



# Comparative effectiveness of second line glucose lowering drug treatments using real world data: emulation of a target trial

Yihong Deng,<sup>1,2</sup> Eric C Polley,<sup>3</sup> Joshua D Wallach ,<sup>4</sup> Jeph Herrin,<sup>5</sup> Joseph S Ross ,<sup>6,7,8</sup> Rozalina G McCoy <sup>1,9</sup>

► Additional supplemental material is published online only. To view, please visit the journal online (<http://dx.doi.org/10.1136/bmjmed-2022-000419>).

For numbered affiliations see end of article.

Correspondence to: Dr Rozalina G McCoy, Division of Community Internal Medicine, Geriatrics, and Palliative Care, Department of Medicine, Mayo Clinic, Rochester, MN 55905, USA; [McCoy.Rozalina@mayo.edu](mailto:McCoy.Rozalina@mayo.edu)

Cite this as: *BMJMED* 2023;2:e000419. doi:10.1136/bmjmed-2022-000419

Received: 25 October 2022  
Accepted: 7 July 2023

## ABSTRACT

**OBJECTIVE** To build on the recently completed GRADE (Glycemia Reduction Approaches in Diabetes: A Comparative Effectiveness Study) randomised trial examining the comparative effectiveness of second line glucose lowering drugs in achieving and maintaining glycaemic control in adults with type 2 diabetes.

**DESIGN** Emulation of a target trial.

**SETTING** Medical and pharmacy claims data from the OptumLabs Data Warehouse, a de-identified US national dataset of beneficiaries of commercially insured and Medicare Advantage plans, 29 March 2013 to 30 June 2021.

**PARTICIPANTS** Adults (≥18 years) with type 2 diabetes who first started taking glimepiride, sitagliptin, liraglutide, insulin glargine, or canagliflozin between 29 March 2013 and 30 June 2021. Participants were treatment naive or were receiving metformin monotherapy at the time of starting the study drug.

**MAIN OUTCOME MEASURES** The main outcomes were time to primary and secondary metabolic

failure of the assigned treatment, calculated as days to haemoglobin A<sub>1c</sub> levels of ≥7.0% and >7.5%, respectively. Secondary metabolic, cardiovascular, and microvascular outcomes were analysed as specified in the GRADE statistical analysis plan. Propensity scores were estimated with the gradient boosting method, and inverse propensity score weighting was used to emulate randomisation to the treatment groups, which were then compared with Cox proportional hazards regression.

**(RESULTS)** The study cohort included participants starting treatment with glimepiride (n=20 511), liraglutide (n=5569), sitagliptin (n=13 039), insulin glargine (n=7262), and canagliflozin (n=5290). The insulin glargine arm was excluded because of insufficient control of confounding. Median times to primary metabolic failure were 439 (95% confidence interval 400 to 489) days in the canagliflozin arm, 439 (426 to 453) days in the glimepiride arm, 624 (567 to 731) days in the liraglutide arm, and 461 (442 to 482) days in the sitagliptin arm. Median time to secondary metabolic failure was also longest in the liraglutide arm. Adults receiving liraglutide had the lowest one year cumulative incidence rate of primary metabolic failure (0.37, 95% confidence interval 0.35 to 0.40) followed by sitagliptin (0.44, 0.43 to 0.45), glimepiride (0.45, 0.44 to 0.45), and canagliflozin (0.46, 0.44 to 0.48). Similarly, the one year cumulative incidence rate of secondary metabolic failure was 0.27 (0.25 to 0.29) in the canagliflozin arm, 0.28 (0.27 to 0.29) in the glimepiride arm, 0.23 (0.21 to 0.26) in the liraglutide arm, and 0.28 (0.27 to 0.29) in the sitagliptin arm. No differences were observed between the study arms in the rates of microvascular and macrovascular complications.

**CONCLUSIONS** In this target trial emulation of an expanded GRADE study framework, liraglutide was more effective in achieving and maintaining glycaemic control as a second line glucose lowering drug than canagliflozin, sitagliptin, or glimepiride.

## WHAT IS ALREADY KNOWN ON THIS TOPIC

- ⇒ The GRADE (Glycemia Reduction Approaches in Diabetes: A Comparative Effectiveness Study) randomised trial compared glimepiride (sulphonylurea), sitagliptin (dipeptidyl-peptidase 4 inhibitor), liraglutide (glucagon-like peptide 1 receptor agonist), and insulin glargine (basal analogue insulin) for the ability of these drugs to lower haemoglobin A<sub>1c</sub> (HbA<sub>1c</sub>) in people with moderately raised levels of HbA<sub>1c</sub> receiving metformin monotherapy
- ⇒ Although the GRADE trial found that liraglutide was the most effective among the studied drugs, these findings were limited by the lack of a sodium-glucose cotransporter 2 inhibitor comparator arm and narrow eligibility requirements that excluded most patients in real world practice

## WHAT THIS STUDY ADDS

- ⇒ In this comparative study of the effectiveness of five classes of second line glucose lowering drugs, liraglutide was significantly more effective in maintaining glycaemic control in adults with type 2 diabetes than other second line glucose lowering drugs
- ⇒ Liraglutide was more effective than glimepiride, sitagliptin, and canagliflozin in achieving and maintaining glycaemic control in adults with type 2 diabetes
- ⇒ Insulin glargine was not included in the comparisons because of insufficient control of confounding with propensity score weighting

## HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE, OR POLICY

- ⇒ This study suggests that observational data and methods can be used to emulate clinical trials and to examine the comparative effectiveness and safety of interventions in routine care

## Introduction

Type 2 diabetes is a common and serious chronic health condition.<sup>1</sup> Timely control of hyperglycaemia, most often measured as serum levels of haemoglobin A<sub>1c</sub> (HbA<sub>1c</sub>), is necessary to prevent complications of diabetes and reduce the risk of death.<sup>2–8</sup> Most clinical practice guidelines recommend targeting HbA<sub>1c</sub> levels to <7% for most non-pregnant adults.<sup>9</sup> Metformin is recommended as the first line glucose

lowering drug because of its efficacy, tolerability, and low cost.<sup>10–13</sup> Less certainty exists for the optimal second line glucose lowering treatment, however, when metformin is no longer sufficient, is contraindicated, or cannot be tolerated. This uncertainty is partly because of the scarcity of evidence directly comparing currently available second line drug treatments. Clinical practice guidelines advise that the choice of second line treatment should be informed by clinical and situational considerations specific to each person.<sup>10–13</sup> Robust evidence exists indicating the preferential use of specific drug classes in the presence of cardiovascular and kidney comorbidities; however, how these drugs compare with each other in their ability to lower HbA<sub>1c</sub> levels is less known.

GRADE (Glycemia Reduction Approaches in Diabetes: A Comparative Effectiveness Study) is a recently completed pragmatic, randomised, parallel arm clinical trial that compared, head to head, four second line glucose lowering drugs for their ability to achieve and maintain glycaemic control in adults with moderately uncontrolled type 2 diabetes receiving metformin monotherapy.<sup>14–16</sup> GRADE found that liraglutide (a glucagon-like peptide 1 receptor agonist) and insulin glargine (a basal analogue of insulin) were significantly more effective in achieving and maintaining glycaemic control than glimepiride (a sulphonylurea) and sitagliptin (a dipeptidyl-peptidase 4 inhibitor), which were least effective.<sup>16</sup>

The design of the GRADE trial had important limitations, however, that reduced its relevance to contemporary clinical practice. GRADE did not include sodium-glucose cotransporter 2 inhibitors, agents increasingly used in clinical practice<sup>17–19</sup> and the preferred treatment for patients with heart failure and chronic kidney disease.<sup>9</sup> Also, the eligibility requirements in the GRADE trial of low baseline levels of HbA<sub>1c</sub> and the start of second line treatment in the context of baseline metformin monotherapy, resulted in participants representing only 9.1% of adults with diabetes living in the US.<sup>20</sup> Both of these factors highlight the need for more timely and generalisable data that are pertinent to the contemporary management of adults with type 2 diabetes.

