Canada's Drug and Health Technology Agency

Supplemental Materials (DRAFT)

Sodium-Glucose Cotransporter-2 Inhibitors in Type 2 Diabetes Mellitus

Streamlined Drug Class Review

Date: January 4, 2024
DRAFT

Note: this version of the report has not been copy-edited. The final report will be posted concurrently with the final CADTH Formulary Management Expert Committee recommendation.

Canada's Drug and Health Technology Agency

Table of Contents

List of Tables

List of Figures

Abbreviations

AE	adverse events
AMSTAR2	A MeaSurement Tool to Assess systematic Reviews 2
CI	confidence interval
CUA	cost utility analysis
DPP-4	dipeptidyl peptidase-4
FMEC	Formulary Management Expert Committee
GLP-1	glucagon-like peptide-1
GRADE	Grading of Recommendations Assessment, Development and Evaluation
GRIPP2	Guidance for Reporting Involvement of Patients and the Public 2
NMA	network meta-analysis
NPDUIS	National Prescription Drug Utilization Information System
OR	odds ratio
pCPA	pan-Canadian Pharmaceutical Alliance
QoL	quality of life
RCT	randomized controlled trial
RR	risk ratio
SAE	serious adverse event
SGLT2	sodium glucose cotransporter-2
SMD	standardized mean difference
SR	systematic review

102

Appendix 1: Literature Search Strategy

Clinical Literature Search

Overview

Interface: Ovid

Databases

- MEDLINE All (1946-present)
- Embase (1974-present)

Note: Subject headings and search fields have been customized for each database. Duplicates between databases were removed in Ovid.

Date of search: August 31, 2023

Search filters applied: Systematic reviews; meta-analyses; network meta-analyses; health technology assessments.

Limits

- Publication date limit: 2016-present
- Language limit: English
- Conference abstracts: excluded

Published literature was identified by searching the following bibliographic databases: MEDLINE via Ovid and Embase via Ovid. All Ovid searches were run simultaneously as a multi-file search. Duplicates were removed using Ovid deduplication for multi-file searches, followed by manual deduplication in EndNote. The search strategy was comprised of both controlled vocabulary, such as the National Library of Medicine's MeSH (Medical Subject Headings), and keywords. Search concepts were developed based on the elements of the PICOS framework and research questions. The main search concepts were Type 2 diabetes and Sodium-Glucose Transporter 2 Inhibitors, including specific drug names as well as general terms for these drugs.

CADTH-developed search filters were applied to limit retrieval to health technology assessments, systematic reviews, meta-analyses, or indirect treatment comparisons. Conference abstracts were excluded from the search results. See Appendix 1 for the detailed search strategies.

Table S1: Syntax Guide

Syntax	Description
/	At the end of a phrase, searches the phrase as a subject heading
exp	Explode a subject heading
*****	Before a word, indicates that the marked subject heading is a primary topic; or, after a word, a truncation symbol (wildcard) to retrieve plurals or varying endings
adj#	Requires terms to be adjacent to each other within # number of words (in any order)
.ti	Title
.ot	Original title
.ab	Abstract
.hw	Heading word; usually includes subject headings and controlled vocabulary
.kf	Keyword heading word
.dq	Candidate term word (Embase)
.pt	Publication type
.mp	Mapped term
.rn	Registry number
.nm	Name of substance word (MEDLINE)
.yr	Publication year

Syntax	Description
.jw	Journal title word (MEDLINE)
medall	Ovid database code: MEDLINE All, 1946 to present, updated daily
oemezd	Ovid database code; Embase, 1974 to present, updated daily

133
134

Multi-Database Strategy

#	Searches
1	diabetes mellitus/ or diabetes mellitus, type 2/ or diabetes mellitus, lipoatrophic/
2	(familial partial lipodystroph* or berardinelli-seip congenital lipodystroph* or dunnigan syndrome* or koberling-dunnigan syndrome* or MODY* or NIDDM or T2DM or T2D or DM2 or DMT2).ti,kf.
3	(Type* adj4 ("2" or "II" or two*) adj4 (diabete* or diabeti* or DM)).ti,kf.
4	((Type2 or T2 or TII) adj4 (diabete* or diabeti* or DM)).ti,kf.
5	((Maturit* or adult* or slow*) adj4 onset* adj4 (diabete* or diabeti* or DM)).ti,kf.
6	((Non-insulin* or Noninsulin*) adj4 depend* adj4 (diabete* or diabeti* or DM)).ti,kf.
7	or/1-6
8	(empagliflozin* or Jardiance* or Jardianz* or Glimpacare* or Gibtulio* or Dzhardins* or Diacurimap* or Synjardy* or Trijardy*).ti,ab,kf,ot,hw,rn,nm.
9	(dapagliflozin* or forxiga* or farxiga* or edistride* or Ebymect* or Qternmet* or Xigduo*).ti,ab,rn,nm,kf,ot,hw.
10	(canagliflozin* or canagliflocin* or Invokana* or Invokamet* or Vokanamet * or canaglu* or sulisent*).ti,ab,rn,nm,kf,ot,hw.
11	*Sodium-Glucose Transporter 2 Inhibitors/
12	((SGLT2* adj2 inhibitor*) or gliflozin*).ti,kf.
13	(sodium adj3 glucose adj2 (transporter* or co-transporter* or cotransporter*) adj2 inhibitor*).ti,kf.
14	or/8-13
15	7 and 14
16	15 use medall
17	diabetes mellitus/ or non insulin dependent diabetes mellitus/ or lipoatrophic diabetes mellitus/
18	(familial partial lipodystroph* or berardinelli-seip congenital lipodystroph* or dunnigan syndrome* or koberling-dunnigan syndrome* or MODY* or NIDDM or T2DM or T2D or DM2 or DMT2).ti,kf.
19	(Type* adj4 ("2" or "II" or two*) adj4 (diabete* or diabeti* or DM)).ti,kf.
20	((Type2 or T2 or TII) adj4 (diabete* or diabeti* or DM)).ti,kf.
21	((Maturit* or adult* or slow*) adj4 onset* adj4 (diabete* or diabeti* or DM)).ti,kf.
22	((Non-insulin* or Noninsulin*) adj4 depend* adj4 (diabete* or diabeti* or DM)).ti,kf.
23	or/17-22
24	*Empagliflozin/ or *empagliflozin plus metformin/
25	(empagliflozin* or Jardiance* or Jardianz* or Glimpacare* or Gibtulio* or Dzhardins* or Diacurimap* or Synjardy* or Trijardy*).ti,ab,kf,dq.
26	*dapagliflozin/ or *dapagliflozin plus metformin/
27	(dapagliflozin* or forxiga* or farxiga* or edistride* or Ebymect* or Qternmet* or Xigduo*).ti,ab,kf,dq.
28	*canagliflozin/ or *canagliflozin plus metformin/
29	(canagliflozin* or canagliflocin* or Invokana* or Invokamet* or Vokanamet* or canaglu* or sulisent*).ti,ab,kf,dq.
30	*sodium glucose cotransporter 2 inhibitor/
31	((SGLT2* adj2 inhibitor*) or gliflozin*).ti,kf.
32	(sodium adj3 glucose adj2 (transporter* or co-transporter* or cotransporter*) adj2 inhibitor*).ti,kf.
33	or/24-32

#	Searches
34	23 and 33
35	(conference abstract or conference review).pt.
36	34 not 35
37	16 or 36
38	network meta-analysis/
39	(meta-analysis/ or meta-analysis as topic/ or "meta analysis (topic)"/) and network.ti,ab,kf.
40	((indirect or indirect treatment or mixed treatment or bayesian) adj3 comparison*).ti,ab,kf.
41	(network* adj3 (meta-analy* or metaanaly*)).ti,ab,kf.
42	(multi* adj3 treatment adj3 comparison*).ti,ab,kf.
43	(mixed adj3 treatment adj3 (meta-analy* or metaanaly*)).ti,ab,kf.
44	umbrella review*.ti,ab,kf.
45	nma.ti,ab,kf.
46	(Multi* adj2 paramet* adj2 evidence adj2 synthesis).ti,ab,kf.
47	(Multiparamet* adj2 evidence adj2 synthesis).ti,ab,kf.
48	(multi-paramet* adj2 evidence adj2 synthesis).ti,ab,kf.
49	MPES.ti,ab,kf.
50	or/38-49
51	37 and 50
52	(systematic review or meta-analysis).pt.
53	meta-analysis/ or systematic review/ or systematic reviews as topic/ or meta-analysis as topic/ or "meta analysis (topic)"/ or "systematic review (topic)"/ or exp technology assessment, biomedical/ or network meta-analysis/
54	((systematic* adj3 (review* or overview*)) or (methodologic* adj3 (review* or overview*))).ti,ab,kf.
55	((quantitative adj3 (review* or overview* or synthes*)) or (research adj3 (integrati* or overview*))).ti,ab,kf.
56	((integrative adj3 (review* or overview*)) or (collaborative adj3 (review* or overview*)) or (pool* adj3 analy*)).ti,ab,kf.
57	(data synthes* or data extraction* or data abstraction*).ti,ab,kf.
58	(handsearch* or hand search*).ti,ab,kf.
59	(mantel haenszel or peto or der simonian or dersimonian or fixed effect* or latin square*).ti,ab,kf.
60	(met analy* or metanaly* or technology assessment* or HTA or HTAs or technology overview* or technology appraisal*).ti,ab,kf.
61	(meta regression* or metaregression*).ti,ab,kf.
62	(meta-analy* or metaanaly* or systematic review* or biomedical technology assessment* or bio-medical technology assessment*).mp,hw.
63	(medline or cochrane or pubmed or medlars or embase or cinahl).ti,ab,hw.
64	(cochrane or (health adj2 technology assessment) or evidence report).jw.
65	(comparative adj3 (efficacy or effectiveness)).ti,ab,kf.
66	(outcomes research or relative effectiveness).ti,ab,kf.
67	((indirect or indirect treatment or mixed-treatment or bayesian) adj3 comparison*).ti,ab,kf.
68	[(meta-analysis or systematic review).md.]

#	Searches
69	(multi* adj3 treatment adj3 comparison*).ti,ab,kf.
70	(mixed adj3 treatment adj3 (meta-analy* or metaanaly*)).ti,ab,kf.
71	umbrella review*.ti,ab,kf.
72	(multi* adj2 paramet* adj2 evidence adj2 synthesis).ti,ab,kf.
73	(multiparamet* adj2 evidence adj2 synthesis).ti,ab,kf.
74	(multi-paramet* adj2 evidence adj2 synthesis).ti,ab,kf.
75	or/52-74
76	37 and 75
77	51 or 76
78	limit 77 to yr="2016 -Current"
79	limit 78 to english language

Grey Literature

Search dates: August 17-31, 2023

Keywords: canagliflozin, invokana, canagliflozin-metformin, invokamet, empagliflozin, jardiance, emplagliflozin-metformin, synjardy, dapagliflozin, forxiga, dapagliflozin-metformin, xigduo, sodium-glucose cotransporter-2 (SGLT2) inhibitors), type 2 diabetes

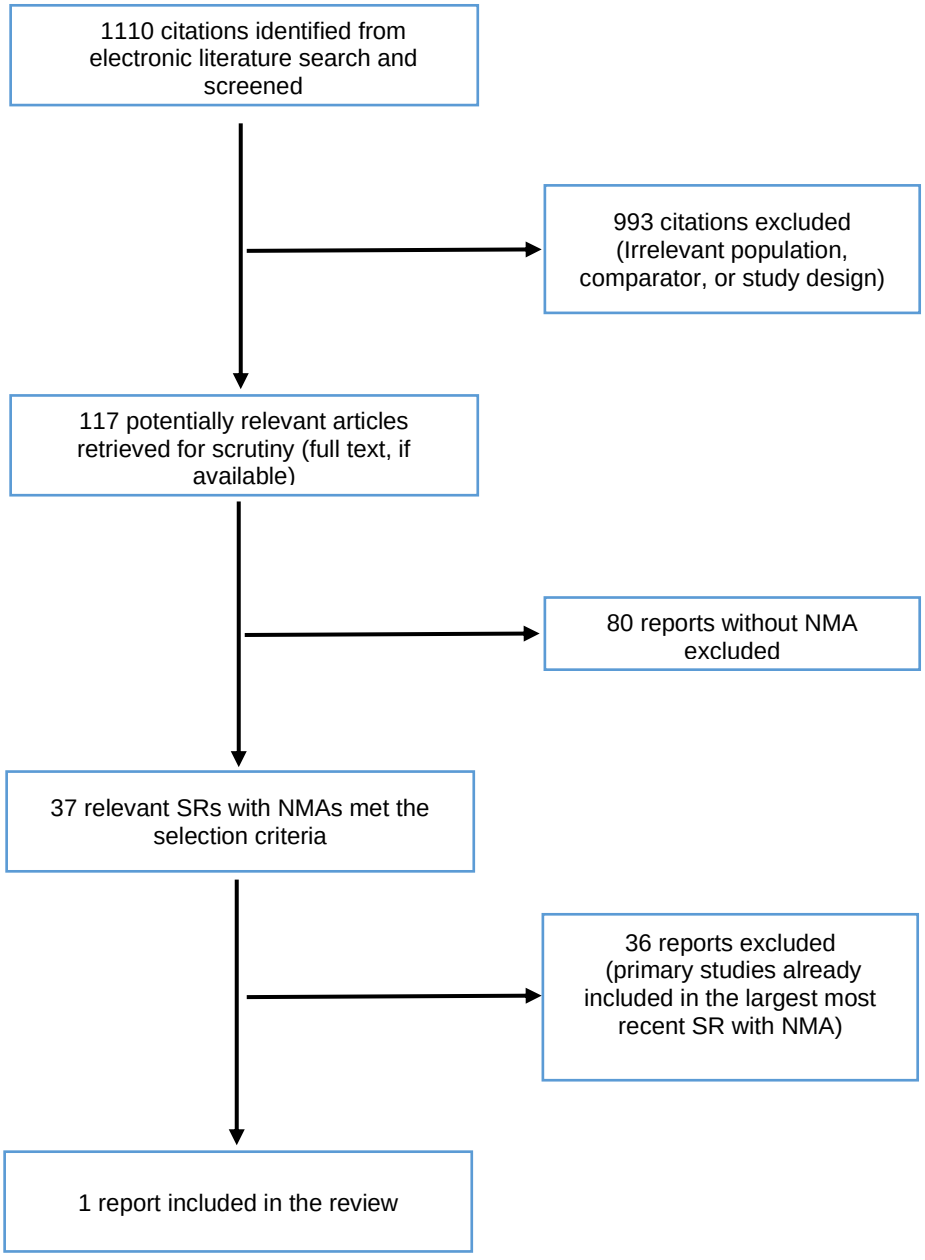
Limits: Publication years: 2016-present, English language

Relevant websites from the following sections of the CADTH grey literature checklist *Grey Matters: A Practical Tool for Searching Health-Related Grey Literature* were searched:

- Health Technology Assessment Agencies
- Health Economics
- Clinical Practice Guidelines
- Drug and Device Regulatory Approvals
- Advisories and Warnings
- Drug Class Reviews
- Databases (free)
- Health Statistics
- Internet Search
- Open Access Journals

Appendix 2: Selection of Included Studies

Figure S1: Flowchart of Selected Reports



Appendix 3: List of Excluded Publications

Table S2: Characteristics of Excluded Systematic Reviews and Network Meta-Analyses

Reference	Number of included studies	Number of studies in NMA	Number of included drug classes	Number of patients	Population	Outcomes
Yang et al. 2023	27	27	7	50237	T2DM and CKD	Cardiorenal
Sabouret et al. 2023	11	0	2	98572	T2DM	Mortality, Cardiorenal
Nguyen et al. 2023	29	0	3	50938	T2DM and CKD	Cardiorenal
Ghosal et al. 2023	16	0	3	NR	T2DM	Renal
Brondal et al. 2023	NR	NR	4	NR	T2DM	Mortality, Cardiorenal
Zhang et al. 2022	18	0	3	51496	T2DM and CKD	Mortality, Cardiorenal
Yang et al. 2022	98	0	3	186335	T2DM	Renal
Tornyos et al. 2022	29	0	1	88418	T2DM	Mortality, Cardiovascular
Tian et al. 2022	10	0	1	68723	T2DM	Mortality, Cardiorenal
Teo et al. 2022	111	0	2	103922	T1DM or T2DM	Cardiovascular, HbA1C, Safety
Qiu et al., 2022	N/A	0	2	NR	T2DM	Mortality, Cardiorenal
Li et al., 2022	36	0	2	85701	T2DM	A fib event
Guigliano et al. 2022	23	0	3	181143	T2DM or no DM	Mortality, Cardiorenal
Wei et al. 2021	NR	NR	2	NR	T2DM	Mortality, Cardiorenal
Tsapas et al. 2021	424	0	9	276336	T2DM	Body weight, Blood Pressure
Tager et al. 2021	64	0	1	74874	T2DM	Mortality, Cardiovascular
Qiu et al. 2021	NR	0	2	NR	T2DM	Mortality, Cardiovascular
Palmer et al. 2021	764	0	2	421346	T2DM	Mortality, Cardiorenal, Safety
Mannucci et al. 2021	NR	0	At least 5	NR	T2DM	HbA1C, body weight, hypoglycemia
Lin et al. 2021	21	0	3	170930	CHF and CKD	Mortality, Cardiorenal
Hu et al. 2021	15	0	2	125796	T2DM	Mortality, Cardiorenal
Duan et al. 2021	14	0	2	NR	T2DM	Mortality, Cardiorenal

