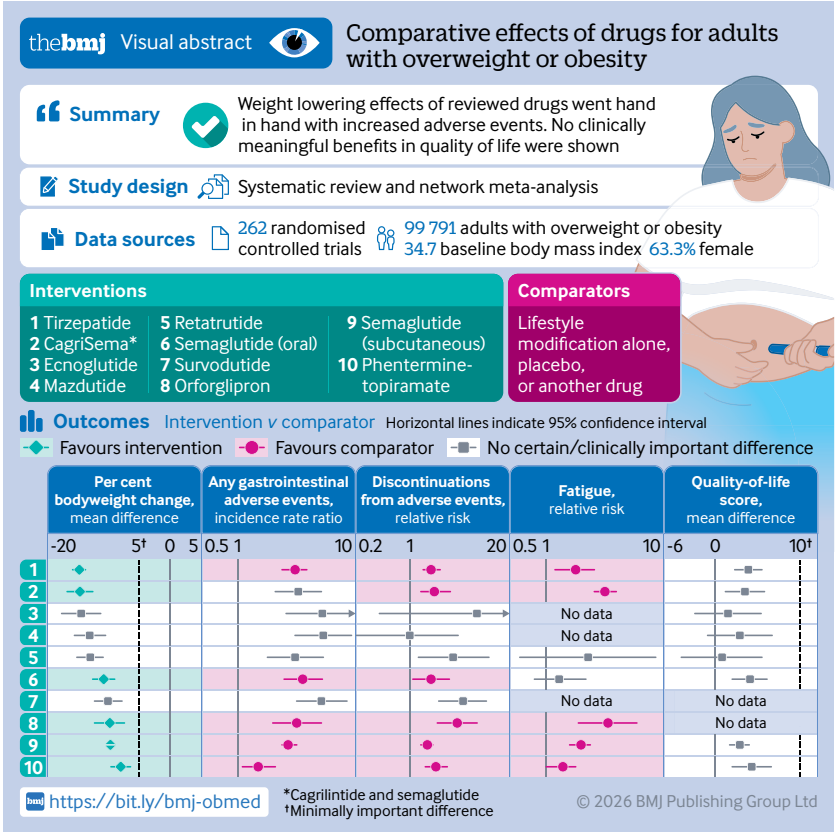
Introduction

Obesity affects more than 890 million adults worldwide and contributes to approximately 4 million

deaths each year.[1] The treatment landscape has changed rapidly with the introduction of highly effective pharmacotherapies, including glucagon-like peptide-1 (GLP-1) receptor agonists and newer multi-agonists such as tirzepatide.[2] Uptake has been substantial, accompanied by sharply rising healthcare expenditures. Additional agents, including the fixed combination cagrilintide-semaglutide (CagriSema) and the GLP-1/glucagon receptor agonist mazdutide, have shown promising results in phase 3 trials and are under regulatory consideration.[3-7]

Obesity is increasingly recognised as a complex chronic disease and reliance on weight loss alone as a measure of treatment success may oversimplify benefits and harms while reinforcing stigma.[8] Patients and other stakeholders have emphasised the importance of outcomes beyond weight, including mortality, cardiovascular and kidney events, quality of life, treatment burden, and functional status, outcomes that are also central to health technology assessment and reimbursement decisions.[8 9] Potential harms such as lean mass loss and treatment discontinuation are also receiving greater attention.[9] However, the comparative benefits and risks of available and emerging drugs across this broader set of outcomes remain uncertain, particularly given the high costs and long term use of these treatments.

Direct head-to-head trials between most drugs are lacking, leaving clinicians, patients, policymakers, and payers without clear evidence to guide treatment selection and reimbursement decisions, as exemplified by recent authoritative clinical practice guidelines.[10 11] Network meta-analysis enables simultaneous comparison of several interventions, including those not directly compared in trials. While several network meta-analyses have reported on the comparative effectiveness of obesity drugs, a contemporary overview including transparent assessment of certainty was lacking.[12 13] Therefore, we conducted a systematic review and network meta-analysis to compare the benefits and harms of all available obesity drugs in adults with overweight or obesity across a comprehensive set of patient important outcomes using a GRADE (Grading of Recommendations Assessment, Development, and Evaluation) based framework.



Methods

This systematic review was conducted as part of the *BMJ Rapid Recommendations* initiative.[14] This collaboration between *The BMJ* and the MAGIC Evidence Ecosystem Foundation aims to respond to complex and practice changing evidence with trustworthy, accessible, and actionable recommendations for global use.[15] The network meta-analysis results will inform a global guideline panel, including patient partners, multidisciplinary clinicians, and methodologists without conflict of interest, in moving from evidence to risk stratified recommendations, as done in living guidelines for diabetes type 2 drugs supported by a living network meta-analysis.[16 17]

This systematic review followed Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA) 2020 and PRISMA-NMA (network meta-analysis) reporting guidelines (appendix 1) and was registered in PROSPERO (CRD42024507993).[18 19]

Data sources and searches

We searched Ovid Medline, Embase, and Cochrane Library (Central) from inception to 12 November 2025. Search terms, informed by our previous systematic review and network meta-analysis,[12] incorporated newly available drugs including triple glucose dependent insulinotropic polypeptide (GIP)/GLP-1/glucagon receptor agonists, dual GIP/GLP-1 receptor agonists, and dual GIP/glucagon receptor agonists (appendix 2). We cross checked and supplemented the electronic database search by screening reference lists from other systematic reviews and meta-analyses.

Eligibility criteria and study selection

Eligible randomised controlled trials enrolled adults with overweight or obesity, with or without cardiovascular or metabolic complications; compared a candidate obesity drug (appendix 3.1) with lifestyle modification alone, placebo, or an alternative candidate drug; and involved a drug treatment duration of at least 12 weeks (chosen as the shortest duration in which the drugs are most likely to result in important weight reduction). Eligible trials reported one or more prespecified outcomes of interest (appendix 3.2). We excluded trials with crossover or quasi-randomised designs; and those evaluating multidrug combinations

doi: 10.1136/bmj-2026-372161 | *BMJ* 2026;394:e372161 | thebmj

that are not formulated as a fixed dose combination (ie, phentermine-topiramate, bupropion-naltrexone, and CagriSema remained eligible); enrolled participants primarily on the basis of psychiatric diagnoses (eg, schizophrenia, depression, and eating disorders) or pregnancy; that randomised participants after weight reduction periods; or that were published in a language other than English.

Paired reviewers (KN, YW, HZ, XX, QF, TL, YG, and HD) independently screened identified titles and abstracts, and subsequently full texts of potentially eligible articles. Reviewers resolved discrepancies through discussion or by consultation with a senior reviewer (SL).

Data collection and outcomes

For each eligible study, paired reviewers (KN, YW, HZ, YM, XP, XZ, XX, QF, CW, and QJ) used a standardised extraction form to independently extract data of study characteristics, participants, and interventions and outcomes (appendix 3.3). A senior reviewer (SL) resolved discrepancies. We assessed benefit outcomes (changes in percentage body weight, percentage fat mass, quality-of-life score, and absolute waist circumference; number of participants achieving body weight reductions of at least 5%, 10%, 15%, and 20%) and safety outcomes (change in percentage lean mass; number with death from any cause, myocardial infarctions, stroke, heart failure, hospital admission for heart failure, kidney failure, kidney disease progression, suicide events, pancreatitis, thyroid cancer, gallbladder related disorders, erectile dysfunction, any gastrointestinal adverse event, fatigue, and discontinuation because of adverse events). Appendix 3.2 presents definitions of outcomes of interest. When absent, we estimated the percentage body weight change and proportion of participants with at least 5%, 10%, 15%, and 20% weight loss using the baseline body weight, the absolute change, or the end of follow-up measure, and assuming a normal distribution (details in appendix 3.4).

Risk of bias within individual studies

Paired reviewers (YW, YG, XX, and QF) independently assessed the risk of bias for each included study at the outcome level using the revised Cochrane risk-of-bias tool for randomised trials (RoB-2).[20] Judgments were made in five domains addressing the randomisation process, deviations from intended interventions, missing outcome data, outcome measurement, and selection of reported results. Each domain contains several questions that lead to a final judgment: "low risk of bias," "some concerns," or "high risk of bias." The domain specific judgments informed an overall risk of bias assessment for each trial. A third reviewer (KN) checked the paired assessments and summarised the final results. Discrepancies were resolved by a senior reviewer (SL).

Data synthesis and analysis

We reported results for binary outcomes of first events such as all cause death and cardiovascular events as risk ratios and those of recurrence in which participants may have more than one event as incidence rate ratios. We reported results for continuous outcomes as mean differences with 95% confidence intervals (CIs). We pooled quality-of-life score changes using standardised mean differences, then transformed these back to mean differences by multiplying them by the standard deviation for a specific questionnaire (ie, 36-Item Short Form Survey, which is the most commonly used scale across the included studies).[21]

We performed a frequentist network meta-analysis for the outcomes without a hypothesised dose-response relation and a bayesian partial dose-response network meta-regression for the outcomes with a hypothesised dose-response relation. We determined the dose-response outcomes including percentage body weight change, absolute waist circumference change, number of participants achieving body weight reductions of at least 5%, 10%, 15%, and 20%, discontinuation because of adverse events, and any gastrointestinal adverse event.

