Universitat de Girona
Facultat de Medicina

Hospital Universitari de Girona
Doctor Josep Trueta

FINAL DEGREE PROJECT

FRAGILITY INDEX AND RISK OF BIAS IN CARDIOVASCULAR OUTCOME TRIALS FOR NEW ANTIDIABETIC DRUGS

Alba Dalmau Vila

Clinical tutors: **Dr Mercè Fernández-Balsells, Dr Wifredo Ricart Engel**

Methodological tutor**: Dr Teresa Puig**

Department of Endocrinology
Dr. Josep Trueta University Hospital
Universitat de Girona, Faculty of Medicine
November 2020

I would like to express my most sincere gratitude to Mercè Fernández-Balsells and Wifredo Ricart for being so supportive and encouraging. Leading by example, they have taught me beyond papers and I will be forever grateful for that.

I would also like to thank Teresa Puig for her methodological advices and for being so involved with us, students.

At last but not least, I would like to thank my beloved family and friends for always believing in me and for being there every step of the way. You are more than I could ever ask for.

ABSTRACT

Background: The alarm set by rosiglitazone, a thiazolidindione that seemed to enhance the existing two to three-fold increased risk of cardiovascular disease comprised in T2DM, induced that regulatory agencies issued a guidance for proving cardiovascular safety for new antidiabetic therapies. Even though thiazolidindiones' safety was proven afterwards, cardiovascular outcome trials (CVOTs) have been conducted since 2008, using mainly a non-inferiority trial design. These aim for the rejection of their null hypothesis; that the new drug is not unacceptably worse than the standard treatment. In order to achieve that, the upper bound of the 95% confidence interval (CI) cannot exceed a HR ≤1,3 for the primary end point of combined cardiovascular death, nonfatal stroke and nonfatal myocardial infarction. Surprisingly, some of these trials happen to find a cardiovascular outcome HR lower than 1, meaning that cardiovascular events are lower in the experimental group than in the control group. And therefore, are the mainstay of the pretended cardiovascular risk reduction of some new antidiabetic agents such as SGLT-2i or GLP-1RA, ending in the inclusion of these drugs in recent widely used clinical practice recommendations' documents.

Objectives: The aim of this study is to assess the robustness and methodological quality of CVOTs conducted for new antidiabetic drugs after the 2008 FDA guidance, that justify changing clinical practice recommendations.

Methods: In order to accomplish that aim, this narrative review applied the Fragility Index (FI) and Fragility Quotient (FQ) to eligible trials, together with the assessment of the Cochrane "Risk of Bias" Tool RoB 2.0 for all CVOTs conducted since 2008 as a measure of internal validity. Duplicate revision is on the way, so that preliminary results are reported.

Results: FI was applied to 7 eligible CVOTs claiming for statistical superiority, obtaining a median FI of 50 (IQR 19-62) and a median FQ of 1,1% (IQR 0.2-1.3), value that was surpassed in 100% of the trials by the median number of patients lost to follow-up and rate of premature discontinuation. Sub-analyses conducted for 3 CVOTs claiming for statistical superiority that were not eligible for FI, showed no statistical significance. RoB assessment conducted for 21 CVOTs resulted in some concerns as the overall appraisal for 75% of them.

Conclusions: CVOTs conducted after the 2008 FDA guidance for new antidiabetic drugs for T2DM do not provide enough statistical robustness and do have some concerns regarding their methodological quality to justify changing clinical practice recommendations. Hence, creating some hesitation on the reliability of the results and whether is ethical to treat with these additional drug therapies that group of patients. Our findings demonstrate that there is a need for critical review for CVOTs, which would probably lead to the revision of treatment recommendations.

KEYWORDS: Type 2 Diabetes mellitus, antidiabetic, Fragility Index, Fragility Quotient, CVOT, RoB 2.0

INDEX

1. ABBREVIATIONS.....2

2. FIGURES AND TABLES.....4

3. INTRODUCTION.....5

3.1 DIABETES OVERVIEW.....5

3.2 DIAGNOSTIC CRITERIA.....7

3.3 CHRONIC COMPLICATIONS 8

3.4 TREATMENT.....11

3.4.1 Lifestyle Modification.....11

3.4.2 Glycaemic Targets.....13

3.4.3 Glycaemic Pharmacological treatment.....14

3.4.4 Cardiovascular Outcome Trials.....17

4. JUSTIFICATION.....18

5. HYPOTHESIS.....19

6. OBJECTIVES.....19

7. METHODS.....20

7.1 IDENTIFICATION OF STUDIES AND DATA COLLECTION.....20

7.2 FRAGILITY INDEX AND FRAGILITY QUOTIENT.....21

7.3 RISK OF BIAS ASSESSMENT.....22

8. RESULTS.....24

8.1 SELECTION OF TRIALS.....24

8.2 FRAGILITY INDEX AND FRAGILITY QUOTIENT RESULTS.....25

8.3 RISK OF BIAS ASSESSMENT RESULTS 31

9. DISCUSSION.....35

10. STRENGHTS AND LIMITATIONS.....38

11. CONCLUSIONS.....39

12. COROLLARY 40

13. REFERENCES.....41

14. ANNEX.....46

1. ABBREVIATIONS

ADA	American Diabetes Association
ASC	Acute Coronary Syndrome
ASCVD	Atherosclerotic Cardiovascular Disease
BMI	Body Mass Index
CARMELINA	Cardiovascular and Renal Microvascular Outcome Study with Linagliptin
CAROLINA	Cardiovascular Outcome Study of Linagliptin vs Glimperide in T2DM
CHF	Congestive Heart Failure
CI	Confidence Interval
CKD	Chronic Kidney Disease
CV	Cardiovascular
CVD	Cardiovascular Disease
CVOT	Cardiovascular Outcome Trials
DASH	Stop Hypertension Style Diet
DKD	Diabetic Kidney Disease
DM	Diabetes mellitus
DPP4-i	Dipeptidyl Peptidase 4 inhibitor
DSMES	Diabetes Self-Management Education and Support
eGFR	Estimated Glomerular Filtration Rate
EXAMINE	Examination of Cardiovascular Outcomes with Alogliptin vs Standard of Care
FDA	Food and Drug Administration
FI	Fragility Index
FQ	Fragility Quotient
GI	Gastrointestinal
GLP-1RA	Glucagon-Like Peptide 1 Receptor Agonist
HDLc	High-Density Lipoprotein Cholesterol
HF	Heart Failure

HR	Hazard Ratio
IFG	Impaired Fasting Glucose
IGT	Impaired Glucose Tolerance
IQR	Interquartile Range
ITT	Intention-to-Treat
LDLc	Low-Density Lipoprotein Cholesterol
MACE	Major Adverse Cardiac Event
MI	Myocardial Infarction
MITT	Modified Intention-to-Treat
MNT	Medical NutritionTherapy
NA	Not applicable
NI	Non-inferior
OGTT	Oral Glucose Tolerance Test
P	p-value
PCOS	Polycystic Ovary Syndrome
PG	Plasma Glucose
PP	Per-Protocol
RCT	Randomised Controlled Trial
RF	Risk Factors
RoB	Risk of Bias
RR	Relative Risk
SAVOR-TIMI 58	Saxagliptin Assessment of Vascular Outcomes Recorded in Patients with Diabetes Mellitus - Thrombolysis in Myocardial Infarction
SC	Standard Care
SLGT-2i	Inhibitors of Sodium-Glucose Cotransporter 2
SU	Sulfonylurea
T1DM	Type 1 Diabetes mellitus
T2DM	Type 2 Diabetes mellitus
TECOS	Trial Evaluating Cardiovascular Outcomes with Sitagliptin
TZD	Thiazolidinedione
UACR	Urinary Albumin-Creatinine Ratio
Y.o	Year-Old

2. FIGURES AND TABLES

FIGURES

Figure 1. Estimated total number of adults (20-79 years) with diabetes in 2019……..……6

Figure 2. Diagnostic criteria for Diabetes mellitus and Prediabetes ……..7

Figure 3. Chronic Diabetes - related complications ……..……8

Figure 4. Cardiovascular disease risk factors ……..9

Figure 5. Main types of Cardiovascular disease in DM……..10

Figure 6. Integral management of T2DM……11

Figure 7. Lifestyle modification items and their impact on T2DM treatment ……....12

Figure 8. Individual factors that should be taken into account in order to stablish a Glycemic target.13

Figure 9. Glucose-lowering medication in type 2 diabetes: overall approach……16

Figure 10. Forest Plot with Differences and 2-Slided CIs Showing 8 Hypothetical Outcomes from a Non-inferiority Trial ……17

Figure 11. Completed and ongoing CVOTs ……..20

Figure 12. Inherent limitations of Fragility Index……..21

Figure 13. CVOT selection algorithm ……24

Figure 14. Results for the primary outcome of the EMPA-REG Outcome and its subgroups ……26

Figure 15. Results for the primary outcome of the CANVAS Program and its subgroups …..……26

Figure 16. Results for the primary outcome of the SUSTAIN-6 trial and its subgroups..……..…26

Figure 17. Risk of Bias of analyzed CVO trials……..31

Figure 18. Summary of Risk of Bias assessment in % ……..32

TABLES

Table 1. Risk Factors for Type 2 Diabetes mellitus ……..5

Table 2. Criteria for testing for Prediabetes or T2DM in asymptomatic adults ……..13

Table 3. Glucose-lowering pharmacological agents available for T2DM ……..…15

Table 4. Cardiovascular Outcome Trials for DPP4-i after the issuance of the FDA 2008 Guidelines …27

Table 5. Cardiovascular outcome trials for SGLT-2i after the issuance of the FDA 2008 Guidelines …28

Table 6. Cardiovascular outcome trials for GLP-1RA after the issuance of the FDA 2008 Guidelines .30

Table 7. Summary of Risk of Bias assessment with RoB 2.0……..32

3. INTRODUCTION

3.1 DIABETES OVERVIEW: DEFINITION, EPIDEMIOLOGY AND RISK FACTORS

Type 2 Diabetes mellitus (T2DM) refers to a group of chronic metabolic disorders that are the result of a complex interaction between genetics and environmental factors, ending up in hyperglycaemia, which is the common phenotype for all types of Diabetes mellitus (DM) (1). What differentiates T2DM from the other types of DM is its risk factors, represented in Table 1, in addition to its pathogenic process, which involves a progressive impaired insulin secretion and/or insulin resistance, added to excessive hepatic glucose production and abnormal fat metabolism (1, 2).

Table 1. Risk Factors for Type 2 Diabetes mellitus

Obesity, years with excess body fat or central adipose tissue	Excess weight itself causes some degree of insulin resistance due to glycolipotoxicity, which is enhanced in visceral/central obesity. Obesity (≥30kg/$m^{2)}$ is **the most important risk factor**, as it is present in 85% of T2DM cases. This epidemic is being referred to as *Diabesity*.
Metabolic syndrome	Presence of 3 of the following 5 factors: visceral adipose tissue, hypertriglyceridemia, reduced high-density lipoprotein (HDL) cholesterol levels, hypertension and glucose intolerance. Its relevance relies on the linkage to 1.5 rate in all-cause mortality and 2-fold increase of CVD outcomes in T2DM.
Sedentarism	Contributes to excess body weight, but also exercise produces an enhanced insulin sensitivity response; leading to a lesser probability of T2DM development.
Smoking	Induces insulin resistance and compensatory insulin-secretion responses, making it more likely to have central fat accumulation.
Socioeconomical status	Low socio-economical status involves higher stress, hopelessness, material deprivation, limited access to healthy food and exercise facilities that lead to excess body weight and T2DM.
Medication	Certain statins, corticoids and beta-blockers can promote T2DM.
Ageing	Ageing itself impairs insulin secretion and enhances insulin resistance through obesity and sarcopenia, which are related to ageing, that has increased due to life expectancy.
Family history	First degree relatives, specially those with earlier age onset, underly unknown polygenic predisposition. Being this association stronger for T2DM than for T1DM.
Ethnicity	**I**nvolves genetic predisposition and enhanced susceptibility, specially for those having a non-white ancestry. Black people, Pima Indians and Hawaiians have the highest prevalence for T2DM worldwide.
Unhealthy eating or dietary habits	Western diet, which consists of a high caloric diet rich in processed red meat, sugar-sweetened beverages and alcohol, together with a low consumption of fruits, vegetables, high-fibre and whole grains, confers a high risk for developing obesity and insulin resistance, and so T2DM.

T2DM is the most common type of DM, accounting for up to 90-95% of the total 463 million DM cases worldwide, representing 9% of adult population, with an incidence of 11.6 cases/1000 people/year in Spain (3,4). This is an estimate that also includes undiagnosed people, which stand for 50%, meaning that 232 out of the 463 million people are unaware that they have the condition, thus remaining untreated and at a higher risk of complications (3,4,5).

Apart from the extensive proportion of undiagnosed and untreated people, another concerning factor is the increase in both incidence and prevalence, which is expected to reach 700 million people by 2045. Thereby, turning T2DM out as one of the most important health issues from XXI century (4). This rapid increase in both prevalence and incidence can be explained one hand, due to improved health care and effective treatments, which have lead to a longer life expectancy, hence, maintaining a high prevalence (4). And on the other hand, due to an increase of its risk factors, being the most relevant: the rise in obesity, sedentarism, high caloric diet and ageing of population (2,3,4,6).

The socioeconomical and geographical distribution that takes place with T2DM, illustrates well how these risk factors have a major impact on its development. For instance, up to 80% of individuals with diabetes live in low or medium-income countries (4). This is thought to be due to limited access to healthcare services (4), but also for the impact that lifestyle changes produce on their own genetic predisposition (3). Thereupon, creating a geographical distribution as seen in the map below (Figure 1), where certain Pacific Islands, the Middle East, India and the United States have the highest incidence (4).

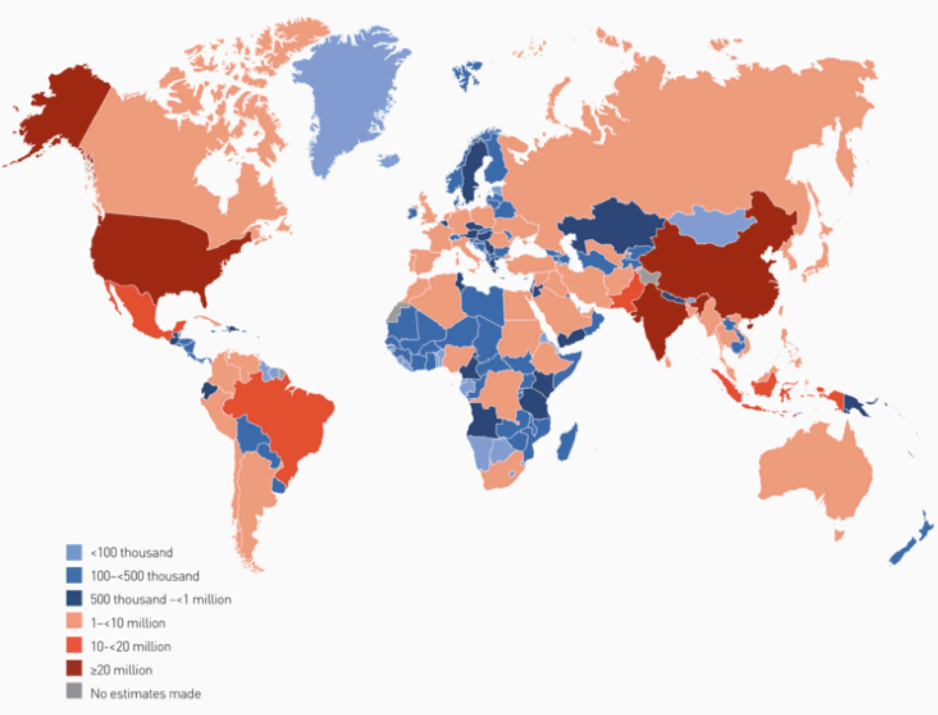


Figure 1. Estimated total number of adults (20-79 years) with diabetes in 2019 (4). Certain Pacific Islands, the Middle East, India and the United States have the highest incidence. The colour of the country or territory in the map relates to the total number of adults aged 20-79 years living with DM in the area.