Reference	Number of included studies	Number of studies in NMA	Number of included drug classes	Number of patients	Population	Outcomes
Bae et al. 2021	17	0	2	87263	T2DM	Renal
Tsapas et al. 2020	453	0	9	NR	T2DM	Mortality, Cardiorenal, HbA1c
Hussein et al. 2020	64	0	2	31384	T2DM	HbA1c, Body Weight, Blood Pressure, Safety
Wang et al. 2019	29	0	1	11999	T2DM	Change in weight
Kanter et al. 2019	21	0	2	NR	T2DM	HbA1c, weight, blood pressure
Hussein et al. 2019	8	0	2	60082	T2DM	Mortality, Cardiorenal
Fei et al. 2019	14	0	3	121047	T2DM	Mortality, Cardiorenal
Alfayez et al. 2019	9	0	3	87162	T2DM	Mortality, Cardiorenal
Zhang et al. 2018	236	0	3	176310	T2DM	Mortality, Cardiorenal
Kramer et al. 2018	9	0	3	87162	T2DM	Heart Failure Hospitalization
Fei et al. 2018	7	0	3	62268	T2DM	Mortality, Cardiovascular
Wang et al. 2017	8	0	At least 4	NR	T2DM	HbA1c, Triglycerides, Safety
Min et al. 2017	14	0	3	6980	T2DM	HbA1c, body weight, glucose, safety
Lee et al. 2017	73	0	5	101183	T2DM	Mortality, Cardiovascular

HbA1C = glycated hemoglobin; NMA = network meta-analysis; NR = not reported; T2DM = Type 2 Diabetes Mellitus

Appendix 4: Critical Appraisal

Table S3 AMSTAR 2: a critical appraisal tool for systematic reviews that include randomized or non-randomized studies of health care interventions or both [1]

For study by Shi et al. 2023[2]

1. Did the research questions and inclusion criteria for the review include the components of PICO?		
For Yes: ☐ Population ☐ Intervention ☐ Comparator group ☐ Outcome	Optional (recommended) ☐ Timeframe for follow-up	**Yes** No
2. Did the report of the review contain an explicit statement that the review methods were established prior to the conduct of the review and did the report justify any significant deviations from the protocol?		
For Partial Yes: The authors state that they had a written protocol or guide that included ALL the following: ☐ review question(s) ☐ a search strategy ☐ inclusion/exclusion criteria ☐ a risk of bias assessment	For Yes: As for partial yes, plus the protocol should be registered and should also have specified: ☐ a meta-analysis/synthesis plan, if appropriate, *and* ☐ a plan for investigating causes of heterogeneity ☐ justification for any deviations from the protocol ***Page 3 Methods: A protocol detailing predefined eligibility criteria, which differed slightly from the previously published network meta-analysis,2 was registered with PROSPERO (CRD42022325948).***	**Yes** Partial Yes No
3. Did the review authors explain their selection of the study designs for inclusion in the review?		
For Yes, the review should satisfy ONE of the following: ☐ Explanation for including only RCTs ☐ OR explanation for including only NRSI ☐ OR explanation for including only RCTs and NRSI		**Yes** No
4. Did the review authors use a comprehensive literature search strategy?		
☐ searched at least 2 databases (relevant to research question) ☐ provided key word and/or search strategy ☐ justified publication restrictions (e.g. language) ***Page 6- Search strategy and information sources***	For Yes, should also have (all the following): ☐ searched the reference lists / bibliographies of included studies ☐ searched trial/study registries ☐ included/consulted content experts in the field ☐ where relevant, searched for grey literature ☐ conducted search within 24 months of completion of the review	Yes **Partial Yes** No

5. Did the review authors perform study selection in duplicate?		
For Yes, either ONE of the following: ☐ at least two reviewers independently agreed on selection of eligible studies and achieved consensus on which studies to include ☐ OR two reviewers selected a sample of eligible studies and achieved good agreement (at least 80 percent), with the remainder selected by one reviewer. ***Page 6- Study selection: Pairs of reviewers (QS, KNo, QF, ZQ, and FY) independently screened identified hits at the title and abstract and full text levels, with discrepancies resolved by a senior reviewer (SL).***		**Yes** No
6. Did the review authors perform data extraction in duplicate?		
For Yes, either ONE of the following: ☐ at least two reviewers achieved consensus on which data to extract from included studies ☐ OR two reviewers extracted data from a sample of eligible studies and achieved good agreement (at least 80 percent), with the remainder extracted by one reviewer. ***Page 6- Data collection and data items: Using a standardised extraction form, the paired trained reviewers (QS, KNo, YM, QF, ZQ, XZ, XC, ZC, XL, and SH) independently extracted the following data***		**Yes** No
7. Did the review authors provide a list of excluded studies and justify the exclusions?		
For Partial Yes: ☐ provided a list of all potentially relevant studies that were read in full-text form but excluded from the review	For Yes, must also have: ☐ Justified the exclusion from the review of each potentially relevant study	Yes Partial Yes **No**
8. Did the review authors describe the included studies in adequate detail?		
For Partial Yes (ALL the following): ☐ described populations ☐ described interventions ☐ described comparators ☐ described outcomes ☐ described research designs	For Yes, should also have ALL the following: ☐ described population in detail ☐ described intervention in detail (including doses where relevant) ☐ described comparator in detail (including doses where relevant) ☐ described study's setting ☐ timeframe for follow-up	**Yes** Partial Yes No

	All the information provided in supplemental appendix	
9. The review authors use a satisfactory technique for assessing the risk of bias (RoB) in individual studies that were included in the review?		
RCTs For Partial Yes, must have assessed RoB from ☐ unconcealed allocation, *and* ☐ lack of blinding of patients and assessors when assessing outcomes (unnecessary for objective outcomes such as all-cause mortality) ***Cochrane RoB was used***	For Yes, must also have assessed RoB from: ☐ allocation sequence that was not truly random, *and* ☐ selection of the reported result from among multiple measurements or analyses of a specified outcome ***(unclear)***	Yes Partial Yes No Includes only NRSI
NRSI For Partial Yes, must have assessed RoB: ☐ from confounding, *and* ☐ from selection bias	For Yes, must also have assessed RoB: ☐ methods used to ascertain exposures and outcomes, *and* ☐ selection of the reported result from among multiple measurements or analyses of a specified outcome	Yes Partial Yes No Includes only RCTs
10. Did the review authors report on the sources of funding for the studies included in the review?		
For Yes ☐ Must have reported on the sources of funding for individual studies included in the review. Note: Reporting that the reviewers looked for this information but it was not reported by study authors also qualifies		**Yes** No
11. If meta-analysis was performed did the review authors use appropriate methods for statistical combination of results?		
RCTs For Yes: ☐ The authors justified combining the data in a meta-analysis ☐ AND they used an appropriate weighted technique to combine study results and adjusted for heterogeneity if present. ☐ AND investigated the causes of any heterogeneity ***Page 7 Data synthesis: methods for meta-analyses reported (include justification of approach, assessment of heterogeneity, transitivity and other assumptions prior to conducting the NMA)***		**Yes** No No meta-analysis conducted
For NRSI For Yes: ☐ The authors justified combining the data in a meta-analysis ☐ AND they used an appropriate weighted technique to combine study results, adjusting for heterogeneity if present ☐ AND they statistically combined effect estimates from NRSI that were adjusted for confounding, rather than combining raw data, or justified combining raw data when adjusted effect estimates were not available ☐ AND they reported separate summary estimates for RCTs and NRSI separately when both were included in the review		Yes No No meta-analysis conducted
12. If meta-analysis was performed, did the review authors assess the potential impact of RoB in individual studies on the results of the meta-analysis or other evidence synthesis?		
For Yes: ☐ included only low risk of bias RCTs ☐ OR, if the pooled estimate was based on RCTs and/or NRSI at variable RoB, the authors performed		**Yes** No

analyses to investigate possible impact of RoB on summary estimates of effect. ***Sensitivity analysis was performed excluding studies with high RoB***	No meta-analysis conducted
13. Did the review authors account for RoB in individual studies when interpreting/ discussing the results of the review?	
For Yes: ☐ included only low risk of bias RCTs OR, if RCTs with moderate or high RoB, or NRSI were included the review provided a discussion of the likely impact of RoB on the results	**Yes** No
14. Did the review authors provide a satisfactory explanation for, and discussion of, any heterogeneity observed in the results of the review?	
For Yes: ☐ There was no significant heterogeneity in the results ☐ OR if heterogeneity was present the authors performed an investigation of sources of any heterogeneity in the results and discussed the impact of this on the results of the review	**Yes** No
15. If they performed quantitative synthesis did the review authors carry out an adequate investigation of publication bias (small study bias) and discuss its likely impact on the results of the review?	
For Yes: ☒ performed graphical or statistical tests for publication bias and discussed the likelihood and magnitude of impact of publication bias ***Page 7– data analysis: Comparison adjusted funnel plots evaluated global small study effects, which could reflect publication bias. Page 8: The evidence did not suggest global publication bias and intransitivity for any outcome***	**Yes** No No meta-analysis conducted
16. Did the review authors report any potential sources of conflict of interest, including any funding they received for conducting the review?	
For Yes: ☐ The authors reported no competing interests OR ☐ The authors described their funding sources and how they managed potential conflicts of interest	**Yes** No

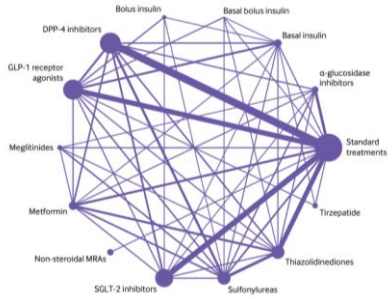
Shea BJ, Reeves BC, Wells G, Thuku M, Hamel C, Moran J, Moher D, Tugwell P, Welch V, Kristjansson E, Henry DA. AMSTAR 2: a critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions, or both. BMJ. 2017 Sep 21;358:j4008.

Table S4 ISPOR Questionnaire to assess relevance and credibility of Network Meta-Analysis Study[3] (for Shi et al. 2023)

For Shi et al. 2023[2]

Network Meta-analysis

Relevance: the extent to which the results of the NMA apply to the setting of interest to the decision maker *Assess this first. If deemed relevant, move forward with credibility.*	**Yes (strength) / No (weakness) / Can't answer (unclear)**
Is the population relevant?	**Yes**
• Should sufficiently match the population of interest to the decision maker • E.g., specific disease of interest; disease stage; severity; comorbidities; treatment history; race; age; sex; other demographic characteristics • Check study selection criteria, which can help inform a judgment • Evidence tables with inclusion criteria and baseline patient characteristics may be helpful, as well as exclusion criteria	• Yes, include only Type 2 DM population. Also some results are analyzed by risk strata that may provide additional context when reviewing the evidence
Are any relevant interventions missing?	**No**
• Are the intervention(s) included in the NMA matching with those of interest to the decision maker? • ***Important*** Are all relevant comparators considered? Note that the inclusion of comparators that are not of interest to the decision maker **does not** compromise relevance. • Consider the dose and schedule of the drug; mode of administration; background treatment; whether the drug is used as induction or maintenance treatment; whether the procedure or technique in the trials is the same as the procedure or technique that is of interest to the decision maker	• No, all comparators/interventions included in our PICA are included in the NMA.
Are any relevant outcomes missing?	**No**
• Are the outcomes relevant to the decision maker? Are they relevant to patients or the healthcare system? • Consider the feasibility of measuring relevant outcomes; the predictive relationship between surrogate outcomes and final outcomes; and what kind of evidence will be considered "good enough" given the patient population, burden of disease, and availability of alternative treatments • Consider the timing of the outcome assessment (e.g., longer follow-up may be more relevant than shorter follow-up)	• No missing outcomes. Decision maker has requested to see additional outcome on HbA1C which will be evaluated by including a supplemental NMA. • Follow up of 24 weeks or longer
Is the context (settings and circumstances) applicable?	**Yes**
• Is the setting in the included RCTs relevant to the setting and circumstances that the decision maker is interested in? • E.g., the year when the included RCTs were performed (if the standard of care has changed dramatically over time)	• Yes data sources include up to 14 October 2022

• Sometimes trials aimed at measuring efficacy, thus the setting or circumstances may be different from the real-world intent (considering compliance, adherence, etc.)	
Credibility: the extent to which the NMA or ITC accurately or validly answers the question it is designed to answer *Encompasses internal validity, reporting quality, transparency, interpretation, conflicts of interest*	
Were the outcomes for the NMA pre-specified (e.g., in a protocol or registry)?	**Yes**
• In the context of a NMA, outcomes should be pre-specified regardless of the number of interventions the review intends to compare or the number of studies the review is able to include	•
Did the researchers attempt to identify and include all relevant RCTs?	**Yes**
• ***Important*** The exclusion of specific direct comparisons without a rationale may introduce bias in the analysis; generally, RCTs are preferable to non-randomized designs, and combining randomized with observational studies in NMA is not recommended • Did the search strategy target RCTs between all interventions of interest? • Were multiple databases searched (e.g., MEDLINE, EMBASE, Central)? • Would review selection criteria admit all RCTs of interest? • Consider whether trial registers were searched	• Target RCTs between all interventions • Multiple databases were searched (MEDLINE, EMBASE, Cochrane Central)
Do the trials for the interventions of interest form one connected network of RCTs?	**Yes**
• To allow comparisons of treatment effects across all interventions in the NMA, the evidence base used should correspond to a connected network, i.e., any two treatments can be compared **directly and indirectly**; the ability of an NMA to incorporate indirect evidence means that inclusion of interventions that are not of direct interest to the authors might provide additional information in the network (e.g., excluding placebo could result in ignoring a considerable amount of indirect evidence) • The specific set of interventions of direct interest are called the **decision set**. The **supplementary set** refers to interventions (e.g., placebo) that are included in the NMA for the purpose of improving interest among interventions in the decision set. The full set of interventions (decision set + supplementary set) has been called the **synthesis comparator set**. • Supplementary interventions should be added when their value outweighs the risk of violating the transitivity assumption (e.g., in sparse networks with few trials per comparison, precision could be increased); there is little evidence to indicate how far one should go in constructing the network evidence base • ***Important*** If some interventions of interest are not part of the same network, then it is not possible to perform an indirect comparison of treatment effects of these interventions without a substantial risk of bias	 Fig 2 | Network plot for all included studies, by drug treatments. Drug treatments were grouped by their drug classes. Network plots consist of the drug nodes with node size being proportional to the sample size and the comparison edges with line thickness being proportional to the number of trials. MRA=non-steroidal mineralocorticoid receptor antagonists; GLP-1=glucagon-like peptide-1; SGLT-2=sodium glucose cotransporter-2; DPP-4=dipeptidyl peptidase-4 Figure from: Shi et al. 2023[2]

Is it apparent that poor quality studies were included, thereby leading to bias?	**No**
• The NMA report should have provided summary information on the key study characteristics of each RCT (i.e., a risk of bias appraisal)	• Risk of Bias assessment were conducted at the study level.
Is it likely that bias was induced by selective reporting of outcomes in the studies?	**No**
• An assessment of the likelihood of bias can be made whether there is consistency in the studies used for the NMA with respect to the different outcomes • It is advised to check that no relevant studies were excluded *only* because the outcome of interest was not reported (i.e., publication bias)	• Publication bias assessment was conducted • Global inconsistency, intransitivity and incoherence were all assessed.
Are there systematic differences in treatment effect modifiers (i.e., baseline patient or study characteristics that have an impact on the treatment effects) across the different treatment comparisons in the network?	**No**
• Effect modifiers = study and patient characteristics that affect the difference between the active intervention and the placebo intervention regarding the outcome of interest • Prognostic factors = study and patient characteristics that affect outcomes to the same extent in the active intervention and placebo intervention arms • Randomization does not hold across the set of trials used for the ITC because patients are not randomized to different trials; as a result, systematic differences in the distribution of patient characteristics across trials can ensue • The validity of an indirect comparison requires that the different sets of RCTs are similar, on average, in all important factors other than the intervention comparison being made; this is called the **transitivity assumption** – transitivity requires that all competing interventions of the SR are jointly randomizable (can imagine all interventions being compared simultaneously in a single multi-arm RCT) • ***Important*** Imbalanced distributions of effect modifiers threaten the plausibility of the transitivity assumption and the validity of the indirect comparison (i.e., there is ***intransitivity***); in practice, this requires effect modifiers to be known and measured	• The authors reported that the evidence did not suggest intransitivity for any outcome.
If there are systematic differences in treatment effect modifiers, were these imbalances in effect modifiers across the different treatment comparisons identified before comparing individual study results?	**Not applicable**
• Researchers undertaking the NMA should begin by generating a list of potential treatment effect modifiers for the interventions of interest on the basis of previous knowledge or reported subgroup results within individual studies before comparing results between studies; study and patient characteristics that are determined to be likely effect modifiers should be compared across studies to identify imbalances between the different types of indirect comparisons in the network	•
Analysis	