The frequentist network meta-analysis was conducted as a random effect network meta-analysis using restricted maximum likelihood estimators for between study variance.[22] The treatment nodes in the network represent specific drugs, except for sodium-glucose cotransporter 2 (SGLT-2) inhibitors, which are represented as a drug class. Global heterogeneity was estimated using the generalised method of moments and assessed using the design based decomposition of Cochran's Q statistic.[23] Indirect estimates were obtained using node splitting and back calculation methods.[24] Comparison adjusted funnel plots were used to assess global publication bias.[25] We assessed intransitivity by comparing the distributions of patients' baseline characteristics across direct comparisons, including mean age, mean body mass index, female proportion, proportion of type 2 diabetes, and follow-up duration.[26]

A bayesian partial dose-response network meta-regression was conducted using a two step approach.[27 28] A ratio based dose analysis was used to determine the dose response. We chose the approved maximal dose as the standard or reference dose for those approved drugs and the maximal dose in phase 3 trials for unapproved drugs. Other doses were calculated as ratios to standard dose. In the first step, we used a series of pairwise dose-response meta-regressions for all direct comparisons to determine the drugs with a significant dose-response relation using a Knapp-Hartung test. In the second step, we conducted a partial dose-response network meta-regression with all drugs. For drugs with a dose-response relation at the first step, we modelled them as a log linear regression on dose ratios; for others without a dose-response relation, we modelled them as a standard random effect network meta-analysis in the same network meta-regression model. We derived the final results based on the standard dose for all drugs.

Effect modifiers and additional analyses

We conducted three prespecified subgroup analyses based on the baseline severity of obesity (overweight with body mass index of 25-29.9 *v* mild obesity with body mass index of 30-34.9 *v* moderate to severe obesity with body mass index ≥35),[29] type 2 diabetes status (with *v* without), and body composition assessment methods (computed tomography or magnetic resonance imaging *v* dual energy x ray absorptiometry *v* bioelectrical impedance analysis). We conducted two post hoc subgroup analyses for the presence of baseline cardiovascular disease (with *v* without) and timeframe of trial follow-up (12-26 weeks *v* 27-52 weeks *v* >52 weeks). We hypothesised larger treatment effects in those with moderate to severe obesity, without previous diabetes, with previous cardiovascular disease, with longer trial durations, and when body composition was assessed by computed tomography or magnetic resonance imaging. We assessed the credibility of subgroup effects using the Instrument to assess the Credibility of Effect Modification Analyses (ICEMAN) tool.[30] In the absence of evident moderate to highly credible subgroup effects, we inferred consistent relative effects and pooled results across populations.

We conducted seven sensitivity analyses for all outcomes, including a random effect bayesian network meta-analysis[31]; excluding studies with fewer than 100 participants; excluding studies with overall high risk of bias; excluding unblinded trials; excluding studies with treatment duration less than 24 weeks; restricting data to the 52 week (±4 week) assessment; and including only drugs approved by the United States Food and Drug Administration. Additional analyses excluded studies that did not report percentage body weight change from baseline to end of follow-up, excluded studies that used the cumulative probability distribution to impute missing outcomes for achieving at least 5%, 10%, 15%, and 20% weight loss, and compared quality-of-life results from different scales. A fixed effect network meta-analysis for sparse data was applied to all cause death, myocardial infarction, and stroke.[32]

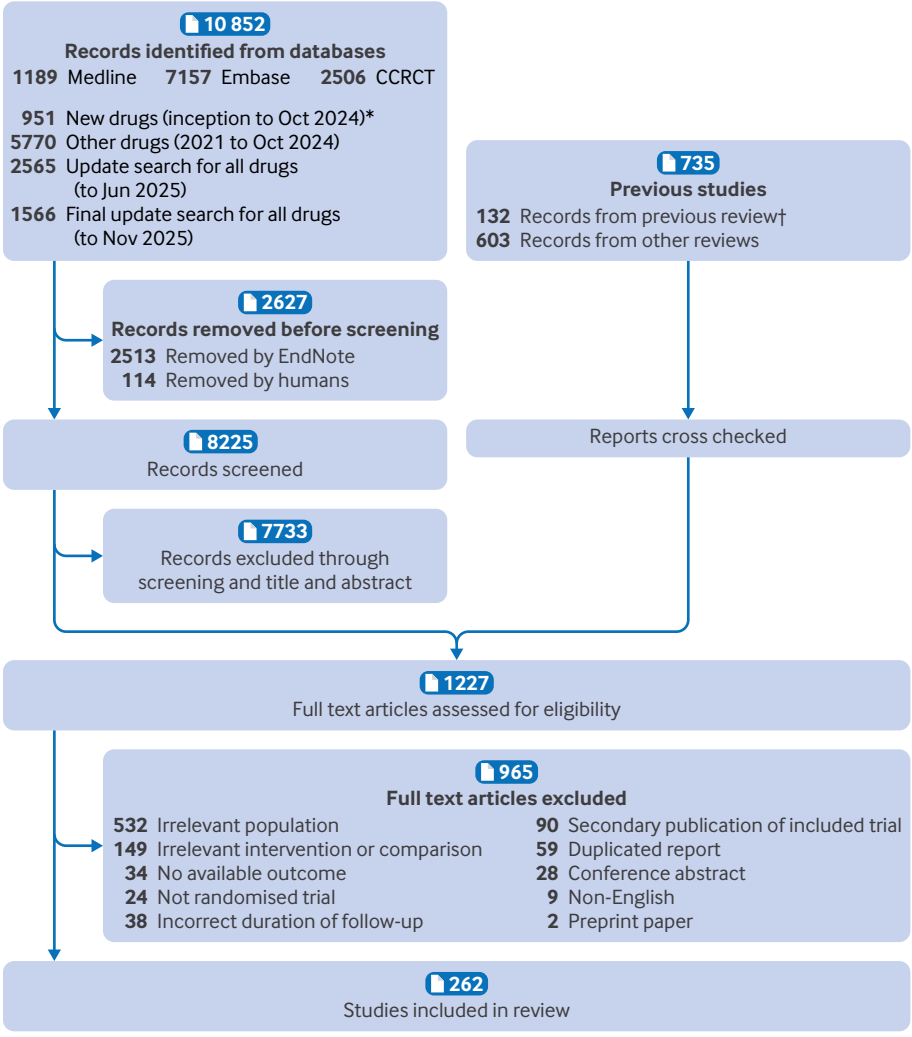


Fig 1 | Flow diagram for trial screen and selection. *New drugs include triple glucose dependent insulinotropic polypeptide (GIP)/glucagon-like peptide-1 (GLP-1)/glucagon receptor agonists, dual GIP/GLP-1 receptor agonists, and dual GIP/glucagon receptor agonists. †Previous review refers to Shi and colleagues.[12] CCRCT=Cochrane Central Register of Controlled Trials

GRADE certainty of evidence assessment

We followed the GRADE framework to rate the certainty of evidence.[33 34] Overall certainty was rated as high, moderate, low, or very low. For emerging drugs supported by limited trial evidence, we typically rated down to low certainty evidence owing to indirectness (selected populations in trials and lack of use in clinical practice). Appendix 3.9 presents details of GRADE approaches. A minimally contextualised framework was adopted so that the magnitude of drug effects and certainty of evidence informed classifications of interventions into categories from among the best to among the worst.[35] In general, we classified interventions based on whether differences in outcomes were greater than prespecified minimally important differences. Lifestyle modification alone served as the reference standard (comparator) for these judgments.

We estimated the baseline risk by pooling the effect of the control arms across included trials with clearly stated lifestyle modification only and then calculated absolute effects by multiplying measures of relative effects (eg, risk ratios with 95% CIs) with estimates of baseline risks. The timeframe for estimating absolute effects was chosen to be one year, given existing trials with limited follow-up duration. Appendix 3 presents details of the methods.

For rating imprecision of absolute effects, we derived minimally important differences from available studies as follows: 5% for percentage body weight change, percentage fat mass change, and percentage lean mass change[36]; 278, 140, 45, and 17 per 1000 people over one year for the proportion of participants achieving at least 5%, 10%, 15%, and 20% weight loss, respectively (ie, twice the proportion reaching the weight loss target with lifestyle modification alone); 10 point change from baseline in 36-Item Short Form Survey score[37]; 4 cm change from baseline in absolute waist circumference.[38]

In the absence of empirical evidence to inform our judgments, we pragmatically defined estimates of minimally important differences to rate imprecision as follows: two per 1000 people over one year for all cause death, myocardial infarction, stroke, heart failure, hospital admission for heart failure, kidney failure, kidney disease progression, pancreatitis, suicide, and thyroid cancer; four per 1000 people over one year for discontinuations because of adverse events, erectile dysfunction, and gallbladder related disorders; and eight per 1000 people over one year for fatigue and any gastrointestinal adverse events.