3.2 DIAGNOSTIC CRITERIA

Diagnostic of DM relies on several parameters related to glucose homeostasis, which are included in the latest updated version of the American Diabetes Association (ADA) guideline as shown in Figure 2:

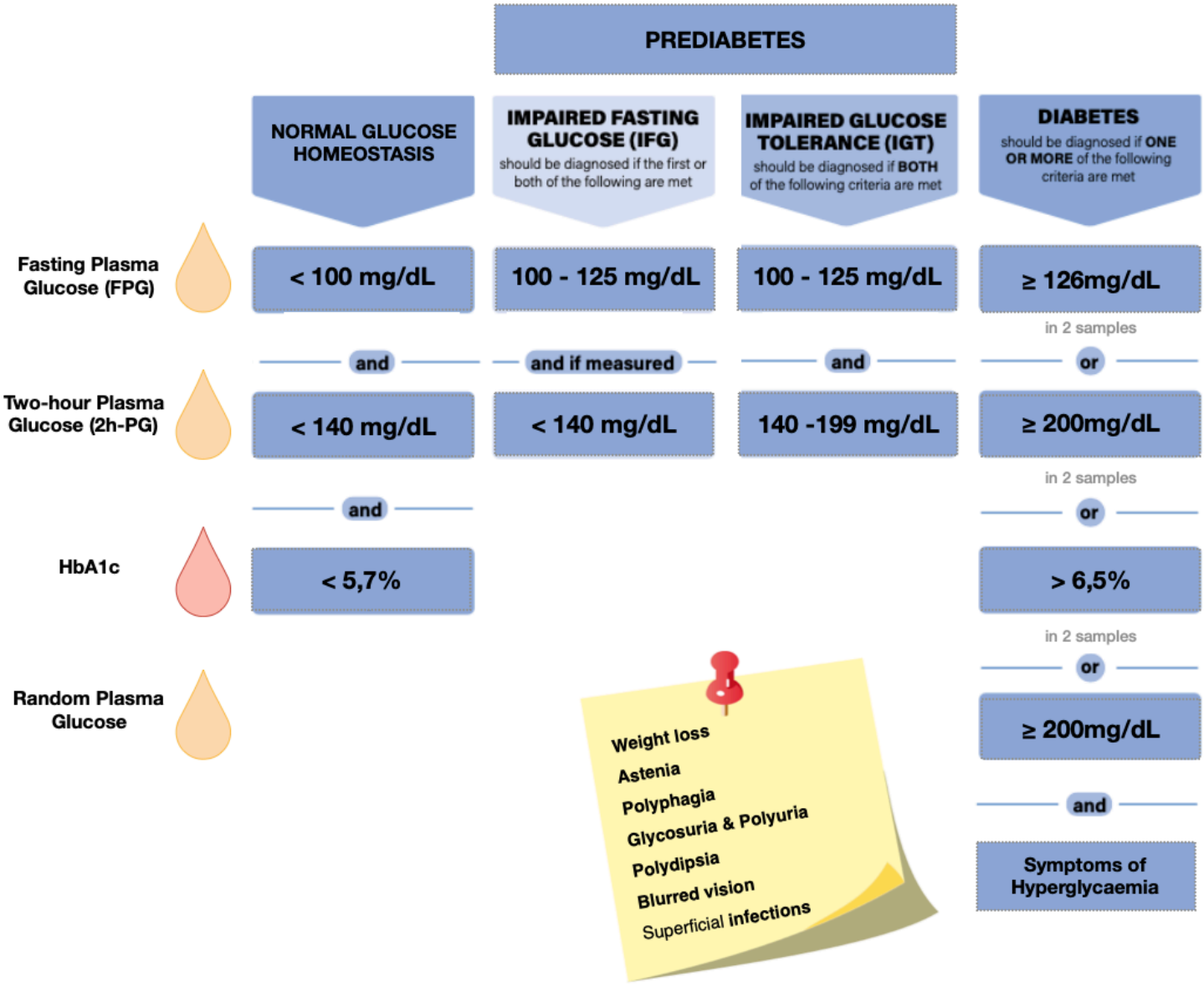


Figure 2. Diagnostic criteria for Diabetes mellitus and Prediabetes, which includes Impaired Glucose Tolerance and Impaired Fasting Glucose**. Fasting plasma glucose (FPG)** defined as plasma glucose levels without food consumption for a minimum of 8 hours. **2-h Plasma Glucose (2-h PG)** defined as plasma glucose levels with a minimum of 12 hours of fasting, followed with an intake of 75g of oral glucose in 5-10 minutes, meanwhile blood extractions are done in several time lapses: 0, 60, 90 and 120 minutes. Glucose intake challenge is also known as oral glucose tolerance test (OGTT). **HbA1c** refers to glycated haemoglobin, which consists of an indirect measure of the average blood glucose levels during the last 120 days. Adapted from (4).

3.3 CHRONIC COMPLICATIONS

Diabetes - related chronic complications are one of the most common issues of DM due to a long hyperglycaemic asymptomatic period before its diagnosis, but also because of a poor glycemic control in some treated patients (4,7). Regardless the cause, persistent high blood glucose levels lead to several multi-organic complications (3,8), turning DM out as one of the main causes of cardiovascular disease, blindness, non-traumatic amputation of the lower-limbs, kidney failure, cancer and death (3). For instance, more than 4 million people passed away in 2019 due to diabetes, which is the equivalent to one death every eight seconds (4).

These complications altogether lead to an important quality of life disturbance, reduced life expectancy and an estimated global expenditure of 645 billion € annually (4,6).

CLASSIFICATION

Chronic complications can be classified in vascular and non vascular complications, as shown in Figure 3, depending on the structure that is being affected by hyperglycaemia.

A) <u>Non vascular complications</u>: are not very prevalent, but are expected to increase in the years that follow as a result of increased incidence and declined mortality. This will lead to more years living with DM and enough time for new emerging non vascular complications to develop as: cancer, infections, major depressive disorder, anxiety, eating disorders and dementia (7,8).

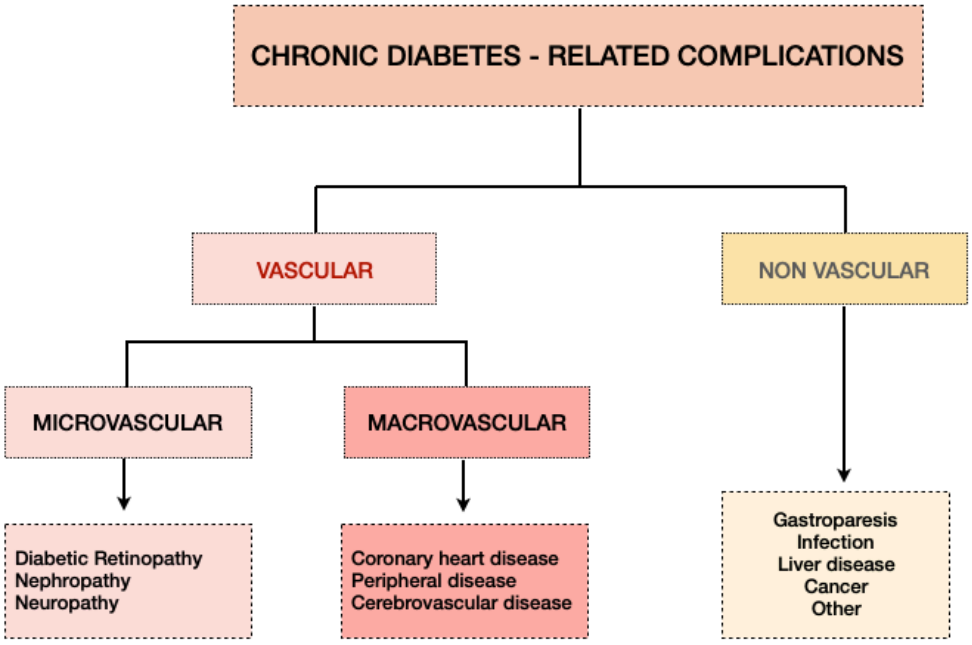


Figure 3. Chronic Diabetes - related complications. Non vascular and vascular complications can be differentiated, being the latter the most frequent. Vascular complications can be further divided into microvascular and macrovascular complications.

B) Vascular complications: are more prevalent than non vascular complications, and can be further divided into microvascular and macrovascular complications (6,7,8).

- **Microvascular complications:** include diabetic eye disease, diabetic neuropathy and diabetic nephropathy. Their relative risk is at least 10-20 times higher compared to non-diabetic population, being directly related to chronic hyperglycaemia and affecting nearly 50% of patients with DM (global aetiology). For that reason, they are considered as diabetic specific and can be addressed with a good glycemic control (7,4).

- **Macrovascular complications:** affect nearly one-third of patients with DM (6), and while their relative risk is 2-4 times higher compared to non-diabetic population, their relevance is greater than microvascular complications because they are the primary cause of morbidity and mortality in DM (6,7,8).

Macrovascular complications are the result of atherosclerotic cardiovascular disease (ASCVD), which develops due to common risk factors present in DM, represented in Figure 4. While DM itself represents a risk factor for developing ASCVD, it is important to remind that 85% of patients with DM have obesity and a high proportion of them have metabolic syndrome (8,9).

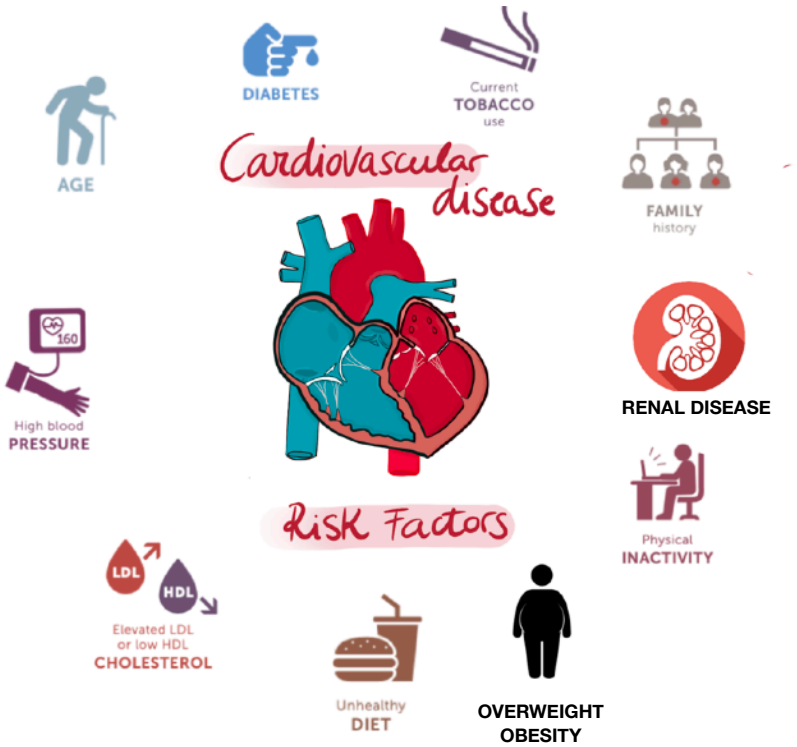


Figure 4. Cardiovascular disease risk factors. These include: smoking, family history, sedentarism, unhealthy diet, elevated LDL or low HDL cholesterol, high blood pressure, ageing, overweight, obesity, albuminuria, chronic kidney disease and diabetes. Adapted from (10).

ASCVD can be manifested as cerebrovascular disease, coronary heart disease or peripheral arterial disease as shown in Figure 5 (8, 10):

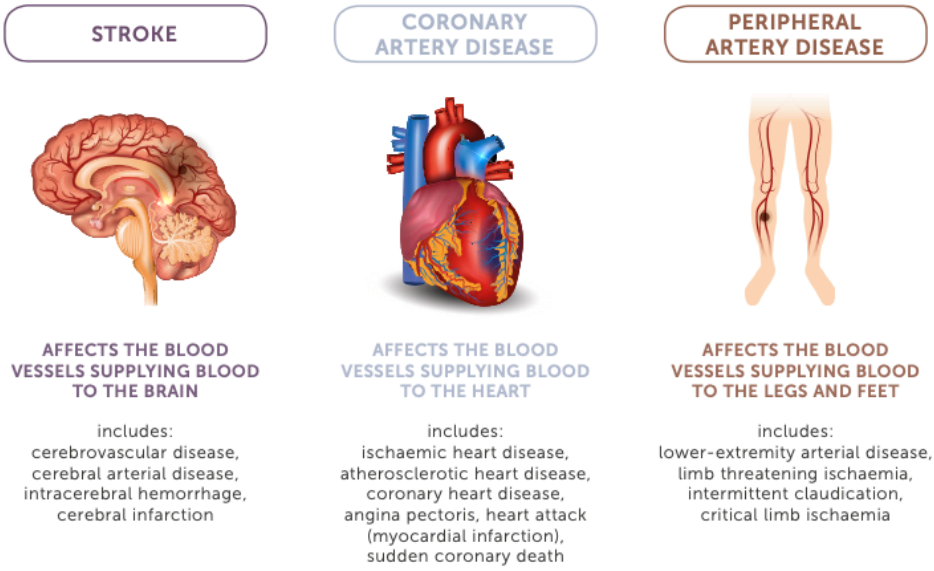


Figure 5. Main types of Cardiovascular disease in DM (10). These include stroke, coronary artery disease and peripheral disease.

These three main types of cardiovascular disease are manifested as a two-to-fourfold increased risk of hospitalisations, procedures, death from acute coronary syndrome, myocardial infarction, ischaemic and hemorrhagic stroke and sudden death. And although all-cause and CVD mortality are decreasing in individuals with DM in high-income countries, due to improved management, it still represents a global burden for this population (6,7).

3.4 TREATMENT

When approaching T2DM treatment, an integrative orientation as shown in Figure 6, has demonstrated to be the most effective way in the long-term management of hyperglycemia and chronic complications. This strategy consists in most of the cases of a combination of lifestyle interventions together with pharmacological treatment. Which help address symptoms and chronic complications by managing hyperglycemia and reducing cardiovascular risk factors (4,10,11).

