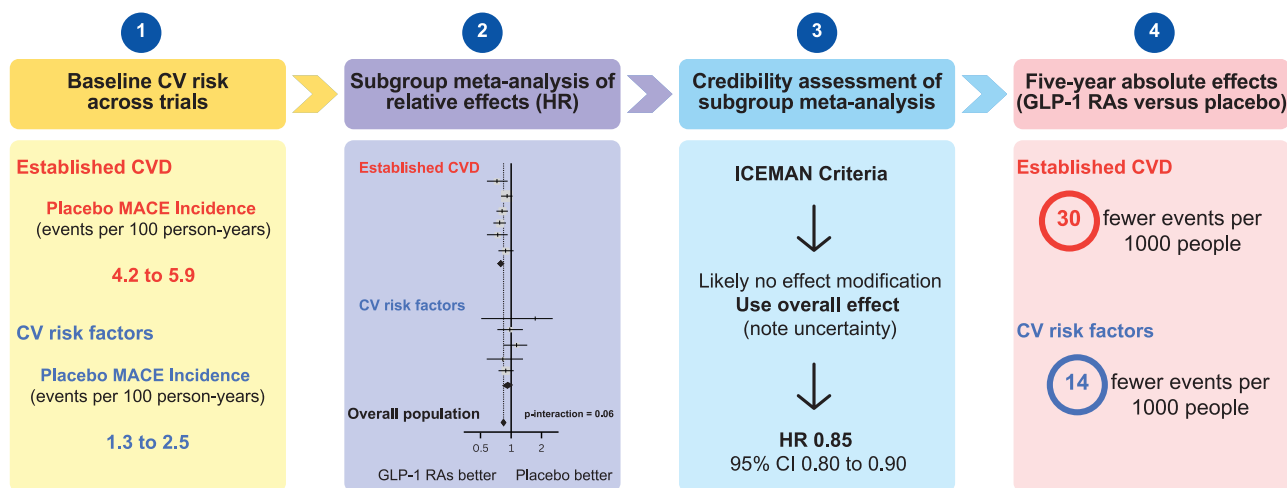


## A Methodological Framework for Meta-analysis and Clinical Interpretation of Subgroup Data: The Case of Major Adverse Cardiovascular Events With GLP-1 Receptor Agonists and SGLT2 Inhibitors in Type 2 Diabetes

Thomas Karagiannis, Apostolos Tsapas, Eleni Bekiari, Konstantinos A. Toulis, and Michael A. Nauck

Diabetes Care 2024;47(2):184–192 | <https://doi.org/10.2337/dc23-0925>

### Methodological Framework for Subgroup Meta-analysis Effect of GLP-1 RAs on MACE based on baseline CV risk



CI: confidence interval, CV: cardiovascular, CVD: cardiovascular disease, GLP-1 RAs: glucagon-like peptide 1 receptor agonists, HR: hazard ratio, ICEMAN: Instrument for assessing the Credibility of Effect Modification Analyses, MACE: major adverse cardiovascular event

### ARTICLE HIGHLIGHTS

• **Why did we undertake this study?**

We present a methodological framework for conducting and interpreting subgroup meta-analyses, using the case of glucagon-like peptide 1 receptor agonists and sodium–glucose cotransporter 2 inhibitors in cardiovascular outcomes trials as a practical example.

• **What is the specific question(s) we wanted to answer?**

Methodological steps comprised evaluation of clinical heterogeneity, credibility assessment of subgroup meta-analysis, and translation of relative into absolute treatment effects.

• **What did we find?**

Absolute benefits for major adverse cardiovascular events were approximately twofold higher in patients with established cardiovascular disease compared with those with indicators of high cardiovascular risk only, reflecting differences in baseline cardiovascular risk.

• **What are the implications of our findings?**

This framework can be applied to subgroup meta-analyses regardless of outcomes or modification variables.



# A Methodological Framework for Meta-analysis and Clinical Interpretation of Subgroup Data: The Case of Major Adverse Cardiovascular Events With GLP-1 Receptor Agonists and SGLT2 Inhibitors in Type 2 Diabetes

Thomas Karagiannis,<sup>1,2</sup>  
Apostolos Tsapas,<sup>1,2,3</sup> Eleni Bekiari,<sup>1,2</sup>  
Konstantinos A. Toulis,<sup>4,5</sup> and  
Michael A. Nauck<sup>6</sup>

*Diabetes Care* 2024;47:184–192 | <https://doi.org/10.2337/dc23-0925>

We present a methodological framework for conducting and interpreting subgroup meta-analyses. Methodological steps comprised evaluation of clinical heterogeneity regarding the definition of subpopulations, credibility assessment of subgroup meta-analysis, and translation of relative into absolute treatment effects. We used subgroup data from type 2 diabetes cardiovascular outcomes trials (CVOTs) with glucagon-like peptide 1 (GLP-1) receptor agonists and sodium–glucose cotransporter 2 (SGLT2) inhibitors for patients with established cardiovascular disease and those at high cardiovascular risk without manifest cardiovascular disease. First, we evaluated the variability in definitions of the subpopulations across CVOTs using major adverse cardiovascular events (MACE) incidence in the placebo arm as a proxy for baseline cardiovascular risk. As baseline risk did not differ considerably across CVOTs, we conducted subgroup meta-analyses of hazard ratios (HRs) for MACE and assessed the credibility of a potential effect modification. Results suggested using the same overall relative effect for each of the two subpopulations (HR 0.85, 95% CI 0.80–0.90, for GLP-1 receptor agonists and HR 0.91, 95% CI 0.85–0.97, for SGLT2 inhibitors). Finally, we calculated 5-year absolute treatment effects (number of fewer patients with event per 1,000 patients). Treatment with GLP-1 receptor agonists resulted in 30 fewer patients with event in the subpopulation with established cardiovascular disease and 14 fewer patients with event in patients without manifest cardiovascular disease. For SGLT2 inhibitors, the respective absolute effects were 18 and 8 fewer patients with event per 1,000 patients. This framework can be applied to subgroup meta-analyses regardless of outcomes or modification variables.

Glucagon-like peptide 1 (GLP-1) receptor agonists and sodium–glucose cotransporter 2 (SGLT2) inhibitors are recommended for patients with type 2 diabetes at increased risk for cardiovascular complications (1,2). These recommendations are based on the findings of placebo-controlled cardiovascular outcomes trials (CVOTs) that have

<sup>1</sup>Clinical Research and Evidence-Based Medicine Unit, Second Medical Department, Aristotle University of Thessaloniki, Thessaloniki, Greece

<sup>2</sup>Diabetes Centre, Second Medical Department, Aristotle University of Thessaloniki, Thessaloniki, Greece

<sup>3</sup>Harris Manchester College, University of Oxford, Oxford, U.K.

<sup>4</sup>Department of Endocrinology, 424 Military Hospital, Thessaloniki, Greece

<sup>5</sup>Institute of Applied Health Research, University of Birmingham, Birmingham, U.K.

<sup>6</sup>Diabetes, Endocrinology, and Metabolism Section, Medical Department I, Katholisches Klinikum Bochum, St. Josef-Hospital, Ruhr University Bochum, Bochum, Germany

Corresponding author: Apostolos Tsapas, [atsapas@auth.gr](mailto:atsapas@auth.gr)

Received 22 May 2023 and accepted 13 September 2023

This article contains supplementary material online at <https://doi.org/10.2337/figshare.24215310>.