To show the feasibility and usefulness of using real world data to emulate randomised controlled trials, and thus generate evidence on the comparative effectiveness and safety of drugs faster, cheaper, and with greater external validity, we previously emulated the GRADE trial with observational claims and electronic health record data before the publication of the GRADE results. Our emulation showed similar findings to the GRADE trial,<sup>21</sup> although efforts to emulate all specifications in the GRADE trial were

hindered because some study conditions (ie, initiation of insulin glargine as a second line treatment at HbA<sub>1c</sub> concentrations of <8.5%) were not adequately represented among large populations in real world practice because they are not aligned with contemporary standards of care.

In this study, we sought to address some of the limitations of the trial design of GRADE and build on our earlier emulation<sup>21</sup> by using the target trial framework<sup>22</sup> and comparing the effectiveness of glimepiride, sitagliptin, liraglutide, insulin glargine, and canagliflozin in achieving and maintaining concentrations of HbA<sub>1c</sub> <7.0% in adults with type 2 diabetes who were naive to these drugs but without further restrictions imposed by the eligibility criteria of the GRADE trial. We also examined the secondary metabolic, microvascular, macrovascular, and safety endpoints planned in GRADE, where feasible using the available claims and electronic health record data. Another prospective randomised controlled trial comparing these drugs with each other for metabolic, microvascular, and macrovascular outcomes is unlikely to be conducted, given the cost, effort, and duration of the original GRADE trial. Hence we expect this emulation to generate important evidence on the comparative effectiveness and safety of these second line glucose lowering drugs in a diverse and generalisable adult population.

## Methods

### Study design

We retrospectively analysed medical and pharmacy claims data from the OptumLabs Data Warehouse, a de-identified national dataset of beneficiaries of commercially insured and Medicare Advantage plans that represents a diverse mixture of ages, ethnic groups, practice settings, and geographic regions across the USA.<sup>23 24</sup> All study data were de-identified consistent with HIPAA (Health Insurance Portability and Accountability Act of 1996) expert de-identification determination. The study is reported according to the Reporting of studies Conducted using Observational Routinely collected Data (RECORD) reporting guideline.

### Study population

We first assembled a cohort of adults (≥18 years) who first started taking glimepiride, sitagliptin, liraglutide, insulin glargine, or canagliflozin between 29 March 2013 (date of approval of canagliflozin by the US Food and Drug Administration; the other study drugs were approved earlier) and 30 June 2021 (online supplemental figure S1 and online supplemental method 1). The index date was set to the date of the first claim for the study drug. People who started two or more study drugs on the index date

were excluded. To ensure consistent and adequate capture of baseline comorbidities and treatment data, participants were required to have six months of continuous enrolment before the index date. We excluded people with prescription fills for any glucose lowering drugs other than metformin, those with type 1 diabetes, those with missing information for age or sex (<1% of the final cohort), pregnant individuals, and those with no available HbA<sub>1c</sub> results during the three months before the index date (baseline HbA<sub>1c</sub>) and in the follow-up period. Laboratory test results are available for a subset of people in the OptumLabs Data Warehouse based on data sharing agreements between OptumLabs and commercial laboratory companies.

### Outcomes

The primary outcome was time to primary metabolic failure of the assigned treatment, calculated as days to HbA<sub>1c</sub> levels  $\geq 7.0\%$ . To assess for potential bias in outcome ascertainment caused by differences in the frequencies and intervals of HbA<sub>1c</sub> testing, we compared the number, frequency, and timing of available HbA<sub>1c</sub> test results and found no difference between the groups (online supplemental table S1). Because testing frequency is guided by baseline HbA<sub>1c</sub> levels, we also examined intervals between sequential HbA<sub>1c</sub> tests grouped by baseline levels of HbA<sub>1c</sub>. No differences were found between the treatment groups (online supplemental table S2). Secondary metabolic, cardiovascular, and microvascular outcomes were analysed as specified in the GRADE statistical analysis plan<sup>15</sup> and detailed previously,<sup>21</sup> if they were feasible to ascertain from claims data (online supplemental table S3).

### Covariates

Age, sex, race or ethnic group, and annual household income of participants were identified from OptumLabs Data Warehouse enrollment files at the time of the index date. Data for sex were taken from information in the OptumLabs Data Warehouse database rather than from participant reported gender. Thresholds were chosen based on clinical relevance and distribution of the data (online supplemental method 2). Comorbidities (determined from all claims during the six months preceding the index date) included retinopathy, nephropathy, neuropathy, coronary artery disease, cerebrovascular disease, peripheral vascular disease, heart failure, and previous severe hypoglycaemia and hyperglycaemia. Baseline drug treatments, included as surrogates for the burden of complications, were identified from pharmacy claims in the six months preceding the index date. Online supplemental tables S4 and S5 list the codes and drugs used to define all covariates. We also operationalised, where feasible, the eligibility criteria of participants, as defined in the

GRADE trial (online supplemental table S6). We included the eligibility criteria as covariates in the propensity score model rather than excluding those participants from the cohort, because our objective was to examine the comparative effectiveness and safety of second line glucose lowering drugs in a generalisable and heterogeneous real world adult population.

### Statistical analysis

Primary analyses followed the intention-to-treat censoring approach, with participants followed until the outcome of interest was reached, the anticipated follow-up duration of the trial (seven years) was achieved, the end of the study period (30 June 2021), the end of insurance coverage, or death (online supplemental figure S2).

Inverse probability of treatment weighting was used to balance the differences in baseline characteristics among the treatment groups. Propensity scores were defined as the probability of receiving each of the treatments given the baseline variables; these propensity score weights were estimated with generalised boosted models, including the baseline variables presented in table 1. Generalised boosted models involve an iterative process with multiple regression trees to capture complex and non-linear relations between treatment assignments and the pretreatment covariates, resulting in the propensity score model that is the best balance among the treatment groups.<sup>25</sup> The number of trees in the final generalised boosted model ensemble was selected with interval 10-fold cross validation to minimise differences between the propensity score weighted treatment groups. Stabilised weights were calculated by dividing the marginal frequency of treatment by the propensity scores of the treatment received.<sup>26</sup>

Online supplemental figure S3 illustrates the distribution of weights. Standardised mean differences were used to assess the balance of covariates after weighting; a standardised mean difference  $\leq 0.1$  was considered a good balance (online supplemental method 3).<sup>27</sup> Before evaluation of the outcomes, weighted sample sizes and ability to account for baseline confounding were examined to determine the feasibility of including each treatment group.