Drugs that were superior to (or inferior to) lifestyle modification alone (ie, point estimate exceeding or falling below the decision threshold) were categorised into the most effective or harmful group. Drugs among the most effective or harmful that are inferior to (or superior to) another drug with high to moderate certainty evidence (ie, 95% CI falling below or

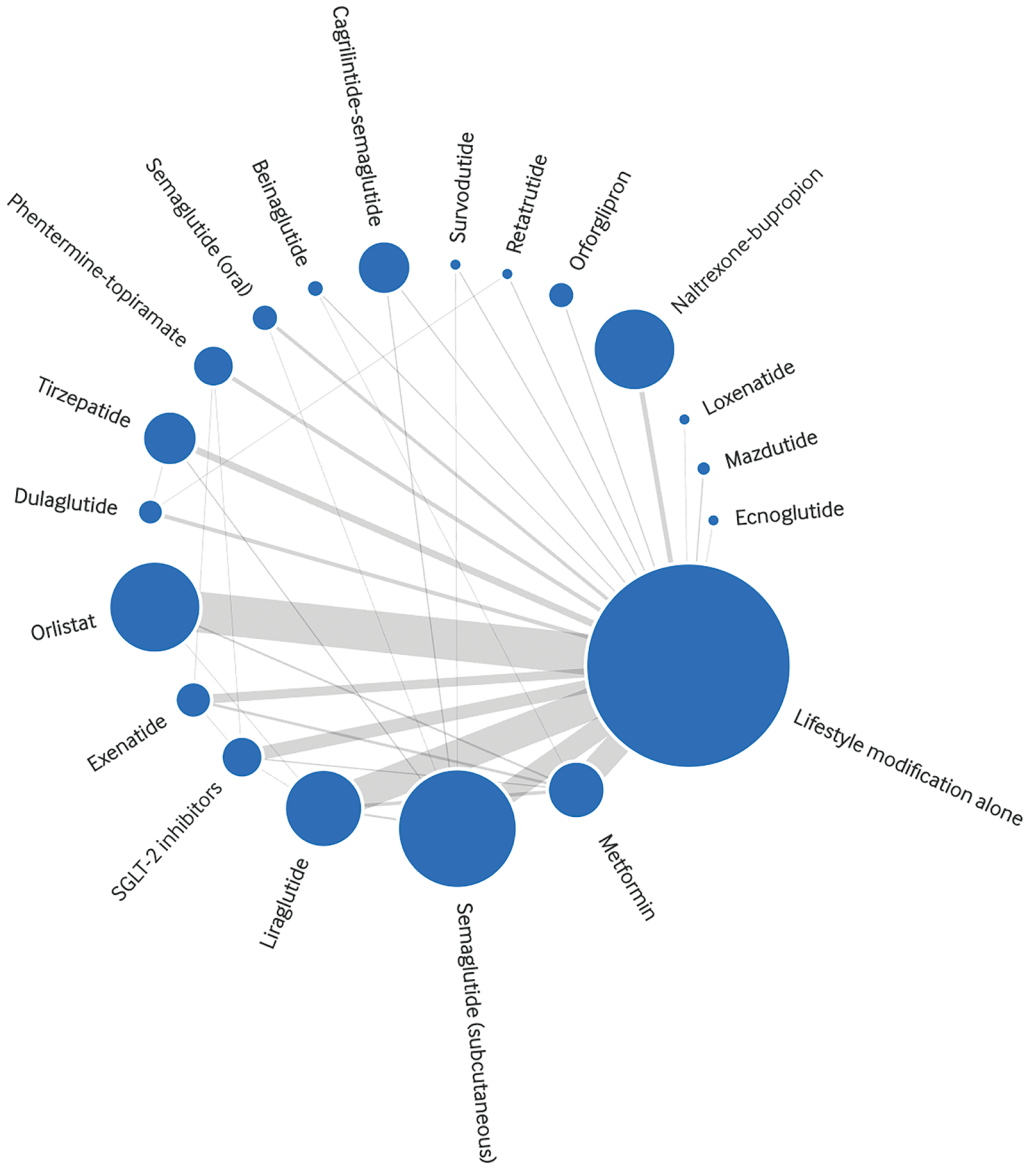
Table 1 | Summary of baseline characteristics of included studies and their participants

Baseline characteristics	Number (%)*
Study settings (of all eligible studies)	
Total No of trials	262
No of participants	99 791
Treatment duration (weeks), median (IQR; range)	26 (24-52; 12-148)
Follow-up (weeks), median (IQR; range)	26 (24-52; 12-172)
Studies involving participants with weight related complications (of all eligible studies)	
Type 2 diabetes†	95 (36.3)
Cardiovascular disease‡	2 (0.8)
Polycystic ovary syndrome‡	20 (7.6)
Hypertension‡	16 (6.1)
Dyslipidemia‡	10 (3.8)
Metabolic dysfunction associated steatotic liver disease‡	6 (2.3)
Obstructive sleep apnea‡	4 (1.5)
Knee osteoarthritis‡	2 (0.8)
Study characteristics (of all participants)	
Age (years), median (IQR; range)	49.0 (43.0-55.0; 23.5-69.3)
Proportion female (%), median (IQR; range)	63.3 (48.0-84.0; 8.3-100.0)
Body mass index, median (IQR; range)	34.7 (32.1-37.1; 26.1-43.9)
Body weight (kg), median (IQR; range)	96.8 (89.0-103.0; 63.9-128.0)
Settings and characteristics of studies involving participants with cardiovascular disease	
No of trials	2
No of participants	26 509
Duration of treatment (weeks), pooled mean (95% CI)¶	72.5 (31.6 to 166.2)
Duration of follow-up (weeks), pooled mean (95% CI)¶	144.3 (118.2 to 166.0)
Age (years), pooled mean (95% CI)¶	61.3 (61.0 to 61.6)
Mean proportion female (%)	37.7
Body mass index, pooled mean (95% CI)¶	35.0 (33.1 to 36.9)
Body weight (kg), pooled mean (95% CI)¶	101.2 (96.1 to 106.6)

CI=confidence interval; IQR=interquartile range.
*Unless stated otherwise.
†Studies classified as involving participants with type 2 diabetes if baseline proportion of participants with type 2 diabetes was ≥60%.
‡Studies classified as involving participants with other weight related complications based on study's reported inclusion criteria.
¶Pooled mean estimated using single arm meta-analyses with a random effect model.

Network plot for all included studies

Network plot shows all potential comparisons that provide comparative evidence for at least one outcome. Consists of treatment nodes with node size proportional to sample size and comparison edges with line thickness proportional to number of trials. Treatment nodes represent specific drugs, except for sodium-glucose cotransporter 2 (SGLT-2) inhibitors, which are represented as drug class



Article DOI: 10.1136/bmj-2026-372161 • Download data

Fig 2 | Network plot for all included studies. Network plots for specific outcomes are shown in appendix 6.3. An interactive version of this graphic and downloadable data are available at https://public.flourish.studio/visualisation/29292919/