Were statistical methods used that preserve within-study randomization? (no naïve comparisons)	**Yes**
• The naïve indirect comparison does not take any differences in study effects across trials into account • With RCTs available that are part of one evidence network, the naïve indirect comparison can be considered a ***fatal flaw***	•
Were the selected grouping variants of an intervention (i.e., nodes) adequately justified?	**Yes**
• The definition of nodes needs careful consideration in situations where variants of one or more interventions are expected to appear in eligible trials; the appropriateness of merging (e.g., different doses of same drug or different drugs in one class) depends to a large extent on the research question • Authors should pre-specify the criteria for how the nodes of an expanded network could be merged; criteria should be formed in such a way that maximizes similarity of the interventions within a node and minimizes similarity across nodes • It is not clear whether more or less expanded networks are more prone to important intransitivity or incoherence	• Nodes by drug interventions were reasonable • Authors reviewed evidence to ensure there is no intraclass difference between grouping all drugs from the same drug class into the same nodes.
If both direct and indirect comparisons are available for pairwise contrasts (i.e., closed loops), was agreement in treatment effects (i.e., consistency) evaluated or discussed?	**Yes**
• ***Important*** In the presence of a closed loop any direct comparisons must be compared with the corresponding indirect comparisons regarding effects size or distribution of treatment effect modifiers; however, statistical tests for inconsistency should not be overinterpreted and should include knowledge of the clinical area • ***Yes!*** If a network has a closed loop, there is both direct and indirect evidence for some treatment contrasts; if there are no systematic differences in treatment effect modifiers across the different direct comparisons that form the loop, then there will be no systematic differences in the direct and indirect estimate for each of the contrasts that are part of the loop. Combining direct estimates with indirect estimates is valid, and the pooled (i.e., mixed) result will reflect a greater evidence base and one with increased precision regarding relative treatment effects. This is called **coherence or consistency assumption**. It implies that the different sources of evidence agree with each other. • Authors should evaluate for coherence; tests for incoherence have low power and therefore may fail to detect incoherence as statistically significant when it is present. Authors should consider the confidence intervals for incoherence factors and decide whether they include values that are sufficiently large to suggest clinically important discrepancies between direct and indirect evidence.	• Global inconsistency was assessed.

• ***No!*** If there are systematic differences in effect modifiers across the different direct comparisons of the network loop, the direct estimates and combining these may be inappropriate; hence, it is important that in the presence of a closed loop, the direct comparisons are compared with the corresponding indirect comparisons regarding effects size or distribution of treatment effect modifiers.	
In the presence of consistency between direct and indirect comparisons, were both direct and indirect evidence included in the NMA?	**Yes**
• If there is a closed loop in an evidence network, the relative treatment effect estimates obtained with direct comparisons are comparable to those obtained with the corresponding indirect comparisons, and there is no (substantial) imbalance in the distribution of effect modifiers, then it is of interest to combine the results of direct and indirect comparisons into a single effect estimate; this is called the **combined or mixed estimate** • The pooled result will be based on a greater evidence base with **increased precision** for relative treatment effects than when only direct evidence for the comparison of interest would be considered	•
With inconsistency or an imbalance in the distribution of treatment effect modifiers across the different types of comparisons in the network of trials, did the researchers attempt to minimize this bias with the analysis?	**Not applicable**
• ***Important*** Generally, if there is an imbalance in the distribution of effect modifiers across the different types of direct comparisons, **transitivity is violated** and the corresponding indirect comparison is based and/or there is inconsistency between direct and indirect evidence • If there are sufficient studies included in the NMA, it may be possible to perform a meta-regression analysis in which the relative treatment effect of each study is a function of not only a treatment comparisons of that study but also an effect modifier (i.e., is adjusted for differences in the level of the effect modifier between studies) • A challenge with meta-regression is low power that depends on the number of studies; as an alternative, some researchers attempt to use models with so-called inconsistency factors; however, the interpretation of the treatment effects with these models is not useful for decision making	•
Was a valid rationale provided for the use of random-effects or fixed-effect models?	**Yes**
• ***Important*** Any argument for the fixed effect model should include a judgment about the similarity of studies according to important effect modifiers and the prior belief, based on experience with the relevant clinical field, that the intervention is likely to have a fixed relative effect irrespective of the populations studied • ***Yes!*** Random effects models are generally advocated since most (if not all) meta-analyses contain studies that are clinically and methodologically diverse; random effects models assume that each study has its own true treatment effect, because study characteristics and the distribution of patient-	• Conducted a random effect network meta-analysis using a frequentist graph theoretical approach

related effect modifiers differ across studies; the study-specific true effects are then assumed to follow a distribution around an overall mean (the meta-analysis mean), and with a variance (between-study heterogeneity) that reflects how direct the true treatment effects between them are • ***No!*** Fixed effects models assume that the true treatment effect is common in all studies comparing the same treatments; this implies that there are no effect modifiers, or that they have the same distribution across all studies in the meta-analysis; this is ***less plausible***	
If a random-effects model was used, were assumptions about heterogeneity explored or discussed?	**Yes**
• ***Important*** In NMA, variants of the random-effects model exist; two common variants differ in their assumptions about between-study heterogeneity for each comparison among treatments – one assumes that between-study heterogeneity is the same for all comparisons, and another allows between-study heterogeneity to differ by comparison • Exploration or, at least, discussion of the choice between random-effects variants is desirable	• The global heterogeneity was evaluated with generalized methods of moments estimate of variance between studies and tested by the design based decomposition of Cochran's Q statistic. •
If there are indications of heterogeneity, were subgroup analyses or meta-regression analysis with prespecified covariates performed?	**Yes**
• ***Important*** Heterogeneity in relative treatment effects can be captured with random-effects models, but the analysis will provide the average relative treatment effect across the different levels of the responsible effect modifier(s); this finding may not be very informative for decision making, especially if there are great differences in relative treatment effects for different levels of effect modifiers • ***Yes!*** It is more informative to estimate relative treatment effects for the different levels of the effect modifier, either with subgroup analysis or with meta-regression analyses in which treatment effects are modeled as a function of the covariate • ***Yes!*** To avoid data dredging, it is strongly recommended that potential treatment effect modifiers are **pre-specified** • This item does not apply if a fixed-effect model was used, or if there was no indication of between-study heterogeneity	• The authors calculated indirect estimates from the network by node splitting and back calculation methods
Reporting Quality and Transparency	
Is a graphical or tabular representation of the evidence network provided with information on the number of RCTs per direct comparison?	**Yes**
• An overview of the included RCTs is required to help understand the findings of a NMA • The evidence base can be summarized with an evidence network in which the available direct comparisons are reflected with edges (i.e., connections) between the different interventions along with the number of RCTs per direct comparison	• Study characteristics and patient characteristics are provided.

• It is recommended that any trial that compares more than 2 interventions (i.e., >2 arms) is highlighted • A table in which studies are presented in the rows, the interventions in the columns, and observed results with each intervention of each study in the cells can be informative as well	
Are the individual study results reported?	
• To assess the face validity of the results of the NMA, the individual study results need to be provided • The presentation of individual study results allows reviewers to compare these with the results of the NMA and facilitates replication	• Yes in the appendex
Are the results of direct comparisons reported separately from results of the indirect comparisons or NMA?	**No**
• To judge whether the assumptions of consistency between direct and indirect evidence holds, estimates of (pooled) direct comparisons can be compared with estimates obtained from the corresponding indirect comparisons; however, this is not a trivial task • A more pragmatic approach is to present (pooled) direct evidence separately from results of the NMA in which direct and indirect evidence for some comparisons are combined; the absence of a difference between these two sets of results does not guarantee there is no inconsistency, but the opposite does hold: if the results based on indirect evidence are systematically different from results based on the combination of direct and indirect evidence, then the indirect evidence has to be inconsistent with the direct evidence	• They are reported together
Are all pairwise contrasts between interventions as obtained with the NMA reported along with measures of uncertainty?	**Yes**
• For decision making, it is important that all possible contrasts are presented • Equally important, for every relative treatment effect that is estimated, measures of uncertainty (i.e., 95% CI, 95% CrI) need to be presented	• In the appendex
Is a ranking of interventions provided given the reported treatment effects and its uncertainty by outcome?	**For some results only**
• In the Bayesian framework, for each outcome of interest, the probability that each treatment ranks first, second, third, and so on out of all interventions compared can be called rank probabilities and are based on the location, spread, and overlap of the posterior distributions of the relative treatment effects • Ranks can be presented in a 'rankogram', bar charts, etc. • ***Important*** Solely presenting the probability of being best can result in erroneous conclusions regarding the relative ranking of treatments because interventions for which there is a lot of uncertainty (i.e., wide CrI) are more likely to be ranked best	•

• The benefit of rank probabilities is that they summarize the distribution of effects, thereby acknowledging both location and uncertainty; other methods, e.g., surface under the cumulative ranking curve, have been proposed	
Is the effect of important patient characteristics on treatment effects reported?	**Yes**
• If it has been determined that patient characteristics are effect modifiers and differ across studies, then it is of interest to report relative treatment effects for different levels of the effect modifier as obtained via meta-regression analysis or subgroup analysis	• Results are reported by risk factors
Interpretation	
Are the conclusions fair and balanced?	**Yes**
• Conclusions should be in line with the reported results of the NMA, the available evidence base, credibility of the analysis methods, and any concerns of bias	•
Conflict of Interest	
Were there any potential conflicts of interest?	**No**
• E.g., an author has financial or personal relationships or affiliations that could affect their decisions, work, or manuscript	•
If yes, were steps taken to address these?	
• To address potential conflicts of interest, all aspects should be noted and the paper should be peer reviewed • The contribution of each author should be clearly noted • A fair and balanced exposition, including the breadth and depth of the study's limitations, should be accurately discussed	•

Matching-Adjusted Indirect Comparisons (MAICs)

Validity
Did the authors present a reasonable rationale for using MAIC?
• The authors should provide a rationale for using MAIC over alternative approaches, specifically NMA: the two primary reasons are **excess heterogeneity in the baseline patient characteristics** between studies and the **lack of a common comparator** • ***Important*** While MAIC technically can also be used for ITCs where an NMA is feasible, an NMA will generally be preferred over a MAIC if the standard assumptions of an NMA are considered reasonable • ***Important*** If heterogeneity across baseline patient characteristics is the primary concern, the rationale should cite specific baseline patient characteristics where excess heterogeneity was of concern; it is okay to accept the rationale at face value but note when evidence of excess heterogeneity is patient characteristics is not clearly provided
Were all potential sources of heterogeneity identified a priori using appropriate methodology?
• The first step in generating weights for MAICs is to identify a list of patient-level characteristics that should be considered as prognostic factors and treatment effect modifiers; these should be identified via a combination of quantitative evidence from external sources, SRs, and consultations with clinical experts
Were potential sources of heterogeneity adequately accounted for in the analysis?
• Some heterogeneity may be attributed to differences in eligibility criteria across trials; one approach to limiting these differences is to "match" the eligibility criteria between studies, such that only patients from the index trial who would have been eligible for the comparator trial are included in the MAIC. Matching eligibility criteria will only be possible when the index trial has more broad inclusion criteria than the comparator trial. • ***Important*** If any sources of heterogeneity are not fully accounted for (e.g., when differences in inclusion and exclusion criteria are not addressed through exclusion of patients) there remains a risk of bias in the estimated relative treatment effects and the conclusions should reflect this uncertainty • ***Important*** MAIC can only adjust for heterogeneity that is directly related to differences in baseline characteristics; any other sources of heterogeneity (e.g., study design, definitions of outcomes) cannot be adjusted for in a MAIC and should be considered a **limitation – the potential for risk of bias should be assessed and conclusions should reflect this uncertainty** • If it is reasonable to account for some sources of heterogeneity through the exclusion of select patients or subgroups, consideration must be given to how these exclusions may impact the generalizability of the results
Were all identified variables included in the weighting process?
• For an anchored comparison, an evaluation of whether **all effect modifiers** have been identified and included in the weighting process is required in consultation with clinical experts • For unanchored comparisons, an evaluation of whether **all prognostic factors and all effect modifiers** have been identified and included in the weighting process is required in consultation with clinical experts (highly unlikely!) • ***Important*** If a key factor has not been included in the weighting process, the risk of bias on the estimated relative treatment effects due to its exclusion must be considered • ***Important*** Results of an unanchored MAIC may be considered to have a high risk of residual bias if there is no reported assessment of residual bias
Were valid methods used to generate weights for the MAIC?
• An effective sample size (ESS) must be reported to assess the loss of precision and level of influence of subsets of patients in the index trial • The ESS should be assessed relative to the original sample size of the index trial after exclusions, where lower ESS indicates greater loss in precision and greater influence of subsets of the patients in the index trial

• ***Important*** The limitations of methods that use goodness-of-fit statistics or data-driven model selection approaches to reduce the number of covariates included in the weighting process should be identified in the critical appraisal • The risk of bias resulting from exclusion of prognostic factors or effect modifiers from the weighting process based on data-driven approaches should be assessed
Generalizability
• The target population of interest for the primary intervention(s) under review must be identified (typically, the target population will be defined based on an approved Health Canada indication for the primary intervention) • Consider how well the study population of the comparator trial (i.e., the trial with aggregate level data) aligns with the target population of interest • When the study population of the comparator trial deviates from the target population of interest, an evaluation of how these deviations impact the generalizability of the results for the target population of interest is required
Special appraisal points for multiple comparisons
• When there are multiple index trials with IPD, separate MAICs should be conducted for each index trial relative to the comparator trial; the results can be combined into a single estimated effect using meta-analysis • When there are multiple comparator trials for a single comparator with aggregate data, effect estimates can be obtained by either conducting separate MAICs between the index trial and each comparator trial and combining the results in meta-analysis; or by combining the populations and effect estimates from the comparator trials using a meta-analysis and conducting a single MAIC for the index trial relative to the combined trial results • When a MAIC is applied in a setting with multiple comparator trials, the reviewer must assess the generalizability of the results by evaluating how well a weighted combination of the populations from the comparator trials corresponds to the target population of interest • When multiple comparators are to be compared to the intervention of interest, separate MAICs must be conducted between the primary intervention and each comparator, and the estimated relative effects from each MAIC cannot be compared across the pairs

Resources

Bryan M. CADTH Methods in Focus. Critical appraisal of matching-adjusted indirect comparisons. October 2020.

Chaimana E et al. Chapter 11: Undertaking network meta-analyses. In: Higgins JPT et al. (editors). Cochrane Handbook for Systematic Reviews of Interventions version 6.2 (updated February 2021). Cochrane, 2021. Available from www.training.cochrane.org/handbook.

Jansen JP, et al. Indirect treatment comparison/network meta-analysis study questionnaire to assess relevance and credibility to inform health care decision making: an ISPOR-AMCP-NCP Good Practice Task Force report. Value Health. 2014;17(2):157-73. doi: 10.1016/j.jval.2014.01.004.

Table S5: ISPOR Questionnaire to assess relevance and credibility of Network Meta-Analysis Study[3] (for Palmer et al. 2021)

Network Meta-analysis

For Study by Palmer et al.[4]

Relevance: the extent to which the results of the NMA apply to the setting of interest to the decision maker *Assess this first. If deemed relevant, move forward with credibility.*	**Yes (strength) / No (weakness) / Can't answer (unclear)**
Is the population relevant?	**Yes**
• Should sufficiently match the population of interest to the decision maker • E.g., specific disease of interest; disease stage; severity; comorbidities; treatment history; race; age; sex; other demographic characteristics • Check study selection criteria, which can help inform a judgment • Evidence tables with inclusion criteria and baseline patient characteristics may be helpful, as well as exclusion criteria	• For adults with type 2 diabetes
Are any relevant interventions missing?	**No**
• Are the intervention(s) included in the NMA matching with those of interest to the decision maker? • ***Important*** Are all relevant comparators considered? Note that the inclusion of comparators that are not of interest to the decision maker **does not** compromise relevance. • Consider the dose and schedule of the drug; mode of administration; background treatment; whether the drug is used as induction or maintenance treatment; whether the procedure or technique in the trials is the same as the procedure or technique that is of interest to the decision maker	• Although main interventions for comparison are SGLT2 inhibitors and GLP-1 receptor agonists. The NMA has included other interventions of interest.
Are any relevant outcomes missing?	**No**
• Are the outcomes relevant to the decision maker? Are they relevant to patients or the healthcare system? • Consider the feasibility of measuring relevant outcomes; the predictive relationship between surrogate outcomes and final outcomes; and what kind of evidence will be considered "good enough" given the patient population, burden of disease, and availability of alternative treatments • Consider the timing of the outcome assessment (e.g., longer follow-up may be more relevant than shorter follow-up)	• Only using this NMA as supplemental to provide results on HbA1C
Is the context (settings and circumstances) applicable?	**Yes**
• Is the setting in the included RCTs relevant to the setting and circumstances that the decision maker is interested in? • E.g., the year when the included RCTs were performed (if the standard of care has changed dramatically over time)	• Including relevant RCTs in Type 2 DM. This is an older NMA but still relevant in our setting.