The cumulative incidences of the primary (time to first HbA<sub>1c</sub> concentration  $\geq 7.0\%$ ) and secondary (time to first HbA<sub>1c</sub>  $> 7.5\%$ ) metabolic failure endpoints within each treatment arm were estimated with the inverse probability of treatment weighting Kaplan-Meier method. We used inverse probability of treatment weighting Cox proportional hazards regression models adjusted by baseline HbA<sub>1c</sub> values to estimate the hazard ratios between treatment groups. Because of the large range of values for baseline HbA<sub>1c</sub>, we used a spline with five degrees of freedom. The at-risk

**Table 1 | Baseline characteristics of weighted cohort**

Characteristics	Canagliflozin (n=4876)	Glimepiride (n=19 838)	Liraglutide (n=4684)	Sitagliptin (n=12 588)	Largest SMD
Mean (SD) age (years)	60.7 (12.5)	61.5 (12.7)	60.5 (12.3)	61.5 (12.7)	0.08
Sex:					0.04
Women	2324 (47.7)	9477 (47.8)	2337 (49.9)	6095 (48.4)	
Men	2552 (52.3)	10 361 (52.2)	2347 (50.1)	6493 (51.6)	
Race or ethnic group:					0.08
White	2989 (61.3)	12 012 (60.6)	2940 (62.8)	7609 (60.4)	
Black	706 (14.5)	3013 (15.2)	689 (14.7)	1873 (14.9)	
Hispanic	777 (15.9)	3179 (16.0)	740 (15.8)	2043 (16.2)	
Asian	258 (5.3)	1062 (5.4)	188 (4.0)	692 (5.5)	
Other, unknown, missing	146 (3.0)	572 (2.9)	128 (2.7)	371 (2.9)	
Annual household income (US\$*):					0.05
<40 000	1349 (27.7)	5692 (28.7)	1384 (29.6)	3634 (28.9)	
40 000-74 999	1426 (29.3)	5803 (29.3)	1330 (28.4)	3630 (28.8)	
75 000-124 999	1219 (25.0)	4949 (24.9)	1174 (25.1)	3140 (24.9)	
125 000-199 999	495 (10.2)	1888 (9.5)	432 (9.2)	1213 (9.6)	
≥200 000	238 (4.9)	808 (4.1)	219 (4.7)	545 (4.3)	
Unknown or missing	148 (3.0)	699 (3.5)	144 (3.1)	427 (3.4)	
Mean (SD) baseline HbA <sub>1c</sub> concentration (%)	8.3 (1.9)	8.3 (1.9)	8.1 (1.8)	8.3 (1.9)	0.07
Baseline HbA <sub>1c</sub> concentration (%):					0.08
≤6.4	682 (14.0)	2531 (12.8)	681 (14.5)	1654 (13.1)	
6.5-6.9	533 (10.9)	2292 (11.6)	552 (11.8)	1456 (11.6)	
7-7.9	1374 (28.2)	5637 (28.4)	1371 (29.3)	3613 (28.7)	
8-8.9	897 (18.4)	3729 (18.8)	876 (18.7)	2302 (18.3)	
9-9.9	530 (10.9)	2048 (10.3)	474 (10.1)	1335 (10.6)	
≥10	861 (17.7)	3601 (18.2)	731 (15.6)	2229 (17.7)	
Mean (SD) baseline creatinine concentration†	0.9 (2.5)	1.0 (3.7)	0.9 (0.3)	1.0 (2.9)	0.04
Baseline comorbidities:					
Nephropathy	589 (12.1)	2792 (14.1)	642 (13.7)	1745 (13.9)	0.06
Retinopathy	274 (5.6)	1037 (5.2)	252 (5.4)	669 (5.3)	0.01
Neuropathy	649 (13.3)	2722 (13.7)	584 (12.5)	1743 (13.8)	0.04
Hyperglycaemia	<11‡	19 (0.1)	<11‡	<11‡	0.04
Hypoglycaemia	0 (0.0)	18 (0.1)	0 (0.0)	<11‡	0.04
Coronary artery disease	615 (12.6)	2733 (13.8)	611 (13.0)	1735 (13.8)	0.03
Chronic kidney disease	262 (5.4)	1360 (6.9)	304 (6.5)	849 (6.7)	0.06
Cerebrovascular disease	186 (3.8)	970 (4.9)	219 (4.7)	642 (5.1)	0.06
Peripheral vascular disease	300 (6.1)	1374 (6.9)	311 (6.6)	885 (7.0)	0.04
Chronic obstructive pulmonary disease	66 (1.4)	409 (2.1)	85 (1.8)	273 (2.2)	0.06
Smoking	333 (6.8)	1423 (7.2)	329 (7.0)	900 (7.2)	0.01
Baseline drug treatments:					
Statin	2929 (60.1)	12 027 (60.6)	2737 (58.4)	7613 (60.5)	0.04
Non-statin lipid lowering drugs	505 (10.4)	2055 (10.4)	515 (11.0)	1333 (10.6)	0.02
ACE inhibitor or angiotensin receptor blocker	2888 (59.2)	12 052 (60.7)	2780 (59.3)	7587 (60.3)	0.03
Sacubitril or valsartan	11 (0.2)	40 (0.2)	<11‡	28 (0.2)	0.03
Warfarin	88 (1.8)	464 (2.3)	121 (2.6)	278 (2.2)	0.05
Direct oral anticoagulant agents	159 (3.3)	626 (3.2)	113 (2.4)	415 (3.3)	0.05
Peripheral neuropathy drugs	571 (11.7)	2464 (12.4)	556 (11.9)	1603 (12.7)	0.03
Index year:					0.13
2013	263 (5.4)	1276 (6.4)	310 (6.6)	816 (6.5)	
2014	461 (9.4)	1962 (9.9)	497 (10.6)	1279 (10.2)	
2015	579 (11.9)	2137 (10.8)	487 (10.4)	1363 (10.8)	
2016	714 (14.6)	2745 (13.8)	651 (13.9)	1702 (13.5)	

Continued



Table 1 Continued

Characteristics	Canagliflozin (n=4876)	Glimepiride (n=19 838)	Liraglutide (n=4684)	Sitagliptin (n=12 588)	Largest SMD
2017	814 (16.7)	3206 (16.2)	772 (16.5)	2029 (16.1)	
2018	655 (13.4)	2778 (14.0)	680 (14.5)	1792 (14.2)	
2019	684 (14.0)	2733 (13.8)	634 (13.5)	1741 (13.8)	
2020	655 (13.4)	2495 (12.6)	568 (12.1)	1565 (12.4)	
2021	52 (1.1)	506 (2.6)	85 (1.8)	301 (2.4)	

Data are number (%), unless otherwise specified.

ACE=angiotensin converting enzyme; HbA<sub>1c</sub>=haemoglobin A<sub>1c</sub>; SMD=standardised mean difference.

\*US\$1=€0.77, €0.89.

†Baseline creatinine concentration was not included in the propensity scores model because of missing values.

‡Cell suppression based on OptumLabs cell size suppression rules. n<11 are masked to protect confidentiality of participants.

time for the proportional hazards model was set as three months after the index date because the primary outcome can only be seen starting at the third month. The proportional hazards assumption was assessed with the Schoenfeld residuals for each model. We found significant violation for proportional hazards assumption in the models of primary and secondary metabolic failure outcomes. To deal with the violation of the proportional hazards assumption, we used three time periods (0-40, 41-365, and >365 days after the index date) for baseline values of HbA<sub>1c</sub> and two periods (0-365 and >365 days after the index date) for the treatment groups for the pairwise comparisons. The boundaries for the time periods were selected based on visual inspection of the Schoenfeld residuals. For the subgroup and falsification endpoint analysis, because no significant proportional hazards violation existed, we used one period for the treatment groups. All pairwise comparisons between the treatment groups were estimated and we applied the Holm method to adjust the P values for multiple testing with an omnibus test for the hazard ratio not equal to one in at least one time period versus one in all time periods. The at-risk start time for modelling secondary metabolic (except for those that are HbA<sub>1c</sub> related), cardiovascular, and microvascular disease outcomes was set at the study index date. Follow-up time by treatment group was estimated with the inverse probability of treatment weighted Kaplan-Meier method for the censoring distribution.