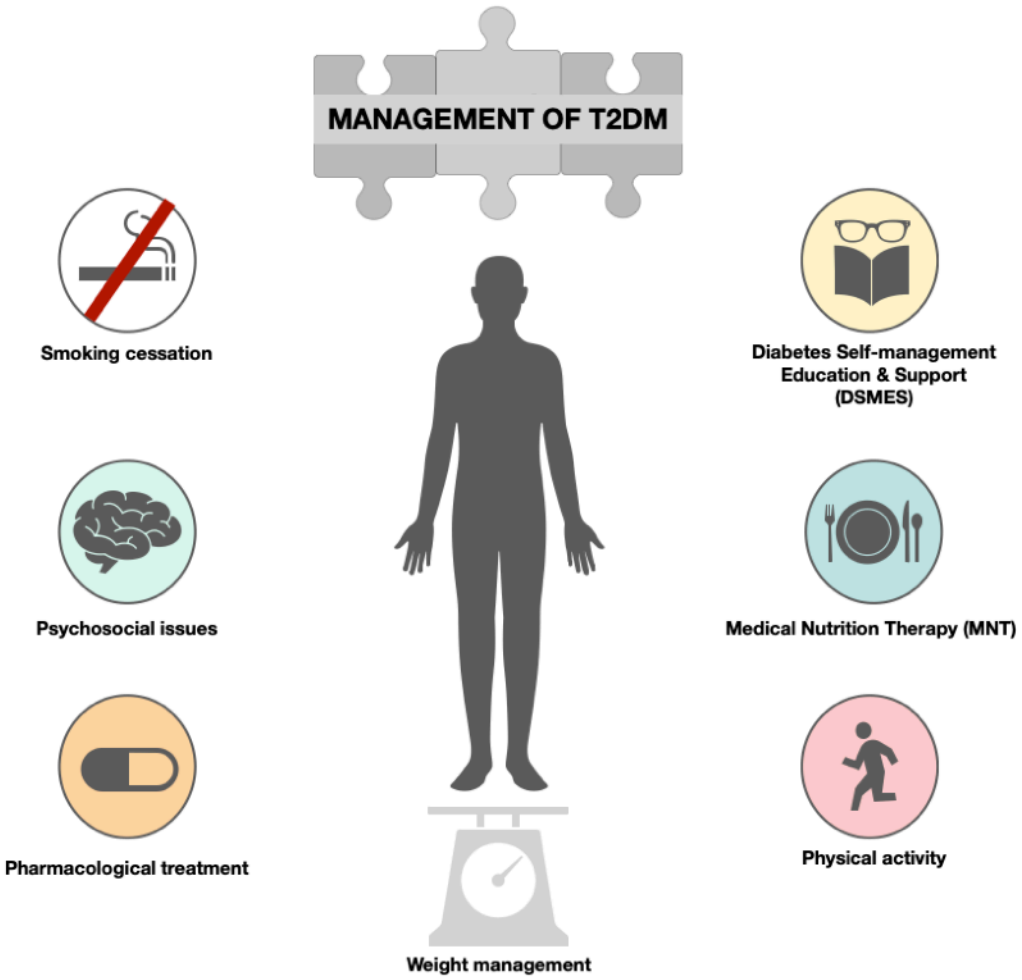


Figure 6. Integral management of T2DM

3.4.1 Lifestyle Modification

Lifestyle modification is imperative in T2DM treatment, however, it requires a high level of engagement from the patient as daily routine and habits need to be modified. These interventions can be arranged in 5 big blocks, as shown in Figure 7**,** which include the following recommendations (9,12):

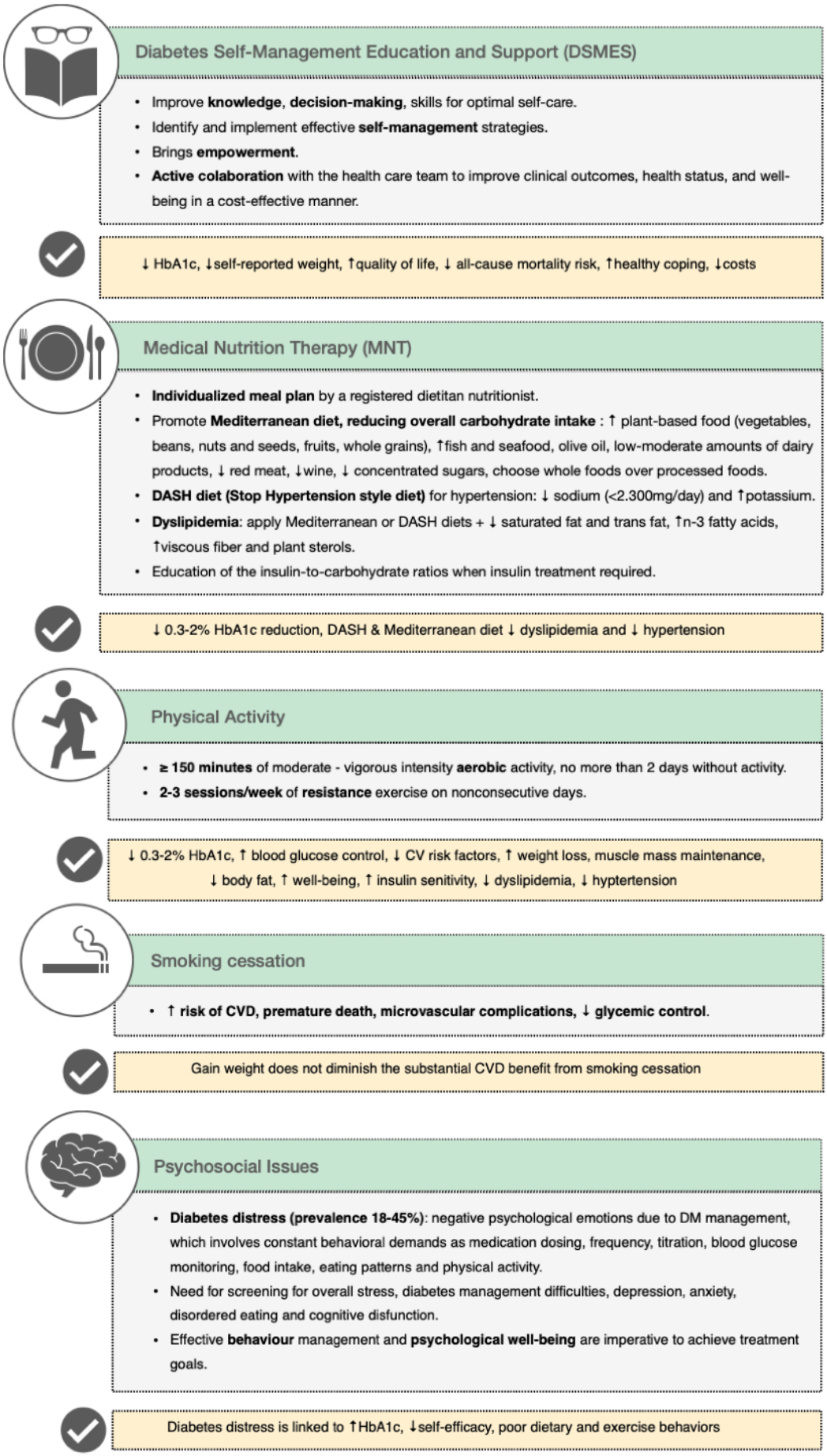


Figure 7. Lifestyle modification items and their impact on T2DM treatment.

Screening and Prevention of T2DM

Strong evidence supports that T2DM can be prevented by addressing cardiovascular risk factors, perviously shown on Figure 4, by applying lifestyle modifications discussed on Figure 7. That is why it is so important to apply the actual prevention guidelines, which put emphasis on maintaining a healthy body weight, obtained in most of the cases by exercising at least 150 minutes per week and eating a healthy diet (4,7,10,13). In addition, screening for prediabetes and T2DM is further recommended in patients who are at high risk, which are defined by ADA as shown in Table 2:

Table 2. Criteria for testing for Prediabetes or T2DM in asymptomatic adults (2)

1. BMI ≥ 25kg/m2 + ≥1 Risk Factor
2. Annual testing in patients with Prediabetes
3. Lifelong testing at least every 3 years in women diagnosed with gestational diabetes mellitus.
4. Initiate testing >45 years in general population, if normal repeat every 3 years.

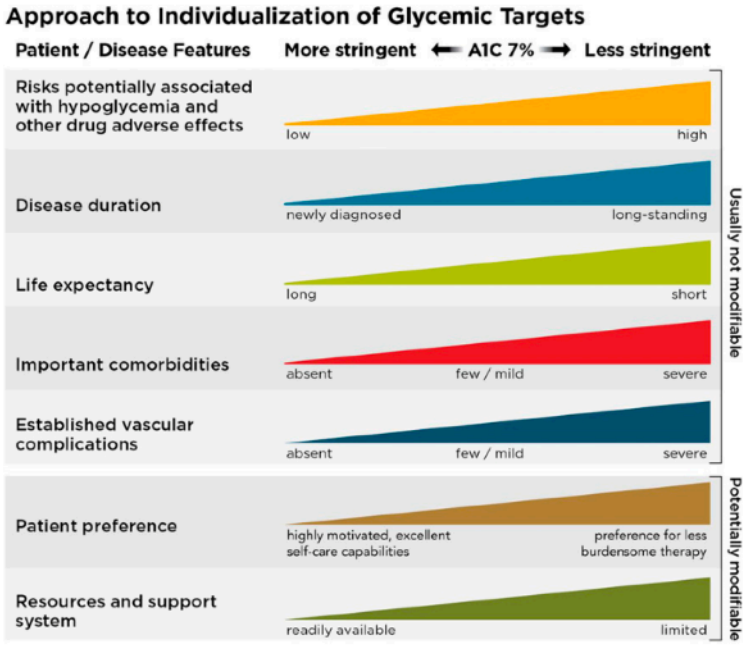
Risk Factors:

- First-degree relative with DM
- Ethnicity related to high risk
- History of CVD
- Hypertension or on treatment
- HLD-colesterol <35mg/dL and/or triglyceride >250mg/dL
- Physical inactivity
- Women with PCOS
- Other clinical conditions associated with insulin resistance

Patients with prediabetes or at high risk should be referred to an intensive behavioural lifestyle intervention program (13).

3.4.2 Glycaemic Targets

As lifestyle interventions are not always enough nor applied correctly, the addition of pharmacological treatment to T2DM management allows handling blood glucose levels in a more efficient manner, lowering glucose to normoglycemia or near-normoglycemia (14).



Approach to glycemic targets should be individualised, taking into account general recommendations which aim for an HbA1c value <7%, together with the patient's characteristics and preferences displayed on Figure 8**.** The end point is to decide with the patient whether to conduct a more stringent effort, aiming for values HbA1c <6,5%, or on the other side deciding a more relaxed approach with values greater than >7%, which are usually preferred in elderly (14).

Figure 8. Individual factors that should be taken into account in order to establish a Glycemic Target (14). Characteristics and predicaments toward left, justify more stringent efforts to lower HbA1c, in contrast to those toward the right which suggest less stringent efforts.

3.4.3 Glycaemic Pharmacological treatment

Glycaemic targets can be accomplished using different pharmacological treatments, which are displayed on Table 3. These drugs are included in the current ADA guidelines, represented on Figure 9, which contain the latest version of T2DM management. The flowchart establishes metformin as the preferred initial pharmacological treatment, which should be administered when T2DM diagnosis is established together with lifestyle modifications. Once metformin is initiated, it should be continued as long as tolerated and not contraindicated, with its effectiveness being reevaluated every 3-6months. This step consists of the baseline for all T2DM treatments (15).

From here, patients are divided into two groups: those who have indicators of high-risk or established atherosclerotic cardiovascular disease (ASCVD), chronic kidney disease (CKD) or heart failure (HF), and those who do not. For the former patients, actual guidelines recommend the addition of another antidiabetic drug in order to obtain a CV benefit, which is not dependent upon HbA1c lowering, preferably SLGT-2i (inhibitors of sodium-glucose cotransporter 2) or GLP-1AR (glucagon-like peptide-1 receptor agonists) (14). These are going to be briefly discussed, as further information can be found on Table 3.

The inclusion of this new recommendation on clinical guidelines, relies on the results that have been obtained from cardiovascular outcome trials (CVOTs) performed for new antidiabetic drugs as a requirement of the FDA. This was issued on 2008, after a published paper brought the possibility of an increased cardiovascular risk of patients assigned to rosiglitazone; a thiazolidenione which proved its safety afterwards (16,17). Since then, several CVOTs have been conducted, some of them demonstrating apparent cardiovascular benefits, and thus leading to their inclusion in clinical practice recommendations (14,15).

From what concerns to SGLT-2i, it has shown a reduction of hyperglycemia in patients with T2DM. The mechanism of action is based on the inhibition of SGLT2, which is a high-capacity and low-affinity glucose transporter expressed exclusively in the luminal membranes of the S1 and S2 segments of the proximal renal tubules. As it is the responsible for the majority of glucose reabsorption, pharmacological treatment with SGLT-2i results in less proximal tubular glucose reabsorption, greater urinary excretion and a loss in glucose that translates in decreased plasma glucose levels (18,19). These include: empagliflozin, canagliflozin, ertugliflozin and dapagliflozin.

GLP-1RA is a family of parenteral glucose-lowering drugs that activate the receptor for the endogenous incretin GLP-1. This translates to lower glucose levels by inhibition of glucagon secretion, promotion of insulin release in response to hyperglycemia, slower gastric emptying and augmented satiety (20). These include: lixisenatide, liraglutide, semaglutide (which also exists as an oral presentation), exanatide, albiglutide and dulaglutide.

Table 3. Glucose-lowering pharmacological agents available for T2DM. DKD: Diabetic Kidney disease, GI: gastrointestinal, CI: contraindicated, eGFR: estimated glomerular filtrate rate, ASCVD: atherosclerotic cardiovascular disease.Adapted from (15).

	Mechanism of Action	Efficacy: ↓HbA1c	Hypoglycemia	Weight Change	CV effects		Renal effects		Additional considerations	Cost
					ASCVD	Heart Failure	DKD	Use considerations		
Biguanides (Metformin) **First-line Therapy**	↓ Hepatic glucose production ↑ peripheral sensitivity ↑ GLP-1	↑high	✗No	= / ↓ Neutral/Loss	↓CV events	= Neutral		**CI:** eGFR **<30mL/ min/1,73m2**	- GI intolerance - Lactic acidosis (rare) - Vitamin B12 deficiency	€ Low
Dipeptidyl peptidase IV (DPP-IVi) inhibitors	Prolong endogenous GLP-1 action	⇅ intermediate	✗No	= Neutral		Potential Risk: **Saxagliptin**	= Neutral	**Adjust:** Sitagliptin, Saxagliptin,Alogliptin. **NO adjustment:** Linagliptin	-Potential risk of **acute pancreatitis** **-Joint pain**	€€€ High
Sulfonylureas	Insulin secretagogue	↑high	✓**Yes**	↑**Gain**	= Neutral			**Glyburide**: not in Renal Failure **Glipizide & glimperide**: initiate conservatively		€ Low
GLP-1 receptor agonists	↑ insulin ↓ glucagon ↓ gastric emptying ↑ satiety	↑high	✗No	↓Loss	**Lixisenatide** = Neutral / Benefit	= Neutral	Benefit of **Liraglutide**	**Adjust:** Exenatide, Lixisenatide. ↑dose: potential risk **acute kidney injury.**	-**Injection** site reactions - Nausea, ↓**GI motility** . **Risk thyroid C-cell tumors (Liraglutide, Albiglutide, Exenatide, Dulaglutide)** Actue **pancreatitis** risk?	€€€ High
Sodium-glucose cotransporter 2 inhibitors (SGLT-2i)	↑ glycosuria by inhibiting SGLT-2 proximal tubule **Insulin independent**	⇅ inter-mediate	✗No	↓Loss	Benefit of **Empagliflozin Canagliflozin**	Benefit of **Empagliflozin, Canagliflozin, Dapagliflozin.**		**Adjust**: Canagliflozin, dapagliflozin, empagliflozin, ertugliflozin.	-**Genitourinary infections** -Risk **volume depletion** - ↑ LDL cholesterol - **Fournier's** gangrene - Risk **amputation** - **Ketoacidosis** - Risk **bone** fractures (Canagliflozin)	€€€ High
Thiazolidinediones (PPAR γ agonists)	↑ glucose utilisation ↓ insulin resistance	↑high	✗No	↑**Gain**	Benefit of **Pioglitazone**	↑**Risk**	= Neutral	No adjustment required **Not recommended** in renal impairment	- ↓ **insulin requirements** - ↑ LDL cholesterol (Rosiglitazone) -Peripheral **edema** - Risk **bone** fracture - Macular Edema	€ Low
Insulin	↑ glucose utilisation ↓ hepatic glucose	↑↑↑ highest	✓**Yes**	↑**Gain**	= Neutral			**Adjust:** ↓ eGFR	-Known safety profile -Injection site reactions	€ Low/ €€€ High