• Sometimes trials aimed at measuring efficacy, thus the setting or circumstances may be different from the real-world intent (considering compliance, adherence, etc.)	
Credibility: the extent to which the NMA or ITC accurately or validly answers the question it is designed to answer *Encompasses internal validity, reporting quality, transparency, interpretation, conflicts of interest*	
Were the outcomes for the NMA pre-specified (e.g., in a protocol or registry)?	**Yes**
• In the context of a NMA, outcomes should be pre-specified regardless of the number of interventions the review intends to compare or the number of studies the review is able to include	•
Did the researchers attempt to identify and include all relevant RCTs?	**Yes**
• ***Important*** The exclusion of specific direct comparisons without a rationale may introduce bias in the analysis; generally, RCTs are preferable to non-randomized designs, and combining randomized with observational studies in NMA is not recommended • Did the search strategy target RCTs between all interventions of interest? • Were multiple databases searched (e.g., MEDLINE, EMBASE, Central)? • Would review selection criteria admit all RCTs of interest? • Consider whether trial registers were searched	• The search strategy targeted RCTs comparing SGLT2 or GLP-1 receptor agonists with placebo • Included MEDLINE, EMBASE, Cochrane Central up to August 11 2020
Do the trials for the interventions of interest form one connected network of RCTs?	**Yes**
• To allow comparisons of treatment effects across all interventions in the NMA, the evidence base used should correspond to a connected network, i.e., any two treatments can be compared **directly and indirectly**; the ability of an NMA to incorporate indirect evidence means that inclusion of interventions that are not of direct interest to the authors might provide additional information in the network (e.g., excluding placebo could result in ignoring a considerable amount of indirect evidence) • The specific set of interventions of direct interest are called the **decision set**. The **supplementary set** refers to interventions (e.g., placebo) that are included in the NMA for the purpose of improving interest among interventions in the decision set. The full set of interventions (decision set + supplementary set) has been called the **synthesis comparator set**. • Supplementary interventions should be added when their value outweighs the risk of violating the transitivity assumption (e.g., in sparse networks with few trials per comparison, precision could be increased); there is little evidence to indicate how far one should go in constructing the network evidence base • ***Important*** If some interventions of interest are not part of the same network, then it is not possible to perform an indirect comparison of treatment effects of these interventions without a substantial risk of bias	• See Figure 2 in the publication. • All nodes are connected except for bolus insulin and alpha glucosidase inhibitor which are not interventions of interest in this review.

Is it apparent that poor quality studies were included, thereby leading to bias?	**No**
• The NMA report should have provided summary information on the key study characteristics of each RCT (i.e., a risk of bias appraisal)	• Only included RCT and risk of bias appraisal has been done for each trial.
Is it likely that bias was induced by selective reporting of outcomes in the studies?	**No**
• An assessment of the likelihood of bias can be made whether there is consistency in the studies used for the NMA with respect to the different outcomes • It is advised to check that no relevant studies were excluded *only* because the outcome of interest was not reported (i.e., publication bias)	• Appendix 5: Evaluations of network inconsistency and heterogeneity • Appendix 6: Direct, indirect and network treatment estimates
Are there systematic differences in treatment effect modifiers (i.e., baseline patient or study characteristics that have an impact on the treatment effects) across the different treatment comparisons in the network?	**Yes**
• Effect modifiers = study and patient characteristics that affect the difference between the active intervention and the placebo intervention regarding the outcome of interest • Prognostic factors = study and patient characteristics that affect outcomes to the same extent in the active intervention and placebo intervention arms • Randomization does not hold across the set of trials used for the ITC because patients are not randomized to different trials; as a result, systematic differences in the distribution of patient characteristics across trials can ensue • The validity of an indirect comparison requires that the different sets of RCTs are similar, on average, in all important factors other than the intervention comparison being made; this is called the **transitivity assumption** – transitivity requires that all competing interventions of the SR are jointly randomizable (can imagine all interventions being compared simultaneously in a single multi-arm RCT) • ***Important*** Imbalanced distributions of effect modifiers threaten the plausibility of the transitivity assumption and the validity of the indirect comparison (i.e., there is ***intransitivity***); in practice, this requires effect modifiers to be known and measured	• Evidence presented by risk strata: • Very low risk (no or few than 3 cardiovasculra riks factors) • Low risk (three or more cardivasular risk factors) • Moderate risk (cardiovascular disease) • High risk (chronic kidney disease) • Very high risk (cardiovascular and chronic kidney disease)
If there are systematic differences in treatment effect modifiers, were these imbalances in effect modifiers across the different treatment comparisons identified before comparing individual study results?	
• Researchers undertaking the NMA should begin by generating a list of potential treatment effect modifiers for the interventions of interest on the basis of previous knowledge or reported subgroup results within individual studies before comparing results between studies; study and patient characteristics that are determined to be likely effect modifiers should be compared across studies to identify imbalances between the different types of indirect comparisons in the network	• Appendix 6: Direct, indirect and network treatment estimates • The authors assessed agreement between direct and indirect estimates in every closed loop of evidence using node splitting approaches and for the entire network using a design-by-treatment interaction model.
Analysis	

Were statistical methods used that preserve within-study randomization? (no naïve comparisons)	**Yes**
• The naïve indirect comparison does not take any differences in study effects across trials into account • With RCTs available that are part of one evidence network, the naïve indirect comparison can be considered a ***fatal flaw***	• Appendix 6: Direct, indirect and network treatment estimates •
Were the selected grouping variants of an intervention (i.e., nodes) adequately justified?	**Yes**
• The definition of nodes needs careful consideration in situations where variants of one or more interventions are expected to appear in eligible trials; the appropriateness of merging (e.g., different doses of same drug or different drugs in one class) depends to a large extent on the research question • Authors should pre-specify the criteria for how the nodes of an expanded network could be merged; criteria should be formed in such a way that maximizes similarity of the interventions within a node and minimizes similarity across nodes • It is not clear whether more or less expanded networks are more prone to important intransitivity or incoherence	• Grouping of nodes by drug category seems reasonable
If both direct and indirect comparisons are available for pairwise contrasts (i.e., closed loops), was agreement in treatment effects (i.e., consistency) evaluated or discussed?	**Yes**
• ***Important*** In the presence of a closed loop any direct comparisons must be compared with the corresponding indirect comparisons regarding effects size or distribution of treatment effect modifiers; however, statistical tests for inconsistency should not be overinterpreted and should include knowledge of the clinical area • ***Yes!*** If a network has a closed loop, there is both direct and indirect evidence for some treatment contrasts; if there are no systematic differences in treatment effect modifiers across the different direct comparisons that form the loop, then there will be no systematic differences in the direct and indirect estimate for each of the contrasts that are part of the loop. Combining direct estimates with indirect estimates is valid, and the pooled (i.e., mixed) result will reflect a greater evidence base and one with increased precision regarding relative treatment effects. This is called **coherence or consistency assumption**. It implies that the different sources of evidence agree with each other. • Authors should evaluate for coherence; tests for incoherence have low power and therefore may fail to detect incoherence as statistically significant when it is present. Authors should consider the confidence intervals for incoherence factors and decide whether they include values that are sufficiently large to suggest clinically important discrepancies between direct and indirect evidence.	• Appendix 6: Direct, indirect and network treatment estimates •

• ***No!*** If there are systematic differences in effect modifiers across the different direct comparisons of the network loop, the direct estimates and combining these may be inappropriate; hence, it is important that in the presence of a closed loop, the direct comparisons are compared with the corresponding indirect comparisons regarding effects size or distribution of treatment effect modifiers.	
In the presence of consistency between direct and indirect comparisons, were both direct and indirect evidence included in the NMA?	**Yes**
• If there is a closed loop in an evidence network, the relative treatment effect estimates obtained with direct comparisons are comparable to those obtained with the corresponding indirect comparisons, and there is no (substantial) imbalance in the distribution of effect modifiers, then it is of interest to combine the results of direct and indirect comparisons into a single effect estimate; this is called the **combined or mixed estimate** • The pooled result will be based on a greater evidence base with **increased precision** for relative treatment effects than when only direct evidence for the comparison of interest would be considered	Appendix 5: Evaluations of network inconsistency and heterogeneity •
With inconsistency or an imbalance in the distribution of treatment effect modifiers across the different types of comparisons in the network of trials, did the researchers attempt to minimize this bias with the analysis?	**Yes**
• ***Important*** Generally, if there is an imbalance in the distribution of effect modifiers across the different types of direct comparisons, **transitivity is violated** and the corresponding indirect comparison is based and/or there is inconsistency between direct and indirect evidence • If there are sufficient studies included in the NMA, it may be possible to perform a meta-regression analysis in which the relative treatment effect of each study is a function of not only a treatment comparisons of that study but also an effect modifier (i.e., is adjusted for differences in the level of the effect modifier between studies) • A challenge with meta-regression is low power that depends on the number of studies; as an alternative, some researchers attempt to use models with so-called inconsistency factors; however, the interpretation of the treatment effects with these models is not useful for decision making	Appendix 5: Evaluations of network inconsistency and heterogeneity •
Was a valid rationale provided for the use of random-effects or fixed-effect models?	**Yes**
• ***Important*** Any argument for the fixed effect model should include a judgment about the similarity of studies according to important effect modifiers and the prior belief, based on experience with the relevant clinical field, that the intervention is likely to have a fixed relative effect irrespective of the populations studied • ***Yes!*** Random effects models are generally advocated since most (if not all) meta-analyses contain studies that are clinically and methodologically diverse; random effects models assume that each study has its own true treatment effect, because study characteristics and the distribution of patient-	• The direct comparison of two treatments, the authors conducted a frequentist pairwise meta-analysis using a restricted maximum likehood estimation and reported, with corresponding 95% confidence intervals, odds ratios for cihotomous outcomes, mean differences for continuous outcomes and standardized mean difference for health related QOL.

related effect modifiers differ across studies; the study-specific true effects are then assumed to follow a distribution around an overall mean (the meta-analysis mean), and with a variance (between-study heterogeneity) that reflects how direct the true treatment effects between them are • ***No!*** Fixed effects models assume that the true treatment effect is common in all studies comparing the same treatments; this implies that there are no effect modifiers, or that they have the same distribution across all studies in the meta-analysis; this is ***less plausible***	The authors conducted NMA using frequentist methods with restricted maximum likelihood estimation to quantify network heterogeneity, assuming a common heterogeneity estimate within a network. Agreement between direct and indirect estimates was assessed in every closed loop of evidence using node splitting approaches and for the entire network using a design-by-treatment interaction model.
If a random-effects model was used, were assumptions about heterogeneity explored or discussed?	**Yes**
• ***Important*** In NMA, variants of the random-effects model exist; two common variants differ in their assumptions about between-study heterogeneity for each comparison among treatments – one assumes that between-study heterogeneity is the same for all comparisons, and another allows between-study heterogeneity to differ by comparison • Exploration or, at least, discussion of the choice between random-effects variants is desirable	•
If there are indications of heterogeneity, were subgroup analyses or meta-regression analysis with prespecified covariates performed?	**Yes**
• ***Important*** Heterogeneity in relative treatment effects can be captured with random-effects models, but the analysis will provide the average relative treatment effect across the different levels of the responsible effect modifier(s); this finding may not be very informative for decision making, especially if there are great differences in relative treatment effects for different levels of effect modifiers • ***Yes!*** It is more informative to estimate relative treatment effects for the different levels of the effect modifier, either with subgroup analysis or with meta-regression analyses in which treatment effects are modeled as a function of the covariate • ***Yes!*** To avoid data dredging, it is strongly recommended that potential treatment effect modifiers are **pre-specified** • This item does not apply if a fixed-effect model was used, or if there was no indication of between-study heterogeneity	Appendix 5: Evaluations of network inconsistency and heterogeneity •
Reporting Quality and Transparency	
Is a graphical or tabular representation of the evidence network provided with information on the number of RCTs per direct comparison?	**Yes**
• An overview of the included RCTs is required to help understand the findings of a NMA • The evidence base can be summarized with an evidence network in which the available direct comparisons are reflected with edges (i.e., connections) between the different interventions along with the number of RCTs per direct comparison	• Appendix 6: Direct, indirect and network treatment estimates •

• It is recommended that any trial that compares more than 2 interventions (i.e., >2 arms) is highlighted • A table in which studies are presented in the rows, the interventions in the columns, and observed results with each intervention of each study in the cells can be informative as well	
Are the individual study results reported?	**Yes**
• To assess the face validity of the results of the NMA, the individual study results need to be provided • The presentation of individual study results allows reviewers to compare these with the results of the NMA and facilitates replication	•
Are the results of direct comparisons reported separately from results of the indirect comparisons or NMA?	**Yes**
• To judge whether the assumptions of consistency between direct and indirect evidence holds, estimates of (pooled) direct comparisons can be compared with estimates obtained from the corresponding indirect comparisons; however, this is not a trivial task • A more pragmatic approach is to present (pooled) direct evidence separately from results of the NMA in which direct and indirect evidence for some comparisons are combined; the absence of a difference between these two sets of results does not guarantee there is no inconsistency, but the opposite does hold: if the results based on indirect evidence are systematically different from results based on the combination of direct and indirect evidence, then the indirect evidence has to be inconsistent with the direct evidence	• Appendix 6: Direct, indirect and network treatment estimates •
Are all pairwise contrasts between interventions as obtained with the NMA reported along with measures of uncertainty?	**Yes**
• For decision making, it is important that all possible contrasts are presented • Equally important, for every relative treatment effect that is estimated, measures of uncertainty (i.e., 95% CI, 95% CrI) need to be presented	•
Is a ranking of interventions provided given the reported treatment effects and its uncertainty by outcome?	**No**
• In the Bayesian framework, for each outcome of interest, the probability that each treatment ranks first, second, third, and so on out of all interventions compared can be called rank probabilities and are based on the location, spread, and overlap of the posterior distributions of the relative treatment effects • Ranks can be presented in a 'rankogram', bar charts, etc. • ***Important*** Solely presenting the probability of being best can result in erroneous conclusions regarding the relative ranking of treatments because interventions for which there is a lot of uncertainty (i.e., wide CrI) are more likely to be ranked best	•

• The benefit of rank probabilities is that they summarize the distribution of effects, thereby acknowledging both location and uncertainty; other methods, e.g., surface under the cumulative ranking curve, have been proposed	
Is the effect of important patient characteristics on treatment effects reported?	**Yes**
• If it has been determined that patient characteristics are effect modifiers and differ across studies, then it is of interest to report relative treatment effects for different levels of the effect modifier as obtained via meta-regression analysis or subgroup analysis	•
Interpretation	
Are the conclusions fair and balanced?	**Yes**
• Conclusions should be in line with the reported results of the NMA, the available evidence base, credibility of the analysis methods, and any concerns of bias	•
Conflict of Interest	
Were there any potential conflicts of interest?	**No**
• E.g., an author has financial or personal relationships or affiliations that could affect their decisions, work, or manuscript	•
If yes, were steps taken to address these?	
• To address potential conflicts of interest, all aspects should be noted and the paper should be peer reviewed • The contribution of each author should be clearly noted • A fair and balanced exposition, including the breadth and depth of the study's limitations, should be accurately discussed	•