Results are presented as median times to metabolic failure and the expected proportions of participants with metabolic failure at one and two years by treatment group, and pairwise hazard ratios. P<0.05 was considered significant for all two sided tests. All analyses were performed with SAS 9.4 (SAS Institute, Cary, NC) and R version 4.0.2. (R Foundation, Vienna, Austria). Online supplemental appendix 2 has sample R code for propensity score modelling and outcome comparisons.

### Subgroup analyses

A priori defined subgroup analyses were performed based on baseline HbA<sub>1c</sub> values (<7.0%, 7.0-7.9%, 8.0-8.9%, and ≥9.0%).

### Sensitivity analyses

Firstly, we repeated the analyses for primary and secondary metabolic failure with the per protocol censoring approach, with participants followed until the outcome of interest was reached, the study drug was discontinued (defined as not refilling a prescription within 30 days after the end of the last treatment episode), the anticipated follow-up duration of the trial was reached (seven years), the end of the study period (30 June 2021), the end of insurance coverage, or death (online supplemental figure S2). Secondly, to examine the comparative effectiveness of study drugs while treated only with these drugs and not with any other drug treatments for diabetes, accounting for real world treatment practices, we repeated all analyses with the as treated censoring approach, where participants were followed until the outcome of interest occurred, the study drug was discontinued, any other drug was added, the anticipated follow-up duration of the trial was reached (seven years), the end of the study period (30 June 2021), the end of insurance coverage, or death. Thirdly, we assessed residual confounding by testing a falsification endpoint that was unlikely to be associated with the studied drugs: diagnoses of pneumonia, cholecystitis, and appendicitis (online supplemental table S4) during the follow-up period.

### Patient and public involvement

Patients were not directly involved in the design, conduct, or dissemination of this study, although people living with type 2 diabetes involved in another ongoing study led by the senior author (RGM) stressed the need for contemporary information about the relative effectiveness of glucose lowering drugs on HbA<sub>1c</sub> levels specifically. This study was also informed by the calls by clinicians, professional societies, regulatory bodies, and payors to identify preferred glucose lowering treatment strategies

in the absence of direct comparisons across the studied drugs and to examine whether and how data collected in the process of routine patient care can be used to emulate and augment the evidence obtained through prospective clinical trials. Study findings cannot be shared directly with participants because patients included in OptumLabs Data Warehouse are de-identified; however, results will be disseminated with patient and public communities via scientific, social media, and institutional communication channels.

**Results**

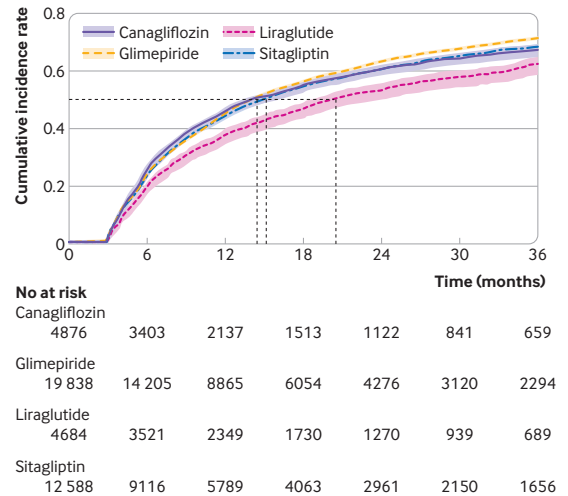
**Study population**

We identified 20 511 adults with type 2 diabetes who started glimepiride, 13 039 who started sitagliptin, 5569 who started liraglutide, 7262 who started insulin glargine, and 5290 who started canagliflozin (online supplemental figure S1). Online supplemental table S7 shows the baseline characteristics of participants before weighting. We found substantial differences (largest standardised mean difference >0.2) in age, race or ethnic group, annual household income, baseline levels of HbA<sub>1c</sub>, and comorbidities across the four treatment groups. Participants in the liraglutide arm were more likely to be younger, white, and have a higher income than those in the other treatment arms. Participants in the insulin glargine arm were most likely to have the lowest income and the highest prevalence of all of the comorbidities examined.

The insulin glargine arm was excluded from all analyses because of the imbalance in baseline variables after applying propensity scores weighting (online supplemental table S7) and statistically significant differences for the falsification endpoint tests for the insulin glargine pairwise comparisons with the other four arms (all pP<0.05; data not shown). After propensity score weighting, mean age was 60.7 (standard deviation 12.5) years in the canagliflozin arm, 61.5 (12.7) years in the glimepiride arm, 60.5 (12.3) years in the liraglutide arm, and 61.5 (12.7) years in the sitagliptin arm (table 1). Women comprised 47.7%, 47.8%, 49.9%, and 48.4% of the canagliflozin, glimepiride, liraglutide, and sitagliptin treatment arms, respectively. White participants comprised 61.3%, 60.6%, 62.8%, and 60.4% of the treatment arms, respectively. Mean baseline HbA<sub>1c</sub> levels were 8.3% (standard deviation 1.9), 8.3% (1.9), 8.1% (1.8), and 8.3% (1.9) in the canagliflozin, glimepiride, liraglutide, and sitagliptin arms, respectively. Online supplemental table S8 summarises the distribution of the exclusion criteria from GRADE across the treatment arms. All standardised mean differences were <0.1, except index year.

**Primary and secondary metabolic failure**

Median follow-up until intention-to-treat censoring was 885 (95% confidence interval 850 to 929) days



**Figure 1 | Cumulative incidence rate of primary metabolic failure by treatment arm (intention-to-treat approach). Primary metabolic failure was defined as time to first haemoglobin A<sub>1c</sub> concentration of ≥7.0%**

in the canagliflozin arm, 871 (846 to 894) days in the glimepiride arm, 853 (810 to 917) days in the liraglutide arm, and 883 (852 to 910) days in the sitagliptin arm (online supplemental figure S4). Median times to primary metabolic failure were 439 (95% confidence interval 400 to 489) days in the canagliflozin arm, 439 (426 to 453) days in the glimepiride arm, 624 (567 to 731) days in the liraglutide arm, and 461 (442 to 482) days in the sitagliptin arm (figure 1 and online supplemental table S9). Median time to secondary metabolic failure was also longest in the liraglutide arm (online supplemental figure S5 and online supplemental table S9). To ensure that we had adequate sample sizes to accommodate subgroup analyses for all metabolic outcomes, we measured the weighted number of events per treatment arm (online supplemental table S10).

In the Kaplan-Meier analysis, liraglutide was more effective in delaying the time to both primary and secondary metabolic failure than the other drugs (table 2). At one year, the estimated cumulative incidence rate of primary metabolic failure was 0.46 (95% confidence interval 0.44 to 0.48) in the canagliflozin arm, 0.45 (0.44 to 0.45) in the glimepiride arm, 0.37 (0.35 to 0.40) in the liraglutide arm, and 0.44 (0.43 to 0.45) in the sitagliptin arm. Similarly, the one year cumulative incidence rate of secondary metabolic failure was 0.27 (0.25 to 0.29) in the canagliflozin arm, 0.28 (0.27 to 0.29) in the glimepiride arm, 0.23 (0.21 to 0.26) in the liraglutide arm, and 0.28 (0.27 to 0.29) in the sitagliptin arm. These trends in cumulative incidence rates persisted at two years for both primary and secondary metabolic failures.