© 2024 by the American Diabetes Association. Readers may use this article as long as the work is properly cited, the use is educational and not for profit, and the work is not altered. More information is available at <https://www.diabetesjournals.org/journals/pages/license>.

demonstrated cardioprotective effects of individual agents within these drug classes. Some of these studies exclusively recruited patients with established cardiovascular disease (3–6), while most CVOTs (7–14) also included participants at high cardiovascular risk but without manifest cardiovascular disease (i.e., with cardiovascular risk factors only or subclinical cardiovascular disease). An important clinical question is whether outcomes of treatment with GLP-1 receptor agonists or SGLT2 inhibitors differ between these two subpopulations. This question has been addressed in individual CVOTs through subgroup analyses comparing the hazard ratio (HR) for participants with established cardiovascular disease with the HR of the subgroup without cardiovascular disease. These data have also been pooled in meta-analyses exploring the presence of an effect modification between the two subgroups (15–17).

Subgroup meta-analyses can provide valuable insights for tailoring treatment decisions to specific patient groups based on clinically important characteristics. However, they should be approached and implemented with caution to prevent potential misinterpretation or unwarranted generalizations. Prior to conducting a subgroup meta-analysis, it is important to evaluate the clinical heterogeneity across included trials with regard to the baseline risk for the specific outcome within each of the two subpopulations of interest. In the case of CVOTs, a strategy to examine the consistency of baseline cardiovascular risk is to compare the incidence of the outcome of interest for placebo (18) for each subpopulation across trials. Moreover, to mitigate clinical heterogeneity, in performing the subgroup meta-analysis it is preferable to use outcome data that refer to consistent definitions of subpopulations across included trials. The interpretation of subgroup analyses typically relies on the  $P$  value for the test of interaction between subgroups ( $P_{\text{interaction}}$ ). However, even though tests for interaction usually lack sufficient power to detect true subgroup differences and can lead to an increased risk for type II errors (false negatives), power calculations for subgroup analyses are rarely reported in randomized controlled trials (RCTs) and meta-analyses (19–22). In addition, even when a statistically significant  $P_{\text{interaction}}$  value is observed, this finding might be a false positive (type I error), making it

crucial to consider additional criteria beyond the  $P_{\text{interaction}}$  to assess the credibility of a potential subgroup effect (19,23–25). Furthermore, even if the observed relative risk reduction is similar for two subpopulations, the absolute risk reduction may differ substantially between the two subpopulations depending on each population's baseline risk. Therefore, it is recommended to calculate and report absolute effects alongside relative effects because absolute effects provide a more clinically relevant estimation of a treatment's benefits (26–28).

The aim of this perspective is to present a methodological framework that addresses important considerations involved in performing and interpreting subgroup meta-analyses. A “methodological framework” has been described as a systematic approach or tool that provides structured practical guidance to the user through a process, using stages or a step-by-step methodology (29). In this context, we used subgroup data from CVOTs with GLP-1 receptor agonists and SGLT2 inhibitors. These trials focused on assessing major adverse cardiovascular events (MACE) (a composite end point of cardiovascular death, stroke, or myocardial infarction) in patients with type 2 diabetes who had either established cardiovascular disease or high cardiovascular risk without manifest cardiovascular disease (cardiovascular risk factors alone or subclinical cardiovascular disease). Within our framework, we address important challenges and considerations, namely, evaluation of between-trial clinical heterogeneity regarding the definition of subpopulations, assessment of statistical power and credibility of subgroup meta-analysis results, and translation of relative treatment effects into clinically meaningful absolute effects. While previous studies (18,30) and meta-analyses (15–17,27) may have included examination of these individual aspects, a comprehensive and systematic approach for integrating all considerations in a unified context is currently lacking. With our proposed methodological framework we aim to fill this gap by providing a stepwise approach that addresses these key questions in a structured manner:

1. How do definitions for the subpopulation with established cardiovascular disease and for the subpopulation with cardiovascular risk factors but

without manifest cardiovascular disease vary across individual CVOTs?

2. Is the baseline cardiovascular risk, as measured by the incidence of MACE in placebo-treated patients, for each of these subpopulations sufficiently consistent across different CVOTs to justify pooling of data in a meta-analysis?
3. Does the relative treatment effect differ between the subpopulation with established cardiovascular disease and the subpopulation at high risk without manifest disease?
4. What are the absolute treatment effects for each subpopulation?

## STEP 1: DEFINITION OF SUBPOPULATIONS ACROSS TRIALS

The first step in evaluating the clinical heterogeneity of the subpopulations across trials is extracting detailed definitions for each subpopulation from each CVOT, to assess consistency of the definitions used across trials. Supplementary Tables 1 and 2 summarize definitions of established cardiovascular disease and high cardiovascular risk used in individual CVOTs of GLP-1 receptor agonists (5,6,9–14) and SGLT2 inhibitors (3,4,7,8) based on their respective eligibility criteria. In all trials, the definition of established cardiovascular disease comprised coronary artery, cerebrovascular, or peripheral artery disease, with the sole exception being Evaluation of Lixisenatide in Acute Coronary Syndrome (ELIXA), which recruited only patients with recent myocardial infarction or unstable angina (5). All trials except three (4,9,10) had a minimum age threshold (30, 40, or 50 years). Liraglutide Effect and Action in Diabetes: Evaluation of Cardiovascular Outcome Results (LEADER), Trial to Evaluate Cardiovascular and Other Long-term Outcomes With Semaglutide in Subjects With Type 2 Diabetes (SUSTAIN-6), and Peptide Innovation for Early Diabetes Treatment (PIONEER) 6 had nearly identical eligibility criteria (12–14). Of note, in these three trials, unlike the other CVOTs, isolated chronic renal impairment or heart failure was included in the definition of established cardiovascular disease. Detailed definitions of coronary artery disease, cerebrovascular disease, and peripheral artery disease were relatively similar, albeit not identical, across trials (Supplementary Tables 1 and 2).

The definition of high cardiovascular risk without manifest cardiovascular disease

(patients with cardiovascular risk factors alone or subclinical cardiovascular disease) varied across the eight CVOTs with inclusion of such participants. Specifically, EXenatide Study of Cardiovascular Event Lowering (EXSCEL) did not include any predefined eligibility criteria for this subpopulation (10), while the remaining seven trials included the presence of subclinical cardiovascular disease (e.g., hypertension with left ventricular hypertrophy or chronic renal impairment) or presence of at least one or two cardiovascular risk factors for patients aged >50, 55, or 60 years (Supplementary Tables 1 and 2). Although risk factors used were not identical across trials, the most used factors included hypertension (all trials), albuminuria (six trials), dyslipidemia (four trials), and current tobacco use (four trials). Notably, for this subpopulation, Researching Cardiovascular Events With a Weekly INcretin in Diabetes (REWIND) included two predefined subsets of criteria to differentiate between subclinical cardiovascular disease and cardiovascular risk factors (Supplementary Table 1).