Canada's Drug and Health Technology Agency

Validity
Did the authors present a reasonable rationale for using MAIC?
• The authors should provide a rationale for using MAIC over alternative approaches, specifically NMA: the two primary reasons are **excess heterogeneity in the baseline patient characteristics** between studies and the **lack of a common comparator** • ***Important*** While MAIC technically can also be used for ITCs where an NMA is feasible, an NMA will generally be preferred over a MAIC if the standard assumptions of an NMA are considered reasonable • ***Important*** If heterogeneity across baseline patient characteristics is the primary concern, the rationale should cite specific baseline patient characteristics where excess heterogeneity was of concern; it is okay to accept the rationale at face value but note when evidence of excess heterogeneity is patient characteristics is not clearly provided
Were all potential sources of heterogeneity identified a priori using appropriate methodology?
• The first step in generating weights for MAICs is to identify a list of patient-level characteristics that should be considered as prognostic factors and treatment effect modifiers; these should be identified via a combination of quantitative evidence from external sources, SRs, and consultations with clinical experts
Were potential sources of heterogeneity adequately accounted for in the analysis?
• Some heterogeneity may be attributed to differences in eligibility criteria across trials; one approach to limiting these differences is to "match" the eligibility criteria between studies, such that only patients from the index trial who would have been eligible for the comparator trial are included in the MAIC. Matching eligibility criteria will only be possible when the index trial has more broad inclusion criteria than the comparator trial. • ***Important*** If any sources of heterogeneity are not fully accounted for (e.g., when differences in inclusion and exclusion criteria are not addressed through exclusion of patients) there remains a risk of bias in the estimated relative treatment effects and the conclusions should reflect this uncertainty • ***Important*** MAIC can only adjust for heterogeneity that is directly related to differences in baseline characteristics; any other sources of heterogeneity (e.g., study design, definitions of outcomes) cannot be adjusted for in a MAIC and should be considered a **limitation – the potential for risk of bias should be assessed and conclusions should reflect this uncertainty** • If it is reasonable to account for some sources of heterogeneity through the exclusion of select patients or subgroups, consideration must be given to how these exclusions may impact the generalizability of the results
Were all identified variables included in the weighting process?
• For an anchored comparison, an evaluation of whether **all effect modifiers** have been identified and included in the weighting process is required in consultation with clinical experts • For unanchored comparisons, an evaluation of whether **all prognostic factors and all effect modifiers** have been identified and included in the weighting process is required in consultation with clinical experts (highly unlikely!) • ***Important*** If a key factor has not been included in the weighting process, the risk of bias on the estimated relative treatment effects due to its exclusion must be considered • ***Important*** Results of an unanchored MAIC may be considered to have a high risk of residual bias if there is no reported assessment of residual bias
Were valid methods used to generate weights for the MAIC?
• An effective sample size (ESS) must be reported to assess the loss of precision and level of influence of subsets of patients in the index trial • The ESS should be assessed relative to the original sample size of the index trial after exclusions, where lower ESS indicates greater loss in precision and greater influence of subsets of the patients in the index trial • ***Important*** The limitations of methods that use goodness-of-fit statistics or data-driven model selection approaches to reduce the number of covariates included in the weighting process should be identified in the critical appraisal

• The risk of bias resulting from exclusion of prognostic factors or effect modifiers from the weighting process based on data-driven approaches should be assessed
Generalizability
• The target population of interest for the primary intervention(s) under review must be identified (typically, the target population will be defined based on an approved Health Canada indication for the primary intervention) • Consider how well the study population of the comparator trial (i.e., the trial with aggregate level data) aligns with the target population of interest • When the study population of the comparator trial deviates from the target population of interest, an evaluation of how these deviations impact the generalizability of the results for the target population of interest is required
Special appraisal points for multiple comparisons
• When there are multiple index trials with IPD, separate MAICs should be conducted for each index trial relative to the comparator trial; the results can be combined into a single estimated effect using meta-analysis • When there are multiple comparator trials for a single comparator with aggregate data, effect estimates can be obtained by either conducting separate MAICs between the index trial and each comparator trial and combining the results in meta-analysis; or by combining the populations and effect estimates from the comparator trials using a meta-analysis and conducting a single MAIC for the index trial relative to the combined trial results • When a MAIC is applied in a setting with multiple comparator trials, the reviewer must assess the generalizability of the results by evaluating how well a weighted combination of the populations from the comparator trials corresponds to the target population of interest • When multiple comparators are to be compared to the intervention of interest, separate MAICs must be conducted between the primary intervention and each comparator, and the estimated relative effects from each MAIC cannot be compared across the pairs

Resources

Bryan M. CADTH Methods in Focus. Critical appraisal of matching-adjusted indirect comparisons. October 2020.

Chaimana E et al. Chapter 11: Undertaking network meta-analyses. In: Higgins JPT et al. (editors). Cochrane Handbook for Systematic Reviews of Interventions version 6.2 (updated February 2021). Cochrane, 2021. Available from www.training.cochrane.org/handbook.

Jansen JP, et al. Indirect treatment comparison/network meta-analysis study questionnaire to assess relevance and credibility to inform health care decision making: an ISPOR-AMCP-NCP Good Practice Task Force report. Value Health. 2014;17(2):157-73. doi: 10.1016/j.jval.2014.01.004.

Appendix 5: Drugs Included in the National Prescription Drug Utilization System Database Search

Table S6: Drugs Included in the National Prescription Drug Utilization System Database Search

ATC Level 4	ATC	NAME
A10AB Insulins and analogues for injection, fast-acting	A10AB01	insulin (human)
A10AB Insulins and analogues for injection, fast-acting	A10AB03	insulin (pork)
A10AB Insulins and analogues for injection, fast-acting	A10AB04	insulin lispro
A10AB Insulins and analogues for injection, fast-acting	A10AB05	insulin aspart
A10AB Insulins and analogues for injection, fast-acting	A10AB06	insulin glulisine
A10AC Insulins and analogues for injection, intermediate-acting	A10AC01	insulin (human)
A10AC Insulins and analogues for injection, intermediate-acting	A10AC03	insulin (pork)
A10AC Insulins and analogues for injection, intermediate-acting	A10AC04	insulin lispro
A10AD Insulins and analogues for injection, intermediate- or long-acting combined with fast-acting	A10AD01	insulin (human)
A10AD Insulins and analogues for injection, intermediate- or long-acting combined with fast-acting	A10AD03	insulin (pork)
A10AD Insulins and analogues for injection, intermediate- or long-acting combined with fast-acting	A10AD04	insulin lispro
A10AD Insulins and analogues for injection, intermediate- or long-acting combined with fast-acting	A10AD05	insulin aspart
A10AE Insulins and analogues for injection, long-acting	A10AE01	insulin (human)
A10AE Insulins and analogues for injection, long-acting	A10AE03	insulin (pork)
A10AE Insulins and analogues for injection, long-acting	A10AE54	insulin glargine and lixisenatide
A10AF Insulins and analogues for inhalation	A10AF01	insulin (human)
A10BA Biguanides	A10BA02	metformin
A10BD Combinations of oral blood glucose lowering drugs	A10BD07	metformin and sitagliptin
A10BD Combinations of oral blood glucose lowering drugs	A10BD10	metformin and saxagliptin
A10BD Combinations of oral blood glucose lowering drugs	A10BD11	metformin and linagliptin
A10BD Combinations of oral blood glucose lowering drugs	A10BD15	metformin and dapagliflozin
A10BD Combinations of oral blood glucose lowering drugs	A10BD20	metformin and empagliflozin
A10BF Alpha glucosidase inhibitors	A10BF01	acarbose
A10BG Thiazolidinediones	A10BG02	rosiglitazone
A10BG Thiazolidinediones	A10BG03	pioglitazone
A10BH Dipeptidyl peptidase 4 (DPP-4) inhibitors	A10BH01	sitagliptin
A10BH Dipeptidyl peptidase 4 (DPP-4) inhibitors	A10BH03	saxagliptin
A10BH Dipeptidyl peptidase 4 (DPP-4) inhibitors	A10BH05	linagliptin
A10BJ Glucagon-like peptide-1 (GLP-1) analogues	A10BJ03	lixisenatide
A10BJ Glucagon-like peptide-1 (GLP-1) analogues	A10BJ06	semaglutide
A10BK Sodium-glucose co-transporter 2 (SGLT2) inhibitors	A10BK01	dapagliflozin
A10BK Sodium-glucose co-transporter 2 (SGLT2) inhibitors	A10BK02	canagliflozin
A10BK Sodium-glucose co-transporter 2 (SGLT2) inhibitors	A10BK03	empagliflozin
A10BX Other blood glucose lowering drugs, excl. insulins	A10BX02	repaglinide

Canada's Drug and Health Technology Agency

Appendix 6: Public Claimants and Expenditures for Antihyperglycemic Agents

Table S7: Claimants for Antihyperglycemic Agents by Class ATC4 (2019-2022)

	2019	2020	2021	2022
Alpha glucosidase inhibitors	6,246	4,520	4,648	4,700
Biguanides	870,625	876,295	913,753	943,245
Combinations of oral blood glucose lowering drugs	194,120	201,066	208,203	215,343
Dipeptidyl peptidase 4 (dpp-4) inhibitors	205,436	200,869	198,507	188,463
Glucagon-like peptide-1 (glp-1) analogues	24,721	68,814	130,696	204,258
Insulins and analogues for injection, fast-acting	177,846	174,115	176,430	174,938
Insulins and analogues for injection, intermediate- or long-acting combined with fast-acting	39,205	33,758	29,786	25,991
Insulins and analogues for injection, intermediate-acting	43,558	36,800	32,884	28,976
Insulins and analogues for injection, long-acting	254,216	261,411	272,632	280,054
Other blood glucose lowering drugs, excl. insulins	10,143	9,373	9,553	9,026
Sodium-glucose co-transporter 2 (sglt2) inhibitors	212,592	256,891	324,151	403,436
Sulfonylureas	317,091	308,301	312,408	312,754
Thiazolidinediones	5,935	4,554	3,589	3,341

Table S8: Expenditures for Antihyperglycemic Agents by Class ATC4 (2019-2022)

	2019 ($)	2020 ($)	2021 ($)	2022 ($)
Alpha glucosidase inhibitors	1,151,949	908,214	676,953	679,987
Biguanides	40,208,916	40,966,518	41,202,115	42,062,929
Combinations of oral blood glucose lowering drugs	182,496,309	194,709,259	203,221,913	207,430,454
Dipeptidyl peptidase 4 (dpp-4) inhibitors	181,510,557	181,050,203	177,921,208	167,601,951
Glucagon-like peptide-1 (glp-1) analogues	12,942,271	111,684,036	216,075,303	356,572,651
Insulins and analogues for injection, fast-acting	76,174,663	76,179,145	75,896,662	74,298,068
Insulins and analogues for injection, intermediate- or long-acting combined with fast-acting	25,182,332	21,597,869	18,636,249	16,496,585
Insulins and analogues for injection, intermediate-acting	12,882,976	10,850,007	9,084,051	7,643,953
Insulins and analogues for injection, long-acting	196,183,647	204,042,669	205,553,289	205,347,755
Other blood glucose lowering drugs, excl. insulins	1,153,219	1,128,338	1,054,828	1,000,272
Sodium-glucose co-transporter 2 (sglt2) inhibitors	157,230,404	200,322,242	250,453,872	312,727,026
Sulfonylureas	23,078,370	22,828,288	22,345,230	21,974,399
Thiazolidinediones	1,828,477	1,312,247	1,139,265	1,045,770

Table S9: Average Cost of Utilization per Beneficiary for Antihyperglycemic Agents by Molecule (2022)

Treatment	Average Annual Cost of Utilization per Beneficiary ($)
Alpha-glucosidase Inhibitors	
ACARBOSE	194
Biguanides	
METFORMIN	83
Combination	
METFORMIN AND LINAGLIPTIN	906
METFORMIN AND SAXAGLIPTIN	888
METFORMIN AND SITAGLIPTIN	1146
METFORMIN AND DAPAGLIFLOZIN	752
METFORMIN AND EMPAGLIFLOZIN	840
DPP-4i	
LINAGLIPTIN	865
SAXAGLIPTIN	629
SITAGLIPTIN	1100
GLP-1 Agonists	
LIXISENATIDE	622
SEMAGLUTIDE	1968
Insulin	
INSULIN (HUMAN)	476
INSULIN (PORK)	959
INSULIN ASPART	577
INSULIN DEGLUDEC	1022
INSULIN DETEMIR	1045
INSULIN GLARGINE	693
INSULIN GLARGINE AND LIXISENATIDE	1348
INSULIN GLULISINE	467
INSULIN LISPRO	564
Insulins and analogues for injection, fast-acting	92
Meglitinides	
REPAGLINIDE	164
SGLT2i	
CANAGLIFLOZIN	1039
DAPAGLIFLOZIN	830
EMPAGLIFLOZIN	900
Sulfonylureas	
GLIBENCLAMIDE	94
GLICLAZIDE	117
GLIMEPIRIDE	527
TZDs	
PIOGLITAZONE	412
ROSIGLITAZONE	804

Appendix 7: Anticipated Absolute Effect for Selected Outcome: Non-Fatal Stroke

Table S10: Anticipated Absolute Effect for Non-Fatal Stroke

Population	Outcome	Intervention	Comparator	Relative Effect	Baseline (5 years)	Anticipated Absolute Effects (5 years)	Grade
Adults with 3 or fewer cardiovascular risk factors	Non-fatal stroke	SGLT2 inhibitors	GLP-1 receptor agonists	1.16 (1.00, 1.35)	26 per 1000 persons	4 more (0 to 9) per 1000 persons	Moderate
Adults with more than 3 cardiovascular risk factors	Non-fatal stroke	SGLT2 inhibitors	GLP-1 receptor agonists	1.16 (1.00, 1.35)	50 per 1000 persons	8 more (0 to 16 more) per 1000 persons	Low
Adults with cardiovascular disease not chronic kidney disease	Non-fatal stroke	SGLT2 inhibitors	GLP-1 receptor agonists	1.16 (1.00, 1.35)	93 per 1000 persons	14 more (0 to 29 more) per 1000 persons	Moderate
Adults with chronic kidney disease but not cardiovascular disease	Non-fatal stroke	SGLT2 inhibitors	GLP-1 receptor agonists	1.16 (1.00, 1.35)	104 per 1000 persons	15 more (0 to 32 more) per 1000 persons	Moderate
Adults with established cardiovascular disease and chronic kidney disease	Non-fatal stroke	SGLT2 inhibitors	GLP-1 receptor	1.16 (1.00, 1.35)	166 per 1000 persons	22 more (0 to 46 more) per 1000 persons	Moderate

Source: Shi Q, et al., Copyright 2023. This work is licensed under the Attribution 4.0 International License. Full text available here: https://www.bmj.com/content/381/bmj-2022-074068

Appendix 8: Re-analysis to compare SGLT2 inhibitors with Semaglutide and / or Dulaglutide: Proposal and Results

Comparisons of efficacy and safety between SGLT2 inhibitors, Semaglutide, or Dulaglutide: proposal and results for a network meta-analysis

Proposal

We performed a frequentist random effect network meta-analysis for drug treatments on adults with type 2 diabetes.

Types of participants

We included trials enrolling adults with type 2 diabetes.

Types of interventions and controls

We included the trials if they compared SGLT2 inhibitors, semaglutide, or dulaglutide with each other or standard treatment with or without placebo. During analysis of scenario 1, semaglutide and dulaglutide were treated as one drug class label as "Semaglutide/Dulaglutide". In analysis of scenario 2, dulaglutide was excluded. SGLT2 inhibitors include Bexagliflozin, Canagliflozin, Dapagliflozin, Empagliflozin, Ertugliflozin, Henagliflozin, Ipragliflozin, Luseogliflozin, Sotagliflozin, and Tofogliflozin. Standard treatments include standard care (i.e., lifestyle modification) and standard drug treatments (e.g., metformin and/or sulfonylureas) other than the drug of interest in the randomised trial.

Types of Outcome

Primary outcomes
1) all-cause death
2) cardiovascular death
3) non-fatal stroke
4) end-stage kidney disease

Secondary outcomes
5) non-fatal myocardial infarction
6) admission to hospital for heart failure
7) health-related quality of life, such as diabetes-related quality of life or SF-36.

Analysis of Scenario 1 included both primary outcomes and secondary outcomes, while Scenario 2 only analysed primary outcomes. We measured the binary outcomes using odds ratios. We measured the quality of life score with standardised mean differences. We adopted the outcome definition reported in the original trials. End-stage kidney disease was defined as one of following criteria: long-term dialysis, kidney transplantation, a sustained eGFR <15 ml per minute per 1.73 m^2, a sustained percent decline in eGFR of at least 40% or a doubling of serum creatinine, or kidney-related death.

Types of studies

Parallel group randomized controlled trials published in English were eligible.

Follow-up and assessment time points

We included trials with at least 24 weeks of follow-up. We assessed the outcomes at maximum follow-up.

Results for Scenario 1

Figure S2: Re-analysis of Scenario 1 with Semaglutide and Dulaglutide: Forest Plot of Binary outcomes

Alt text: Forest Plot representing the relative effects of binary outcomes including all-cause death, cardiovascular death, end-stage kidney disease, hospitalization for heart failure, non-fatal myocardial infarction and non-fatal stroke, when semaglutide and dulaglutide are compared to SGLT2 inhibitors.