Because the proportional hazards assumption was not met, indicating that the hazard ratios of the different drugs changed over time, we performed pairwise comparisons between the drug classes

**Table 2 | Estimated cumulative incidence rates of primary and secondary metabolic failure by treatment arm (intention-to-treat approach)**

Metabolic failure	Cumulative incidence rate (95% CI)	
	1 year	2 years
Primary metabolic failure (HbA <sub>1c</sub> ≥7%)		
Canagliflozin	0.46 (0.44 to 0.48)	0.60 (0.59 to 0.62)
Glimepiride	0.45 (0.44 to 0.45)	0.63 (0.62 to 0.64)
Liraglutide	0.37 (0.35 to 0.40)	0.53 (0.50 to 0.56)
Sitagliptin	0.44 (0.43 to 0.45)	0.60 (0.59 to 0.61)
Secondary metabolic failure (HbA <sub>1c</sub> >7.5%)		
Canagliflozin	0.27 (0.25 to 0.29)	0.40 (0.38 to 0.42)
Glimepiride	0.28 (0.27 to 0.29)	0.44 (0.43 to 0.45)
Liraglutide	0.23 (0.21 to 0.26)	0.37 (0.34 to 0.40)
Sitagliptin	0.28 (0.27 to 0.29)	0.43 (0.42 to 0.44)

CI=confidence interval; HbA<sub>1c</sub>=haemoglobin A<sub>1c</sub>.

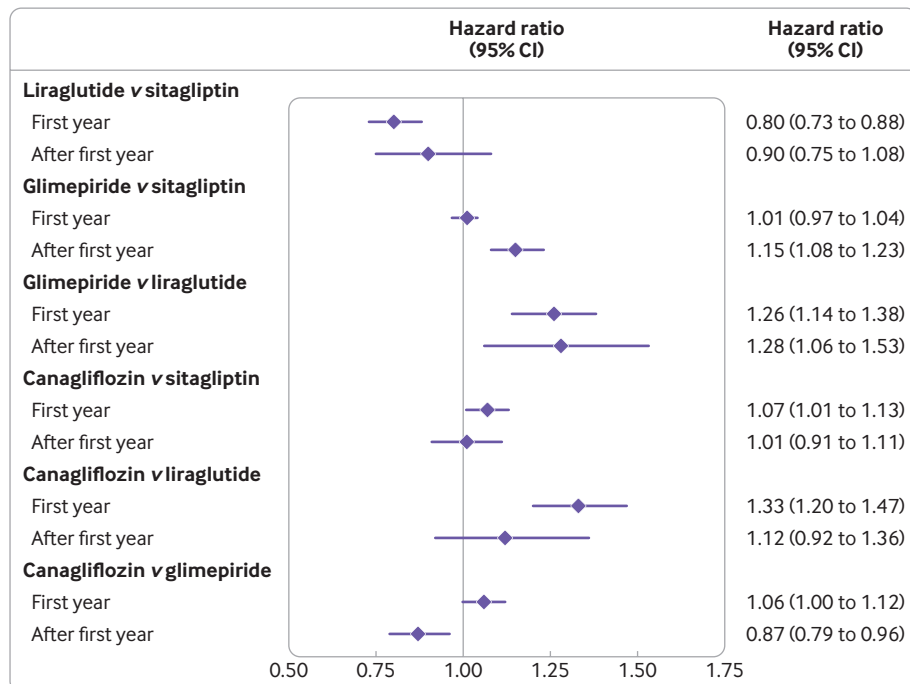
separately for the two time periods (figure 2 and online supplemental table S11). We found that for glimepiride versus sitagliptin, glimepiride was more likely to achieve primary and secondary metabolic failure during later years of treatment, but we found no significant difference between the two drugs during the first year of treatment. For liraglutide versus sitagliptin, liraglutide was less likely to achieve primary and secondary metabolic failure

during the first year of treatment, with no significant difference between the drugs in subsequent years. For canagliflozin versus glimepiride, canagliflozin was less likely to achieve primary and secondary metabolic failure in later years of treatment, with no significant difference between the drugs during the first year of treatment. Canagliflozin and glimepiride were both more likely to achieve primary metabolic failure than liraglutide in the first year, with a consistent effect in the subsequent years for glimepiride and no difference for canagliflozin.

**Other secondary outcomes**

Insulin was started by 410 (8.4%) participants in the canagliflozin arm, 1900 (9.6%) in the glimepiride arm, 679 (14.5%) in the liraglutide arm, and 1307 (10.4%) in the sitagliptin arm. Online supplemental table S11 presents pairwise comparisons for starting insulin (ie, tertiary metabolic failure). Overall, 323 patients/participants had visits to the emergency department or were admitted to hospital for hypoglycaemia during the study period, including <11 individuals in the liraglutide arm, precluding formal statistical analyses.

Online supplemental table S12 shows event rates for all other outcomes. Compared with canagliflozin, glimepiride had a higher risk for major adverse cardiovascular events (online supplemental table S13). Glimepiride had higher risks for all cause mortality and admission to hospital than sitagliptin and canagliflozin. Liraglutide had a lower risk for all cause mortality than sitagliptin. We found no significant differences between the groups for end stage kidney



**Figure 2 | Pairwise comparisons of treatment effects on primary metabolic failure across different time periods. CI=confidence interval**

disease, retinopathy, neuropathy, other cardiovascular events, heart failure, pancreatitis, pancreatic and thyroid cancer, or cancer.

### Subgroup and sensitivity analyses

We examined the comparative risks of primary metabolic failure in subgroups of baseline levels of HbA<sub>1c</sub> and found larger hazard ratios for all pairwise comparisons at lower HbA<sub>1c</sub> levels, with the exception of canagliflozin and sitagliptin where no difference was seen (online supplemental table S14). Online supplemental table S15 shows event rates by subgroup, suggesting that sample sizes were likely adequate for these secondary analyses; formal power calculations were not conducted. Results were mostly similar for secondary metabolic failure.

We also conducted sensitivity analyses with the per protocol (online supplemental figure S6 and online supplemental table S16) and as treated (online supplemental figure S7 and online supplemental table S17) censoring approaches. Another glucose lowering drug was added before discontinuation of the assigned treatment in 727 (15%) participants in the canagliflozin arm, 2618 (13%) in the glimepiride arm, 1374 (29%) in the liraglutide arm, and 2252 (18%) in the sitagliptin arm. Results of the sensitivity analyses were consistent with the primary analyses. We found no significant differences among the treatment groups in the pneumonia, cholecystitis, and appendicitis falsification endpoints (online supplemental table S18).

## Discussion

### Principal findings

We used the target trial framework to emulate an expanded adaptation of the GRADE trial, based on observational data for a diverse population of participants treated under usual care conditions. Comparing the start of treatment with glimepiride, sitagliptin, liraglutide, or canagliflozin as second line agents in achieving and maintaining glycaemic control among adults with type 2 diabetes, we found that liraglutide was associated with a longer time to both primary (HbA<sub>1c</sub> >7.0%) and secondary (HbA<sub>1c</sub> ≥7.5%) metabolic failure than the other agents. We found no difference between the study arms in the rates of most microvascular and macrovascular complications, with two notable exceptions. First, glimepiride was associated with higher risk of major adverse cardiovascular events compared to canagliflozin and with higher risk of all cause mortality and hospital admission compared to either sitagliptin or canagliflozin. Second, liraglutide was associated with lower risk for all cause mortality compared to sitagliptin. Because a prospective randomised controlled trial comparing these drugs head to head is unlikely to be conducted, our findings fill an important knowledge gap in the clinical management of type 2 diabetes and highlight the potential for healthcare data generated as part of

routine medical practice to provide important and timely insights about the comparative effectiveness and safety of commonly used drugs.