## STEP 2: EVALUATION OF VARIABILITY IN BASELINE RISK ACROSS TRIALS

Given the dissimilarities in eligibility criteria among CVOTs, to evaluate the comparability of baseline cardiovascular risk of each subpopulation across CVOTs, we extracted data for MACE incidence (number of patients with event per 100 person-years) in the placebo arm for each subpopulation in each trial. Even though no established statistical metric can determine variability in subpopulations definitions, by calculating MACE incidence we can judge clinical heterogeneity in baseline risk between trials. This pragmatic approach, while subjective, can help in evaluation of whether variability is acceptable for a meaningful subgroup meta-analysis.

In most cases, we retrieved incidence data from the primary publications of CVOTs. In LEADER, incidence was not reported for either of the two subpopulations (14), and, as such, we imputed values by dividing the percentage of participants with MACE in the placebo arm for each respective subgroup by the median trial duration. For EXSCEL, we imputed incidence for the subpopulation without manifest cardiovascular disease based on reported incidence for the

overall trial population (10) and for the subpopulation with established cardiovascular disease (31). For Dapagliflozin Effect on Cardiovascular Events trial (DECLARE-TIMI 58) (7), we used relevant data published in two meta-analyses from an author group that included authors of the original trial publication (15,17).

For CVOTs with GLP-1 receptor agonists, MACE incidence for patients with established cardiovascular disease ranged between 3.9 in PIONEER 6 and 6.3 in ELIXA (Table 1). For patients at high cardiovascular risk without manifest cardiovascular disease, number of events was very low in the three trials that had a relatively smaller sample and shorter duration, namely, SUSTAIN-6, PIONEER 6, and the AMPLITUDE-O trial (Effect of Efglenatide on Cardiovascular Outcomes). Despite these between-trial differences, MACE incidence for this subpopulation did not vary considerably among trials, ranging between 1.3 in AMPLITUDE-O and 2.5 in PIONEER 6 (Table 1). Of note, in a post hoc analysis of SUSTAIN-6 and PIONEER 6, investigators pooled data for semaglutide from both trials and reported results for the subpopulation with preexisting cardiovascular disease and for the subpopulation at high cardiovascular risk, using definitions similar to those used in REWIND for each subpopulation (32). The main difference in these definitions compared with the definitions used in the original reports of SUSTAIN-6 and PIONEER 6 was in classifying participants with transient ischemic attack, chronic renal impairment, or heart failure in the subpopulation at high cardiovascular risk and not in the subpopulation with established cardiovascular disease. In this pooled analysis, MACE incidence in the placebo arm for each subpopulation was 2.5 and 4.8, respectively (32). Moreover, in a post hoc analysis of LEADER, where the subpopulations were redefined with use of a similar rationale, MACE incidence in the placebo arm for those without established cardiovascular disease was also 2.5 when participants with isolated chronic renal impairment or heart failure were categorized in this subpopulation, while the imputed MACE incidence for those with established cardiovascular disease was 4.5 (33).

In the four CVOTs with SGLT2 inhibitors, MACE incidence in the placebo arm for patients with established cardiovascular disease was consistent across trials,

ranging from 4.0 to 4.4 (Table 2). Similarly, for patients at high cardiovascular risk without manifest cardiovascular disease, MACE incidence was comparable between the two trials with recruitment of such participants (Table 2).

Overall, these evaluations indicated a relatively consistent baseline cardiovascular risk within each subpopulation across different CVOTs, supporting the pooling of data for a subgroup meta-analysis.

## STEP 3: CREDIBILITY ASSESSMENT OF SUBGROUP META-ANALYSES

We did inverse-variance random-effects meta-analyses (34) with a DerSimonian-Laird estimator for between-study heterogeneity (35). Our decision for using a random-effects model was based on methodological guidance advocating its use over a fixed-effect(s) model because it considers variability across included studies and allows generalization of the results beyond the studies included in the meta-analysis (19,36). In addition, a random-effects model strengthens a test of interaction because a significant result is usually harder to achieve than when using a fixed-effect(s) model (19,36). Notably, the random-effects model, by granting greater relative weight to smaller studies, can potentially exacerbate small study effects/publication bias (37), particularly when numerous small studies are present (36). In such scenarios, a reasonable approach might entail a sensitivity analysis using a fixed-effects model. However, in our analysis, this consideration was not applicable, as we focused exclusively on large CVOTs (36). We chose to synthesize HRs instead of odds ratios or risk ratios because in assessment of time-to-event outcomes HR accounts not only for the number of events but also for the timing of their occurrence (38).

We used effect estimates from reports with similar definitions for established cardiovascular disease and for high cardiovascular risk (32,33) to reduce clinical heterogeneity of subgroup meta-analysis findings. For GLP-1 receptor agonists, we conducted a sensitivity analysis excluding trials with agents withdrawn from the market (Harmony Outcomes assessing albiglutide) or not currently approved (AMPLITUDE-O assessing efglenatide). We also performed another sensitivity analysis using effect estimates based on the original definitions used in the

**Table 1—MACE frequency and incidence for patients treated with placebo in trials with GLP-1 receptor agonists**

Drug	LEADER Liraglutide	SUSTAIN-6 Semaglutide SC	PIONEER 6 Semaglutide PO	REWIND Dulaglutide	EXSCEL Exenatide	Harmony Outcomes Albiglutide	AMPLITUDE-O Efglenatide	ELIXA Lixisenatide
Median trial duration, years	3.8	2.1	1.3	5.4	3.2	1.6	1.8	2.1
Patients with established cardiovascular disease								
MACE frequency, <i>n</i> patients with event/ <i>n</i> participants (%)	629/3,767 (16.7)	137/1,382 (9.9)	68/1,345 (5.0)	315/1,554 (20.3)	786/5,388 (14.6)	428/4,732 (9.0)	122/1,230 (9.9)	399/3,034 (13.2%)
MACE incidence, <i>n</i> patients with event per 100 person-years	4.4*	5.0	3.9	4.2	5.1	5.9	5.7	6.3
Patients with cardiovascular risk factors or subclinical cardiovascular disease								
MACE frequency, <i>n</i> patients with event/ <i>n</i> participants (%)	65/905 (7.2)	9/267 (3.4)	8/247 (3.2)	317/3,128 (10.1)	119/2,008 (5.9)	NA	3/129 (2.3)	NA
MACE incidence, <i>n</i> patients with event per 100 person-years	1.9*	1.6	2.5	2.0	1.6†	NA	1.3	NA

MACE: cardiovascular death or myocardial infarction or stroke. NA, not applicable; PO, per os; SC, subcutaneous. \*MACE incidence values were imputed based on the following formula: incidence = frequency/median trial duration. †MACE incidence was imputed based on the following formula: (incidence for patients with established cardiovascular disease × number of patients with established cardiovascular disease) ÷ (incidence for patients with risk factors × number of patients with risk factors) = overall incidence × total number of patients.

inclusion criteria of LEADER, SUSTAIN-6, and PIONEER 6. We excluded ELIXA from all analyses because it had a different definition for established cardiovascular disease and for MACE compared with other trials (5).