Intervention	Percentage body weight change MD (95%CI)	Absolute waist circumference change MD (95%CI)	Participants with body weight reduction ≥20% RR (95%CI)	Quality-of-life score MD (95%CI)	Percentage fat mass change MD (95%CI)	Percentage lean mass change MD (95%CI)	Discontinuations because of any adverse events RR (95%CI)	Any gastrointestinal adverse events IRR (95%CI)	Gallbladder related disorders RR (95%CI)	Fatigue RR (95%CI)
Tirzepatide	-14.9 (-16.0 to -13.9)	-11.0 (-12.2 to -9.7)	16.8 (11.4 to 25.0)	3.9 (2.3 to 5.5)	-25.7 (-30.8 to -20.6)	-8.3 (-12.9 to -3.7)	1.9 (1.5 to 2.5)	3.1 (2.4 to 3.9)	1.6 (0.9 to 2.8)	1.8 (1.2 to 2.6)
CagriSema	-14.8 (-16.9 to -12.7)	-10.7 (-13.1 to -8.4)	25.2 (13.0 to 49.8)	3.5 (1.2 to 5.8)	-28.7 (-33.7 to -23.6)	-10.5 (-15.4 to -5.6)	2.1 (1.4 to 3.4)	3.3 (2.1 to 5.2)	2.6 (1.6 to 4.3)	3.2 (2.6 to 4.0)
Ecnoglutide	-14.6 (-17.8 to -11.4)	-8.2 (-11.1 to -5.4)	81.5 (5.0 to 1370.1)	1.5 (-2.4 to 5.4)	-	-	7.6 (0.4 to 132.6)	5.3 (2.6 to 10.8)	1.0 (0.0 to 49.5)	-
Mazdutide	-13.2 (-15.7 to -10.6)	-7.4 (-9.6 to -5.3)	24.6 (5.2 to 122.4)	2.9 (-0.9 to 6.7)	-27.1 (-34.0 to -20.2)	-11.6 (-16.5 to -6.6)	1.0 (0.2 to 4.3)	5.4 (3.1 to 9.4)	3.0 (0.3 to 29.0)	-
Retatrutide	-13.1 (-15.3 to -11.0)	-14.4 (-17.4 to -11.5)	14.2 (4.2 to 50.0)	0.8 (-4.0 to 5.5)	-	-	3.7 (1.3 to 10.9)	3.1 (1.8 to 5.4)	-	2.3 (0.6 to 8.7)
Semaglutide (oral)	-10.9 (-12.7 to -9.1)	-7.8 (-9.8 to -6.0)	10.7 (5.1 to 21.9)	4.1 (2.0 to 6.1)	-16.8 (-23.9 to -9.8)	-3.3 (-9.2 to 2.6)	1.9 (1.1 to 3.3)	3.6 (2.5 to 5.3)	2.5 (1.0 to 6.3)	1.3 (0.8 to 2.2)
Survodutide	-10.2 (-12.4 to -7.9)	-9.8 (-12.6 to -7.1)	14.5 (2.6 to 77.9)	-	-	-	5.0 (2.4 to 10.1)	5.2 (3.2 to 8.6)	0.5 (0.0 to 11.7)	-
Orforglipron	-9.9 (-12.4 to -7.5)	-7.3 (-9.8 to -4.8)	6.9 (3.1 to 14.9)	-	-18.3 (-27.5 to -9.1)	-6.9 (-12.0 to -1.8)	4.2 (2.3 to 7.6)	3.2 (2.0 to 5.2)	2.0 (0.6 to 6.5)	3.4 (1.9 to 6.0)
Semaglutide (subcutaneous)	-9.8 (-10.6 to -9.1)	-7.8 (-8.6 to -6.9)	10.9 (8.2 to 14.2)	2.9 (1.7 to 4.0)	-17.9 (-21.1 to -14.6)	-5.8 (-8.7 to -2.9)	1.7 (1.4 to 2.0)	2.7 (2.4 to 3.2)	1.3 (1.1 to 1.5)	2.0 (1.6 to 2.4)
Phentermine-topiramate	-8.1 (-9.7 to -6.5)	-6.3 (-8.8 to -3.9)	5.9 (3.3 to 10.5)	4.3 (2.0 to 6.6)	-12.8 (-18.8 to -6.8)	-3.1 (-8.8 to 2.5)	2.2 (1.6 to 3.1)	1.5 (1.1 to 2.1)	-	1.4 (1.0 to 1.8)
Liraglutide	-4.7 (-5.4 to -4.0)	-3.2 (-4.1 to -2.3)	2.8 (1.7 to 4.3)	2.1 (0.6 to 3.7)	-8.6 (-11.2 to -6.1)	-1.6 (-3.4 to 0.2)	2.2 (1.8 to 2.9)	2.9 (2.5 to 3.5)	2.1 (1.2 to 3.6)	1.6 (1.3 to 2.0)
Beinaglutide	-4.4 (-6.5 to -2.3)	-2.4 (-4.4 to -0.4)	3.9 (0.3 to 47.0)	-	-11.0 (-16.4 to -5.5)	-2.1 (-7.2 to 3.0)	3.8 (1.1 to 13.2)	2.0 (1.2 to 3.4)	-	2.6 (0.9 to 7.6)
Naltrexone-bupropion	-4.0 (-5.3 to -2.7)	-3.2 (-4.6 to -1.8)	3.9 (1.8 to 8.4)	3.6 (1.8 to 5.4)	-	-	2.3 (1.8 to 2.9)	4.2 (3.0 to 5.8)	1.7 (0.1 to 35.5)	8.9 (2.1 to 37.4)
Loxenatide	-3.8 (-7.3 to -0.4)	-	-	-	-	-	-	1.6 (0.2 to 15.9)	-	-
Orlistat	-3.0 (-3.5 to -2.5)	-2.1 (-2.8 to -1.4)	2.2 (1.4 to 3.4)	1.5 (-2.4 to 5.4)	-3.0 (-4.9 to -1.0)	1.4 (-0.9 to 3.6)	1.6 (1.4 to 1.9)	2.2 (1.9 to 2.5)	0.7 (0.3 to 1.6)	0.2 (0.0 to 1.2)
Exenatide	-2.6 (-3.5 to -1.8)	-2.5 (-4.2 to -0.8)	3.8 (0.1 to 132.2)	-	-5.8 (-11.8 to 0.1)	-2.0 (-7.7 to 3.7)	1.8 (1.2 to 2.7)	2.5 (1.9 to 3.3)	-	0.5 (0.0 to 7.3)
SGLT-2 inhibitors	-2.6 (-3.4 to -1.7)	-1.5 (-2.9 to -0.1)	2.6 (0.3 to 21.0)	2.9 (0.1 to 5.6)	-4.4 (-6.8 to -2.1)	-2.2 (-4.9 to 0.5)	1.3 (0.8 to 2.0)	0.8 (0.5 to 1.1)	-	-
Dulaglutide	-2.4 (-4.2 to -0.5)	-0.8 (-8.1 to 6.3)	0.4 (0.0 to 1239.8)	-	-	-	0.8 (0.6 to 1.3)	2.4 (1.5 to 4.0)	0.3 (0.0 to 6.8)	0.4 (0.1 to 2.5)
Metformin	-1.9 (-2.5 to -1.2)	-0.3 (-1.2 to 0.6)	3.3 (0.3 to 41.9)	0.8 (-4.2 to 5.8)	-3.4 (-6.3 to -0.5)	-0.1 (-2.6 to 2.3)	1.4 (1.0 to 2.0)	1.9 (1.5 to 2.5)	0.5 (0.0 to 4.4)	1.1 (0.2 to 7.2)
Lifestyle modification alone	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 lifestyle modification alone	Possibly not convincingly different from lifestyle modification alone
Among the intermediate harmful	Possibly among the intermediate harmful
Among the most harmful	Possibly among the most harmful

Fig 3 | Summary of relative effects of obesity drugs on benefit and harm outcomes. Drugs superior to (or inferior to) lifestyle modification alone (ie, point estimate exceeding or falling below decision threshold) are categorised into most effective or harmful group. Drugs among most effective or harmful that are inferior to (or superior to) another drug with high to moderate certainty evidence (ie, 95% CI falling below or exceeding null effect completely) are categorised into intermediately effective (or harmful) group. Full version of this figure including all outcomes given in appendix 6.1. CagriSema=cagrilintide and semaglutide; CI=confidence interval; IRR=incidence rate ratio; MD=mean difference; RR=risk ratio; SGLT-2=sodium-glucose cotransporter 2

exceeding null effect completely) are then categorised into the intermediately effective (or harmful) group.

Patient and public involvement

As part of ongoing recruitment of patient partners to the linked clinical practice guideline under development (*BMJ Rapid Recommendations*), we invited six adults facing weight problems from North America, Europe, and Africa to give their input on the systematic review and network meta-analysis at the revision phase of the manuscript. We presented these patient partners with the clinical question including selected outcomes, timeframe (one year), subgroup analyses, and decision thresholds, with their confirmation of relevance of the scope. Patient partners also commented on additional questions, including guideline development.

Results

Study selection and study characteristics

Two hundred and sixty two trials enrolling 99 791 participants proved eligible, including 19 drugs (fig 1; appendices 4 and 11). Table 1 depicts characteristics of trials and participants; the median age across trials was 49.0 years, the median proportion of female participants was 63.3%, the median baseline body mass index was 34.7, and the median length of follow-up was 26 weeks (table 1; appendix 4). One hundred and two trials were judged as being at low risk of bias; trials judged as being at high risk of bias mainly had issues relating to deviations from intended interventions, missing outcome data, and measurement issues for adverse events (appendix 5). Appendix 6.6 shows the assessment of global heterogeneity for the network and between study heterogeneity for each comparison. No concerns regarding global publication bias and intransitivity existed for any outcomes (appendices 6.7 and 6.8).

Figure 2 shows the network plot depicting all included drugs in the network meta-analysis across outcomes. Appendix 6.3 depicts network plots for each outcome and appendix 6.5 presents a summary of the network estimates with certainty of evidence for each comparison for each outcome.

Comparative effects of obesity drugs

Figure 3 presents a GRADE evidence summary of key short term benefits and harms categorised from among the most effective to among the most harmful compared with lifestyle modification alone. Appendix 7 details the assessments of GRADE certainty of evidence for relative estimates of direct, indirect, and network comparisons. Figure 4 and appendix 8 present the anticipated absolute effects across short term outcomes of benefit and harm for all comparisons. An interactive online tool allows navigation of GRADE certainty, relative estimates, and absolute effects for all outcomes (https://nongkl.shinyapps.io/obesitynma_data_visualization/). All subsequent estimates presented are compared with lifestyle modification alone and focus on the effects of drugs supported by moderate to high certainty evidence.

Body weight and waist circumference

Drugs supported by moderate or high certainty evidence compared with lifestyle modification alone reduced body weight by 1.9-14.9%. Among these drugs, tirzepatide (nine trials, 4206 participants; mean difference -14.9%, 95% CI -16.0% to -13.9%) and CagriSema (two trials, 4019 participants; -14.8%, -16.9% to -12.7%) were among the most effective. Oral semaglutide (five trials, 1433 participants; -10.9%, -12.7% to -9.1%), orforglipron (two trials, 1790 participants; -9.9%, -12.4% to -7.5%), subcutaneous semaglutide (31 trials, 27 364 participants; -9.8%, -10.6% to -9.1%), and phentermine-topiramate (six trials, 3639 participants; -8.1%, -9.7% to -6.5%) were among the intermediately effective treatments. Liraglutide, naltrexone-bupropion, orlistat, exenatide, SGLT-2 inhibitors, dulaglutide, and metformin were not convincingly different from lifestyle modification alone (fig 3). Consistent with the percentage body weight changes, tirzepatide probably increased participants achieving ≥10% (728 more) and ≥20% (269 more) weight reduction (moderate certainty). CagriSema probably increased participants achieving ≥10% weight reduction (860 more; moderate certainty). Oral semaglutide increased participants achieving ≥10% weight reduction (584 more; high certainty). Orforglipron probably increased participants achieving ≥10% weight reduction (546 more; moderate certainty). Subcutaneous semaglutide probably increased participants achieving ≥10% (501 more; moderate certainty) and ≥20% (169 more; high certainty). Phentermine-topiramate increased participants achieving ≥10% (731 more; high certainty; fig 4; appendix 6.2).