The six pairwise comparisons of glimepiride, sitagliptin, liraglutide, and canagliflozin showed that liraglutide was the most effective in achieving and maintaining HbA<sub>1c</sub> levels <7.0% and ≤7.5% in both the intention-to-treat and per protocol analyses. The greater effectiveness of liraglutide compared with other glucose lowering drugs is consistent with previous studies.<sup>28–34</sup> Our study's new contribution was the direct comparison of all four commonly used classes of glucose lowering drugs, including the sodium-glucose cotransporter 2 inhibitor canagliflozin. We found that liraglutide was the most effective, canagliflozin and sitagliptin were moderately effective, whereas glimepiride was least effective in delaying both primary and secondary metabolic failure. For the pairwise comparisons, the greater effectiveness of liraglutide compared with the other drugs was most pronounced in the first year of treatment, narrowing with continued use, and at lower baseline HbA<sub>1c</sub> values, narrowing as baseline levels of HbA<sub>1c</sub> increased. Among patients with baseline levels of HbA<sub>1c</sub> ≥9%, where clinical guidelines recommend combination or insulin treatment,<sup>9</sup> glimepiride was associated with earlier time to both primary and secondary metabolic failure than sitagliptin (signalling a potential inadequacy of pancreatic insulin secretion in response to treatment with sulphonylureas at high glucose levels) but no difference between the other three drug classes in reaching HbA<sub>1c</sub> values <7.0% or ≤7.5% at these high levels of HbA<sub>1c</sub>.

The risk of starting insulin (ie, tertiary metabolic failure) was highest in individuals treated with liraglutide, particularly during the first year of treatment, perhaps reflecting the greater potency of liraglutide (such that if people fail liraglutide, they are started on insulin rather than trying other non-insulin drugs) and injectable administration (ie, people already taking injectable drugs might have less hesitation in starting insulin than those treated with oral agents).

We found that glimepiride was associated with a significantly higher risk of death than sitagliptin (23% higher; P=0.002), liraglutide (131% higher; P<0.001), and canagliflozin (56% higher; P=0.008). The cause of these deaths is unknown and may be driven by hypoglycaemia, although rates of severe hypoglycaemic events requiring emergency department or hospital care were low in all groups. Previous data on the risk of mortality associated with the use of sulphonylureas have been inconsistent, with meta-analyses or randomised controlled trials finding no difference in the risk of mortality compared with placebo or active comparators,<sup>35</sup> but data from observational studies demonstrated increased risk.<sup>36–37</sup> Previous studies also did not compare



sulphonylureas with newer glucose lowering drugs, two of which (liraglutide and canagliflozin) have robust evidence suggesting benefits for cardiovascular mortality and all cause mortality. Liraglutide and canagliflozin have important cardiovascular, kidney, and metabolic benefits independent of their effect on glycaemic control, including reductions in cardiovascular events, cardiovascular death, progression of kidney disease, renal death, and hospital admission for heart failure.<sup>13 38</sup> Our findings therefore support clinical guideline recommendations to consider glucagon-like peptide 1 receptor agonists and sodium-glucose cotransporter 2 inhibitors as second line agents for most people with diabetes, with those with cardiovascular and kidney comorbidities particularly benefiting from their use, and cautious use of sulphonylureas if they cannot be avoided.

#### Strengths and limitations of this study

This work builds on the recently completed GRADE study<sup>16</sup> and on our emulation of the GRADE trial based on data from the OptumLabs Data Warehouse,<sup>21</sup> by including a sodium-glucose cotransporter 2 inhibitor comparator arm and broadening the eligibility criteria to ensure greater generalisability of the findings of the study. Sodium-glucose cotransporter 2 inhibitors, in common with glucagon-like peptide 1 receptor agonists, are increasingly recommended as second line, and even first line, glucose lowering agents because of their beneficial effects on the kidney, heart failure, cardiovascular disease, weight, and mortality outcomes.<sup>13</sup> Sodium-glucose cotransporter 2 inhibitors are less costly than glucagon-like peptide 1 receptor agonists, however, are given orally, and lack the gastrointestinal side effect profile of glucagon-like peptide 1 receptor agonists, making them highly suitable drugs in the management of type 2 diabetes. In the absence of head-to-head comparisons of the metabolic effects of sodium-glucose cotransporter 2 inhibitors and glucagon-like peptide 1 receptor agonists, patients and clinicians lack the information necessary to inform their decision making when choosing between these drugs. Similarly, for individuals with no cardiovascular or kidney comorbidities or risk factors, the choice of glucose lowering treatment is often driven by metabolic considerations, and these data had been lacking. Our finding of the greater glycaemic effectiveness of liraglutide relative to other second line glucose lowering drugs therefore highlights the advantages of using liraglutide, and likely other glucagon-like peptide 1 receptor agonists, in the management of type 2 diabetes.

Our study had several limitations. Firstly, despite the larger and more heterogeneous study population than our previous emulation of the GRADE

trial,<sup>21</sup> we could not adequately match people treated with insulin glargine with those treated with other second line drugs and these individuals had to be excluded from the pairwise comparisons. This problem represents one of the biggest limitations of the target trial framework because analyses are limited to interventions that are adequately represented in clinical practice. This limitation, however, does not adversely affect the generalisability of our findings because insulin glargine is not used in this clinical context in routine care (in fact, identifying a population that is equally likely to be treated with insulin glargine and the other agents as second line treatments would likely result in a non-generalisable cohort). Secondly, even with rigorous causal inference analytic methods, observational studies are subject to residual confounding. We sought to check for and mitigate this risk with several falsification endpoint analyses. Thus, although randomised controlled trials are a gold standard for evaluating the comparative efficacy and safety of interventions, observational data can be used to emulate idealised target trials when a randomised controlled trial is not feasible, practical, or ethical.<sup>39 40</sup> Thirdly, the frequency and timing of HbA<sub>1c</sub> tests in routine practice are influenced by many factors, including baseline levels of HbA<sub>1c</sub>, perceived risk of deterioration of glycaemic control, and the person's capacity to complete testing. Fourthly, HbA<sub>1c</sub> results are not available for all people in the OptumLabs Data Warehouse, such that our analyses used a convenience sample of participants who obtained their HbA<sub>1c</sub> test in a commercial laboratory company that provided data to OptumLabs. Also, not all factors that influence glycaemic control can be captured in real world data and therefore could not be accounted for in the analyses. In addition, drugs obtained outside of health insurance benefits (ie, through low cost generic drug programmes, patient assistance programmes, or as drug samples) would have been missed but we expect this practice to be rare in our cohort. Lastly, the study cohort was conducted in Americans with private health plans (both employer sponsored and Medicare Advantage), and the results might not be fully generalisable to people with public health plans, those with no insurance coverage, or those outside of the US.

#### Conclusions

In this target trial emulation of the GRADE study framework, we found that liraglutide was more effective in achieving and maintaining glycaemic control as a second line glucose lowering drug than canagliflozin, sitagliptin, or glimepiride for people with diabetes. Our findings suggest that observational data and methods can be used to emulate clinical trials and examine the comparative effectiveness and safety of interventions in routine care.