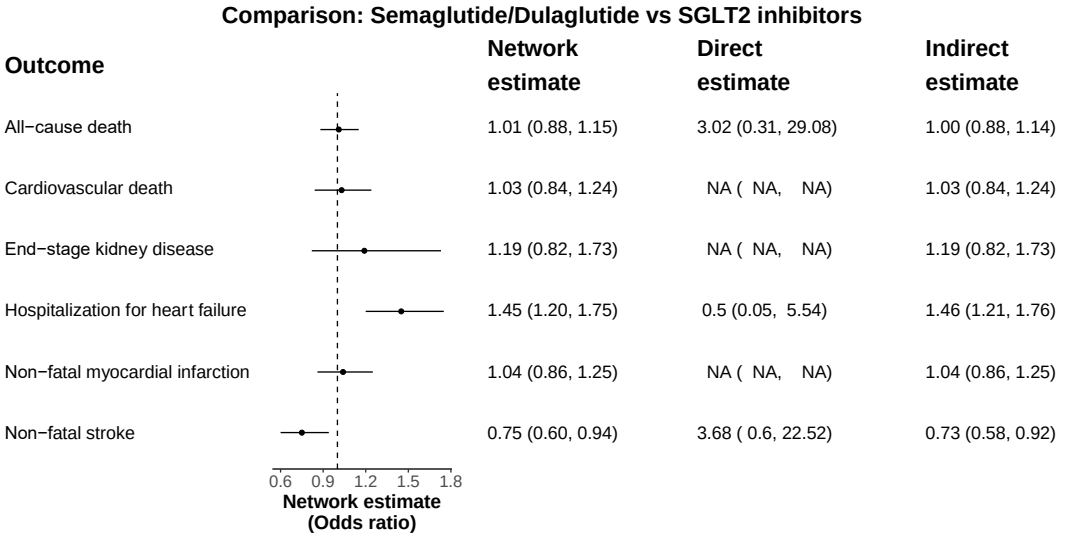
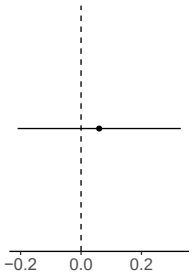


Figure S3: Re-analysis of Scenario 1 with Semaglutide and Dulaglutide: Forest Plot of Health-related quality of life

Alt text: Forest Plot representing the relative effect of health-related quality of life, when semaglutide and dulaglutide are compared to SGLT2 inhibitors.

Comparison: Semaglutide/Dulaglutide vs SGLT2 inhibitors

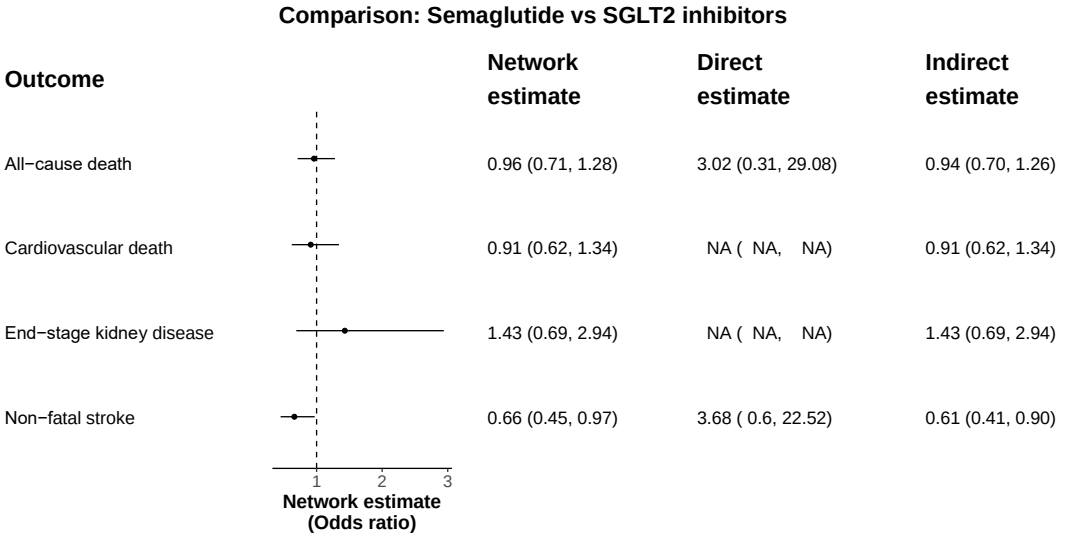
Outcome	Network estimate	Direct estimate	Indirect estimate
Health−related quality of life	0.06 (−0.21, 0.33)	−0.03 (−0.34, 0.28)	0.33 (−0.22, 0.87)



Results for Scenario 2

Figure S4: Re-analysis of Scenario 2 with Semaglutide: Forest Plot of Binary Outcomes

Alt text: Forest Plot representing the relative effects of binary outcomes including all-cause death, cardiovascular death, end-stage kidney disease, and non-fatal stroke, when semaglutide is compared to SGLT2 inhibitors.



Canada's Drug and Health Technology Agency

Appendix 9: Re-analysis to compare SGLT2 inhibitors with Semaglutide and Dulaglutide – Scenario 1: Forest Plots

These forest plots presenting relative effect of individual trial and pooled relative effects of each comparison.

Figure S5: Forest Plot: Scenario 1 for All-cause death

Alt Text: Forest plot presenting the relative effect of individual trial and pooled relative effects of each comparison on all-cause death

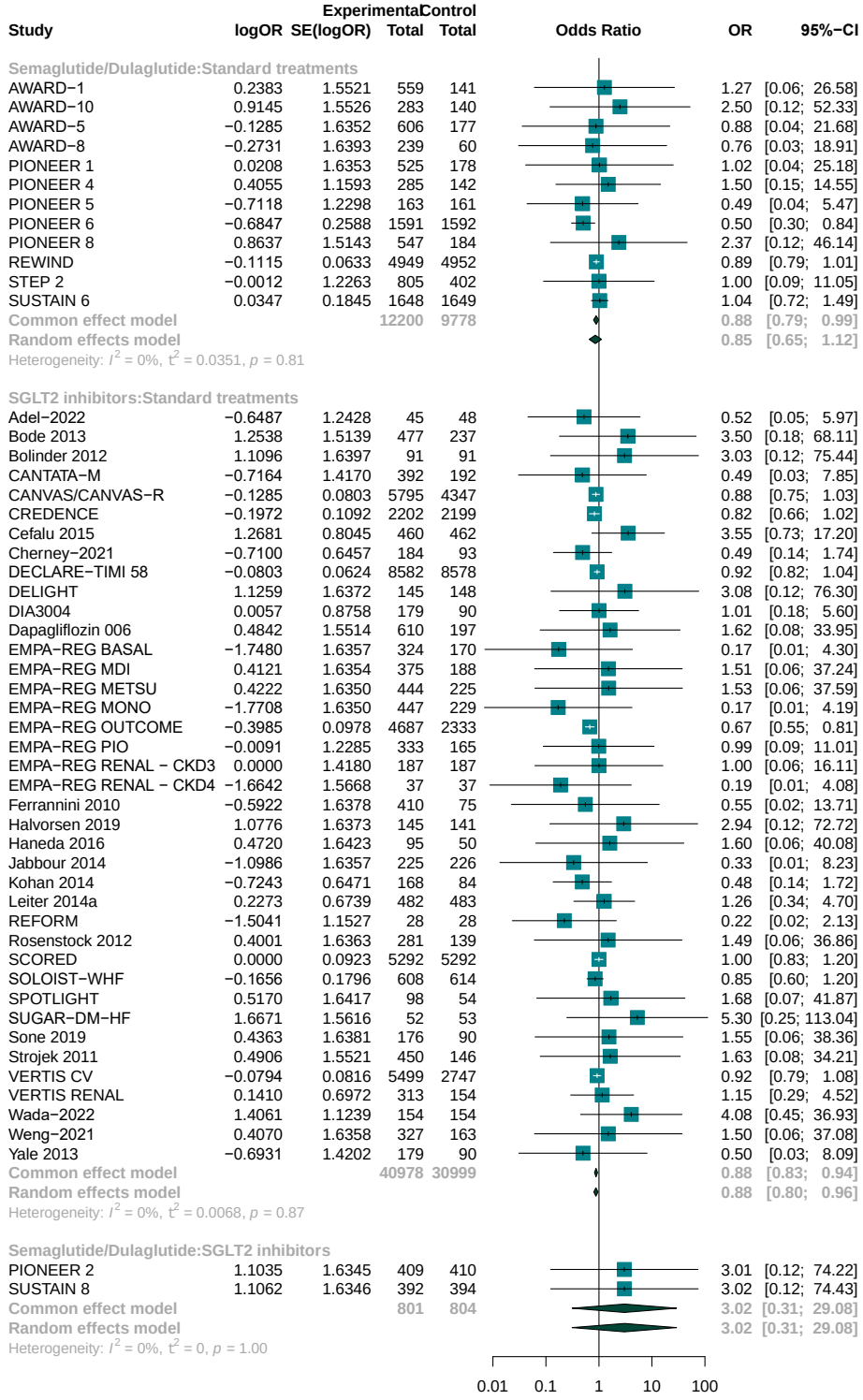

Study
logOR
SE(logOR)
Experimental Total
Control Total
Odds Ratio
OR
95%−CI
Semaglutide/Dulaglutide:Standard treatments
AWARD−1 0.2383 1.5521 559 141 1.27 [0.06; 26.58]
AWARD−10 0.9145 1.5526 283 140 2.50 [0.12; 52.33]
AWARD−5 −0.1285 1.6352 606 177 0.88 [0.04; 21.68]
AWARD−8 −0.2731 1.6393 239 60 0.76 [0.03; 18.91]
PIONEER 1 0.0208 1.6353 525 178 1.02 [0.04; 25.18]
PIONEER 4 0.4055 1.1593 285 142 1.50 [0.15; 14.55]
PIONEER 5 −0.7118 1.2298 163 161 0.49 [0.04; 5.47]
PIONEER 6 −0.6847 0.2588 1591 1592 0.50 [0.30; 0.84]
PIONEER 8 0.8637 1.5143 547 184 2.37 [0.12; 46.14]
REWIND −0.1115 0.0633 4949 4952 0.89 [0.79; 1.01]
STEP 2 −0.0012 1.2263 805 402 1.00 [0.09; 11.05]
SUSTAIN 6 0.0347 0.1845 1648 1649 1.04 [0.72; 1.49]
Common effect model 12200 9778 0.88 [0.79; 0.99]
Random effects model 0.85 [0.65; 1.12]
Heterogeneity: I^2 = 0%, t^2 = 0.0351, p = 0.81
SGLT2 inhibitors:Standard treatments
Adel−2022 −0.6487 1.2428 45 48 0.52 [0.05; 5.97]
Bode 2013 1.2538 1.5139 477 237 3.50 [0.18; 68.11]
Bolinder 2012 1.1096 1.6397 91 91 3.03 [0.12; 75.44]
CANTATA−M −0.7164 1.4170 392 192 0.49 [0.03; 7.85]
CANVAS/CANVAS−R −0.1285 0.0803 5795 4347 0.88 [0.75; 1.03]
CREDENCE −0.1972 0.1092 2202 2199 0.82 [0.66; 1.02]
Cefalu 2015 1.2681 0.8045 460 462 3.55 [0.73; 17.20]
Cherney−2021 −0.7100 0.6457 184 93 0.49 [0.14; 1.74]
DECLARE−TIMI 58 −0.0803 0.0624 8582 8578 0.92 [0.82; 1.04]
DELIGHT 1.1259 1.6372 145 148 3.08 [0.12; 76.30]
DIA3004 0.0057 0.8758 179 90 1.01 [0.18; 5.60]
Dapagliflozin 006 0.4842 1.5514 610 197 1.62 [0.08; 33.95]
EMPA−REG BASAL −1.7480 1.6357 324 170 0.17 [0.01; 4.30]
EMPA−REG MDI 0.4121 1.6354 375 188 1.51 [0.06; 37.24]
EMPA−REG METSU 0.4222 1.6350 444 225 1.53 [0.06; 37.59]
EMPA−REG MONO −1.7708 1.6350 447 229 0.17 [0.01; 4.19]
EMPA−REG OUTCOME −0.3985 0.0978 4687 2333 0.67 [0.55; 0.81]
EMPA−REG PIO −0.0091 1.2285 333 165 0.99 [0.09; 11.01]
EMPA−REG RENAL − CKD3 0.0000 1.4180 187 187 1.00 [0.06; 16.11]
EMPA−REG RENAL − CKD4 −1.6642 1.5668 37 37 0.19 [0.01; 4.08]
Ferrannini 2010 −0.5922 1.6378 410 75 0.55 [0.02; 13.71]
Halvorsen 2019 1.0776 1.6373 145 141 2.94 [0.12; 72.72]
Haneda 2016 0.4720 1.6423 95 50 1.60 [0.06; 40.08]
Jabbour 2014 −1.0986 1.6357 225 226 0.33 [0.01; 8.23]
Kohan 2014 −0.7243 0.6471 168 84 0.48 [0.14; 1.72]
Leiter 2014a 0.2273 0.6739 482 483 1.26 [0.34; 4.70]
REFORM −1.5041 1.1527 28 28 0.22 [0.02; 2.13]
Rosenstock 2012 0.4001 1.6363 281 139 1.49 [0.06; 36.86]
SCORED 0.0000 0.0923 5292 5292 1.00 [0.83; 1.20]
SOLOIST−WHF −0.1656 0.1796 608 614 0.85 [0.60; 1.20]
SPOTLIGHT 0.5170 1.6417 98 54 1.68 [0.07; 41.87]
SUGAR−DM−HF 1.6671 1.5616 52 53 5.30 [0.25; 113.04]
Sone 2019 0.4363 1.6381 176 90 1.55 [0.06; 38.36]
Strojek 2011 0.4906 1.5521 450 146 1.63 [0.08; 34.21]
VERTIS CV −0.0794 0.0816 5499 2747 0.92 [0.79; 1.08]
VERTIS RENAL 0.1410 0.6972 313 154 1.15 [0.29; 4.52]
Wada−2022 1.4061 1.1239 154 154 4.08 [0.45; 36.93]
Weng−2021 0.4070 1.6358 327 163 1.50 [0.06; 37.08]
Yale 2013 −0.6931 1.4202 179 90 0.50 [0.03; 8.09]
Common effect model 40978 30999 0.88 [0.83; 0.94]
Random effects model 0.88 [0.80; 0.96]
Heterogeneity: I^2 = 0%, t^2 = 0.0068, p = 0.87
Semaglutide/Dulaglutide:SGLT2 inhibitors
PIONEER 2 1.1035 1.6345 409 410 3.01 [0.12; 74.22]
SUSTAIN 8 1.1062 1.6346 392 394 3.02 [0.12; 74.43]
Common effect model 801 804 3.02 [0.31; 29.08]
Random effects model 3.02 [0.31; 29.08]
Heterogeneity: I^2 = 0%, t^2 = 0, p = 1.00
0.01
0.1
1
10
100