Drugs supported by moderate or high certainty evidence reduced waist circumference by 0.3-11.0 cm compared with lifestyle modification alone. Among these, tirzepatide (six trials, 3006 participants; mean difference -11.0 cm, 95% CI -12.2 to -9.7) and CagriSema (two trials, 4019 participants; -10.7 cm, -13.1 to -8.4) were among the most effective. Oral semaglutide (four trials, 1126 participants; -7.8 cm, -9.8 to -6.0), subcutaneous semaglutide (22 trials, 25 832 participants; -7.8 cm, -8.6 to -6.9), orforglipron (two trials, 1790 participants; -7.3 cm, -9.8 to -4.8), and

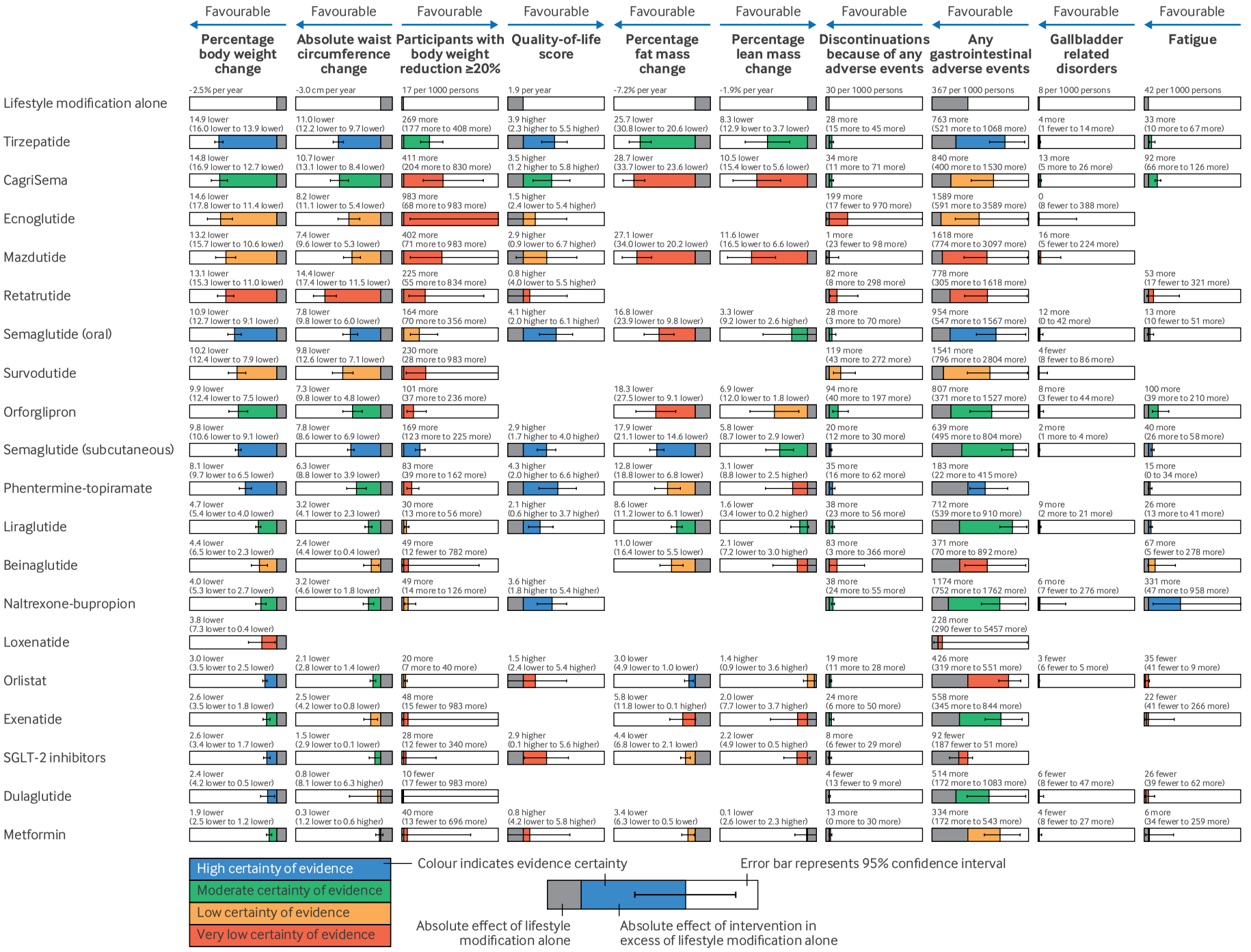


Fig 4 | Summary of absolute effects of obesity drugs on benefit and harm outcomes. Results presented as mean differences or risk differences with corresponding 95% CIs using lifestyle modifications alone as reference group. Absolute effect of lifestyle modification alone was estimated from placebo arm of original studies using hierarchical bayesian model or random effects meta-analysis (see appendix 3.7). Minimally important differences defined as twice the baseline rate for ≥20% weight reduction; 5% change for body weight, fat mass, and lean mass; 4 cm change for waist circumference; 10 point change for quality of life; four per 1000 people over one year for discontinuations because of any adverse events and gallbladder related disorders; and eight per 1000 people over one year for gastrointestinal adverse events and fatigue. Upper confidence limits exceeding plot boundaries truncated for visualisation. Full version of figure including all outcomes given in appendix 6.2. CagriSema=cagrilintide and semaglutide; CI=confidence interval; SGLT-2=sodium-glucose cotransporter 2

phentermine-topiramate (three trials, 2438 participants; -6.3 cm, -8.8 to -3.9) were among the intermediately effective. Naltrexone-bupropion, liraglutide, orlistat, SGLT-2 inhibitors, and metformin were not convincingly different from lifestyle modification alone (fig 3).

For both change in body weight and in waist circumference, no drugs informed by moderate or high certainty evidence were more harmful than lifestyle modification alone (fig 3; appendix 6.1).

Fat mass and lean mass

Moderate or high certainty evidence showed that tirzepatide (one trial, 1273 participants; mean difference -25.7%, 95% CI -30.8% to -20.6%) was among the most effective in reducing fat mass. Subcutaneous semaglutide (three trials, 2145 participants; -17.9%, -21.1% to -14.6%) and liraglutide (five trials, 432 participants; -8.6%, -11.2% to -6.1%) were among the intermediately effective.

Moderate certainty evidence showed that tirzepatide (one trial, 1273 participants; mean difference -8.3%, 95% CI -12.9% to -3.7%) and subcutaneous semaglutide (three trials, 2145 participants; -5.8%, -8.7% to -2.9%) were among the most harmful in reducing lean mass. Moderate certainty evidence showed that liraglutide and oral semaglutide had little or no effect on lean mass loss compared with lifestyle modification alone (fig 3).

All cause death, myocardial infarction, and stroke

The analysis included 59 trials with 57 603 participants and 1045 deaths and a median follow-up duration of 57.5 weeks. Twenty eight trials with 42 016 participants for a median of 52 weeks reported on 715 myocardial infarctions, and 12 trials with 34 536 participants for a median of 44 weeks reported on 373 strokes. Subcutaneous semaglutide was among the most effective in reducing the risk of all cause death (17 trials, 25 264 participants; risk ratio 0.81, 95% CI 0.72 to 0.93, high certainty) and myocardial infarction (six trials, 19 663 participants; 0.72, 0.61 to 0.85, high certainty). No drugs supported by high or moderate certainty evidence proved to be convincingly different from lifestyle modification alone in reducing the risk of stroke. Moderate certainty evidence showed that subcutaneous semaglutide probably had little or no effect on the risk of stroke (appendix 6.1).

Heart failure and hospital admission for heart failure

The analysis included 19 trials with 28 728 participants and 764 events of heart failure and a median follow-up duration of 52 weeks. Four trials with 19 480 participants reported on 295 hospital admissions for heart failure and a median follow-up duration of 52 weeks. Moderate to high certainty evidence showed that subcutaneous semaglutide (six trials, 19 889 participants; risk ratio 0.43, 95% CI 0.21 to 0.84) and tirzepatide (three trials, 2240 participants; 0.49, 0.27 to 0.88) were among the most effective in reducing the risk of heart failure. Tirzepatide was among the most effective in reducing the risk of hospital admission for heart failure (one trial, 731 participants; 0.47, 0.24 to 0.91). All other drugs were supported by low or very low certainty evidence (appendix 6.1).

Kidney failure and kidney disease progression

The analysis included seven trials with 20 082 participants and nine kidney failure events and a median follow-up duration of 52 weeks. Eight trials with 21 021 participants reported on 365 events of kidney disease progression and a median follow-up duration of 52 weeks. Subcutaneous semaglutide probably reduced the risk of kidney disease progression (two trials, 18 133 participants, risk ratio 0.80, 95% CI 0.65 to 0.98, moderate certainty). All other drugs were supported by low or very low certainty evidence (appendix 6.1).

Quality of life

The analysis included 43 trials with 45 663 participants. Using 10 points as the minimally important difference, moderate or high certainty evidence showed that phentermine-topiramate (three trials, 3060 participants; mean difference 4.3, 95% CI 2.0 to 6.6), oral semaglutide (four trials, 1317 participants; 4.1, 2.0 to 6.1), tirzepatide (six trials, 3501 participants; 3.9, 2.3 to 5.5), naltrexone-bupropion (four trials, 3695 participants; 3.6, 1.8 to 5.4), CagriSema (two trials, 4019 participants; 3.5, 1.2 to 5.8), subcutaneous semaglutide (10 trials, 23 828 participants; 2.9, 1.7 to 4.0), and liraglutide (eight trials, 5221 participants; 2.1, 0.6 to 3.7) had little to no effects on quality of life (fig 3).