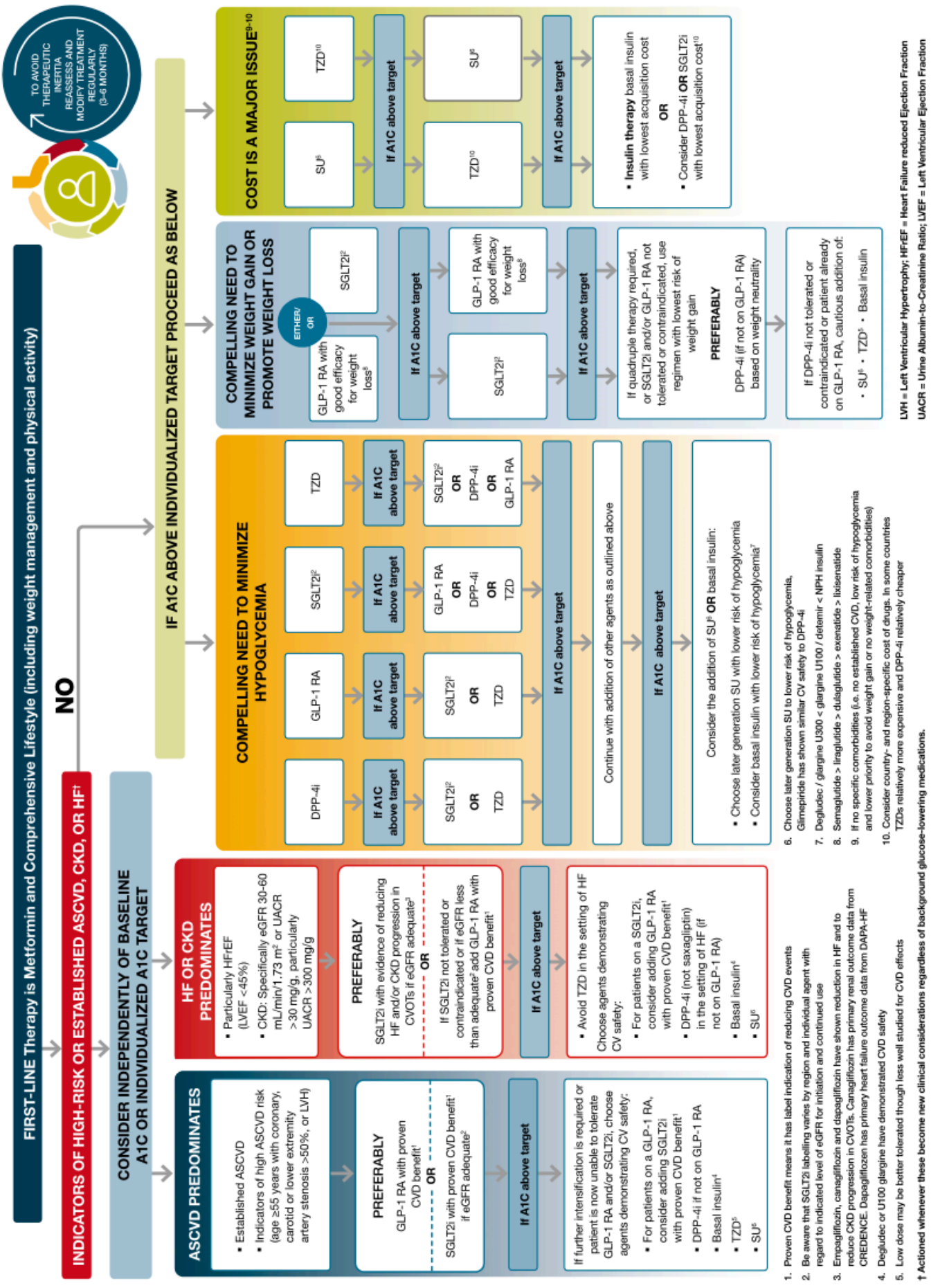


Figure 9. Glucose-lowering medication in type 2 diabetes: overall approach (15). ASCVD : atherosclerotic cardiovascular disease, CKD: chronic kidney disease, CV: cardiovascular, CVD: cardiovascular disease, CVOTs: cardiovascular outcomes trials, DPP-4i: dipeptidyl peptidase 4 inhibitor, eGFR: estimated glomerular filtration rate, GLP-1RA: glucagon-like peptidase 1 receptor agonist, HF: heart failure, SGLT-2i: sodium-glucose cotransporter 2, SU: sulfonylurea, TZD: thiazolidinedione.

3.4.4 Cardiovascular Outcome Trials

From what concerns to CVOTs, most of them are non-inferiority trials, with a design, hypothesis and aims different from those of superiority trials. As a matter of fact, non-inferiority trials seek for the rejection of their null hypothesis, that the experimental treatment is worse than the control group for the primary end point. In order to prove this hypothesis, the new treatment has to show that it is not unacceptably worse that the standard treatment by a predefined non-inferiority margin or Δ. The upper bound of the confidence interval (CI) should not exceed this margin in order to demonstrate non-inferiority (21). This non-inferiority margin has been established by the FDA for CVOTs at a HR of 1,3 for commercialised drugs and of 1,8 for not commercialised drugs (shown in Figure 10.B and C). Once this hypothesis is rejected, non-inferiority is reached. The ambiguous part comes with those trials where a HR of ≤1 is obtained with a confidence interval in a range lower than 1, shown on Figure 10.A, which implies that the proportion of events that are happening in the experimental group are lower than those in the control group. In this situation, superiority is declared. However, it only implies statistical superiority, as the design of the trial has followed a non-inferiority design with a limit of up to a 30% increase with respect to control group in what refers to the number of events (21-23). However, there are some trials that claim for this superiority, pretending to have a positive effect on cardiovascular risk reduction. As a result, these drugs have been included in recent widely used clinical practice recommendations' documents as cardiovascular protective agents (16).

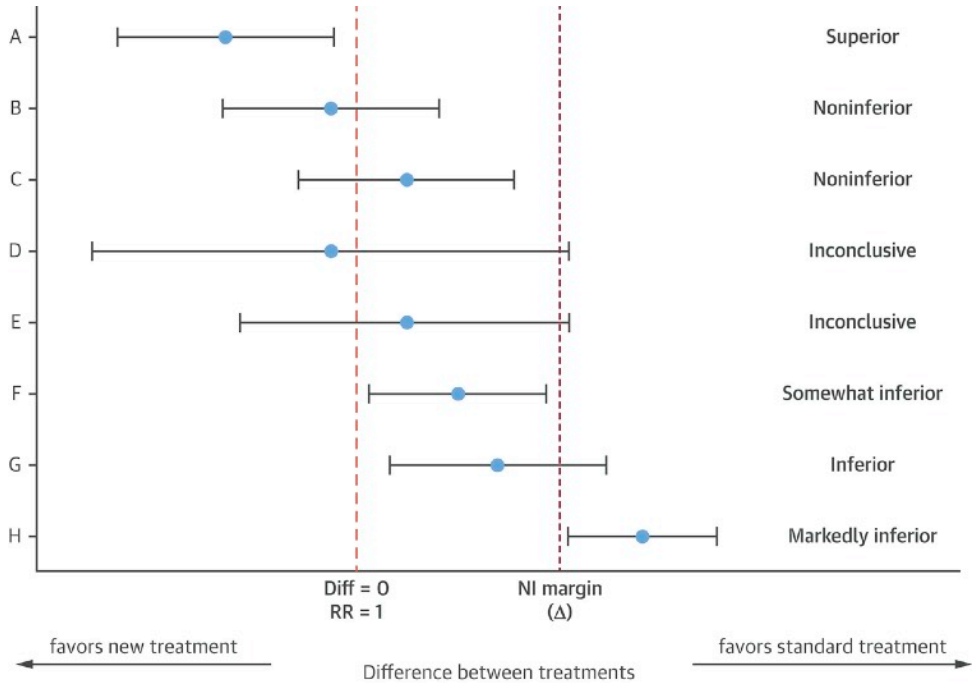


Figure 10. Forest Plot with Differences and 2-Slided CIs Showing 8 Hypothetical Outcomes from a Non-inferiority Trial (23). Orange dashed vertical line indicates zero difference, neutrality. Red dashed vertical line indicates the established non-inferiority margin. CI: confidence interval, diff: difference, NI: non-inferiority, RR: relative risk.

4. JUSTIFICATION

Type 2 Diabetes mellitus (T2DM) is known to be a chronic disease with high blood glucose levels, however, it is also one of the main causes of atherosclerotic cardiovascular disease and death among this population (3,6,7). Whereas addressing known cardiovascular risk factors is a priority in this group, actual clinical guidelines do include the recommendation of treating with additional antidiabetic agents those patients who have risk factors or have indicators of high-risk or established atherosclerotic cardiovascular disease, chronic kidney disease or heart failure. The end point is not to address the glycemic target, but to prevent or delay cardiovascular events (15).

These recommendations are grounded on data from cardiovascular outcome trials (CVOTs), which are mainly non-inferiority trials required for any new antidiabetic treatment to show that it is not unacceptably worse than the standard treatment in what concerns to cardiovascular events. For that aim, a non-inferiority margin or Δ, established as the upper bound of the 95% confidence interval of HR ≤1,3 for the primary end point has to be demonstrated. However, in those cases where a HR with a confidence interval lower than 1 has been achieved, superiority has been declared. Such apparent cardiovascular beneficial effects have resulted in the inclusion of SGLT-2i and GLP-1RA in widely used clinical practice recommendations (14,15). This has raised some concerns, because while declaring significant superiority is not mistaken (24), it is inaccurate to attribute them with a significant effect (21-23).

Moreover, while the inclusion of these drugs in clinical guidelines recommendations relies on their statistical significance, robustness of these trials was only reviewed on 2017 by Kruse (25). His study included 4 CVOTs and showed a lack of statistical robustness. For that reason, the aim of this narrative review is to determine the statistical robustness and methodological quality of the available CVOTs conducted after the 2008 FDA guidance issue, and assess if the actual recommendations grounded on these CVOTs have enough power.

5. HYPOTHESIS

Cardiovascular Outcome Trials conducted after the 2008 FDA guidance for new antidiabetic drugs for T2DM provide enough statistical robustness to justify changing clinical practice recommendations and are of high methodological quality.

6. OBJECTIVES

Objective 1: To evaluate the statistical robustness of CVOTs conducted for new antidiabetic drugs in T2DM since the 2008 FDA Guidance.

Objective 2: To analyse the methodological quality of the same trials, by determining the risk of bias as a measure of the internal validity of these studies.

7. METHODS

7.1 IDENTIFICATION OF STUDIES AND DATA COLLECTION

In order to identify CVOTs conducted after the 2008 FDA guidance publication, we have based our Project on a review of these trials published in 2018 in Diabetes Care (17), which included a useful timeline publication represented schematically in Figure 11.

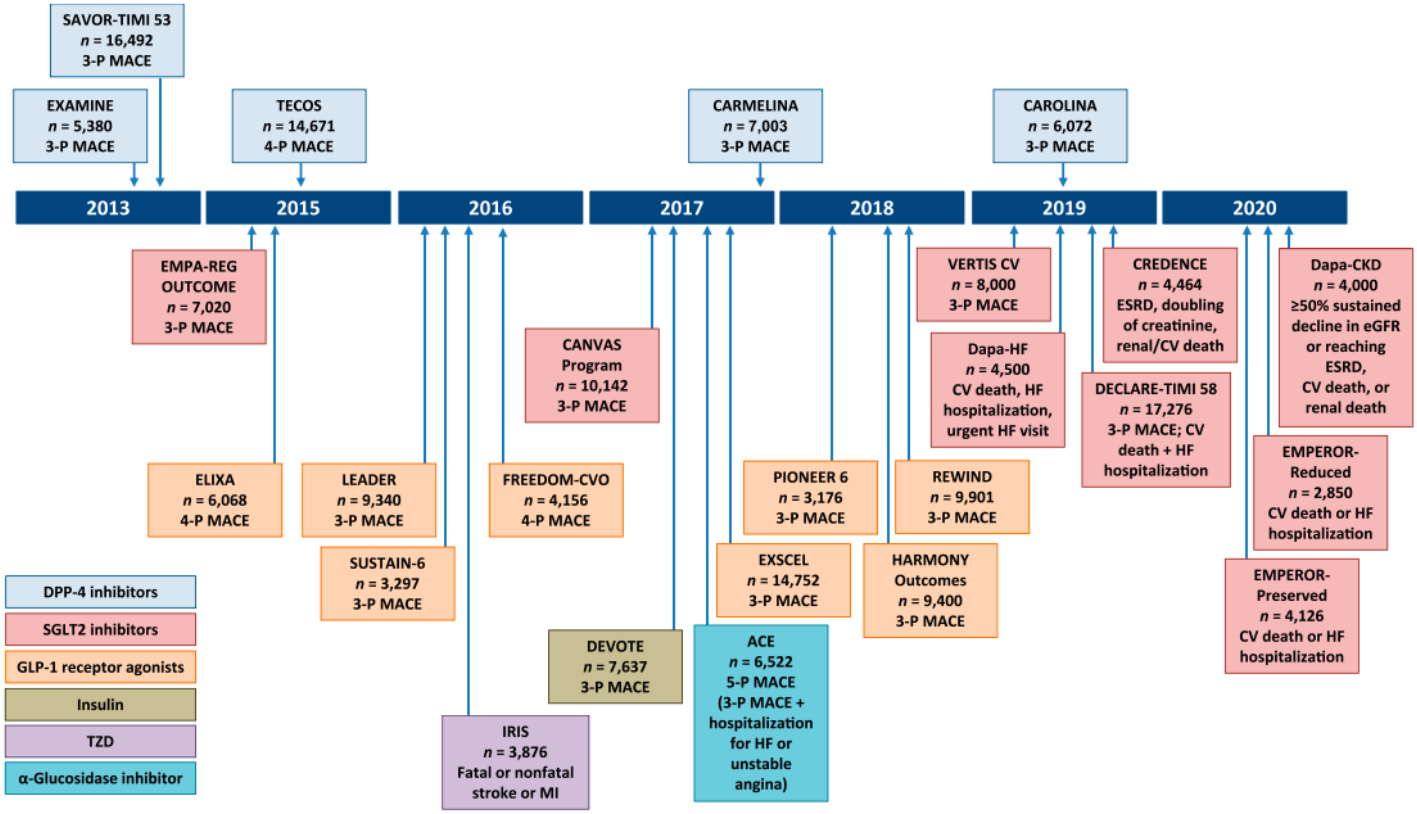


Figure 11. Completed and ongoing CVOTs (17).

Each trial was examined, together with their protocol and supplementary appendix for data extraction. These included: trial identification, year started and reported, median follow-up, total sample size, sample size of each group, intervention, inclusion criteria, HbA1c values, diabetes duration, baseline antidiabetic treatment (if available), prior cardiovascular disease or heart failure, primary outcome, event rates for the primary outcome in each group, statistical significance for the primary outcome, calculus of premature discontinuation, randomisation process, baseline characteristics of both groups, blinding methods, statistical analysis and statistical analysis plan, availability of data related to the primary outcome at the end of the study and methods for measuring the outcome.

7.2 FRAGILITY INDEX AND FRAGILITY QUOTIENT

The Fragility Index (FI) is a metric used to determine the robustness of statistically significant results from randomised trials with dichotomous outcomes. The value obtained with this metric represents the minimum number of patient events that would need to become non-events in order to alter a significant result to a non-significant result ($p \geq 0.05$). The higher the FI value, the bigger the robustness (25).

FI was calculated for trials claiming for statistically superiority using an online calculator at https://clincalc.com/Stats/FragilityIndex.aspx, entering the sample size and the number of events for the primary outcome in each group (25). The program obtains the FI value by converting one patient in the control from non-event to event, and then recalculates a two-sided Fisher's exact test until $p \geq 0.05$ is reached (24,25).

However, FI has some inherent limitations as shown in Figure 12, that should be taken into account for its interpretation (24):

- The use of Fisher's exact test for statistical significance makes it more prone to a type II error.
- It only applies to dichotomous outcomes.
- It cannot be applied to an outcome with continuous variable.
- It is not appropriate for time-to-event outcomes.
- There is no standard FI "cut-off" or lower limit of the FI to classify a study as fragile or robust.
- Value against the number of patients lost to follow-up is important because if the number of patients lost to follow-up is greater than the FI, the study should be considered less robust.

Figure 12. Inherent limitations of Fragility Index.