## AUTHOR AFFILIATIONS

<sup>1</sup>Robert D and Patricia E Kern Center for the Science of Health Care Delivery, Mayo Clinic, Rochester, Minnesota, USA

<sup>2</sup>OptumLabs, Eden Prairie, Minnesota, USA

<sup>3</sup>Department of Public Health Sciences, University of Chicago, Chicago, Illinois, USA

<sup>4</sup>Department of Epidemiology, Emory University School of Public Health, Atlanta, Georgia, USA

<sup>5</sup>Section of Cardiovascular Medicine, Department of Internal Medicine, Yale School of Medicine, New Haven, Connecticut, USA

<sup>6</sup>Department of Internal Medicine, Yale School of Medicine, New Haven, Connecticut, USA

<sup>7</sup>Center for Outcomes Research and Evaluation, Yale-New Haven Hospital, New Haven, Connecticut, USA

<sup>8</sup>Department of Health Policy and Management, Yale School of Public Health, New Haven, Connecticut, USA

<sup>9</sup>Division of Community Internal Medicine, Geriatrics, and Palliative Care, Department of Medicine, Mayo Clinic, Rochester, Minnesota, USA

**Twitter** Joshua D Wallach @JoshuaWallach, Joseph S Ross @jross119 and Rozalina G McCoy @RozalinaMD

**Contributors** YD curated and analysed the data, validated and visualised the results, and reviewed and edited the manuscript. ECP supervised the statistical analyses, co-developed the formal analysis plan, and reviewed and edited the manuscript. JDW, JH, and JSR contributed to study design, and reviewed and edited the manuscript. RGM co-developed the formal analysis plan, drafted the manuscript, secured funding, and supervised the study. YD and RGM had access to the data and are guarantors of this work. The corresponding author attests that all listed authors meet authorship criteria and that no others meeting the criteria have been omitted. Transparency: The lead authors (the guarantors) 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, if relevant, registered) have been explained.

**Funding** RGM is supported by the National Institute of Health (NIH), National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK) grant No K23DK114497. The funders had no involvement in any aspect of the study, including study design; data collection, analysis, or interpretation; manuscript preparation; or decision to submit it for publication. The contents are those of the authors and do not necessarily represent the official views of the NIH.

**Competing interests** All authors have completed the ICMJE uniform disclosure form at [www.icmje.org/disclosure-of-interest/](http://www.icmje.org/disclosure-of-interest/) and declare: support from the National Institute of Health (NIH), National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK), for the submitted work; RGM has received unrelated research support from NIDDK and the Patient-Centered Outcomes Research Institute (PCORI); RGM has also served as a consultant to Emmi on the development of patient education materials related to diabetes; JDW has received research support from the National Institute on Alcohol Abuse and Alcoholism of the NIH, the Food and Drug Administration (FDA), Johnson & Johnson, and the Collaboration for Research Integrity and Transparency (CRIT) at Yale University from the Arnold Ventures; JDW serves as a consultant for Hagens Berman Sobol Shapiro and Dugan Law Firm; JH works under contract to the Centers for Medicare and Medicaid Services on the development and evaluation of measures of provider quality; JH also receives research support from the National Institute on Aging, the National Cancer Institute, PCORI, the American Heart Association, and additional support from Yale University, Johnson & Johnson, and the Department of Defense; JSR currently receives research support through Yale University from Johnson & Johnson to develop methods of clinical trial data sharing, from the Medical Device Innovation Consortium as part of the National Evaluation System for Health Technology (NEST), from the FDA for the Yale-Mayo Clinic Center for Excellence in Regulatory Science and Innovation (CERSI) programme (U01FD005938), from the Agency for Healthcare Research and Quality (R01HS022882), from the National Heart, Lung, and Blood Institute of the NIH (R01HS025164, R01HL144644), and from Arnold Ventures; JSR was an expert witness at the request of relator's attorneys, the Greene Law Firm, in a qui tam suit alleging violations of the False Claims Act and Anti-Kickback Statute against Biogen that was settled September 2022; 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.

**Ethics approval** The study was exempt from approval by the Mayo Clinic institutional review board (IRB No 13-006907) because de-identified data were used in the study. Patient consent could not be obtained because the data were de-identified.

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

**Data availability statement** Data may be obtained from a third party and are not publicly available. This study was based on de-identified data from OptumLabs Data Warehouse and on a linked 100% sample of Medicare fee-for-service claims. These data are third party data owned by OptumLabs and contain sensitive patient information; therefore, the data are only available on request. Interested researchers engaged in HIPAA (Health Insurance Portability and Accountability Act) compliant research may contact [connected@optum.com](mailto:connected@optum.com) for data access requests. Data use requires researchers to pay for rights to use and access the data. These data are subject to restrictions on sharing as a condition of access.

**Supplemental material** This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

**Open access** This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

## ORCID iDs

Joshua D Wallach <http://orcid.org/0000-0002-2816-6905>

Joseph S Ross <http://orcid.org/0000-0002-9218-3320>

Rozalina G McCoy <http://orcid.org/0000-0002-2289-3183>

## REFERENCES

- 1 CDC. National diabetes statistics report. Atlanta: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, 2021.
- 2 Hemmingsen B, Lund SS, Gluud C, *et al*. Intensive glycaemic control for patients with type 2 diabetes: systematic review with meta-analysis and trial sequential analysis of randomised clinical trials. *BMJ* 2011;343:d6898. [10.1136/bmj.d6898](https://doi.org/10.1136/bmj.d6898)
- 3 Rodriguez-Gutierrez R, Gonzalez-Gonzalez JG, Zuñiga-Hernandez JA, *et al*. Benefits and harms of intensive glycaemic control in patients with type 2 diabetes. *BMJ* 2019;367:15887. [10.1136/bmj.15887](https://doi.org/10.1136/bmj.15887)
- 4 Schmittiel J, Uratsu CS, Karter AJ, *et al*. Why don't diabetes patients achieve recommended risk factor targets? Poor adherence versus lack of treatment intensification. *J Gen Intern Med* 2008;23:588–94. [10.1007/s11606-008-0554-8](https://doi.org/10.1007/s11606-008-0554-8)
- 5 Shichiri M, Kishikawa H, Ohkubo Y, *et al*. Long-term results of the Kumamoto study on optimal diabetes control in type 2 diabetic patients. *Diabetes Care* 2000;23:B21–9.
- 6 Turner RC, Cull CA, Frighi V, *et al*. Glycemic control with diet, sulfonylurea, metformin, or insulin in patients with type 2 diabetes mellitus: progressive requirement for multiple therapies (UKPDS 49). *JAMA* 1999;281:2005–12. [10.1001/jama.281.21.2005](https://doi.org/10.1001/jama.281.21.2005)
- 7 Adler AI, Stratton IM, Neil HA, *et al*. Association of systolic blood pressure with macrovascular and microvascular complications of type 2 diabetes (UKPDS 36): prospective observational study. *BMJ* 2000;321:412–9. [10.1136/bmj.321.7258.412](https://doi.org/10.1136/bmj.321.7258.412)
- 8 Holman RR, Paul SK, Bethel MA, *et al*. 10-year follow-up of intensive glucose control in type 2 diabetes. *N Engl J Med* 2008;359:1577–89. [10.1056/NEJMoa0806470](https://doi.org/10.1056/NEJMoa0806470)
- 9 ADA. American Diabetes Association standards of medical care in diabetes—2021. Section 6. Glycemic targets. *Diabetes Care* 2021;44:S73–84. [10.2337/dc21-S006](https://doi.org/10.2337/dc21-S006)
- 10 NICE. National Institute for Health and Care Excellence pathways: managing blood glucose in adults with type 2 diabetes. National Institute for Health and Care Excellence; 2015.
- 11 Garber AJ, Abrahamson MJ, Barzilay JL, *et al*. Consensus statement by the American Association of Clinical Endocrinologists and American