Figure S6: Forest Plot: Scenario 1 for Cardiovascular death

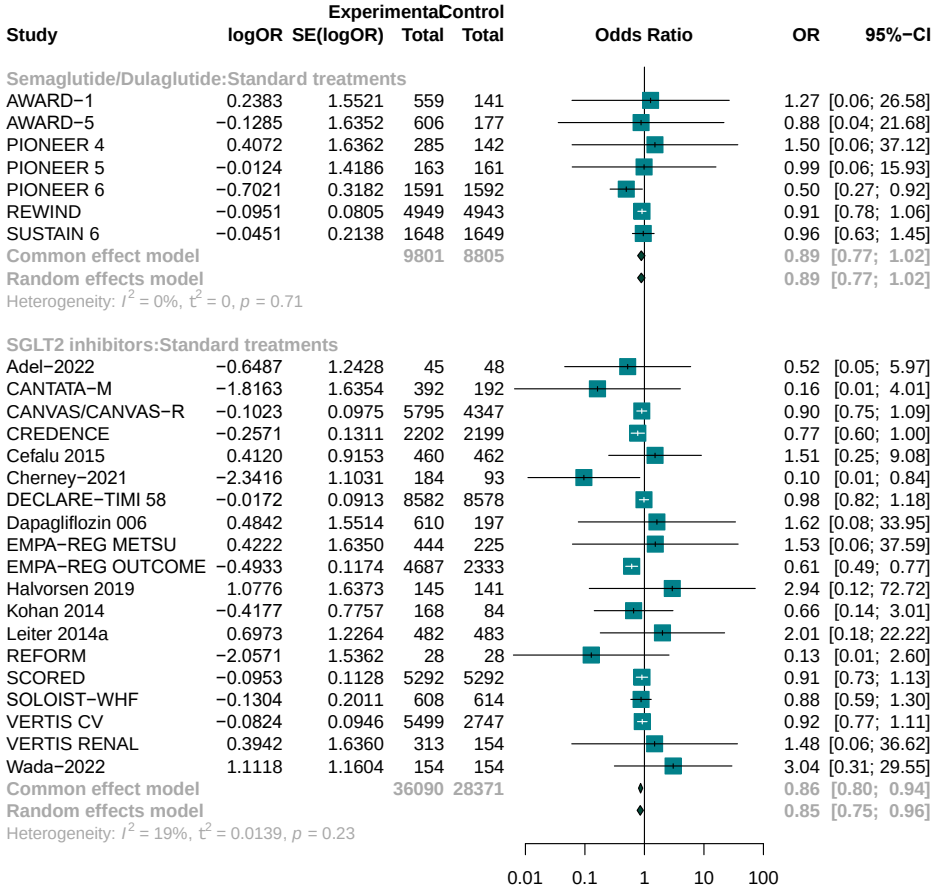


Figure S7: Forest Plot: Scenario 1 for Non-fatal Stroke

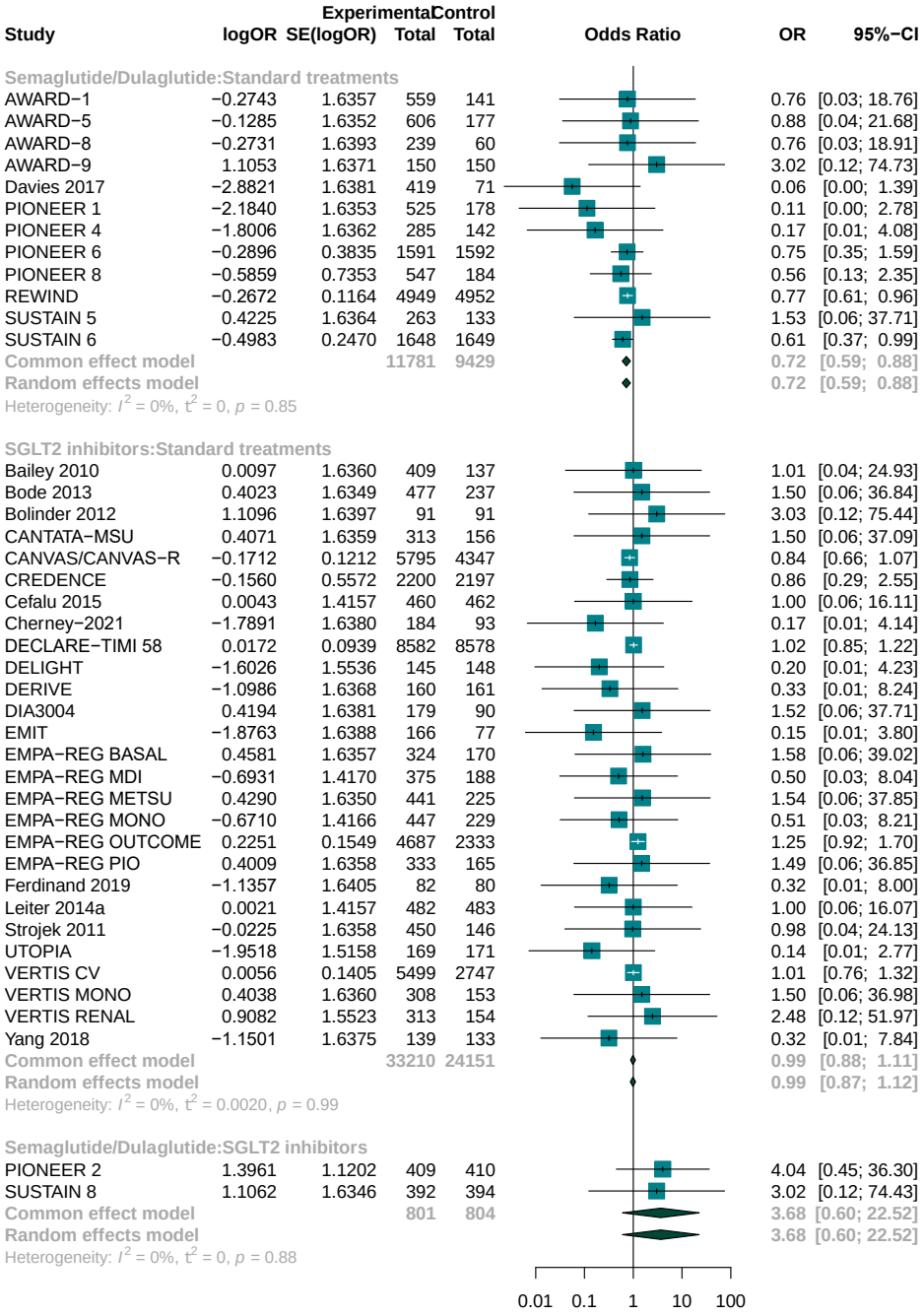


Figure S8: Forest Plot: Scenario 1 for End-stage Kidney Disease

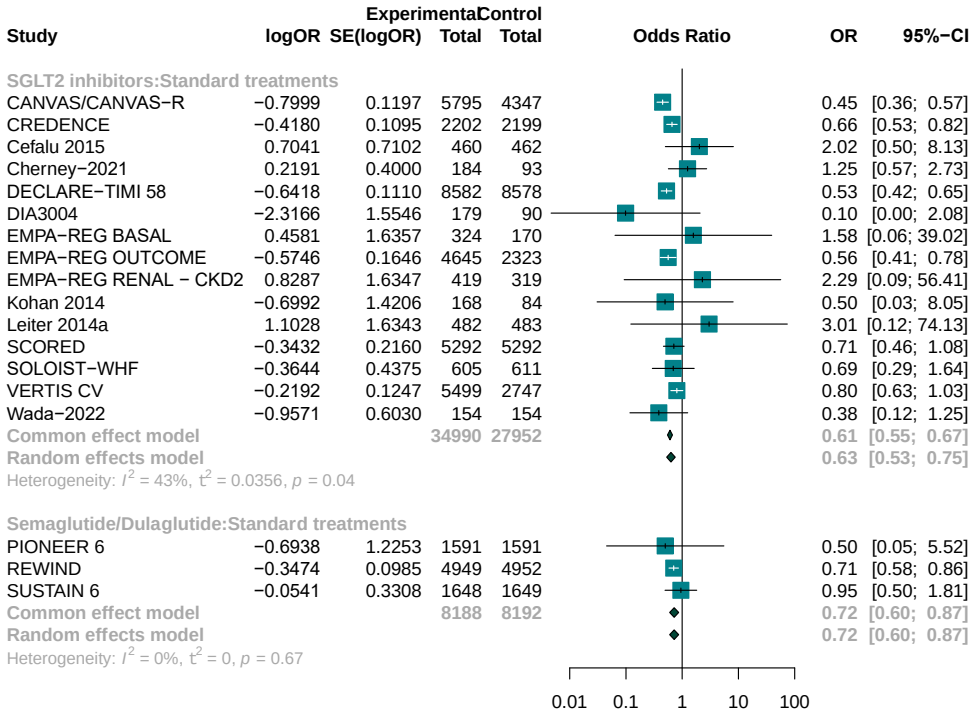


Figure S9: Forest Plot: Scenario 1 for Non-fatal Myocardial Infarction

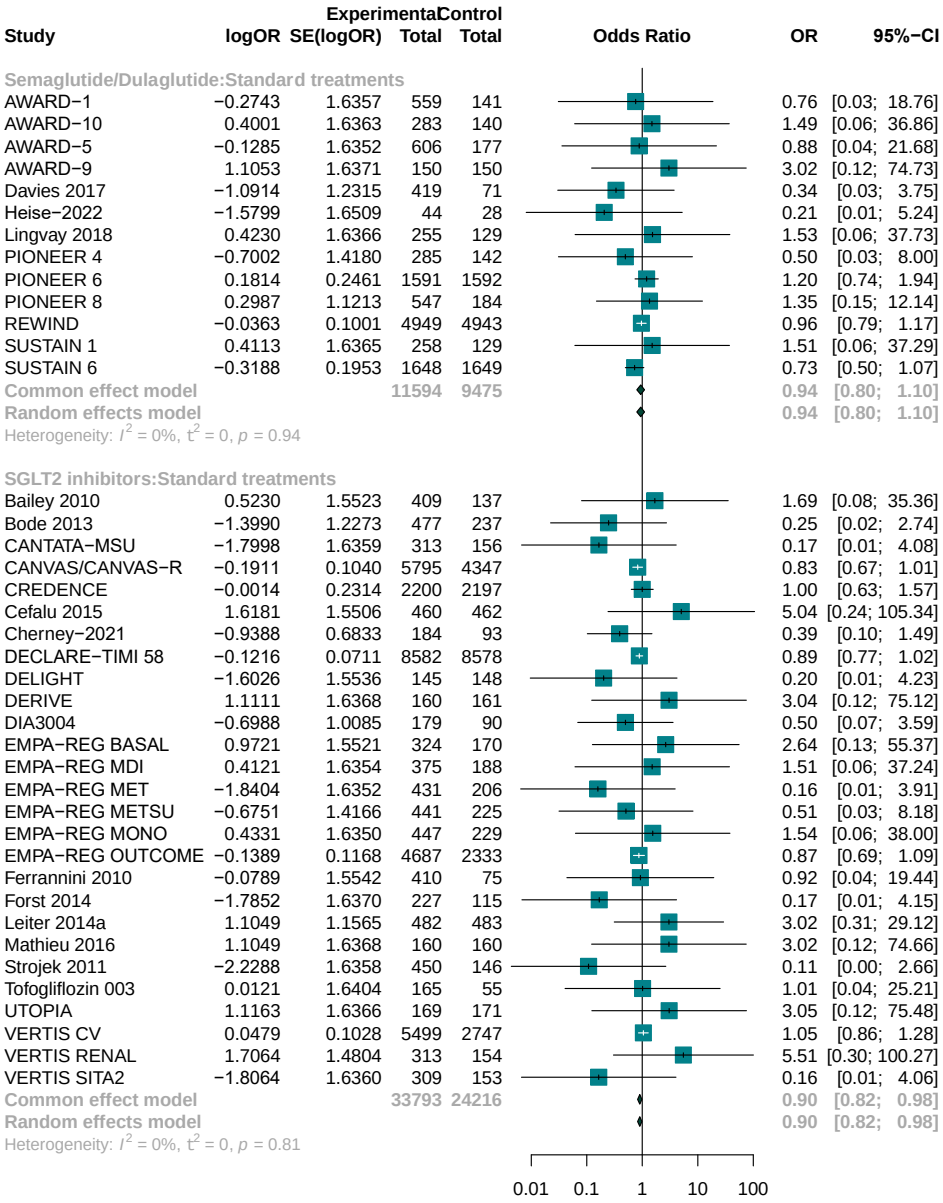


Figure S10: Forest Plot: Scenario 1 for Hospitalization for Heart Failure

Alt Text: Forest plot presenting the relative effect of individual trial and pooled relative effects of each comparison on hospitalization for heart failure

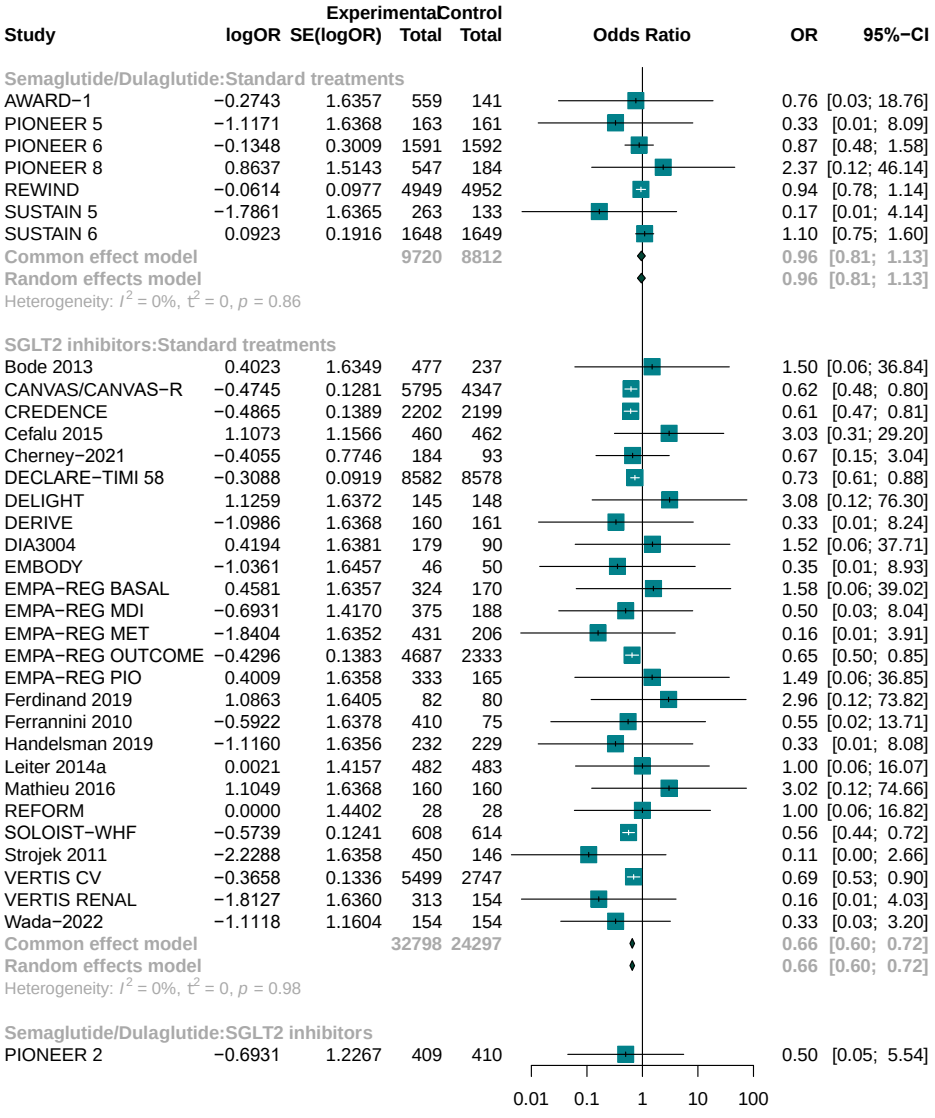
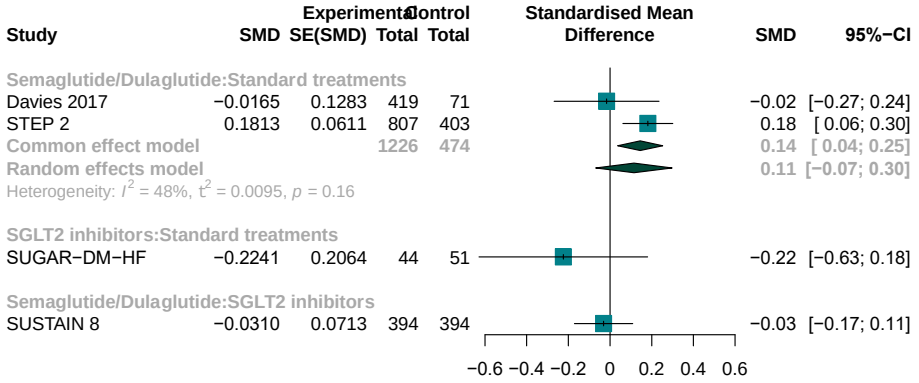


Figure S11: Forest Plot: Scenario 1 for Health-Related Quality of Life

Alt Text: Forest plot presenting the relative effect of individual trial and pooled relative effects of each comparison on health-related quality of life



Appendix 10: Re-analysis to compare SGLT2 inhibitors with Semaglutide – Scenario 2: Forest Plots

These forest plots presenting relative effect of individual trial and pooled relative effects of each comparison.