Indeed, while Fragility Index (FI) has some limitations, it is a metric tool that has been used by some authors in other medical areas as orthopaedic surgery, general medicine, cardiac disease and heart failure in order to review statistical significance in RCT. The main reason of its use is because there has been increasing concern on the lack of knowledge that practitioners have on statistics. Leading them to rely solely on p-value and the distance of the lower boundary of a CI to determine whether a trial presents statistical significance. When in some occasions, it depends on just 1 event to turn a significant to a non-significant intervention. This might have a huge impact on treatment guidelines, as their foundation often start with the decision of whether a treatment effect is believed to exist (23,26).

The Fragility Quotient (FQ) was calculated for each outcome by dividing FI by the sample size of the trial. FQ allows to omit the effect that sample size can have on FI, with a smaller FQ indicating a less robust study outcome (26). FQ values were expressed in percentages in order to ease its interpretation.

Median and interquartile range (IQR) for FI and FQ values were calculated using the online calculator www.alcula.com.

7.3 RISK OF BIAS ASSESSMENT

Risk of bias refers to the systematic error or deviation from the truth that a study may present, over or underestimating the effect of the intervention. Which should be differentiated from quality; because bias can occur in well-conducted studies and not all methodological flaws introduce bias, poor reporting or imprecision. Hence, it focuses on the internal validity, explaining whether the result reflects what the study aims to estimate (27).

In order to assess the risk of bias, the Cochrane "Risk of Bias" Tool 2.0 (RoB 2.0), which runs on Excel, was used together with the guidance of the corresponding manual (28) and a series of videos included in the "RoB 2: Learning webinar series" from the Cochrane Training Website (27).

This tool provides different sets of questions comprised in 5 bias domains, which are all mandatory, plus an overall bias which is used to guide analysis and interpretation (28). Depending on the answer to each question, an algorithm map is created, obtaining 3 levels of risk of bias: low, some concerns or high risk of bias. Detailed information about domains with their corresponding items, guidance and response options are included in the Annex.

The 5 assessed bias Domains, together with a brief description, are the ones that follow (27):

Domain 1. Bias arising from the randomisation process.
The aim of this domain is mainly to assess whether allocation sequence was random and if there were baseline differences between groups.

Domain 2. Bias due to deviations from the intended interventions.
This Domain includes blinding of both participants and trial personnel, which is essential in trials that aim to eliminate placebo effects and isolate specific effects of protocol interventions. It also includes deviations from intended interventions, and finally the used analysis in order to estimate the effect of assignment to intervention.

Deviations from intended intervention include 3 relevant aspects:
1. Additional given interventions that are inconsistent with the trial protocol (non-protocol interventions).
2. Failure by the trial staff to implement the intervention as intended.
3. Non-adherence to assigned intervention by trial participants.

It is important to consider that RoB 2.0 tool considers changes to intervention included in the "package of care" or "usual care" due to drug toxicities or disease progress that are consistent with trial protocol, even if not written down, as *not contributing to increment the risk of bias*. This is justified by stating that the conducted interventions are meant to be implemented in the clinical practice where the standard care is going to be applied. However, it does include the experience of side effects or toxicities that are specific to one of the interventions as high risk of bias or some concerns.

Finally, from what concerns to the method used for the analysis, it is important to recognise that the RoB 2.0 tool considers as appropriate both “intention-to-treat” (ITT) and its modified version (mITT), which are both used when the effect of interest is that of assignment to intervention. ITT includes all randomised participants regardless of the intervention received, measuring the outcome data for all of them. Whether if the effect of adhering to intervention is the aim, a “Per-Protocol” (PP) analysis should be conducted. However, both “Per-Protocol” in addition to “As treated” (trial participants grouped according to the intervention that they received, rather than to the assigned intervention) should be considered as inappropriate when applying RoB 2.0 tool assessment.

Domain 3. Bias due to missing outcome data.

RoB 2.0 states that the availability of data from 95% of the participants at the end of the study is sufficient and should be considered as low risk of bias.

Domain 4. Bias in measurement of the outcome.

In randomised trials, outcome measurement is usually performed similarly in both groups, however 2 specific situations may arise:

- Different outcome assessors: the outcome assessor should be an observer not directly involved in the intervention provided, however, there are some situations where a participant-reported outcome is performed, which may raise some concerns.
- “Diagnostic detection bias”: takes place in those situations where the number of visits and complementary tests differ between groups because of the intervention and/or adverse events. This may lead to the assumption from the practitioner of which practice is being held on that patient.

Domain 5. Bias in selection of the reported result.

For a particular domain, multiple possible results can be generated, leading to choose the most favourable for the analysis. This can lead to selective reporting, meaning that results and data are reported in order to favour results, for example not reporting a result because $p>0.05$.

For this domain is important to consider the Pre-specified analysis plan, which can be found in the trial protocol or its statistical analysis plan (SAP).

8. RESULTS

8.1 SELECTION OF TRIALS

From the 25 potentially eligible CVOTs, as represented in Figure 11, 3 of them (IRIS, DEVOTE and ACE) were excluded because they were not initiated as a direct result of the FDA guidance (17). Another one, the FREEDOM CVO, was not included either because its results have not been published yet. This lead to 21 selected CVOTs included in the study, which were all assessed with the RoB tool. From what concerns to Fragility Index, it could be applied to 10 of them as shown on Figure 13, because they followed a 1:1 randomisation and dichotomous outcome.

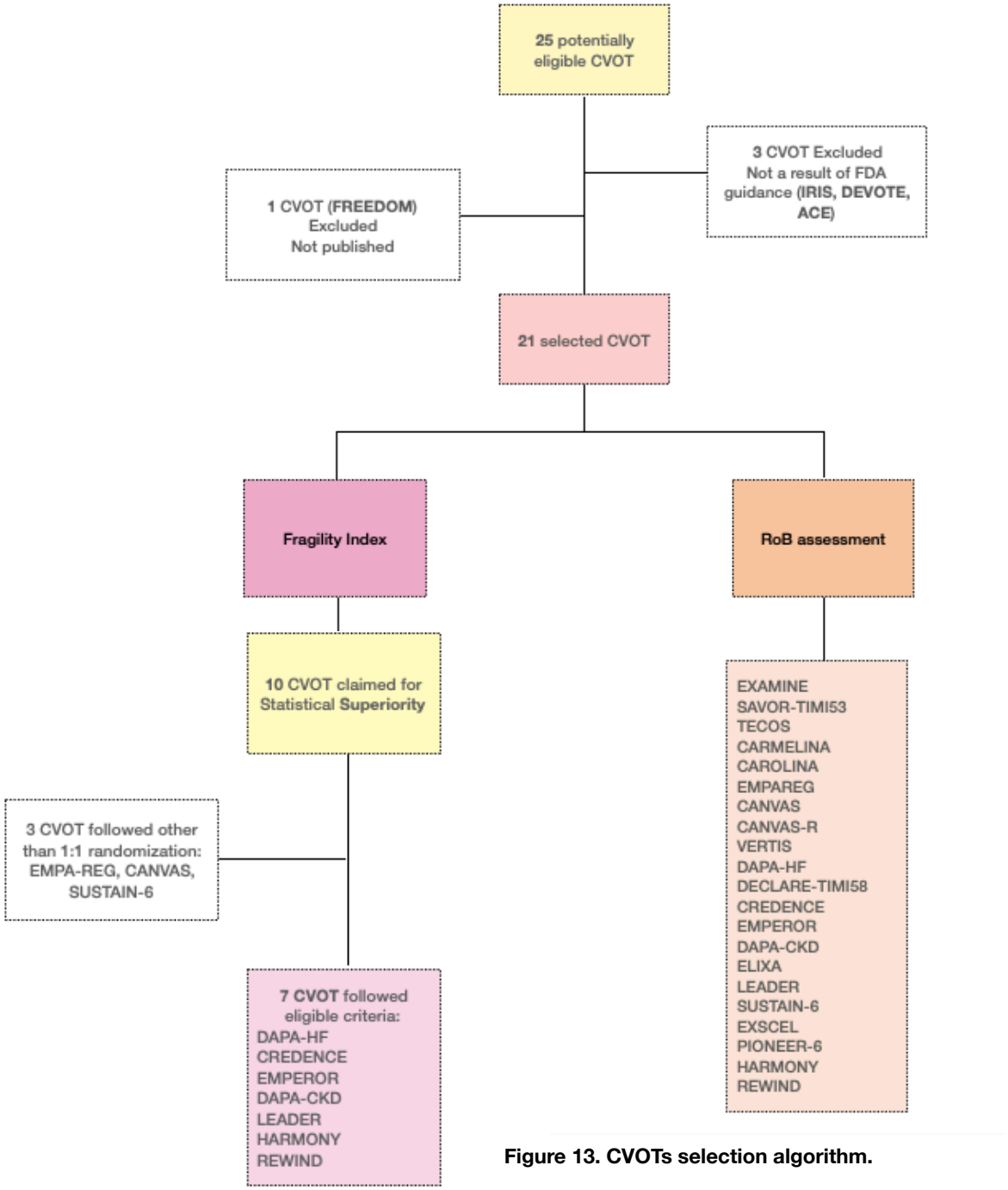


Figure 13. CVOTs selection algorithm.

8.2 FRAGILITY INDEX AND FRAGILITY QUOTIENT RESULTS

As shown in the CVOTs selection algorithm in Figure 13, from the 21 selected CVOTs, 10 of them claimed for statistical superiority, and from those, Fragility Index (FI) could only be applied to 7 of them, as they followed eligible criteria for FI. These were: DAPA-HF (29-31), CREDENCE (32-34), EMPEROR (35-37), DAPA-CKD (38-40), LEADER (41-43), HARMONY (44) and REWIND (45) trials.

The median FI value obtained was of 50 (IQR 19-62), ranging from 4 for the REWIND study to 71 for DAPA-CKD study (45,38-40). Additionally, in order to omit the effect that sample size can have on FI, Fragility Quotient (FQ) was applied obtaining a median FQ of 1,1% (IQR 0.2-1.3). Being in 3 of the 7 CVOTs a value nearly 0, with 0,2% in the LEADER trial, 0,3% in HARMONY and 0,04% in the REWIND study. And in the remaining 4 CVOTs, a value around 1%.

Results can be found on Tables 4 to 6, which present relevant information about each trial, grouped within their corresponding drug family. Apart from FI and FQ values, it also includes the number of patients lost to follow-up, with a median of 22 patients (IQR 4-57) for the 7 CVOTs included in the FI analysis. Being the lowest for the DAPA-HF study with just 2 patients, and the highest value for the REWIND study with 287 patients lost to follow-up (29-31). FI value was surpassed by the number of patients lost to follow up in 3 out of the 7 included CVOTs.

As to premature discontinuation, for those trials claiming statistical superiority and eligible for FI, a median value of 303 patients (IQR 249-1140) for all experimental groups and a median of 335 patients (IQR 258-1297) for control groups was obtained. From what concerns to the proportion of discontinuation a median of 8.9% (IQR 5.35-14) of premature discontinuation was obtained. Being the lowest of 1.4% for the LEADER trial (41-43) and the highest of 21.9% for the REWIND study (45).

If the whole data for all 21 CVOTs were to be analysed, the median value of premature discontinuation would be of 554 patients (IQR 250-1253) for the experimental group and a median value of 632 patients (IQR 302-1212) for the control group (18,29-75). Representing a proportion of discontinuation for both control and experimental groups of 10% (IQR 5.25-13.3), being the lowest of 1.4% for the LEADER trial (42-43) and the highest of 29.9% for the CANVAS program (47-49).

Finally, the remaining 3 trials claiming statistical superiority were the EMPA-REG outcome trial (18,46), the CANVAS program (47-49) and SUSTAIN-6 (50-52), which followed a randomisation other than 1:1 ratio. Specifically a 1:1:1 ratio for EMPA-REG and CANVAS program, and 1:1:1:1 in the SUSTAIN-6 trial. Thus figuring as “Not Applicable” (NA*). However, if a “modified” FI were to be applied for each subgroup, it could not had been carried out because as shown in Figures 14 to 16, when comparing the subgroups for each study, the obtained p-value is not significant, and thus FI cannot be applied.

EMPA-REG Outcome: as represented in Figure 14, the main analysis for the primary outcome obtained a 0.86 HR and CI (0.74-0.99) with a p-value= 0.04 for superiority. Values obtained for each subgroup are represented on the right, next to each intervention. With a 0.85 HR and CI (0.72-1.01) with a p-value= 0.07 for the 10 mg empagliflozin group and 0.86 HR and CI (0.73-1.02) with a p-value= 0.09 for the 25 mg empagliflozin group (18,46).

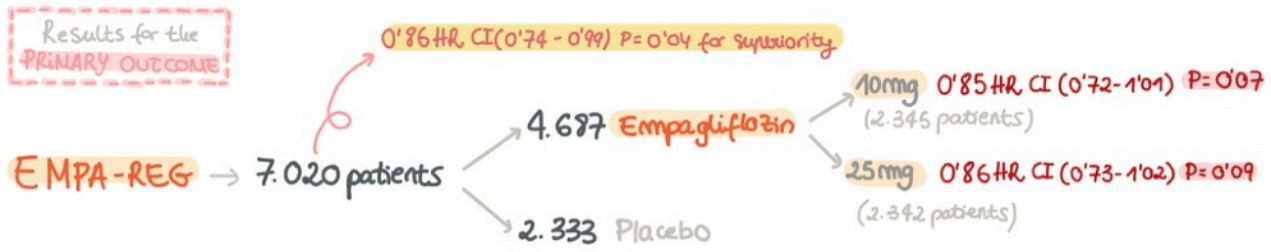


Figure 14. Results for the primary outcome of the EMPA-REG Outcome and its subgroups. HR: Hazard Ratio, CI: confidence interval, P: p-value.

CANVAS Program: as represented in Figure 15, the main analysis for the primary outcome obtained a 0.86 HR and CI (0.75-0.97) with a p-value= 0.02 for superiority. Values obtained for each subgroup are represented on the right, next to each intervention. The CANVAS trial obtained a value of 0.88 HR and CI (0.72-1.03), and the CANVAS-R a 0.82 HR and CI (0.66-1.01). P value was presented as a unique value for both trials, with a p-value= 0.59 for homogeneity (47-49).

Figure 15. Results for the primary outcome of the CANVAS program and its subgroups. HR: Hazard Ratio, CI: confidence interval, P: p-value.

SUSTAIN-6 trial: as represented in Figure 16, the main analysis for the primary outcome obtained a 0.74 HR and CI (0.58-0.95) with a p-value= 0.02 for superiority. Values obtained for each subgroup are represented on the right, next to each intervention. With a 0.77 HR and CI (0.55-1.08) with a p-value= 0.13 for the 0,5 mg semaglutide group and 0.71 HR and CI (0.49-1.02) with a p-value= 0.06 for the 1 mg semaglutide group (50-52).

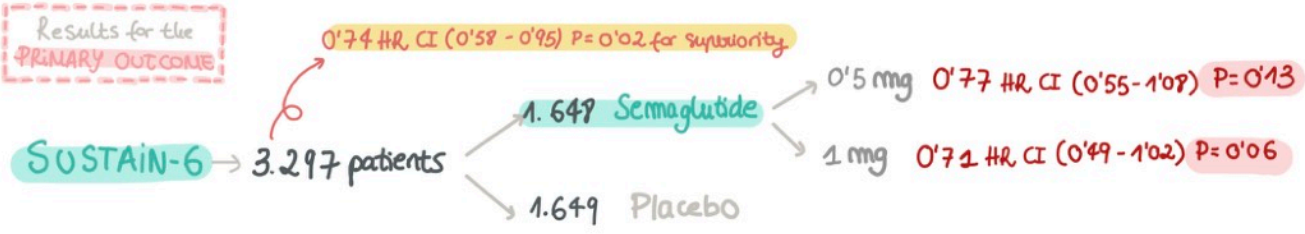


Figure 16. Results for the primary outcome of the SUSTAIN-6 trial and its subgroups. HR: Hazard Ratio, CI: confidence interval, P: p-value.