As advocated in pertinent methodological guidance, if  $P_{\text{interaction}} \geq 0.1$  we assumed that the overall HR is consistent for both subpopulations (25). However, if  $P_{\text{interaction}} < 0.1$ , it is recommended to further assess the credibility of a potential subgroup effect using a valid method, such as the Instrument for Assessing the Credibility of Effect Modification Analyses (ICEMAN) (19). ICEMAN offers a structured approach for assessing subgroup effect credibility, incorporating multiple parameters beyond  $P_{\text{interaction}}$ . It consists of eight core questions with four response options, indicating increasing credibility from left to right. The final visual analog scale categorizes credibility into very low, low, moderate, and high, reflecting probabilities of <25%, 25–50%, 51–75%, and >75%, respectively, for the existence of subgroup effect modification (19). One ICEMAN parameter specifically focuses on the  $P_{\text{interaction}}$  value; a  $P_{\text{interaction}}$  value  $\leq 0.005$  indicates increased credibility and  $P_{\text{interaction}}$  value  $< 0.1$  and  $> 0.05$  suggests that chance is a very likely explanation for the subgroup effect (decreased credibility), while the remaining two categories fall between these two extremes (19). According to our overall ICEMAN assessment, we made a decision on whether to use the overall HR for both subpopulations or each subpopulation's respective HR in the calculation of absolute treatment estimates (19,25). All analyses were done with R, version 4.0.5 (R Core Team, Vienna, Austria), and the statistical packages meta and dmetar. Additionally, we calculated the power of analyses for subgroup differences based on the method described by Hedges and Pigott (39), using the power.analysis.subgroup function in R, which is included in the dmetar package. The exact R script used for the power calculation can be found in Supplementary Table 3.

Figure 1 shows meta-analysis results for CVOTs with GLP-1 receptor agonists versus placebo based on the presence or absence of established cardiovascular disease. In the overall population, GLP-1 receptor agonists reduced the risk of MACE by 15% (HR 0.85, 95% CI 0.80–0.90). The impact of statistical

**Table 2—MACE frequency and incidence for patients treated with placebo in trials with SGLT2 inhibitors**

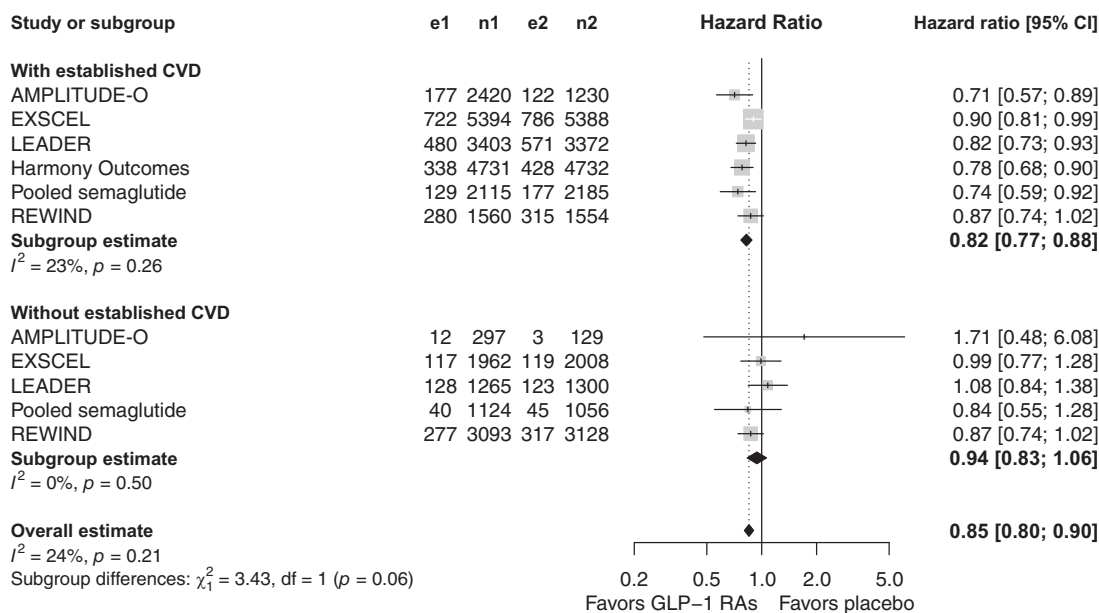
Drug	CANVAS program	DECLARE-TIMI 58	EMPA-REG OUTCOME	VERTIS CV
	Canagliflozin	Dapagliflozin	Empagliflozin	Ertugliflozin
Median trial duration, years	2.4	4.2	3.1	3.0
Patients with established cardiovascular disease				
MACE frequency, <i>n</i> patients with event/ <i>n</i> participants (%)	NR	537/3,500 (15.3)	282/2,333 (12.1)	327/2,745 (11.9)
MACE incidence, <i>n</i> patients with event per 100 person-years	4.1	4.1	4.4	4.0
Patients with cardiovascular risk factors or subclinical cardiovascular disease				
MACE frequency, <i>n</i> patients with event/ <i>n</i> participants (%)	NR	266/5,078 (5.2)	NA	NA
MACE incidence, <i>n</i> patients with event per 100 person-years	1.6	1.3	NA	NA

MACE: cardiovascular death or myocardial infarction or stroke. CANVAS, Canagliflozin Cardiovascular Assessment Study; EMPA-REG OUTCOME, BI 10773 (Empagliflozin) Cardiovascular Outcome Event Trial in Type 2 Diabetes Mellitus Patients; NA, not applicable; NR, not reported; VERTIS CV, Evaluation of Ertugliflozin Efficacy and Safety Cardiovascular Outcomes Trial.

heterogeneity, as suggested by the  $I^2$  statistic, was low for the overall population ( $I^2 = 24%$ ) and for the two subgroups. The power of the subgroup difference test was 48.5% (39), and the  $P_{interaction}$  value was  $<0.1$  ( $P_{interaction} = 0.06$ ), warranting further exploration of an effect modification with ICEMAN (19,25). ICEMAN credibility assessment is summarized in Supplementary Table 4. All trials except one provided within-trial subgroup information, and the effect modification (as measured by the ratio of HRs between subgroups in each trial; HR of one subgroup divided by the HR of the

other subgroup) was similar from trial to trial. We reasoned that there is no sound rationale for expecting a priori the relative effect to differ between the two subgroups.  $P_{interaction}$  was  $>0.05$ , suggesting that chance is a very likely explanation. We used a random-effects model to increase credibility, as advocated by the developers of ICEMAN. We reduced credibility due to low power (48.5%) of the subgroup difference test and because  $P_{interaction}$  increased to 0.2 and 0.09 in the two sensitivity analyses (Supplementary Fig. 1 and Supplementary Fig. 2). Two ICEMAN questions were not applicable

because they were specifically related to continuous effect modifiers and between-trial data comparisons, which were not used in this particular scenario. We also omitted the question related to the number of effect modifiers assessed, as we deliberately focused solely on one specific effect modifier. Based on all assessments, we deemed the overall credibility of the subgroup analysis low, meaning that there is not enough evidence to claim an effect modification and, as such, it is reasonable to use the same overall relative effect (HR 0.85, 95% CI 0.80–0.90) for each of the two subpopulations (Supplementary Table 4).