Figure S12: Forest Plot: Scenario 2 for All-cause Death

Alt Text: Forest plot presenting the relative effect of individual trial and pooled relative effects of each comparison on all-cause death

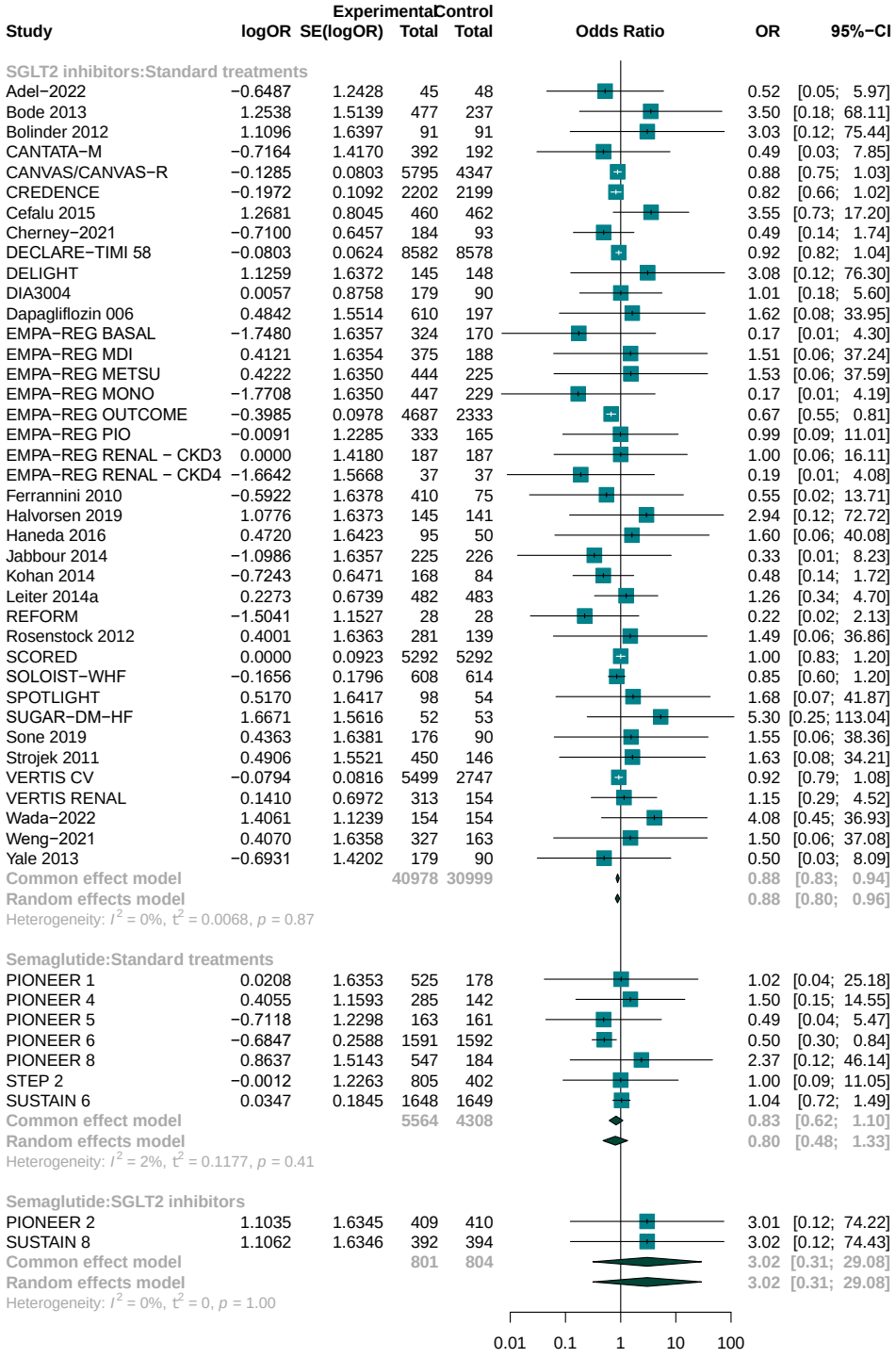

Study
logOR
SE(logOR)
Experimental Total
Control Total
Odds Ratio
OR
95%−CI
SGLT2 inhibitors:Standard treatments
Adel−2022 −0.6487 1.2428 45 48 0.52 [0.05; 5.97]
Bode 2013 1.2538 1.5139 477 237 3.50 [0.18; 68.11]
Bolinder 2012 1.1096 1.6397 91 91 3.03 [0.12; 75.44]
CANTATA−M −0.7164 1.4170 392 192 0.49 [0.03; 7.85]
CANVAS/CANVAS−R −0.1285 0.0803 5795 4347 0.88 [0.75; 1.03]
CREDENCE −0.1972 0.1092 2202 2199 0.82 [0.66; 1.02]
Cefalu 2015 1.2681 0.8045 460 462 3.55 [0.73; 17.20]
Cherney−2021 −0.7100 0.6457 184 93 0.49 [0.14; 1.74]
DECLARE−TIMI 58 −0.0803 0.0624 8582 8578 0.92 [0.82; 1.04]
DELIGHT 1.1259 1.6372 145 148 3.08 [0.12; 76.30]
DIA3004 0.0057 0.8758 179 90 1.01 [0.18; 5.60]
Dapagliflozin 006 0.4842 1.5514 610 197 1.62 [0.08; 33.95]
EMPA−REG BASAL −1.7480 1.6357 324 170 0.17 [0.01; 4.30]
EMPA−REG MDI 0.4121 1.6354 375 188 1.51 [0.06; 37.24]
EMPA−REG METSU 0.4222 1.6350 444 225 1.53 [0.06; 37.59]
EMPA−REG MONO −1.7708 1.6350 447 229 0.17 [0.01; 4.19]
EMPA−REG OUTCOME −0.3985 0.0978 4687 2333 0.67 [0.55; 0.81]
EMPA−REG PIO −0.0091 1.2285 333 165 0.99 [0.09; 11.01]
EMPA−REG RENAL − CKD3 0.0000 1.4180 187 187 1.00 [0.06; 16.11]
EMPA−REG RENAL − CKD4 −1.6642 1.5668 37 37 0.19 [0.01; 4.08]
Ferrannini 2010 −0.5922 1.6378 410 75 0.55 [0.02; 13.71]
Halvorsen 2019 1.0776 1.6373 145 141 2.94 [0.12; 72.72]
Haneda 2016 0.4720 1.6423 95 50 1.60 [0.06; 40.08]
Jabbour 2014 −1.0986 1.6357 225 226 0.33 [0.01; 8.23]
Kohan 2014 −0.7243 0.6471 168 84 0.48 [0.14; 1.72]
Leiter 2014a 0.2273 0.6739 482 483 1.26 [0.34; 4.70]
REFORM −1.5041 1.1527 28 28 0.22 [0.02; 2.13]
Rosenstock 2012 0.4001 1.6363 281 139 1.49 [0.06; 36.86]
SCORED 0.0000 0.0923 5292 5292 1.00 [0.83; 1.20]
SOLOIST−WHF −0.1656 0.1796 608 614 0.85 [0.60; 1.20]
SPOTLIGHT 0.5170 1.6417 98 54 1.68 [0.07; 41.87]
SUGAR−DM−HF 1.6671 1.5616 52 53 5.30 [0.25; 113.04]
Sone 2019 0.4363 1.6381 176 90 1.55 [0.06; 38.36]
Strojek 2011 0.4906 1.5521 450 146 1.63 [0.08; 34.21]
VERTIS CV −0.0794 0.0816 5499 2747 0.92 [0.79; 1.08]
VERTIS RENAL 0.1410 0.6972 313 154 1.15 [0.29; 4.52]
Wada−2022 1.4061 1.1239 154 154 4.08 [0.45; 36.93]
Weng−2021 0.4070 1.6358 327 163 1.50 [0.06; 37.08]
Yale 2013 −0.6931 1.4202 179 90 0.50 [0.03; 8.09]
Common effect model 40978 30999 0.88 [0.83; 0.94]
Random effects model 0.88 [0.80; 0.96]
Heterogeneity: I^2 = 0%, t^2 = 0.0068, p = 0.87
Semaglutide:Standard treatments
PIONEER 1 0.0208 1.6353 525 178 1.02 [0.04; 25.18]
PIONEER 4 0.4055 1.1593 285 142 1.50 [0.15; 14.55]
PIONEER 5 −0.7118 1.2298 163 161 0.49 [0.04; 5.47]
PIONEER 6 −0.6847 0.2588 1591 1592 0.50 [0.30; 0.84]
PIONEER 8 0.8637 1.5143 547 184 2.37 [0.12; 46.14]
STEP 2 −0.0012 1.2263 805 402 1.00 [0.09; 11.05]
SUSTAIN 6 0.0347 0.1845 1648 1649 1.04 [0.72; 1.49]
Common effect model 5564 4308 0.83 [0.62; 1.10]
Random effects model 0.80 [0.48; 1.33]
Heterogeneity: I^2 = 2%, t^2 = 0.1177, p = 0.41
Semaglutide:SGLT2 inhibitors
PIONEER 2 1.1035 1.6345 409 410 3.01 [0.12; 74.22]
SUSTAIN 8 1.1062 1.6346 392 394 3.02 [0.12; 74.43]
Common effect model 801 804 3.02 [0.31; 29.08]
Random effects model 3.02 [0.31; 29.08]
Heterogeneity: I^2 = 0%, t^2 = 0, p = 1.00
0.01
0.1
1
10
100

Figure S13: Forest Plot: Scenario 2 for Cardiovascular Death

Alt Text: Forest plot presenting the relative effect of individual trial and pooled relative effects of each comparison on cardiovascular death

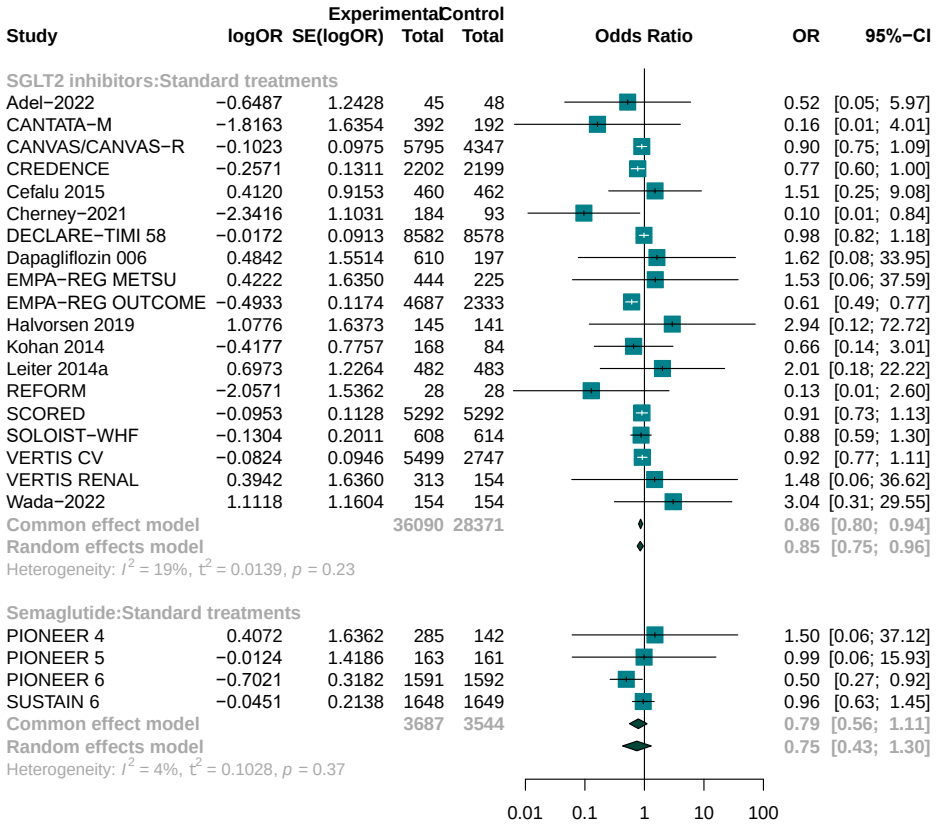


Figure S14: Forest Plot: Scenario 2 for Non-fatal Stroke

Alt Text: Forest plot presenting the relative effect of individual trial and pooled relative effects of each comparison on non-fatal stroke

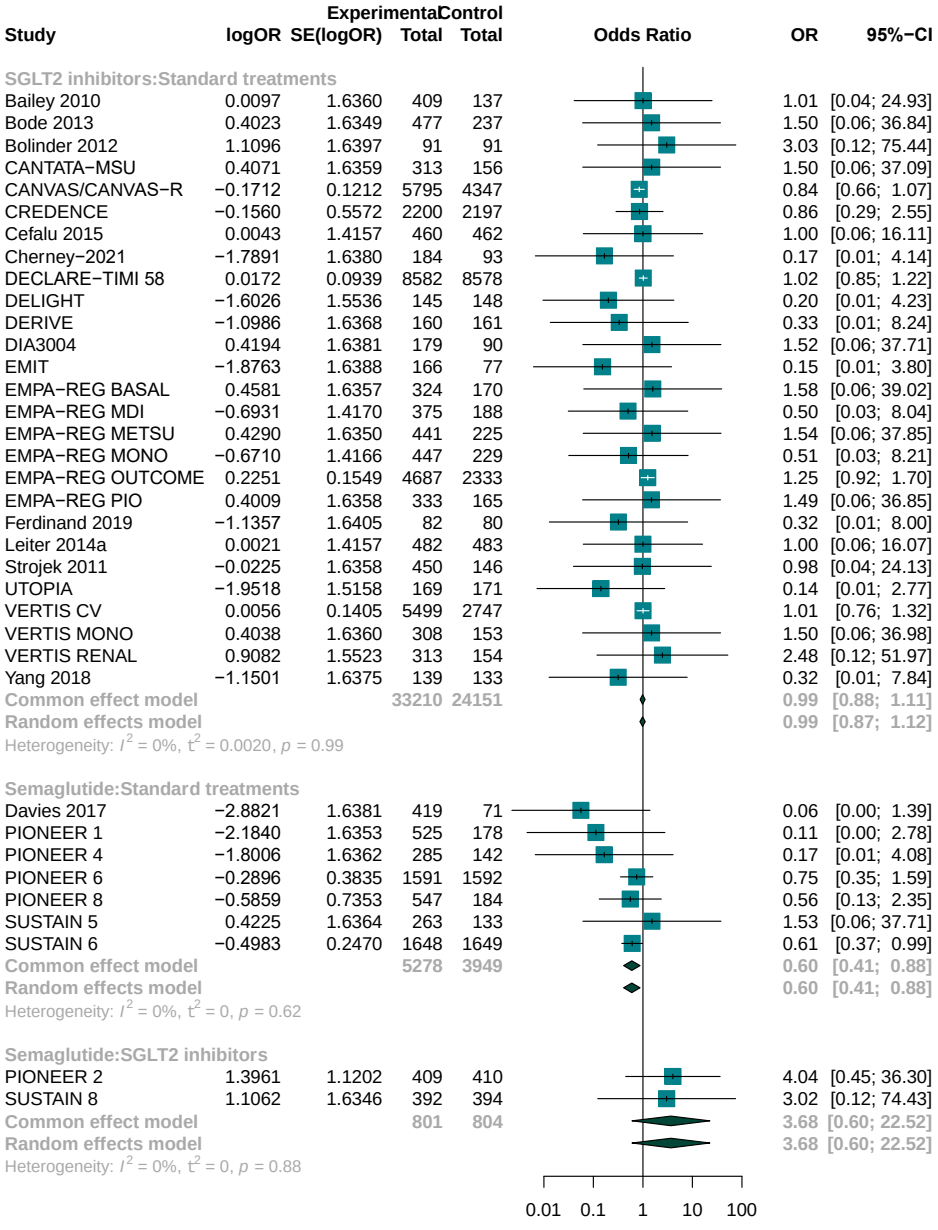
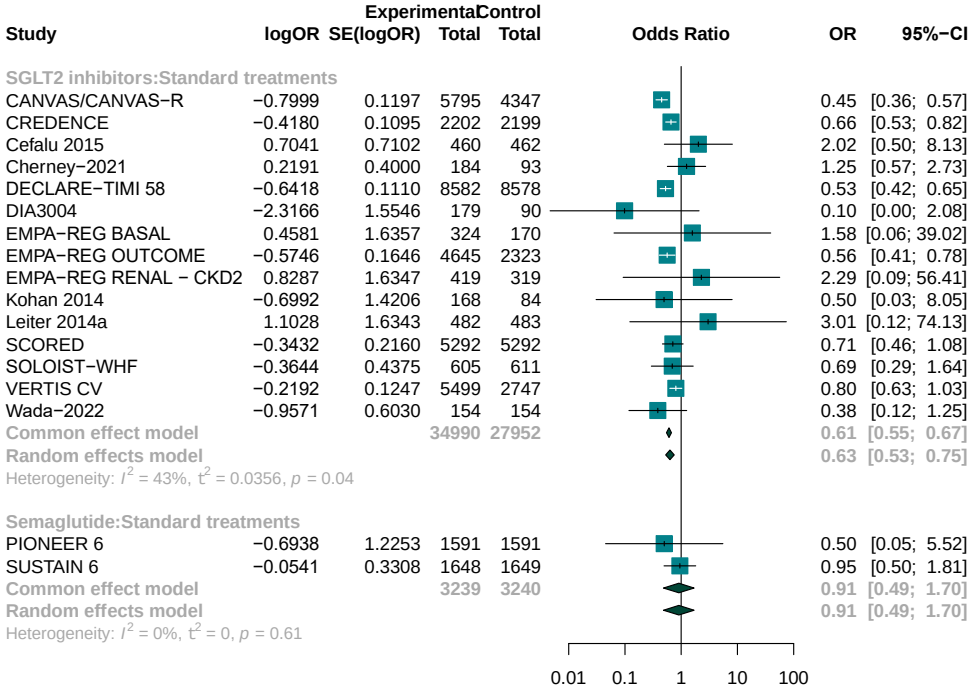


Figure S15: Forest Plot: Scenario 2 for End-stage Kidney Disease

Alt Text: Forest plot presenting the relative effect of individual trial and pooled relative effects of each comparison on end-stage kidney disease



References

1. Shea BJ, Reeves BC, Wells G, et al. AMSTAR 2: a critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions, or both. *BMJ.* 2017;358:j4008.
2. Shi Q, Nong K, Vandvik PO, et al. Benefits and harms of drug treatment for type 2 diabetes: systematic review and network meta-analysis of randomised controlled trials. *BMJ.* 2023;381:e074068.
3. Jansen JP, Trikalinos T, Cappelleri JC, et al. Indirect treatment comparison/network meta-analysis study questionnaire to assess relevance and credibility to inform health care decision making: an ISPOR-AMCP-NPC Good Practice Task Force report. *Value Health.* 2014;17(2):157-173.
4. 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.