Table 4. Cardiovascular Outcome Trials for DPP4-i after the issuance of the FDA 2008 Guidelines. EXAMINE (71,72): Examination of Cardiovasclar Outcomes with Alogliptin versus Standard of Care, SAVOR-TIMI 58 (73,74): Saxagliptin Assessment of Vascular Outcomes Recorded in Patients with Diabetes Mellitus - Thrombolysis in Myocardial Infarction, TECOS (75): Trial Evaluating Cardiovascular Outcomes with Sitagliptin, CARMELINA (66-68): Cardiovascular and Renal Microvascular Outcome Study with Linagliptin, CAROLINA (69-70): Cardiovascular Outcome Study of Linagliptin vs Glimperide in T2DM. A1c: glycated haemoglobin, ASC: acute coronary syndrome, ASCVD: atherosclerotic cardiovascular disease, CHF: congestive heart failure, CI: confidence interval, CV: cardiovascular, CVD: cardiovascular disease, DPP4-i: dipeptidyl peptidase 4, eGFR: estimated glomerular filtration rate, GLP-1RA: glucagon-like peptide 1, HF: heart failure, HR: hazard ratio, LDLc: low-density lipoprotein cholesterol, MACE: major adverse cardiac event, MI: myocardial infarction, RF: risk factors, SC: standard diabetes care, T2DM: type 2 Diabetes mellitus, UACR: urinary albumin-creatinine ratio, y.o: year-old.NA* Not applicable because superiority was not claimed.

	DPP-4i	EXAMINE (NCT00968708)	SAVOR - TIMI 53 (NCT01107886)	TECOS (NCT00790205)	CARMELINA (NCT01897532)	CAROLINA (NCT01243424)
	Year started / reported	2009/2013	2010/2013	2008/2015	2013/2018	2010/2020
	Median follow-up (years)	1,5	2,1	3	2,2	6,3
	Total patients (n)	5380	16492	14671	6979	6042
	Intervention added to Standard Care	25mg or 12,5mg or 6,25mg **alogliptin/day** vs 25mg or 12,5mg or 6,25mg **placebo/day**	5mg **saxagliptin/day** vs 5mg **placebo/day**	100mg **sitagliptin/day** vs 100mg **placebo/day**	5mg **linagliptin/day** vs 5mg **placebo/day**	5mg **linagliptin/day** vs 1-4mg **Glimperide/day**
	Inclusion criteria	**T2DM + ASCVD** (*acute myocardial infarction or unstable angina requiring hospitalization*) 15-90 days before randomization	**>40 y.o. T2DM + CVD** (*coronary, cerebrovascular or peripheral vascular system*) OR **>55 y.o. (men) or >60 y.o. (women) + >1 Risk factors** (*dyslipidemia, hyptertension or active smoking*)	**>50 y.o. T2DM + preexisting CVD** (*major coronary artery disease, ischemic cerebrovascular disease, atherosclerotic peripheral arterial disease*)	**T2DM + high CVD risk** (*history of coronary artery disease, stroke or peripheral vascular disease, micro/ macroalbuminuria*) OR **high Renal risk** (*eGFR 45-75mL/min/1,73m2 and UACR >200mg/g OR eGFR 15-45*)	**T2DM + high CV risk: ≥70y.o.** OR **ASCVD** (*ischemic heart disease, cerebrovascular disease or peripheral artery disease*) OR **>2 RF** (*+10years with T2DM, systolic blood pressure >140mmHg or receiving at least 1 blood-pressure-lowering treatment, current smoker, LDLc ≥135mg/dL or receiving lipid-lowering treatment*) OR **microvascular complications** (*eGFR 30-59mL/min/1,73m2, UACR ≥30microg/mg or proliferative retinopathy*)
	A1C inclusion (%)	6,5 -11	6,5 - 12	6,5 -8	6,5 - 10	6,5 - 8,5
	Mean baseline A1C (%)	8	8	7,2	7,9	7,2
	Diabetes duration (years)	7,1	10,3	11,6	14,7	6,3
Baseline Antidiabetics	Metformin (%)	66	70	82	54,8	83
Baseline Antidiabetics	Insulin (%)	30	42	23	58	exclusion criteria
Baseline Antidiabetics	Thiazolidinediones (%)	2,5	6	2,7	unknown	exclusion criteria
Baseline Antidiabetics	Sulfonylureas (%)	46	40	45	32	28
	Prior CVD/CHF (%)	100/28	78/13	74/18	57/26,8	34/4
	Primary outcome	**3-point MACE:** CV death, nonfatal myocardial infarction or nonfatal stroke	**3-point MACE**: CV death, myocardial infarction or ischemic stroke	**4-point MACE:** CV death, nonfatal myocardial infarction,nonfatal stroke,hospitalization unstable angina	**3-point MACE:** CV death, nonfatal myocardial infarction or nonfatal stroke	
	Nº events Primary Outcome	**305 alogliptin vs 316 control**	**613 saxagliptin vs 609 control**	**839 sitagliptin vs 851 control**	**434 linagliptin vs 420 control**	**356 linagliptin vs 362 control**
	HR and CI for the Primary outcome	**0.96 (≤1.16) p<0.001 noninferiority**	**1 (0.89-1.12) p<0.001 noninferiority**	**0.98 (0.89-1.08) p<0.001 noninferiority**	**1.02 (0.89-1.17) p<0.01 noninferiority**	**0.98 (0.84-1.14) p<0.001 noninferiority**
	Fragility Index	**NA***	**NA***	**NA***	**NA***	**NA***
	Rate of premature discontinuation % (patients)	**10,4% (564) alogliptin 11,2% (606) placebo**	**9,2% (1527) saxagliptin 10,3% (1705) placebo**	**1,7% (251) sitagliptin 2,05% (302) placebo**	**11,9% (834) linagliptin 13,6% (955) placebo**	**2% (124) linagliptin 1,9% (115) glimperide**
	Lost to Follow-up	**25 patients**	**28 patients**	**367 patients**	**21 patients**	**42 patients**

Table 5. Cardiovascular Outcome Trials of SGLT-2i after the issuance of the FDA 2008 Guidelines (Part 1). EMPA-REG OUTCOME (18,46): Empagliflozin Cardiovascular Outcome Event Trial in Type 2 diabetes Mellitus Patients, CANVAS (47-49): Canagliflozin Cardiovascular Assessment Study, CANVAS-R: Canagliflozin Cardiovascular Assessment Study Renal, DECLARE-TIMI 58 (60-62): Dapagliflozin Effect on Cardiovascular Events-Thrombolysis in Myocardial Infarction 58, VERTIS CV (57-59): Evaluation of Ertugliflozin Efficacy and Safety Cardiovascular Outcomes Trials. A1c: glycated haemoglobin, ASC: acute coronary syndrome, ASCVD: atherosclerotic cardiovascular disease, CHF: congestive heart failure, CI: confidence interval, CV: cardiovascular, CVD: cardiovascular disease, DPP4-i: dipeptidyl peptidase 4, eGFR: estimated glomerular filtration rate, GLP-1RA: glucagon-like peptide 1, HDLc: high-density lipoprotein cholesterol, HF: heart failure, HR: hazard ratio, MACE: major adverse cardiac event, RF: risk factors, SC: standard diabetes care, SGLT-2i: inhibitors of sodium-glucose cotransporter 2, T2DM: type 2 Diabetes mellitus, y.o: year-old. NA* Not applicable because superiority was not claimed. NA** Not applicable because of randomisation different to 1:1.

			CANVAS PROGRAM			
	SGLT-2i	**EMPA-REG (NCT01131676)**	**CANVAS (NCT01032629)**	**CANVAS-R (NCT01989754)**	**DECLARE -TIMI 58 (NCT01730534)**	**VERTIS CV (NCT01986881)**
	Year started / reported	2010/2015	2009/2017	2013/2018	2013/2018	2013/2020
	Median follow-up (years)	3,1	5,7	2,1	4,2	3,5
	Total patients (n)	7020	4330	5812	17160	8238
	Intervention added to Standard Care	10mg or 25mg **empagliflozin/day** vs 10mg or 25mg **placebo/day**	100mg or 300mg **canagliflozin/day** vs 100mg or 300mg **placebo/day**	100mg - 300mg **canagliflozin/day** vs 00mg - 300mg **placebo/day**	10mg **dapagliflozin/day** vs 10mg **placebo/day**	5mg or 15mg **ertugliflozin/day** vs 5mg or 15mg **placebo/day**
	Inclusion criteria	**T2DM** (*BMI≤45kg/m2 and eGFR≥30mL/min/1,73m2*) + **CVD** (*myocardial infarction, coronary or peripheral artery disease, stroke*)	**T2DM + ≥30 y.o. with ASCVD** (*stroke, myocardial infarction, hospital admission for unstable angina, coronary artery bypass graft, percutaneous coronary intervention, peripheral revascularization, carotid or peripheral vascular disease, amputation secondary to vascular disease*) OR **T2DM + ≥50y.o. with <2 CVD RF** (*T2DM for +10 years, systolic blood pressure >140mmHg, current smoker, micro/macroalbuminuria, HDLc <39mg/dL*)		**≥40 y.o. with T2DM** (*with eGFR≥60mL/min/1,73m2*) **+ ASCVD** OR **men ≥55 y.o. or women ≥60 y.o. with T2DM + Multiple RF** (*hyptertension, dyslipidemia, use of lipid-lowering therapies, current smoker*)	**≥40 y.o. with T2DM + ASCVD**
Baseline Antidiabetics	**A1C inclusion (%)**	7 - 10	7 - 10,5		6,5 - 12	7 -10,5
	Mean baseline A1C (%)	8,1	8,2		8,3	8,2
	Diabetes duration (years)	10	13,5		11	13
	Metformin (%)	74	77		82	76
	Insulin (%)	48	50		41	47
	Thiazolidinediones (%)	4	unknown		exclusion criteria	exclusion criteria
	Sulfonylureas (%)	42	43		42	41
	DPP4-i (%)	11	12,4		17	11
	GLP1-RA (%)	3	4		4	3
	Prior CVD/CHF (%)	99/100	65,6/14,4		40/10	100/23
	Primary outcome	**3-point MACE:** CV death, nonfatal myocardial infarction or nonfatal stroke	**3-point MACE**: CV death, myocardial infarction or nonfatal stroke	**Progression to albuminuria**	**3-point MACE**: CV death, myocardial infarction or ischemic stroke	**3-point MACE:** CV death, nonfatal myocardial infarction or nonfatal stroke
	N° events Primary Outcome	**490 empagliflozin vs 282 control**	**1558 canagliflozin vs 1369 control**		**756 dapagliflozin vs 803 Control**	**653 ertugliflozin vs 327 control**
	HR and CI for the Primary outcome	**0.86 (0.74 - 0.99) p=0.04 for superiority**	**0.86 (0.75 - 0.97) p=0.02 superiority**		**0.93 (0.84-1.03) p<0.17**	**0.97 (0.85-1.11) p<0.001 noninferiority**
	Fragility Index	**NA ****	**NA ****		**NA***	**NA ****
	Fragility Quotient					
	Rate of premature discontinuation % (patients)	**15,6% (1097) empagliflozin 9,7% (683) placebo**	**29,2% (2961) canagliflozin 29,9% (3032) placebo**		**10,5% (1811) dapagliflozin 12,5% (2151) placebo**	**15,6% (1291) ertugliflozin 9,3% (767) placebo**
	Lost to Follow-up	**56 patients**	**40 patients**		**30 patients**	**57 patients**

Table 5 (continued). Cardiovascular Outcome Trials of SGLT-2i after the issuance of the FDA 2008 Guidelines. CREDENCE (32-34): Canagliflozin and Renal Events in Diabetes with Established Nephropathy Clinical Evaluation, DAPA-HF (29-31): Dapagliflozin and Prevention of Adverse Outcomes in Heart Failure, DAPA-CKD (38-40): Dapagliflozin and Prevention of Adverse Outcomes in Chronic Kidney Disease, EMPEROR-Reduced (35-37): Empagliflozin Outcome Trial in Patients with Chronic Heart Failure and a Reduced Ejection Fraction. A1c: glycated haemoglobin, CHF: congestive heart failure, CI: confidence interval, CV: cardiovascular, CVD: cardiovascular disease, DPP4-i: dipeptidyl peptidase 4, eGFR: estimated glomerular filtration rate, GLP-1RA: glucagon-like peptide 1, HR: hazard ratio, IV: intravenous, MACE: major adverse cardiac event, NYHA: New York Heart Association, RF: risk factors, SC: standard diabetes care, SGLT-2i: inhibitors of sodium-glucose cotransporter 2, T2DM: type 2 Diabetes mellitus, UACR: urinary albumin-creatinine ratio, y.o: year-old. Unknown*: not available data, requested to trials.

	SGLT-2i	CREDENCE (NCT02065791)	DAPA HF (NCT03036124)	DAPA CKD (NCT03036150)	EMPEROR - Reduced (NCT03057977)
	Year started / reported	2014/2019	2017/2020	2017/2020	2017/2020
	Median follow-up (years)	2,62	1,5	2,4	
	Total patients (n)	4401	4744	4304	3730
	Intervention added to Standard Care	100mg **canagliflozin/day** vs 100mg **placebo/day**	10mg **dapagliflozin/day** vs 10mg **placebo/day**	10mg **dapagliflozin/day** vs 10mg **placebo/day**	10mg **empagliflozin/day** vs 10mg **placebo/day**
	Inclusion criteria	**≥30y.o. with T2DM + albuminuric chronic kidney disease** (*eGFR 30-90mL/min/1,73m2 and UACR 200-5000*)	Established **heart failure + reduced ejection fraction (HRrEF) ≤40%** (NYHA II-IV) **with or without T2DM**	**Chronic kidney disease** (*eGFR 25-75mL/min/1,73m2 and UACR 200-5000*) **with or without T2DM**	Established **heart failure + reduced ejection fraction (HRrEF) ≤40%** (NYHA II-IV) **with or without T2DM**
Baseline Antidiabetics	A1C inclusion (%)	6,5 - 12	–	UNKNOWN*	
	Mean baseline A1C (%)	8,3	unknown		
	Diabetes duration (years)	15,8	unkwnon		
	Metformin (%)	57,8	51		
	Insulin (%)	65,5	27		
	Thiazolidinediones (%)	3,1	–		
	Sulfonylureas (%)	28,8	22		
	DPP4-i (%)	17,1	16		
	GLP-1RA (%)	4,2	1		
	Prior CVD/CHF (%)	50/14	100/100	37/10	100/100
	Primary outcome	**End-stage kidney disease** (dialysis, transplantation or eGFR<15mL/min/1,73m2), **doubling serum creatinine level or death from renal or CV causes.**	**Worsening of HF** (hospitalization or an urgent visit resulting in IV therapy) or **CV death**	Sustained decline in **eGFR ≥50%, end-stage kidney disease, death from renal or CV causes.**	**CV death or hospitalization** for sorsening heart failure
	N° events Primary Outcome	**245 canagliflozin vs 340 control**	**386 dapagliflozin vs 502 control**	**197 dapagliflozin vs 312 control**	**361 empagliflozin vs 462 control**
	HR and CI for the Primary outcome	**0.70 (0.59 - 0.82) p=0.00001**	**0.74 (0.65 - 0.85) p<0.001**	**0.61 (0.51-0.72) p<0.001**	**0.75 (0.65 - 0.86) p<0.001**
	Fragility Index	**50**	**62**	**71**	**50**
	Fragility Quotient (%)	**1,1 %**	**1,3 %**	**1,6 %**	**1,13 %**
	Rate of premature discontinuation % (patients)	**12,3% (543) canagliflozin 14,9%% (658) placebo**	**5,2% (249) dapagliflozin 5,4% (258) placebo**	**6,3% (274) dapagliflozin 7,1% (309) placebo**	**8,12% (303) empagliflozin 8,9% (335) placebo**
	Lost to Follow-up	**4 patients**	**2 patients**	**13 patients**	**22 patients**

Table 6. Cardiovascular Outcome Trials of GLP-1RA after the issuance of the FDA 2008 Guidelines. ELIXA (56): Evaluation of Lixisenatide in Acute coronary Syndrome, EXSCEL (53-55): Exenatide Study of Cardiovascular Event Lowering, LEADER (41-43): Liraglutide Effect and Action in Diabetes, SUSTAIN-6 (50-52): Trial to Evaluate Cardiovascular and Other Long-term Outcomes with Semaglutide in Subjects with Type 2 Diabetes, PIONEER-6 (63-65): Peptide Innovation for Early Diabetes Treatment, HARMONY (44): Albiglutide and Cardiovascular Outcomes, REWIND (45): Researching Cardiovascular Events with a Weekly Increatin in Diabetes. A1c: glycated haemoglobin, ASCVD: atherosclerotic cardiovascular disease, CHF: congestive heart failure, CI: confidence interval, CV: cardiovascular, CVD: cardiovascular disease, DPP4-i: dipeptidyl peptidase 4, eGFR: estimated glomerular filtration rate, GLP-1RA: glucagon-like peptide 1, HR: hazard ratio, IV: intravenous, MACE: major adverse cardiac event, NYHA: New York Heart Association, RF: risk factors, SC: standard diabetes care, SGLT-2i: inhibitors of sodium-glucose cotransporter 2, T2DM: type 2 Diabetes mellitus, y.o: year-old. NA* Not applicable because superiority was not claimed. NA** Not applicable because of randomisation different to 1:1.