**Figure 1—Meta-analysis results for MACE from trials with GLP-1 receptor agonists vs. placebo for subgroups based on presence of cardiovascular disease.** The subpopulation without established CVD includes participants who have cardiovascular risk factors or subclinical cardiovascular disease but do not have manifest cardiovascular disease. CVD, cardiovascular disease; GLP-1 RAs, GLP-1 receptor agonists; e1, number of patients with event in GLP-1 receptor agonist arm; df, degrees of freedom; e2, number of patients with event in placebo arm; n1, number of patients in GLP-1 receptor agonist arm; n2, number of patients in placebo arm. The boldface type indicates the pooled (meta-analysis) effect estimate both for the overall population (“Overall estimate”) and for the two subgroups (“Subgroup estimate”).

Meta-analysis results for SGLT2 inhibitors are shown in Supplementary Fig. 3. In the overall population, SGLT2 inhibitors reduced the risk of MACE by 9% (HR 0.91, 95% CI 0.85–0.97). The impact of statistical heterogeneity was low for the overall population ( $I^2 = 8\%$ ) and for the two subgroups. The power of the subgroup difference test was 30.1%, and the  $P_{\text{interaction}}$  was  $\geq 0.1$  ( $P_{\text{interaction}} = 0.14$ ), warranting no further assessment of a potential subgroup effect using ICEMAN (25). As such, based on available evidence, it is reasonable to use the overall relative effect (HR 0.91, 95% CI 0.85–0.97) for each of the two subpopulations.

#### STEP 4: ABSOLUTE TREATMENT EFFECTS FOR EACH SUBPOPULATION

Subgroup meta-analyses suggested that there is not sufficient evidence to claim that the relative treatment effect (HR) with GLP-1 receptor agonists or SGLT2 inhibitors differs between patients with established cardiovascular disease and patients at high cardiovascular risk but without manifest cardiovascular disease. Therefore, it is suggested to use the overall relative effect for both subpopulations (19,25). However, absolute treatment effects differ between the two subpopulations because they are directly influenced by patients' baseline cardiovascular risk. To calculate a 5-year cardiovascular risk, we extrapolated the frequency of MACE in the placebo arm of each subgroup to 5 years for each CVOT, assuming a constant annual risk (40). Based on data from CVOT reports with use of similar definitions for each respective subpopulation, we computed an average absolute risk over 5 years for the subpopulation with

established cardiovascular disease (5-year risk 22.3%) and for the subpopulation at high cardiovascular risk (5-year risk 9.4%) (Supplementary Table 5). Finally, using the GRADEpro Guideline Development Tool (GRADEpro GDT) (41), we applied the overall relative treatment effect on these absolute risks to produce a 5-year anticipated absolute effect estimate (number of fewer patients with MACE per 1,000 patients) with respective 95% CI for each subpopulation after treatment with a GLP-1 receptor agonist or an SGLT2 inhibitor. We also calculated number-needed-to-treat values (NNTs) for each subpopulation based on the reciprocal of the respective absolute risk reduction.

Table 3 displays the results. For patients with type 2 diabetes and established cardiovascular disease (i.e., coronary, cerebrovascular, or peripheral artery disease), GLP-1 receptor agonists reduced MACE compared with placebo (30 fewer patients with event per 1,000 patients in 5 years, 95% CI 20 fewer to 40 fewer), while treatment with SGLT2 inhibitors resulted in 18 fewer patients with MACE per 1,000 patients in 5 years (95% CI 6 fewer to 30 fewer). For patients with type 2 diabetes and high cardiovascular risk but without manifest cardiovascular disease, GLP-1 receptor agonists reduced MACE compared with placebo (14 fewer patients with event per 1,000 patients in 5 years, 95% CI 9 fewer to 18 fewer), while treatment with SGLT2 inhibitors resulted in 8 fewer patients with MACE per 1,000 patients in 5 years (95% CI 3 fewer to 14 fewer).

#### DISCUSSION

We have presented a methodological framework to calculate absolute treatment

effects on MACE for two subpopulations of patients with type 2 diabetes, following subgroup meta-analysis of CVOTs with GLP-1 receptor agonists and SGLT2 inhibitors. Consecutive steps included the following: 1) extracting data on detailed definitions of subpopulations used in individual trials based on their respective recruitment criteria, 2) evaluating variability across trials in terms of participants' baseline cardiovascular risk by comparing MACE incidence in the placebo-treated arm across different trials for each subpopulation, 3) conducting subgroup meta-analyses of relative effects (HRs) and applying ICEMAN criteria in case of a potential subgroup effect to decide which HR to use (either the HR for the overall population or the respective HR of each subpopulation) for computing absolute treatment estimates, and 4) calculating the baseline cardiovascular risk of each subpopulation and applying the HR to this risk to generate absolute risk reductions over a clinically meaningful time frame of 5 years. Within our framework we aimed to address critical challenges and considerations associated with subgroup meta-analyses in a comprehensive and systematic manner. Unlike previous studies and meta-analyses that focused on individual aspects of these challenges (15–18,27,30), our stepwise approach integrates all considerations in a stepwise and rigorous manner. Through this unified approach, we provide structured practical guidance to researchers and clinicians, enabling a more meaningful conduct and interpretation of subgroup meta-analyses.

Even though there is no established, formal statistical metric to quantify the comparability of subgroup definitions

**Table 3—Five-year anticipated absolute effects and NNTs for MACE in comparing treatments in patients with type 2 diabetes**

Comparison	Subgroup population	Relative effect, HR (95% CI)	Five-year absolute effect (95% CI)	NNT (95% CI)
GLP-1 receptor agonists vs. placebo	Established cardiovascular disease	0.85 (0.80–0.90)	30 fewer per 1,000 (from 20 fewer to 40 fewer)	33 (25–50)
GLP-1 receptor agonists vs. placebo	Cardiovascular risk factors or subclinical cardiovascular disease	0.85 (0.80–0.90)	14 fewer per 1,000 (from 9 fewer to 18 fewer)	71 (56–111)
SGLT2 inhibitors vs. placebo	Established cardiovascular disease	0.91 (0.85–0.97)	18 fewer per 1,000 (from 6 fewer to 30 fewer)	56 (33–167)
SGLT2 inhibitors vs. placebo	Cardiovascular risk factors or subclinical cardiovascular disease	0.91 (0.85–0.97)	8 fewer per 1,000 (from 3 fewer to 14 fewer)	125 (71–333)

MACE: cardiovascular death or myocardial infarction or stroke.