- College of Endocrinology on the comprehensive type 2 diabetes management algorithm – 2019 executive summary. *Endocr Pract* 2019;25:69–100. 10.4158/CS-2018-0535
- 12 Conlin PR, Colburn J, Aron D, *et al.* Synopsis of the 2017 U.S. Department of Veterans Affairs/U.S. Department of Defense Clinical Practice guideline: management of type 2 diabetes mellitus. *Ann Intern Med* 2017;167:655–63. 10.7326/M17-1362
  - 13 ADA. American Diabetes Association standards of medical care in diabetes—2021. Section 9. *Diabetes Care* 2021;44:S111–24. 10.2337/dc21-S009
  - 14 Nathan DM, Buse JB, Kahn SE, *et al.* Rationale and design of the glycemia reduction approaches in diabetes: a comparative effectiveness study (GRADE). *Diabetes Care* 2013;36:2254–61. 10.2337/dc13-0356
  - 15 ClinicalTrials.gov. A comparative effectiveness study of major glycemia-lowering medications for treatment of type 2 diabetes (GRADE). ClinicalTrials.gov Identifier: NCT01794143. U.S. National Library of Medicine. National Institutes of Health. Department of Health and Human Services; 2021.
  - 16 The GRADE Study Research Group. Glycemia reduction in type 2 diabetes—glycemic outcomes. *N Engl J Med* 2022;387:1063–74. 10.1056/NEJMoa2200433
  - 17 Fang M, Wang D, Coresh J, *et al.* Trends in diabetes treatment and control in U.S. adults. *N Engl J Med* 2021;384:2219–28. 10.1056/NEJMs2032271
  - 18 McCoy RG, Van Houten HK, Karaca-Mandic P, *et al.* Second-line therapy for type 2 diabetes management: the treatment/benefit paradox of cardiovascular and kidney comorbidities. *Diabetes Care* 2021;44:2302–11. 10.2337/dc20-2977
  - 19 McCoy RG, Dykhoff HJ, Sangaralingham L, *et al.* Adoption of new glucose-lowering medications in the U.S. - the case of SGLT2 inhibitors: nationwide cohort study. *Diabetes Technol Ther* 2019;21:702–12. 10.1089/dia.2019.0213
  - 20 Wexler DJ, Krause-Steinrauf H, Crandall JP, *et al.* Baseline characteristics of randomized participants in the glycemia reduction approaches in diabetes: a comparative effectiveness study (GRADE). *Diabetes Care* 2019;42:2098–107. 10.2337/dc19-0901
  - 21 Deng Y, Polley EC, Wallach JD, *et al.* Emulating the GRADE trial using real world data: retrospective comparative effectiveness study. *BMJ* 2022;379:e070717. 10.1136/bmj-2022-070717
  - 22 Hernán MA, Robins JM. Using big data to emulate a target trial when a randomized trial is not available. *Am J Epidemiol* 2016;183:758–64. 10.1093/aje/kwv254
  - 23 Wallace PJ, Shah ND, Dennen T, *et al.* Building a novel node in the learning health care system. *Health Aff (Millwood)* 2014;33:1187–94. 10.1377/hlthaff.2014.0038
  - 24 OptumLabs. OptumLabs and OptumLabs Data Warehouse (OLDW) descriptions and citation. Cambridge, MA, 2019.
  - 25 McCaffrey DF, Griffin BA, Almirall D, *et al.* A tutorial on propensity score estimation for multiple treatments using generalized boosted models. *Stat Med* 2013;32:3388–414. 10.1002/sim.5753
  - 26 Xu S, Ross C, Raebel MA, *et al.* Use of stabilized inverse propensity scores as weights to directly estimate relative risk and its confidence intervals. *Value Health* 2010;13:273–7. 10.1111/j.1524-4733.2009.00671.x
  - 27 Austin PC. Balance diagnostics for comparing the distribution of baseline covariates between treatment groups in propensity-score matched samples. *Stat Med* 2009;28:3083–107. 10.1002/sim.3697
  - 28 Khunti K, Charbonnel B, Cooper A, *et al.* Associations between second-line glucose-lowering combination therapies with metformin and HbA1c, body weight, quality of life, hypoglycaemic events and glucose-lowering treatment intensification: the DISCOVER study. *Diabetes Obes Metab* 2021;23:1823–33. 10.1111/dom.14400
  - 29 Tsapas A, Avgerinos I, Karagiannis T, *et al.* Comparative effectiveness of glucose-lowering drugs for type 2 diabetes: a systematic review and network meta-analysis. *Ann Intern Med* 2020;173:278–86. 10.7326/M20-0864
  - 30 Tran S, Retnakaran R, Zinman B, *et al.* Efficacy of glucagon-like peptide-1 receptor agonists compared to dipeptidyl peptidase-4 inhibitors for the management of type 2 diabetes: a meta-analysis of randomized clinical trials. *Diabetes Obes Metab* 2018;20:68–76. 10.1111/dom.13137
  - 31 Palmer SC, Mavridis D, Nicolucci A, *et al.* Comparison of clinical outcomes and adverse events associated with glucose-lowering drugs in patients with type 2 diabetes: a meta-analysis. *JAMA* 2016;316:313–24. 10.1001/jama.2016.9400
  - 32 Lingvay I, Capehorn MS, Catarig A-M, *et al.* Efficacy of once-weekly semaglutide vs empagliflozin added to metformin in type 2 diabetes: patient-level meta-analysis. *J Clin Endocrinol Metab* 2020;105:e4593–604. 10.1210/clinem/dgaa577
  - 33 Jia S, Wang Z, Han R, *et al.* Incretin mimetics and sodium-glucose co-transporter 2 inhibitors as monotherapy or add-on to metformin for treatment of type 2 diabetes: a systematic review and network meta-analysis. *Acta Diabetol* 2021;58:5–18. 10.1007/s00592-020-01542-4
  - 34 Patoulias D, Katsimardou A, Kalogirou M-S, *et al.* Glucagon-like peptide-1 receptor agonists or sodium-glucose cotransporter-2 inhibitors as add-on therapy for patients with type 2 diabetes? A systematic review and meta-analysis of surrogate metabolic endpoints. *Diabetes Metab* 2020;46:272–9. 10.1016/j.diabet.2020.04.001
  - 35 Varvaki Rados D, Catani Pinto L, Reck Remonti L, *et al.* The association between sulfonylurea use and all-cause and cardiovascular mortality: a meta-analysis with trial sequential analysis of randomized clinical trials. *PLoS Med* 2016;13:e1001992. 10.1371/journal.pmed.1001992
  - 36 Azoulay L, Suissa S. Sulfonylureas and the risks of cardiovascular events and death: a methodological meta-regression analysis of the observational studies. *Diabetes Care* 2017;40:706–14. 10.2337/dc16-1943
  - 37 Forst T, Hanefeld M, Jacob S, *et al.* Association of sulphonylurea treatment with all-cause and cardiovascular mortality: a systematic review and meta-analysis of observational studies. *Diab Vasc Dis Res* 2013;10:302–14. 10.1177/1479164112465442
  - 38 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. 10.1136/bmj.m4573
  - 39 Danaei G, García Rodríguez LA, Cantero OF, *et al.* Electronic medical records can be used to emulate target trials of sustained treatment strategies. *J Clin Epidemiol* 2018;96:12–22. 10.1016/j.jclinepi.2017.11.021
  - 40 Wang SV, Schneeweiss S, RCT-DUPLICATE Initiative, *et al.* Emulation of randomized clinical trials with nonrandomized database analyses: results of 32 clinical trials. *JAMA* 2023;329:1376–85. 10.1001/jama.2023.4221
- Additional supplemental material is published online only. To view, please visit the journal online (<http://dx.doi.org/10.1136/bmjmed-2022-000419>).