GLP-1RA	ELIXA (NCT01147250)	EXSCEL (NCT01144338)	LEADER (NCT01179048)	SUSTAIN-6 (NCT01720446)	PIONEER6 (NCT02692716)	HARMONY (NCT02465515)	REWIND (NCT01394952)
Year started / reported	2010/2015	2010/2017	2010/2016	2012/2016	2016/2019	2015/2018	2011/2019
Median follow-up (years)	2,1	3,2	3,8	2,1	1,3	1,6	5,4
Total patients (n)	6068	14752	9340	3297	3183	9463	9901
Intervention added to Standard Care	10-20µg **lixisenatide/day** vs 10-20µg **placebo/day**	2mg **exenatide/week** vs 2mg **placebo/week**	1,8mg **liraglutide/day** vs 1,8mg **placebo/day**	0.5mg or 1mg **semaglutide/ week** vs 0.5mg or 1mg **placebo/w.**	14mg **semaglutide/day** vs 14mg **placebo/day**	30-50mg **albiglutide/week** vs 30-50mg **placebo/week**	1,5mg **dulaglutide/week** vs 1,5mg **placebo/week**
Inclusion criteria	**≥30 y.o. T2DM + ASCVD** (*acute myocardial infarction or unstable angina requiring hospitalization*) 180days before randomization	**T2DM + ASCVD 70% or not ASCVD 30%**	**≥ 50y.o T2DM + ASCVD** (*coronary heart, cerebrovascular or peripheral vascular disease, CKD stage ≥3, NYHA II-III*) OR **≥ 60y.o T2DM + >1 CVD RF**			**≥40 y.o. T2DM + ASCVD**	**≥50 y.o. T2DM + CVD** OR **≥60 y.o. T2DM + ≥RF** (*current smoker, dyslipidaemia, hypertension or abdominal obesity*)
A1C inclusion (%)	5,5 - 11	6,5 - 10	≥ 7	≥ 7	≥ 6,5	≥ 7	≤ 9,5
Mean baseline A1C (%)	7,7	8	8,7	8,7	8,2	8,7	7,4
Diabetes duration (years)	9,3	12	12,8	13,9	14,9	13,8	10,5
Metformin (%)	66	76	76	73	77	73,6	81
Insulin (%)	39	46	44	58	60	59	24
Thiazolidinediones (%)	1,5	4	6	2,3	3,7	2	1,8
Sulfonylureas (%)	33	36	50	42	32	28	46
DPP4-i (%)	–	15	exclusion criteria	exclusion criteria	exclusion criteria	15,5	5,5
SGLT-2i (%)	–	1	unknown	0,2	9,6	6,5	unknown
Prior CVD/CHF (%)	100/22	73,1/16,2	81/18	60/24	84/23	100/20	32/9
Primary outcome	**4-point MACE:** CV death, myocardial infarction, stroke or hospitalization for unstable angina	**3-point MACE**: CV death, nonfatal myocardial infarction or nonfatal stroke.					
N° events Primary Outcome	**406 lixisenatide vs 309 control**	**839 exenatide vs 905 control**	**608 liraglutide vs 694 control**	**108 semaglutide vs 146 control**	**61 semaglutide vs 76 control**	**338 albiglutide vs 428 control**	**549 dulaglutide vs 663 control**
HR and CI for the Primary outcome	**1.02 (0.89 - 1.17) p<0.001 noninferiority**	**0.91 (0.83 - 1) p<0.001 noninferiority**	**0.87 (0.78-0.97) p=0.01 superiority**	**0.74 (0.58-0.95) p<0.001 noninferiority**	**0.79 (0.57 - 1.11) p<0.001 noninferiority**	**0.78 (0.68 - 0.90) p=0.0006 superiority**	**0.88 (0.79 - 0.99) p<0.026 superiority**
Fragility Index	**NA***	**NA***	**19**	**NA ******	**NA***	**36**	**4**
Fragility Quotient %			**0,2 %**			**0,3 %**	**0,04 %**
Rate of premature discontinuation % (patients)	**13,72%(833) lixisenatide 11,9% (727) placebo**	**1,7% (262) exenatide 2% (303) placebo**	**1,4%(139) liraglutide 1,7% (159) placebo**	**10,6%(350) semaglutide 9,4% (310) placebo**	**7,6% (244) semaglutide 4,9% (156) placebo**	**12% (1140) albiglutide 13,7% (1297) placebo**	**21,1% (2092) dulaglutide 21,9% (2171) placebo**
Lost to Follow-up	**146 patients**	**177 patients**	**28 patients**	**13 patients**	**0 patients**	**57 patients**	**287 patients**

8.3 RISK OF BIAS ASSESSMENT RESULTS

Regarding the Risk of Bias assessment, 21 CVOTs were analysed as shown in the algorithm of Figure 13. However, as the CANVAS and CANVAS-R were grouped as the CANVAS Program (47-49), results feature a total of 20 studies. For each study the risk of bias in each domain was obtained, which is represented in Figure 17, making evident that most of the studies present low risk of bias in combination with some concerns in some domains that are going to be discussed in the pages that follow.

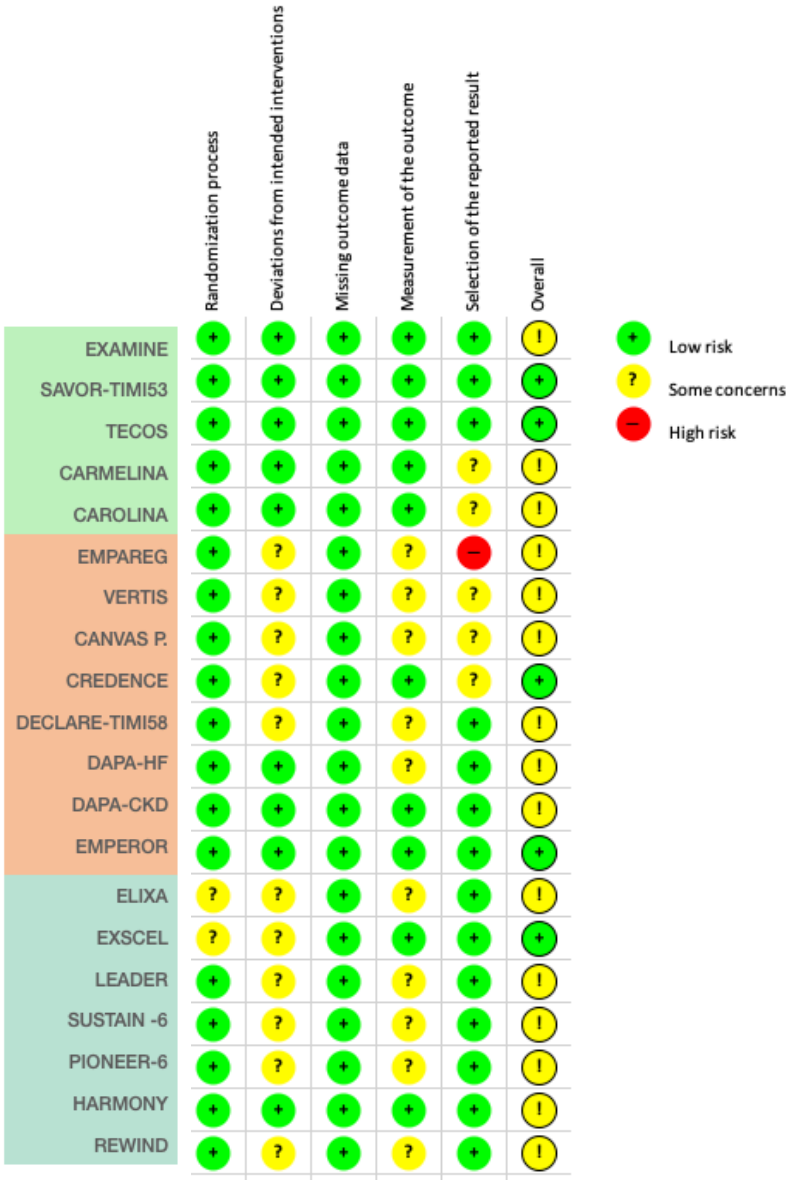


Figure 17. Risk of Bias of analysed CVO trials. RoB for the 5 Domains included in the RoB 2.0 Tool: randomisation process, deviations from the intended interventions, missing outcome data, selection of the reported result and overall bias. Low risk of bias for the **first domain** was found for all trials, with the exception of the ELIXA and EXSCEL trials. **Domain 2** presented some concerns for the EMPA-REG, VERTIS, CANVAS, CREDENCE, DECLARE-TIMI58, ELIXA, EXSCEL, LEADER, SUSTAIN-6, PIONEER-6 and the REWIND studies. Low risk of bias was assessed for all studies concerning **Domain 3**. For **Domain 4** there were some concerns for EMPA-REG, VERTIS, CANVAS, DECLARE-TIMI58, DAPA-HF, ELIXA, LEADER, SUSTAIN-6, PIONEER-6 and REWIND. The **5th Domain** presented some concerns for the CARMELINA, CAROLINA, VERTIS, CANVAS and CREDENCE studies, while for EMPAREG there was a high risk of bias. Finally, overall bias presented a mix of various results, being some concerns the most common result obtained in this assessment.

Table 7. Summary of Risk of Bias assessment with RoB 2.0. Represents percentage of articles in each domain whether low, some concerns or high risk was assessed.

	Randomization process	Deviations from intended interventions	Mising outcome data	Measurement of the outcome	Selection of the reported result	Overall Bias
Assignment to intervention (the 'intention-to-treat' effect)						
Total number of study = 20						
Low risk	90	45	100	50	70	25
Some concerns	10	55	0	50	25	75
High risk	0	0	0	0	5	0

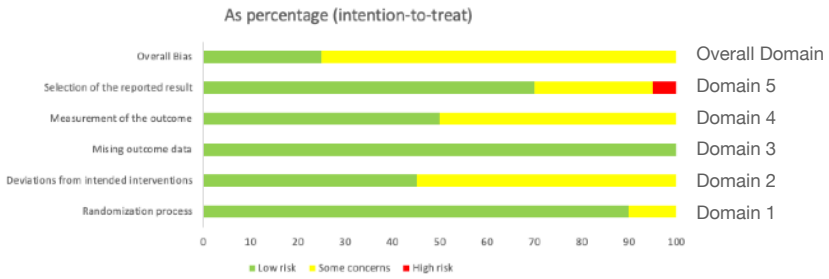


Figure 18. Summary of Risk of Bias assessment in percentage

RESULTS:

- ***Domain 1***. *Bias arising from the randomisation process:* 90% of the studies, as shown in Table 7 and its corresponding Figure 18, had a low risk of bias, performing in most of the cases a randomised allocation via an interactive computerised telephone or web response system. In addition, they were well balanced with respect to baseline characteristics. However, there is a 10% figuring as some concerns, which corresponds to the EXSEL trial (53-55), as seen in Figure 17, in which clinical characteristics of patients did not differ significantly between groups, with the exception of lipid-lowering medications and SGLT-2i inhibitors. And the remaining 5% stands for the ELIXA trial (56), which presented nominally significant between-group differences in 4 of the 35 baseline comparisons regarding to: age, eGFR, glycated haemoglobin and prior stroke.

- ***Domain 2.*** *Bias due to deviations from the intended interventions:* as shown in Table 7 and Figure 18, 45% of the studies presented a low risk of bias and 55% some concerns, which refers to the following trials, as represented in Figure 17: EMPA-REG (18,46), VERTIS (57-59), CANVAS (47-49), CREDENCE (32-34), DECLARE-TIMI58 (60-62), ELIXA (56), EXSCEL (53-55), LEADER (41-43), SUSTAIN-6 (50-52), PIONEER-6 (63-65) and REWIND studies (45). The risk of bias in this domain was attributed mainly due to known specific analytic parameter modifications from the given intervention, as a lower HbA1c value and a greater weight loss as compared to the placebo group. Those modifications in addition to specific adverse events in some patients, may rose some assumptions to whether the patient was on the intervention or on the placebo arm.

In SGLT-2i trials, which consisted of the EMPA-REG, VERTIS, CANVAS, CREDENCE and DECLARE-TIMI58 trials, patients receiving the intervention reported an increase of genital infections (due to a higher glucose excretion via urination) plus a reduction in HbA1c and weight that was higher

when compared to placebo (18,32-34,46-49,57-62). In addition, there are some specific adverse events for each drug, like a higher risk of amputation of toes, feet or legs when comparing canagliflozin (CANVAS (47-49)) and ertugliflozin with placebo (VERTIS (57-59)). Or the higher hematocrits and lower median NT-proBNP obtained in the EMPEROR study for empagliflozin (35-37).

On the other hand, the use of GLP-1RA in the ELIXA, EXSCEL, LEADER, SUSTAIN-6, PIONEER-6 and the REWIND studies, also induced similar adverse events with a greater reduction in glycated haemoglobin and weight loss (41-43,45,50-56,63-65).

- ***Domain 3.*** *Bias due to missing outcome data:* as shown in Table 7 and Figure 18, 100% of the trials had a low risk of bias in this domain because they had >95% of the final vital status data at the end of the study.

- ***Domain 4.*** *Bias in measurement of the outcome:* as shown in Table 7 and Figure 18, 50% presented a low risk of bias and 50% some concerns.

As shown in Figure 17, there were some concerns for the EMPA-REG, VERTIS, CANVAS, DECLARE-TIMI58, DAPA-HF, ELIXA, LEADER, SUSTAIN-6, PIONEER-6 and REWIND trials. The risk of bias in this domain was heightened mainly by the “Diagnostic Detection Bias” which for the SGLT-2i trials was attributable to genitourinary infections, leading to more visits to their doctor and additional tests (18,29-31,41-43,45-52,56-65). There are some specific additional adverse effects, as the ones shown on the CANVAS trial (47-49), with canagliflozin presenting a higher risk of amputation (also seen in the VERTIS trial with ertugliflozin), bone fractures and volume depletion (57-59). Or dapagliflozin, in the DECLARE-TIMI58 trial, which presented a higher rate of diabetic ketoacidosis when compared to placebo (60-62).

Furthermore, in the DAPA-HF the application of a self reported Kansas City Cardiomyopathy Questionnaire may had lead to the knowledge of the assigned intervention because the dapagliflozin group had an improvement in that aspect (29-31).