from trial to trial, it is important that judgments regarding clinical heterogeneity across trials are based on a transparent method to allow readers to make informed judgments about the validity and generalizability of meta-analysis results. With the initial steps of our framework we aim to evaluate clinical heterogeneity among trials by examining subpopulation definitions and baseline risk for the outcome of interest within each of the two subpopulations. While no formal quantitative threshold exists for categorization of the extent of clinical variability/heterogeneity as significant or nonsignificant, this subjective assessment facilitates informed interpretations in a transparent manner and can affect the overall certainty assessment of the meta-analysis effect estimate within the Grading of Recommendations Assessment, Development and Evaluation (GRADE) framework, particularly the domain of indirectness (42). Investigators of previous studies have also extracted definitions of eligibility criteria and placebo incidence rates, pointing out differences among CVOTs (18,30). However, we also evaluated the definition and baseline risk separately for each subpopulation of interest and used outcome data from post hoc reports of trials that used more consistent definitions for cardiovascular risk. This enabled us to evaluate more accurately the clinical heterogeneity across trials, as the placebo incidence rates were more similar from trial to trial based on these definitions compared with those used in the original trial publications. This approach also allowed us to derive more reliable estimates of treatment effects for each subpopulation, in comparison with previous meta-analyses with use of data based on the original (less consistent) definitions and eligibility criteria of the CVOTs (16,17). In addition, we performed a power calculation for the  $P_{\text{interaction}}$ , which suggested that both analyses had a low power to detect subgroup differences, hence highlighting the degree of uncertainty in interpreting findings of subgroup analyses. We also used ICEMAN for the credibility assessment and interpretation of a potential modification effect in the analysis for GLP-1 receptor agonists (19). Use of ICEMAN helps reduce overreliance on the  $P_{\text{interaction}}$ , illustrating that it is just one of many parameters to be considered and that credibility of subgroup effects is on a continuum rather than a binary matter (19,23,25). As such, use of ICEMAN

provides a more comprehensive assessment of subgroup effects, with consideration of multiple parameters and minimizing the risk of false-positive (type I error) interpretations. Of note, the 2022 consensus statement by the American Diabetes Association (ADA) and the European Association for the Study of Diabetes (EASD) on the management of hyperglycemia in type 2 diabetes, also included use of ICEMAN for credibility assessment of published meta-analyses to support clinical practice recommendations (2), while researchers in fields other than type 2 diabetes have also showcased its usage (43,44).

We also emphasize the importance of reporting both relative and absolute treatment effects. Absolute effects should be reported alongside relative effects because making treatment choices in clinical practice involves focusing on trading of absolute, rather than relative, effects (28). In their meta-analysis for GLP-1 receptor agonists, Sattar et al. (16) calculated NNTs for MACE only for the overall CVOTs population and not separately for participants with and without established cardiovascular disease. In another meta-analysis investigators produced NNTs for the overall population using a different methodological approach, calculating an NNT for each CVOT and subsequently pooling these NNTs in a meta-analysis producing an overall meta-NNT for each drug class (27). However, calculating absolute treatment effects for the overall population has limited clinical applicability, as these estimates depend on patients' underlying cardiovascular risk, which varies between subpopulations with and without preexisting cardiovascular disease. In our methodological approach we calculate separate absolute estimates for each subpopulation after deciding which relative estimate to use. Moreover, in addition to NNTs, we calculated natural frequencies (patients with event per 1,000 patients) as a more interpretable measure of absolute effects to facilitate clinical decision-making (28).

Even though the relative treatment effect of GLP-1 receptor agonists and SGLT2 inhibitors on MACE did not differ between the two subpopulations, the anticipated 5-year absolute benefits were much lower (approximately one-half) in patients without manifest cardiovascular disease. This difference in absolute effects between the two subpopulations is not surprising and is attributed to the different

baseline cardiovascular risk of each subpopulation. Subpopulations with higher baseline cardiovascular risk gain greater absolute benefits, while subpopulations with lower baseline risk exhibit lower absolute benefits, even when having a similar relative treatment effect. This phenomenon underscores the importance of considering both relative and absolute effects in the interpretation of subgroup meta-analyses. The clinical interpretation of our findings is that it is reasonable to support a strong recommendation for using these medications to reduce MACE in people with type 2 diabetes and established cardiovascular disease, while they may be considered for patients at high cardiovascular risk but without manifest cardiovascular disease, given the lower absolute benefits in the latter subpopulation. This aligns with the recent ADA *Standards of Care in Diabetes* and ADA/EASD consensus statement for the management of hyperglycemia in type 2 diabetes (1,2). However, clinical practice recommendations should include consideration of multiple critical outcomes in addition to MACE (45), while treatment decisions in clinical practice should be further individualized based on each patient's personal values, preferences, and characteristics (46,47). Moreover, practice recommendations should ideally involve a comprehensive evaluation of the overall certainty of meta-analysis estimates across all important outcomes with the application of the GRADE approach (48). GRADE includes consideration of various parameters such as risk of bias, inconsistency, indirectness, and imprecision to assess the overall quality of evidence (48). Our framework can contribute to a nuanced assessment of the domain of inconsistency within the GRADE approach for meta-analyses of subgroups. In fact, the GRADE guidance for addressing inconsistency has been recently updated to include use of ICEMAN (25).

Limitations should be acknowledged. Our assumption of a constant annual risk over a 5-year time frame, while considered acceptable (28) and having also been used in other meta-analyses in type 2 diabetes (40), may not accurately reflect the dynamic nature of cardiovascular risk, which is expected to change over time in a real-world setting. Additionally, we used CVOT data to estimate the baseline cardiovascular risk for each subpopulation, which could potentially limit the generalizability of our findings to broader populations,



given that RCTs are conducted with specific eligibility criteria and the characteristics of participants in these trials may not fully represent the entire indicated population. Alternatively, baseline risks can be obtained through other sources, such as high-quality long-term observational studies or individualized risk prediction models, instead of the actual trials included in the meta-analysis (28,40,49). Moreover, our framework could be enhanced by providing effect estimates with 95% prediction intervals in addition to conventional 95% CIs (50). We might have also considered augmenting our approach with a supportive meta-regression analysis exploring the relationship between the proportion of participants with established cardiovascular disease and the effect estimate in each CVOT (51). This rationale has been adopted in a previous meta-analysis, in which authors used meta-regression analysis to reinforce the interpretation of their subgroup meta-analyses (52). However, such a meta-regression analysis should be complementary and supportive and could not replace the primary subgroup meta-analysis and the subsequent credibility assessment with ICEMAN, which provide a transparent presentation of treatment effects in each subgroup for each included trial and allow for estimation of absolute treatment effects. Finally, we did not register a protocol for our meta-analysis because our scope was to showcase the stepwise approach and practical implementation of the proposed framework rather than doing a formal systematic review. However, we recognize that protocol registration is highly recommended in planning to conduct subgroup meta-analyses, as it can impact the validity of ICEMAN assessment. For instance, some questions in ICEMAN are related to issues that should ideally be determined a priori in the protocol, such as the formulation of a hypothesis on the direction of effect modification, the number of effect modifiers to be assessed, and the pre-defined cutoff values for subgroup analyses with continuous effect modifiers.