Discontinuation because of any adverse events

The analysis included 208 trials with 93 388 participants and 8392 discontinuations. Moderate or high certainty evidence showed that orforglipron (two trials, 1790 participants; risk ratio 4.2, 95% CI 2.3 to 7.6), naltrexone-bupropion (seven trials, 13 250 participants; 2.3, 1.8 to 2.9), liraglutide (36 trials, 9881 participants; 2.2, 1.8 to 2.9), phentermine-topiramate (six trials, 3631 participants; 2.2, 1.6 to

3.1), CagriSema (two trials, 4019 participants; 2.1, 1.4 to 3.4), and oral semaglutide (five trials, 1433 participants; 1.9, 1.1 to 3.3) were among the most harmful in increasing the risk of drug discontinuation, followed by tirzepatide (10 trials, 4302 participants; 1.9, 1.5 to 2.5), exenatide (12 trials, 2264 participants; 1.8, 1.2 to 2.7), subcutaneous semaglutide (28 trials, 27 171 participants; 1.7, 1.4 to 2.0), and orlistat (43 trials, 15 456 participants; 1.6, 1.4 to 1.9; fig 3; appendix 6.1). The absolute risk increase for discontinuation per 1000 people over one year was 94 for orforglipron, 38 for naltrexone-bupropion, 38 for liraglutide, 35 for phentermine-topiramate, 34 for CagriSema, 28 for oral semaglutide, 28 for tirzepatide, 24 for exenatide, 20 for subcutaneous semaglutide, and 19 for orlistat. All other drugs were supported by low or very low certainty evidence (fig 3, fig 4).

Any gastrointestinal adverse events

The analysis included 204 trials with 92 078 participants and 55 213 gastrointestinal adverse events. High or moderate certainty evidence supported that naltrexone-bupropion (six trials, 12 965 participants; incidence rate ratio 4.2, 95% CI 3.0 to 5.8), oral semaglutide (five trials, 1433 participants; 3.6, 2.5 to 5.3), orforglipron (two trials, 1790 participants; 3.2, 2.0 to 5.2), and tirzepatide (nine trials, 3723 participants; 3.1, 2.4 to 3.9) were among the most harmful in increasing the risk of gastrointestinal adverse events. Liraglutide (39 trials, 10 097 participants; 2.9, 2.5 to 3.5), subcutaneous semaglutide (28 trials, 27 169 participants; 2.7, 2.4 to 3.2), exenatide (13 trials, 2293 participants; 2.5, 1.9 to 3.3), dulaglutide (five trials, 926 participants; 2.4, 1.5 to 4.0), and phentermine-topiramate (six trials, 3627 participants; 1.5, 1.1 to 2.1) were among the intermediately harmful. All other drugs were supported by low or very low certainty evidence (fig 3). The absolute risk increase for the incidence of gastrointestinal adverse events per 1000 people over one year was 1174 for naltrexone-bupropion, 954 for oral semaglutide, 807 for orforglipron, 763 for tirzepatide, 712 for liraglutide, 639 for subcutaneous semaglutide, 558 for exenatide, 514 for dulaglutide, and 183 for phentermine-topiramate (fig 4).

Gallbladder related disorders

The analysis included 55 trials with 46 513 participants and 899 gallbladder related disorder events. Moderate certainty evidence showed that CagriSema (two trials, 4019 participants; risk ratio 2.6, 95% CI 1.6 to 4.3), oral semaglutide (three trials, 1175 participants; 2.5, 1.0 to 6.3), liraglutide (14 trials, 6508 participants; 2.1, 1.2 to 3.6), and tirzepatide (six trials, 3501 participants; 1.6, 0.9 to 2.8) increased gallbladder related disorders the most (fig 3). The absolute risk increase for gallbladder related disorders per 1000 people over one year was 13 for CagriSema, 12 for oral semaglutide, nine for liraglutide, and four for tirzepatide (fig 4). All other drugs were supported by low or very low certainty evidence (fig 3).

Fatigue

The analysis included 47 trials with 52 940 participants and 1872 fatigue events. High or moderate certainty evidence supported that naltrexone-bupropion (two trials, 9410 participants; risk ratio 8.9, 95% CI 2.1 to 37.4), orforglipron (two trials, 1790 participants; 3.4, 1.9 to 6.0), and CagriSema (two trials, 4019 participants; 3.2, 2.6 to 4.0) were among the most harmful for increasing the risk of fatigue, followed by subcutaneous semaglutide (15 trials, 24 413 participants; 2.0, 1.6 to 2.4), tirzepatide (two trials, 675 participants; 1.8, 1.2 to 2.6), liraglutide (10 trials, 6124 participants; 1.6, 1.3 to 2.0), and phentermine-topiramate (three trials, 3232 participants; 1.4, 1.0 to 1.8). All other drugs were supported by low or very low certainty evidence (fig 3). With our best estimate of baseline risk (derived from control arm of trials), the absolute risk increase for fatigue per 1000 people over one year was 331 for naltrexone-bupropion, 100 for orforglipron, 92 for CagriSema, 40 for subcutaneous semaglutide, 33 for tirzepatide, 26 for liraglutide, and 15 for phentermine-topiramate (fig 4).

Subgroup analyses and sensitivity analyses

Subcutaneous semaglutide probably resulted in larger reductions in body weight and waist circumference in trials with longer follow-up durations. The difference in mean difference between the 12-26 week subgroup and the >52 week subgroup was 4.5% (95% credible interval 2.8 to 6.0) for body weight and 2.3 cm (0.1 to 4.5) for waist circumference (all moderate credibility). We did not find other credible subgroup effects with at least moderate credibility. All sensitivity analyses confirmed the robustness of the estimates.

Discussion

Principal findings

In this network meta-analysis of 262 randomised trials involving nearly 100 000 participants, we found moderate to high certainty evidence that several newer drugs produce substantial weight loss, with tirzepatide and CagriSema showing the largest effects. Oral and subcutaneous semaglutide, orforglipron, and phentermine-topiramate also achieved meaningful reductions. Emerging agents—including retatrutide, ecnoglutide, and mazdutide—show large effects on weight loss but are supported by low or very low certainty evidence. Greater weight loss was consistently accompanied by higher rates of adverse events and treatment discontinuation, indicating a clear benefit-harm trade-off. Despite substantial weight reduction, no drug produced clinically important improvements in quality of life at one year. Evidence for some potential harms, such as erectile dysfunction, was very uncertain. Subcutaneous semaglutide was the only drug associated with reduced all cause mortality, myocardial infarction, and progression of kidney disease. Semaglutide and tirzepatide reduced heart failure risk, but only tirzepatide showed evidence of reduced hospital admission for heart failure.

Strengths and limitations

This review provides a comprehensive and up-to-date comparison of currently available and emerging obesity drugs across a broad set of outcomes important to patients, clinicians, and policymakers.[32] Other strengths include the large evidence base, rigorous methods (eg, GRADE approach to assess certainty), inclusion of benefits and harms, and extensive sensitivity analyses showing the robustness of findings. Incorporation of dose-response data and duration of treatment analyses further enhance applicability to clinical practice.

Several limitations warrant consideration. Most trials had relatively short follow-up, with only a small number extending beyond two years, limiting conclusions about long term safety, quality of life, and effects on cardiovascular and kidney outcomes. The absence of individual participant data precluded more credible exploration of treatment effects across key subgroups, including age, sex, and comorbidity burden. Evidence for several newer agents was sparse and of low certainty, reflecting early stage development rather than absence of effect. Our approach for categorising benefits and harms of treatment options from the least to most effective (fig 3) was pragmatic and limited by single thresholds for categorical outcomes. Finally, trial populations may not fully represent real world patients, who often have greater comorbidity and lower adherence.

Implications for policy and practice

Treatment decisions for obesity should be individualised, balancing expected benefits, harms, treatment burden, costs, availability, and patient preferences.[10 11 39] Some patients may prioritise maximal weight loss, whereas others may place greater value on evident reductions in mortality or cardiovascular risk.[10 11] Among currently available drugs, subcutaneous semaglutide is the only agent with moderate to high certainty evidence of reduced death and major cardiovascular events.[40] Drug availability also varies across jurisdictions—for example, phentermine-topiramate is approved only in the US—which may influence policy and reimbursement decisions.

High costs represent a major barrier for patients and healthcare systems and are likely to limit widespread use.[10] These constraints underscore the importance of prioritising treatment for people most likely to benefit clinically, particularly those at high risk of cardiovascular or kidney complications. Rigorous risk stratification—similar to approaches used for type 2 diabetes—may help identify patients in whom treatment yields the greatest health gains and cost effectiveness.[16 17] Such prioritisation has been emphasised in international guidance, including the World Health Organization (WHO) living guidelines recommendations for obesity management.[10]

For patients primarily seeking weight reduction, tirzepatide and CagriSema offer the greatest likelihood of achieving large weight loss, with approximately a third to a half of treated people reaching ≥20% reduction within one year compared with about 2% with lifestyle modification alone. These effects approach those reported after some bariatric procedures. However, substantial benefits must be weighed against adverse effects, treatment burden and costs, and the high likelihood of gastrointestinal symptoms and fatigue, which frequently lead to discontinuation. Real world data suggest that roughly half of patients discontinue treatment within the first year.[41 42]

Weight regain after discontinuation of pharmacotherapy is substantial.[43] A systematic review of 37 studies found that participants regained weight at an average rate of about 0.4 kg per month after stopping treatment, with a projected return to baseline weight within approximately 1.7 years, accompanied by loss of cardiometabolic improvements.[43] Therefore, on the whole, benefits do not appear to be sustained without continued treatment, highlighting the importance of long term adherence, tolerability, and ongoing clinical support.[43] For many patients, pharmacotherapy may need to be continued long term, with implications for safety monitoring, affordability, and health system sustainability.[10 44]

Loss of lean mass represents an additional potential concern, particularly for older adults or those at risk of frailty.[45] Tirzepatide and subcutaneous semaglutide were associated with notable reductions in lean mass, although long term functional consequences remain uncertain. Current guidelines recommend combining pharmacotherapy with structured aerobic and resistance exercise to help preserve muscle mass, with inherent limitations concerning the heterogeneity of interventions, burden of treatment, and available evidence.[10 46]

Living evidence and guideline development

The rapidly evolving landscape of obesity pharmacotherapy makes this topic well suited to a living evidence approach, with dynamic and rapid updating.[47] Living network meta-analyses have successfully supported living guidelines in rapidly changing fields, including covid-19, where continuously updated evidence informed global recommendations from WHO, the National Institute for Health and Care Excellence, and others.[17 48 49] When linked to rigorous guideline processes using the GRADE approach, such living syntheses—as also shown for diabetes type 2 drugs—can provide a reliable foundation for recommendations that can be adapted to national contexts, reducing duplication of effort, and supporting more efficient use of resources.[16 17] Establishing and maintaining this approach would require sustained infrastructure and funding, but could enable clinicians, patients, and policymakers to make decisions based on the most current evidence as new treatments continue to emerge.