It is important to emphasize that ICEMAN was designed for evaluation of claims of potential subgroup effects and not for making claims of the absence of a subgroup effect (19,25). In fact, the developers of ICEMAN specifically report that its use is not warranted when the  $P_{\text{interaction}}$  value is  $\geq 0.1$  (25), as was the case in our meta-analysis for SGLT2

inhibitors. In such situations, making clinical inferences based on the absence of a significant  $P_{\text{interaction}}$  should be done with great caution, considering the test's low statistical power to detect a true difference in effect between subgroups (22). This limitation has been aptly summarized as "absence of evidence is not evidence of absence" (53), and in the context of subgroup analyses it means that a nonsignificant  $P_{\text{interaction}}$  ( $\geq 0.1$ ) should not be interpreted as conclusive evidence of absence of a subgroup effect. Therefore, our subgroup meta-analysis for SGLT2 inhibitors should be interpreted as indicating that the currently available data from CVOTs do not provide sufficient evidence to support the presence of an effect modification between the two subpopulations rather than suggesting that there is evidence of the absence of a subgroup effect. Our meta-analysis for GLP-1 receptor agonists resulted in a significant  $P_{\text{interaction}}$ , prompting us to conduct further credibility assessment using ICEMAN, which suggested that there is likely no effect modification. It is important to acknowledge that even after use of ICEMAN, this interpretation still carries a level of uncertainty (19,25). This uncertainty is inherent in all subgroup analyses, as they are purely observational in nature even when conducted in the context of a meta-analysis of RCTs. Subgroup meta-analyses, like subgroup analyses of individual RCTs, serve as hypothesis-generating rather than hypothesis-testing tools, offering insights into the presence of a potential subgroup effect rather than establishing causality between a modification variable and a treatment effect (36,54).

Overall, we present a practical framework for performing and interpreting subgroup meta-analyses in the context of CVOTs with GLP-1 receptor agonists and SGLT2 inhibitors. In addressing important considerations and challenges associated with the evaluation of clinical heterogeneity, power calculation, and credibility assessment of subgroup meta-analysis, and translation of relative estimates into clinically meaningful absolute estimates, this framework provides a stepwise guide allowing for a robust interpretation of treatment effects across different subpopulations. The methodological steps demonstrated can be applied for the interpretation of any subgroup meta-analysis irrespective

of the outcome or the modification variable of interest.

**Funding and Duality of Interest.** This research was partially funded by the European Foundation for the Study of Diabetes (EFSD), 2019 EFSD Future Leaders Mentorship Programme for Clinical Diabetologists supported by an unrestricted educational grant from AstraZeneca. A.T. has received research support from Boehringer Ingelheim and consulting fees from Boehringer Ingelheim and Novo Nordisk. E.B. has received research support and consulting fees from Novo Nordisk. M.A.N. has participated on advisory boards with or has received consulting fees from Boehringer Ingelheim, BERLIN-CHEMIE/MENARINI, Eli Lilly & Co., Merck Sharp & Dohme, Novo Nordisk, Pfizer, ShouTi/Gasherbrum Bio, Regor Pharmaceuticals, Structure Therapeutics, Sun Pharma, and Inventiva. He has received grant support from Merck Sharp & Dohme. He has served on the speakers' bureau of BERLIN-CHEMIE/MENARINI, Eli Lilly & Co., Medical Learning Institute, Medscape, Merck Sharp & Dohme, Novo Nordisk, Sanofi, and Sun Pharma. He has received payment for expert testimony from Allen & Overy/Novo Nordisk. No other potential conflicts of interest relevant to this article were reported.

## References

1. ElSayed NA, Aleppo G, Aroda VR, et al.; American Diabetes Association. 9. Pharmacologic approaches to glycemic treatment: *Standards of Care in Diabetes—2023*. *Diabetes Care* 2023;46 (Suppl. 1):S140–S157
2. Davies MJ, Aroda VR, Collins BS, et al. Management of hyperglycemia in type 2 diabetes, 2022. A consensus report by the American Diabetes Association (ADA) and the European Association for the Study of Diabetes (EASD). *Diabetes Care* 2022; 45:2753–2786
3. Cannon CP, Pratley R, Dagogo-Jack S, et al.; VERTIS CV Investigators. Cardiovascular outcomes with ertugliflozin in type 2 diabetes. *N Engl J Med* 2020;383:1425–1435
4. Zinman B, Wanner C, Lachin JM, et al.; EMPA-REG OUTCOME Investigators. Empagliflozin, cardiovascular outcomes, and mortality in type 2 diabetes. *N Engl J Med* 2015;373:2117–2128
5. Pfeffer MA, Claggett B, Diaz R, et al.; ELIXA Investigators. Lixisenatide in patients with type 2 diabetes and acute coronary syndrome. *N Engl J Med* 2015;373:2247–2257
6. Hernandez AF, Green JB, Janmohamed S, et al.; Harmony Outcomes committees and investigators. Albiglutide and cardiovascular outcomes in patients with type 2 diabetes and cardiovascular disease (Harmony Outcomes): a double-blind, randomised placebo-controlled trial. *Lancet* 2018;392:1519–1529
7. Wiviott SD, Raz I, Bonaca MP, et al.; DECLARE-TIMI 58 Investigators. Dapagliflozin and cardiovascular outcomes in type 2 diabetes. *N Engl J Med* 2019;380:347–357
8. Neal B, Perkovic V, Mahaffey KW, et al.; CANVAS Program Collaborative Group. Canagliflozin and cardiovascular and renal events in type 2 diabetes. *N Engl J Med* 2017;377:644–657
9. Gerstein HC, Sattar N, Rosenstock J, et al.; AMPLITUDE-O Trial Investigators. Cardiovascular

- and renal outcomes with efglenatide in type 2 diabetes. *N Engl J Med* 2021;385:896–907
10. Holman RR, Bethel MA, Mentz RJ, et al.; EXSCEL Study Group. Effects of once-weekly exenatide on cardiovascular outcomes in type 2 diabetes. *N Engl J Med* 2017;377:1228–1239
11. Gerstein HC, Colhoun HM, Dagenais GR, et al.; REWIND Investigators. Dulaglutide and cardiovascular outcomes in type 2 diabetes (REWIND): a double-blind, randomised placebo-controlled trial. *Lancet* 2019;394:121–130
12. Husain M, Birkenfeld AL, Donsmark M, et al.; PIONEER 6 Investigators. Oral semaglutide and cardiovascular outcomes in patients with type 2 diabetes. *N Engl J Med* 2019;381:841–851
13. Marso SP, Bain SC, Consoli A, et al.; SUSTAIN-6 Investigators. Semaglutide and cardiovascular outcomes in patients with type 2 diabetes. *N Engl J Med* 2016;375:1834–1844
14. Marso SP, Daniels GH, Brown-Frandsen K, et al.; LEADER Steering Committee; LEADER Trial Investigators. Liraglutide and cardiovascular outcomes in type 2 diabetes. *N Engl J Med* 2016;375:311–322
15. Zelniker TA, Wiviott SD, Raz I, et al. Comparison of the effects of glucagon-like peptide receptor agonists and sodium-glucose cotransporter 2 inhibitors for prevention of major adverse cardiovascular and renal outcomes in type 2 diabetes mellitus. *Circulation* 2019;139:2022–2031
16. Sattar N, Lee MMY, Kristensen SL, et al. Cardiovascular, mortality, and kidney outcomes with GLP-1 receptor agonists in patients with type 2 diabetes: a systematic review and meta-analysis of randomised trials. *Lancet Diabetes Endocrinol* 2021;9:653–662
17. McGuire DK, Shih WJ, Cosentino F, et al. Association of SGLT2 inhibitors with cardiovascular and kidney outcomes in patients with type 2 diabetes: a meta-analysis. *JAMA Cardiol* 2021;6:148–158
18. Melo M, Gavina C, Silva-Nunes J, Andrade L, Carvalho D. Heterogeneity amongst GLP-1 RA cardiovascular outcome trials results: can definition of established cardiovascular disease be the missing link? *Diabetol Metab Syndr* 2021;13:81
19. Schandlmaier 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–E906
20. Brookes ST, Whitley E, Peters TJ, Mulheran PA, Egger M, Davey Smith G. Subgroup analyses in randomised controlled trials: quantifying the risks of false-positives and false-negatives. *Health Technol Assess* 2001;5:1–56
21. Alosch M, Huque MF, Bretz F, D’Agostino RB Sr. Tutorial on statistical considerations on subgroup analysis in confirmatory clinical trials. *Stat Med* 2017;36:1334–1360
22. Brookes ST, Whitley E, Egger M, Smith GD, Mulheran PA, Peters TJ. Subgroup analyses in randomized trials: risks of subgroup-specific analyses; power and sample size for the interaction test. *J Clin Epidemiol* 2004;57:229–236
23. Guyatt GH, Oxman AD, Kunz R, et al.; GRADE Working Group. GRADE guidelines: 7. Rating the quality of evidence— inconsistency. *J Clin Epidemiol* 2011;64:1294–1302
24. Sun X, Ioannidis JPA, Agoritsas T, Alba AC, Guyatt G. How to use a subgroup analysis: users’ guide to the medical literature. *JAMA* 2014;311:405–411
25. Guyatt G, Zhao Y, Mayer M, et al. GRADE guidance 36: updates to GRADE’s approach to addressing inconsistency. *J Clin Epidemiol* 2023;158:70–83
26. Ferrannini E, Rosenstock J. Clinical translation of cardiovascular outcome trials in type 2 diabetes: is there more or is there less than meets the eye? *Diabetes Care* 2021;44:641–646
27. Kuss O, Akbulut C, Schlesinger S, et al. Absolute treatment effects for the primary outcome and all-cause mortality in the cardiovascular outcome trials of new antidiabetic drugs: a meta-analysis of digitalized individual patient data. *Acta Diabetol* 2022;59:1349–1359
28. Guyatt GH, Oxman AD, Santesso N, et al. GRADE guidelines: 12. Preparing summary of findings tables— binary outcomes. *J Clin Epidemiol* 2013;66:158–172
29. McMeekin N, Wu O, Germeni E, Briggs A. How methodological frameworks are being developed: evidence from a scoping review. *BMC Med Res Methodol* 2020;20:173
30. Davies MJ, Kloecker DE, Webb DR, Khunti K, Zaccardi F. Number needed to treat in cardiovascular outcome trials of glucagon-like peptide-1 receptor agonists: a systematic review with temporal analysis. *Diabetes Obes Metab* 2020;22:1670–1677
31. Mentz RJ, Thompson VP, Aguilar D, et al. Effects of once-weekly exenatide on clinical outcomes in patients with preexisting cardiovascular disease. *Circulation* 2018;138:2576–2578
32. Verma S, Fainberg U, Husain M, et al. Applying REWIND cardiovascular disease criteria to SUSTAIN 6 and PIONEER 6: an exploratory analysis of cardiovascular outcomes with semaglutide. *Diabetes Obes Metab* 2021;23:1677–1680
33. Verma S, Poulter NR, Bhatt DL, et al. Effects of liraglutide on cardiovascular outcomes in patients with type 2 diabetes mellitus with or without history of myocardial infarction or stroke. *Circulation* 2018;138:2884–2894
34. Borenstein M, Hedges LV, Higgins JPT, Rothstein HR. A basic introduction to fixed-effect and random-effects models for meta-analysis. *Res Synth Methods* 2010;1:97–111
35. DerSimonian R, Laird N. Meta-analysis in clinical trials revisited. *Contemp Clin Trials* 2015;45(Pt A):139–145
36. Borenstein M, Higgins JPT. Meta-analysis and subgroups. *Prev Sci* 2013;14:134–143
37. Sterne JAC, Egger M. Funnel plots for detecting bias in meta-analysis: guidelines on choice of axis. *J Clin Epidemiol* 2001;54:1046–1055
38. Tierney JF, Stewart LA, Ghersi D, Burdett S, Sydes MR. Practical methods for incorporating summary time-to-event data into meta-analysis. *Trials* 2007;8:16
39. Hedges LV, Pigott TD. The power of statistical tests for moderators in meta-analysis. *Psychol Methods* 2004;9:426–445
40. 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
41. *GRADEpro GDT: GRADEpro Guideline Development Tool*. McMaster University and Evidence Prime, 2022. Accessed 20 May 2023. Available from [grade.pro](http://grade.pro)
42. Guyatt GH, Oxman AD, Kunz R, et al.; GRADE Working Group. GRADE guidelines: 8. Rating the quality of evidence— indirectness. *J Clin Epidemiol* 2011;64:1303–1310
43. Kilpeläinen TP, Tikkinen KAO, Guyatt GH, Vernooij RWM. Evidence-based urology: subgroup analysis in randomized controlled trials. *Eur Urol Focus* 2021;7:1237–1239
44. Myles P, Kasza J, Turner T. Credibility of subgroup findings in clinical trials and meta-analyses. *Br J Anaesth* 2021;127:11–14
45. Guyatt G, Oxman AD, Sultan S, et al. GRADE guidelines: 11. Making an overall rating of confidence in effect estimates for a single outcome and for all outcomes. *J Clin Epidemiol* 2013;66:151–157
46. Tsapas A, Matthews DR. N of 1 trials in diabetes: making individual therapeutic decisions. *Diabetologia* 2008;51:921–925
47. Karagiannis T, Avgerinos I, Toumpalidou M, et al. Patients’ and clinicians’ preferences on outcomes and medication attributes for type 2 diabetes: a mixed-methods study. *J Gen Intern Med*. 2 January 2020 [Epub ahead of print]. DOI: 10.1007/s11606-019-05608-0
48. Balshe H, Helfand M, Schünemann HJ, et al. GRADE guidelines: 3. Rating the quality of evidence. *J Clin Epidemiol* 2011;64:401–406
49. Basu S, Sussman JB, Berkowitz SA, Hayward RA, Yudkin JS. Development and validation of Risk Equations for Complications Of type 2 Diabetes (RECODE) using individual participant data from randomised trials. *Lancet Diabetes Endocrinol* 2017;5:788–798
50. Int’Hout J, Ioannidis JPA, Rovers MM, Goeman JJ. Plea for routinely presenting prediction intervals in meta-analysis. *BMJ Open* 2016;6:e010247
51. Sohn M, Dietrich JW, Nauck MA, Lim S. Characteristics predicting the efficacy of SGLT-2 inhibitors versus GLP-1 receptor agonists on major adverse cardiovascular events in type 2 diabetes mellitus: a meta-analysis study. *Cardiovasc Diabetol* 2023;22:153
52. Schandlmaier S, Kaushal A, Lytvyn L, et al. Low intensity pulsed ultrasound for bone healing: systematic review of randomized controlled trials. *BMJ* 2017;356:j656
53. Alderson P. Absence of evidence is not evidence of absence. *BMJ* 2004;328:476–477
54. Kent DM, Rothwell PM, Ioannidis JPA, Altman DG, Hayward RA. Assessing and reporting heterogeneity in treatment effects in clinical trials: a proposal. *Trials* 2010;11:85