Conclusion

Tirzepatide and CagriSema provide the greatest reductions in body weight among available drugs for adults with overweight or obesity, regardless of diabetes status. Subcutaneous semaglutide currently has the strongest evidence for reductions in mortality

and major cardiovascular outcomes. Larger benefits are generally accompanied by greater harms, treatment burden, and discontinuation, and improvements are not sustained after treatment cessation. Given high costs and the likelihood of long term treatment, careful patient selection and risk stratification are essential to maximise clinical benefit and cost effectiveness. Emerging drugs with substantial weight lowering effects require large, long term trials to determine their impact on cardiovascular, kidney, and other patient important outcomes. Pharmacotherapy should be integrated with lifestyle interventions and strategies to preserve lean mass.

AUTHOR AFFILIATIONS

[1]Department of Endocrinology and Metabolism, West China Hospital, Sichuan University, Chengdu, China

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

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

[4]Department of Endocrinology, Key Laboratory of Endocrinology of the National Health Commission, Peking Union Medical College Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China

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

[6]MAGIC Evidence Ecosystem Foundation, Oslo, Norway

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

[8]Diabetes Research Centre, University of Leicester, Leicester, UK

[9]Diabetes Complications Research Centre, UCD Conway Institute of Biomedical and Biomolecular Research, School of Medicine, University College Dublin, Dublin, Ireland

[10]Department of Community Medicine, Faculty of Medicine Universitas Indonesia, Jakarta, Indonesia

[11]Department of Anaesthesiology, West China Hospital, Sichuan University, Chengdu, Sichuan, China

[12]Laboratory of Anaesthesia and Critical Care Medicine, National-Local Joint Engineering Research Centre of Translational Medicine of Anaesthesiology, State Key Laboratory of Biotherapy, West China Hospital, Sichuan University, Chengdu, Sichuan, China

[13]Sports Medicine Centre, West China Hospital, Sichuan University, Chengdu, China

[14]Department of Paediatrics, University of Antioquia, Medellin, Colombia

[15]Paediatric Intensive Care Unit, Clínica Las Américas-AUNA, Medellin, Colombia

[16]School of Rehabilitation Science, McMaster University, Hamilton, ON, Canada

[17]Department of Internal Medicine, TidalHealth, Salisbury, MD, USA

[18]Department of Neurology, Lehigh Valley Health Network, Allentown, PA, USA

[19]Department of Endocrinology and Metabolism, Affiliated Hospital of Shandong Second Medical University, Weifang, China

[20]School of Rehabilitation Science, McMaster University, Hamilton, ON, Canada

[21]Department of Epidemiology and Biostatistics, School of Public Health, Peking University Health Science Centre, Beijing, 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]Division of General Internal Medicine, University Hospitals of Geneva, Geneva, Switzerland

[24]Department of Behavioural Medicine, Institute of Basic Medical Sciences, Faculty of Medicine, University of Oslo, Oslo, Norway

KN is supported by the Doctoral Student Programme of the Young S&T Talents Cultivation Project, CAST. KK is supported by the National Institute for Health Research (NIHR) Applied Research Collaboration East Midlands (ARC EM), NIHR Global Research Centre for Multiple Long Term Conditions, NIHR Cross NIHR Collaboration for Multiple Long Term Conditions, NIHR Leicester Biomedical Research Centre (BRC) and the British Heart Foundation (BHF) Centre of Excellence.

Contributors: KN, QS, XX, and YW evenly shared their contributions to this work. SL and POV evenly shared their contribution to this work and co-supervised this study as senior researchers. KN, QS, SL, POV, GG, YW, and HT conceived and designed the study. KN, YW, XX, HZ, QF, TL, YG, and HD screened and selected the articles. KN, YW, XX, HZ, YM, XP, XZ, QF, CW, XS, and QJ extracted the data. KN, YW, YG, XX, and QF assessed the risk of bias of newly included trials. KN analysed the data. QS, YM, QH, FS, IDF, and TA provided methodological consultation. KN, QS, SL, GG, POV, and AA rated and revised the GRADE certainty of evidence. SL, POV, GG, KN, AA, KK, CWLR, ISW, IDF, JL, and ZZ interpreted the results. KN, SL, GG, POV, AA, QS, and XX drafted the manuscript. SL, GG, POV, KN, AA, QS, XX, KK, CLR, ISW, TA, IDF, and ZZ critically revised the manuscript. To accelerate the revision process, POV used ChatGPT for English language polishing during the revision of the manuscript. All AI facilitated contents have been circulated, reviewed, and approved by all authors. All authors contributed to revising the manuscript. All authors had fully assessed all the data in the study and had final responsibility for the decision to submit for publication. SL 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 Noncommunicable Chronic Diseases-National Science and Technology Major Project (2025ZD0550604), Chengdu Science and Technology Bureau Key R&D Support Plan (2025-YF09-00033-SN), National Natural Science Foundation of China (72342014), and 1.3.5 project for disciplines of excellence, West China Hospital, Sichuan University (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 www.icmje.org/disclosure-of-interest/ and declare: support from Noncommunicable Chronic Diseases-National Science and Technology Major Project, Chengdu Science and Technology Bureau Key R&D Support Plan, National Natural Science Foundation of China, and 1.3.5 project for disciplines of excellence, West China Hospital, Sichuan University for the submitted work. SL received grants from Noncommunicable Chronic Diseases-National Science and Technology Major Project (2025ZD0550604), Chengdu Science and Technology Bureau Key R&D Support Plan (2025-YF09-00033-SN), National Natural Science Foundation of China (72342014), and 1.3.5 project for disciplines of excellence, West China Hospital, Sichuan University (ZYYC24001). KK has acted as a consultant, speaker or received grants for investigator initiated studies for Abbott, Astra Zeneca, Bayer, Novo Nordisk, Sanofi-Aventis, Servier, Lilly and Merck Sharp & Dohme, Boehringer Ingelheim, Oramed Pharmaceuticals, Pfizer, Roche, Daiichi-Sankyo, Applied Therapeutics, Embecta, and Nestle Health Science. CLRoux received grants from Irish Research Council, Health Research Board, Science Foundation Ireland, and Anabio; consulting fees from Novo Nordisk, Eli Lilly, and Johnson & Johnson, Boehringer Ingelheim, GI Dynamics, Herbalife, Altimmune, Irish Life Health, Amgen, Arrowhead, Roche, AstraZeneca, Keyron, Gila Pharmaceuticals, Metsera, Nymble, AbbVie, and Olympus; payments for presentations from NovoNordisk, Herbalife, Johnson & Johnson, Eli Lilly, Boehringer Ingelheim, Rhythm Pharmaceuticals, and Currax Pharmaceuticals; support for attending meetings and/or travel from Novo Nordisk, Herbalife, Johnson & Johnson, Eli Lilly, and Boehringer Ingelheim; payment for scientific advisory board contributions from Metsera and Nymble. CWLR has acted as an unpaid leadership or fiduciary role in Irish Society for Nutrition and Metabolism and provided clinical obesity care for co-owner obesity clinics that are My Best Weight and Beyond BMI. KK and CWLR did not have a role in assessing the risk of bias of the included studies or in the extraction. TA is chair and board member of the MAGIC Evidence Ecosystem Foundation (www.magicevidence.org), a non-profit organisation providing the authoring and publication platform (MAGICapp) to guideline organisations and producing *BMJ Rapid Recommendations* in partnership with *The BMJ*. TA has not derived any income from this work. POV is board member and chief scientist in MAGIC and chief scientist, with a part-time income. All other authors declare no competing interests.

Ethical approval: Not required.

Data sharing: Please contact the corresponding author (SL) for data sharing on reasonable terms. The R code for network meta-analysis and forest plots is available in the supplementary files (appendix 3.11).

Transparency: The manuscript's guarantor (SL) affirms 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: This study will support a *BMJ Rapid Recommendation* (https://www.bmj.com/rapid-recommendations).

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

1 World Health Organization. Obesity and overweight. https://www.who.int/news-room/fact-sheets/detail/obesity-and-overweight
2 Wadden TA, Chao AM, Moore M, et al. The role of lifestyle modification with second-generation anti-obesity medications: comparisons, questions, and clinical opportunities. *Curr Obes Rep* 2023;12:453-73. doi:10.1007/s13679-023-00534-z
3 Davies MJ, Bajaj HS, Broholm C, et al. Cagrilintide-semaglutide in adults with overweight or obesity and type 2 diabetes. *N Engl J Med* 2025;393:648-59. doi:10.1056/NEJMoa2502082
4 Garvey WT, Blüher M, Osorto Contreras CK, et al. Coadministered cagrilintide and semaglutide in adults with overweight or obesity. *N Engl J Med* 2025;393:635-47. doi:10.1056/NEJMoa2502081
5 Ji L, Jiang H, Bi Y, et al. Once-weekly mazdutide in Chinese adults with obesity or overweight. *N Engl J Med* 2025;392:2215-25. doi:10.1056/NEJMoa2411528
6 Novo Nordisk A/S. Investor presentation for first three months of 2025. 2025. https://www.novonordisk.com/content/dam/nncorp/global/en/investors/pdfs/financial-results/2025/Q1-2025-investor-presentation.pdf
7 Innoventbio. Phase 3 clinical study of mazdutide in Chinese adults with overweight or obesity (GLORY-1) published in The New England Journal of Medicine. https://www.innoventbio.com/InvestorsAndMedia/PressReleaseDetail?key=530
8 Rubino F, Cummings DE, Eckel RH, et al. Definition and diagnostic criteria of clinical obesity. *Lancet Diabetes Endocrinol* 2025;13:221-62. doi:10.1016/S2213-8587(24)00316-4
9 Lingvay I, Cohen RV, Roux CWL, et al. Obesity in adults. *Lancet* 2024;404:972-87. doi:10.1016/S0140-6736(24)01210-8
10 Celletti F, Farrar J, De Regil L. World Health Organization guideline on the use and indications of glucagon-like peptide-1 therapies for the treatment of obesity in adults. *JAMA* 2026;335:434-8. doi:10.1001/jama.2025.24288
11 Pedersen SD, Manjoo P, Dash S, et al. Pharmacotherapy for obesity management in adults: 2025 clinical practice guideline update. *CMAJ* 2025;197:E797-809. doi:10.1503/cmaj.250502
12 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
13 McGowan B, Ciudin A, Baker JL, et al. A systematic review and meta-analysis of the efficacy and safety of pharmacological treatments for obesity in adults. *Nat Med* 2025;31:3317-29. doi:10.1038/s41591-025-03978-z
14 Siemieniuk RA, Agoritsas T, Macdonald H, et al. Introduction to BMJ Rapid Recommendations. *BMJ* 2016;354:i5191. doi:10.1136/bmj.i5191
15 Seterelv SS, Gupta S, Kouri A, et al. From international guidelines to practice: a multimethod evaluation of BMJ Rapid Recommendations focusing on suitability for adaptation and implementation. *BMJ Open* 2025;15:e114609. doi:10.1136/bmjopen-2025-114609
16 Nong K, Jeppesen BT, Shi Q, et al. Medications for adults with type 2 diabetes: a living systematic review and network meta-analysis. *BMJ* 2025;390:e083039. doi:10.1136/bmj-2024-083039
17 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
18 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
19 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
20 Sterne JAC, Savović J, Page MJ, et al. RoB 2: a revised tool for assessing risk of bias in randomised trials. *BMJ* 2019;366:l4898. doi:10.1136/bmj.l4898
21 Murad MH, Wang Z, Chu H, et al. When continuous outcomes are measured using different scales: guide for meta-analysis and interpretation. *BMJ* 2019;364:k4817. doi:10.1136/bmj.k4817
22 Rücker G. Network meta-analysis, electrical networks and graph theory. *Res Synth Methods* 2012;3:312-24. doi:10.1002/jrsm.1058
23 Jackson D, White IR, Riley RD. Quantifying the impact of between-study heterogeneity in multivariate meta-analyses. *Stat Med* 2012;31:3805-20. doi:10.1002/sim.5453
24 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
25 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
26 Brignardello-Petersen R, Tomlinson G, Florez I, et al. Grading of recommendations assessment, development, and evaluation concept article 5: addressing intransitivity in a network meta-analysis. *J Clin Epidemiol* 2023;160:151-9. doi:10.1016/j.jclinepi.2023.06.010
27 Viechtbauer W. Conducting meta-analyses in R with the metafor Package*J Stat Software* 2010;36. doi:10.18637/jss.v036.i03
28 Phillippo DM, Dias S, Ades AE, et al. Multilevel network meta-regression for population-adjusted treatment comparisons. *J R Stat Soc Ser A Stat Soc* 2020;183:1189-210. doi:10.1111/rssa.12579
29 Garvey WT, Mechanick JI, Brett EM, et al. American Association of Clinical Endocrinologists and American College of Endocrinology comprehensive clinical practice guidelines for medical care of patients with obesity*Endocrine Practice* 2016;22:1-203. doi:10.4158/EP161365.GL
30 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
31 Dias S, Sutton AJ, Ades AE, et al. Evidence synthesis for decision making 2: a generalized linear modeling framework for pairwise and network meta-analysis of randomized controlled trials. *Med Decis Making* 2013;33:607-17. doi:10.1177/0272989X12458724
32 Brignardello-Petersen R, Murad MH, Walter SD, et al. GRADE approach to rate the certainty from a network meta-analysis: avoiding spurious judgments of imprecision in sparse networks. *J Clin Epidemiol* 2019;105:60-7. doi:10.1016/j.jclinepi.2018.08.022
33 Izcovich A, Chu DK, Mustafa RA, et al. A guide and pragmatic considerations for applying GRADE to network meta-analysis. *BMJ* 2023;381:e074495. doi:10.1136/bmj-2022-074495
34 Guyatt G, Zeng L, Brignardello-Petersen R, et al. Core GRADE 2: choosing the target of certainty rating and assessing imprecision. *BMJ* 2025;389:e081904. doi:10.1136/bmj-2024-081904
35 Brignardello-Petersen R, Florez ID, Izcovich A, et al. 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 U.S. Food and Drug Administration. Obesity and Overweight: Developing Drugs and Biological Products for Weight Reduction. 2025. https://www.fda.gov/regulatory-information/search-fda-guidance-documents/obesity-and-overweight-developing-drugs-and-biological-products-weight-reduction
37 Maruish ME. *User's Manual for the SF-36v2 Health Survey* . Quality Metric Incorporated, 2011.
38 NHLBI Obesity Education Initiative Expert Panel on the Identification, Evaluation, and Treatment of Obesity in Adults (US). *Clinical Guidelines on the Identification, Evaluation, and Treatment of Overweight and Obesity in Adults: The Evidence Report*. National Heart, Lung, and Blood Institute, 1998.
39 American Diabetes Association Professional Practice Committee for Obesity. Pharmacologic treatment of obesity in adults: standards of care in overweight and obesity. *BMJ Open Diabetes Res Care* 2026;13:e005729. doi:10.1136/bmjdrc-2025-005729
40 Lincoff AM, Brown-Frandsen K, Colhoun HM, et al. Semaglutide and cardiovascular outcomes in obesity without diabetes. *N Engl J Med* 2023;389:2221-32. doi:10.1056/NEJMoa2307563
41 Rodriguez PJ, Zhang V, Gratzl S, et al. Discontinuation and reinitiation of dual-labeled GLP-1 receptor agonists among US adults with overweight or obesity. *JAMA Netw Open* 2025;8:e2457349. doi:10.1001/jamanetworkopen.2024.57349

42 Khunti K, Papamargaritis D, Aroda VR, et al. Re-evaluating the concept of remission in type 2 diabetes: a call for patient-centric approaches. *Lancet Diabetes Endocrinol* 2025;13:615-34. doi:10.1016/S2213-8587(25)00134-2
43 West S, Scragg J, Aveyard P, et al. Weight regain after cessation of medication for weight management: systematic review and meta-analysis. *BMJ* 2026;392:e085304. doi:10.1136/bmj-2025-085304
44 Kar P. Partha Kar: Weight loss drugs have opened the door to a two tier health service. *BMJ* 2025;389:r923. doi:10.1136/bmj.r923
45 Prado CM, Phillips SM, Gonzalez MC, et al. Muscle matters: the effects of medically induced weight loss on skeletal muscle. *Lancet Diabetes Endocrinol* 2024;12:785-7. doi:10.1016/S2213-8587(24)00272-9
46 Nadolsky K, Garvey WT, Agarwal M, et al. American Association of Clinical Endocrinology consensus statement: algorithm for the evaluation and treatment of adults with obesity/adiposity-based chronic disease - 2025 update. *Endocr Pract* 2025;31:1351-94. doi:10.1016/j.eprac.2025.07.017
47 Elliott J, Lawrence R, Minx JC, et al. Decision makers need constantly updated evidence synthesis. *Nature* 2021;600:383-5. doi:10.1038/d41586-021-03690-1
48 Munn Z, Twaddle S, Service D, et al. Developing guidelines before, during, and after the COVID-19 pandemic. *Ann Intern Med* 2020;M20-4907.
49 Agarwal A, Hunt B, Stegemann M, et al. A living WHO guideline on drugs for covid-19. *BMJ* 2020;370:m3379. doi:10.1136/bmj.m3379

Supplementary files: PRISMA checklist and search strategy; Methodology; Study characteristics and risk of bias assessments; Other main network meta-analysis results; GRADE assessments and relative and absolute effects; Subgroup analyses; Sensitivity and additional analyses; Reference list of included studies