On the other hand, the use of GLP-1RA in the ELIXA, LEADER, SUSTAIN-6, PIONEER-6 and REWIND trials presented a major proportion of patients experiencing gastrointestinal adverse events, and thus, creating a “Diagnostic Detection Bias” (41-43,45,50-52,56,63-65). Some specific adverse effects were the ones shown on the LEADER trial, where liraglutide presented a higher proportion of acute gallstone disease compared to placebo (41-43). Or the higher rate of retinopathy complications found on SUSTAIN-6 with semaglutide (50-52).

- ***Domain 5.*** *Bias in selection of the reported result:* as shown in Table 7 and Figure 18, 70% of the studies reported low risk, 25% some concerns and 5% high risk of bias. As seen in Figure 16, the 25% with some concerns belongs to CARMELINA, CAROLINA, VERTIS, CANVAS and CREDENCE studies, while for EMPAREG is the remaining 5% standing for a high risk of bias.

The CARMELINA, CAROLINA, VERTIS and CANVAS trials, brought some concerns in this domain because there were some amendments in the protocol which raised some questions. For the CARMELINA, CAROLINA and VERTIS studies, there was a change in the protocol while the trial was still being conducted from 4-point MACE to 3-point MACE, because of the beneficial results obtained in the EMPA-REG study (57-59,66-70). While for the CREDENCE trial, statistical analysis was done by the sponsor or under the authority of the sponsor, which may raise some concerns (32-34).

In respect to the EMPA-REG OUTCOME, two Lilly employees committed a breach of confidentiality before the 4th version of the protocol was published, whereby several substantial modifications were made. For instance, in the third amendment to the protocol acute myocardial infarction was excluded from the primary composite outcome, which lead to statistical superiority. Subsequently, in the fourth amendment the statistical analysis plan raised the minimum number of events. Finally, changes to the SAP were last modified in May 2015, when the study had already been completed and was about to be published. These events, generate a higher risk of bias for this trial, as blinding and bias in measurement of the outcome cannot be guaranteed (18,22,46).

- ***Overall bias.*** As shown in Table 7 and Figure 18, 25% of the studies were at low risk of bias and 75% had some concerns. This last point of bias is the sum of bias for the previous domains plus the reviewer's point of view.

9. DISCUSSION

The positive results obtained from CVOTs conducted after the 2008 FDA guidance for new antidiabetic drugs for T2DM, have lead to the inclusion of SGLT-2i and GLP-1RA in current clinical practice guidelines. The recommendation consists of adding these agents to basal antihyperglycemic treatment in those patients with risk factors or that have indicators of high-risk or established atherosclerotic cardiovascular disease, chronic kidney disease or heart failure in order to prevent or delay cardiovascular events (15). However, robustness and methodological quality has not yet been assessed. Results obtained in our Study indicate that CVOTs that ground clinical guideline modifications do not provide enough statistical robustness and do have some concerns regarding their methodological quality to justify changing clinical practice recommendations.

Our study found that Fragility Index (FI) for the eligible CVOTs is fragile, with a median value of 50. Thus, meaning that in order to reverse a significant result, 50 events would need to be adjudicated from events to non-events in order to lose statistical significance. However, when Fragility Quotient (FQ) was applied as to omit the effect that sample size can have on FI, a median FQ value of 1,1% (IQR 0.2-1.3) was obtained. This is meaningful, as it leads to just one event per 100 patients included in the study to be adjudicated from event to non-event in order to nullify the significance of the study. What is more, 42% CVOTs presented a FQ value near 0, meaning that less than one event per 100 patients included in the study is required to go from event to non-event to turn a significant result to a nonsignificant result, thus showing a clear fragility for the stated results.

Additionally, our Study found that this value was surpassed by either patients lost to follow-up and patients with premature discontinuation. Likewise, it is noteworthy that in these CVOTs an intention-to-treat analysis was performed, meaning that all randomised patients were included in the analysis. This implies that for trials supporting these drugs, a total of 20.285 patients presented premature discontinuation for both control and intervention groups, with the median value of 1.844 patients per trial. With a total number participants of 77.222 patients, it represents a rate of premature discontinuation of 26%, which should be taken into account. Not only because it surpasses the FQ value, suggesting that the change to event to non-event may had taken place, providing enough data to sway the reported statistical significance of a trial (25), but also because it demonstrates an important bias in how was administered the experimental drug versus placebo.

Our findings are in line with those of Chase and Matt's, who also applied the FI and FQ to 35 randomised controlled trials (RCTs), which included 4 of the 21 CVOTs that have been analysed in our Project (LEADER, TECOS, EXAMINE and SAVOR-TIMI53) (25). While their median FI was of 16 (IQR 8-29) compared to our median FI of 50, when FQ was applied a FQ of 0.7% (IQR 0.3-14) was obtained in their study compared to our median FQ of 1,1% (IQR 0.2-1.3) (25). In such a way, meaning that studies included in both projects are fragile.

On the other hand, the RoB assessment was performed in order to assess methodological quality. Results for the majority of the CVOTs obtained some concerns chiefly due to the impact on the risk of bias in mainly domains 2, 4 and 5 that can be summarised in the brief ideas that follow.

Our study found **some concerns on the quality of the care received by the patients** included in these trials, which claimed that local guidelines were applied to all patients. This is because in most of these trials glycemic targets were lower in the intervention group when compared to the control group, which was the one requiring additional antihyperglycemic therapy. This statement should be taken carefully, as if antihyperglycemic treatment were to be applied correctly, both groups should acquire similar HbA1c levels. In this situation, some conjectures can be done, as for example that the control group was under-treated and thus it is logic that due to a poorer glycemic control more complications were obtained in the control group compared to the experimental group.

Aforementioned, some trials had a period at the beginning of the trial in which antidiabetic treatment could not be modified in both groups. Thus worsening the situation that has been just mentioned. For instance, the EMPA-REG OUTCOME stated 12 weeks (18,46) and the VERTIS CV trial stated 18 weeks (57-59). And what is more, some trials included treatment with other SGLT-2i or GLP-1RA, thus masking obtained results.

These issues create some hesitation on whether standard care treatment was conducted correctly or if there was an interest in order to favour the experimental group's results in pursuance of achieving non-inferiority and superiority.

Additionally, our study found **some concerns on methodological procedures on data pooling and analyses**. On one hand, when a sub-analysis was conducted as shown in Figures 14-16 for those trials claiming for statistical superiority as EMPA-REG OUTCOME, CANVAS and SUSTAIN-6; results for the primary outcome did not result in statistical significance (18,46-52). And on the other hand, there were some trials that included cohorts that had a significant difference in time to follow-up and different designs that were pooled together leading to a favorable result. For instance, in the CANVAS program the length to follow-up was of 295.9 in the CANVAS and 109 weeks in the CANVAS-R (47-49), with designs and patients that had different profiles. Another example would be the VERTIS CV, being the duration of follow-up of 4.3 years for cohort 1 and 2,7 years for cohort 2 (57-59). Thus questioning the validity of the results for the pooled intervention groups vs placebo, which may lead to think that pooled data was used in order to increase the number of participants and events and thus, obtaining statistical significance that otherwise would not had been obtained.

More on that is, how the CARMELINA ,CAROLINA and DECLARE-TIMI studies underwent a protocol change going from 4-point MACE to 3-point MACE for the primary outcome (60-62,66-70). In the same line, the EMPA-REG OUTCOME performed several protocol amendments after a breach of

confidentiality. In these cases, it is thought that amendments were done in order to favour significant statistical results (16,46).

Finally, our study found **some concerns on reporting approach**, as some trials claimed for superiority when using a non-inferiority design. This is a line of reasoning that has been used by 5 trials that were reviewed for this Project: EMPA-REG OUTCOME, CANVAS PROGRAM, LEADER, SUSTAIN-6 and HARMONY. And while declaring superiority is not mistaken, it can lead to confusion (25), as by stating that results reported are superior to placebo is alluring. However, if a non-inferiority design has been used, for its design and hypothesis only statistical superiority can be claimed (21-23). This statement is in most of the cases dismissed, hence, leading to a misconception of the results and thus promote their acceptance in the medical community for ultimately being included in clinical guidelines.

To our knowledge this is the first study that has focused on analysing the robustness of specifically the 21 CVOTs after the 2008 FDA guidance. Our Study ascertained that CVOTs grounding clinical guideline modifications do not provide enough statistical robustness and do have some concerns regarding their methodological quality to justify changing clinical practice recommendations. On one hand due to high fragility, which implies that patients are treated on recommendations grounded on data coming from studies where in the best of the cases, less than 1 patient in 100 patients in the study is needed to change from event to non-event in order to achieve non-significance. Hence, being somehow questionable in what ethics concerns, as for what it has been demonstrated in our study, more harm than benefits could be attained. On the other hand, the risk of bias in the reported studies manifest that there are important issues with regard to how are trials being conducted and how is data being analysed, which may had lead to favourable results. Additionally, these factors may have had influenced the effect of the intervention that was obtained, thus raising some concerns on whether the results reflected what the study aimed to estimate.

Further research is required in this field in order to determine an acceptable threshold for establishing “good” or “bad” FI and FQ values. This fact could help comparing different CVOTs easily if the recommendations listed in the standards of medical care came together with FI and FQ values, which would acknowledge their robustness. Furthermore, it would be interesting if FDA guidelines were more strict and less inclusive in what concerns on clinical design and data analysis, as it could be a way to address the different factors implied in the risk of bias assessment. The end point with that would be to promote high quality research and thus improve clinical performance on treating patients the best possible way.

10. STRENGTHS AND LIMITATIONS

To our knowledge, this is the first study that has focused on analysing the robustness of specifically the 21 CVOTs conducted following FDA requirements, so that it brings more insights and information about these trials than other studies. These are preliminary results of a duplicate review of both Fragility Index and Risk of Bias of the included studies. This is particularly important for the evaluation of risk of bias due to its qualitative nature. When completed, this will be one of the most important strengths of this study.

Fragility Index (FI) can aid our understanding of the robustness of clinical trials' results, however, it has some limitations, as for example the impossibility to calculate it on relevant cardiovascular outcome trials (CVOTS) that do not follow a randomised, 2-group parallel design with a dichotomous outcome, that when analysed in sub-groups, statistical superiority was not found.

Additionally, as reviewed in the discussion section, FI can be highly influenced by sample size leading to misinterpretation when comparing different trials. This can be addressed by applying the Fragility Quotient (FQ). Furthermore, patient drop out and/or loss to follow-up are important data that should be taken into account too, as they could provide enough data to sway the reported statistical significance of a trial (25,26,76,77).

In spite of its limitations, FI is still a powerful and useful metric that could shed a light on result interpretation and decision making. That is why we would encourage its application, as the use of FI in CVOTs supporting treatment guidelines is limited and further research is required in order to determine acceptable thresholds (25,76).

Finally, Risk of Bias (RoB) analyses have offered important information about how are being trials conducted, generating some doubts on whether patients and data are being manipulated in order to obtain a favourable result. This fact highlights the importance of statistical training in the medical community, and on the other hand for the need for more astringent requirements from the FDA in their Industry Guidelines.

11. CONCLUSIONS

CVOTs conducted after the 2008 FDA guidance for new antidiabetic drugs for T2DM do not provide enough statistical robustness and do have some concerns regarding their methodological quality to justify changing clinical practice recommendations. Hence, demonstrating that there is a need for critical review for CVOTs, which would probably lead to the revision of treatment recommendations.

12. COROLLARY

- The provision of high-quality, evidence-based clinical care in T2DM treatment requires a foundation of robust clinical research evidence. The routine inclusion of FI and FQ scores could provide valuable additional data for guideline recommendations, helping clinicians understand the robustness of the individual trials that underpin specific recommendations, thus avoiding relying solely on p-values and CI. By in the end, aiding practitioners deciding the best available treatment options. (25,26,76).

- Our findings suggest that there is room for improvement in the reporting of CVOTs, specially when stating for superiority. This would avoid important misinterpretations in the scientific community, which may be due to a lack of statistical training (26,76).

13. REFERENCES

1. Quintanilla Rodriguez BS, Correa R. Rosiglitazone. [Updated 2020 Jul 5]. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2020 Jan-. Available from: https://www.ncbi.nlm.nih.gov/books/NBK544230/.

2. Classification and Diagnosis of Diabetes: Standards of Medical Care in Diabetes-2020. Suppl SS.2 4.Vol. 43, Diabetes care. NLM (Medline); 2020. p. S14–31.

3. Rojo-Martínez G, Valdés S, Soriguer F, Vendrell J, Urrutia I, Pérez V, et al. Incidence of diabetes mellitus in Spain as results of the nation-wide cohort di@bet.es study. Scientific Reports. 2020 Dec 1;10(1).

4. International Diabetes Federation. IDF Diabetes Atlas, 9th edn. Brussels, Belgium: International Diabetes Federation, 2019.

5. Kyrou I, Tsigos C, Mavrogianni C, Cardon G, van Stappen V, Latomme J, et al. Sociodemographic and lifestyle-related risk factors for identifying vulnerable groups for type 2 diabetes: A narrative review with emphasis on data from Europe. BMC Endocrine Disorders [Internet]. 2020;20(Suppl 1):1–13. Available from: http://dx.doi.org/10.1186/s12902-019-0463-3

6. Zheng Y, Ley SH, Hu FB. Global aetiology and epidemiology of type 2 diabetes mellitus and its complications. Vol. 14, Nature Reviews Endocrinology. Nature Publishing Group; 2018. p. 88–98.

7. Harding JL, Pavkov ME, Magliano DJ, Shaw JE, Gregg EW. Global trends in diabetes complications: a review of current evidence. Diabetologia. 2019;62(1):3–16.

8. Fowler MJ. Microvascular and macrovascular complications of diabetes. Clinical Diabetes. 2011;29(3):116–22.

9. Cardiovascular Disease and Risk Management: Standards of Medical Care in Diabetes-2020. Suppl SS. 10.Vol. 43, Diabetes care. NLM (Medline); 2020. p. S111–34.

10. International Diabetes Federation. Diabetes and Cardiovascular Disease Report. Brussels, Belgium: International Diabetes Federation, 2019.

11. Comprehensive Medical Evaluation and Assessment of Comorbidities: Standards of Medical Care in Diabetes—2020. Diabetes Care [Internet]. 2020 Jan 20;43(Supplement 1):S37–47. Available from: http://care.diabetesjournals.org/lookup/doi/10.2337/dc20-S004

12. Facilitating behavior change and well-being to improve health outcomes: Standards of Medical Care in Diabetes-2020. Diabetes Care. Suppl SS. 5.2020;43(January):S48–65.

13. Prevention or delay of type 2 diabetes: Standards of medical care in diabetesd 2020. Diabetes Care. Suppl SS. 3. 2020;43(January):S32–6.

14. Glycemic targets: Standards of medical care in diabetes-2020. Diabetes Care. 2020 Suppl SS 7 ;43(January):S66–76.

15. Pharmacologic Approaches to Glycemic Treatment: Standards of Medical Care in Diabetes-2020. Diabetes care. Suppl SS. 9. 2020;43(January):S98–110.

16. Eckel RH, Farooki A, Henry RR, Koch GG, Leiter LA. Cardiovascular Outcome Trials in Type 2 Diabetes: What Do They Mean for Clinical Practice? Clinical Diabetes. 2019;37(4):316–37.

17. Cefalu WT, Kaul S, Gerstein HC, Holman RR, Zinman B, Skyler JS, et al. Cardiovascular outcomes trials in type 2 diabetes: Where Do We Go From Here? Ref lections From a Diabetes Care Editors' Expert Forum. Diabetes Care. 2018 Jan 1;41(1):14–31.

18. Steiner S. Empagliflozin, cardiovascular outcomes, and mortality in type 2 diabetes. Vol. 13, Zeitschrift fur Gefassmedizin. Krause und Pachernegg GmbH; 2016. p. 17–8.

19. Neal B, Perkovic V, Matthews DR, Mahaffey KW, Fulcher G, Meininger G, et al. Rationale, design and baseline characteristics of the CANagliflozin cardioVascular Assessment Study–Renal (CANVAS-R): A randomized, placebo-controlled trial.Diabetes,Obesity and Metabolism.2017 Mar1;19(3):387–93.

20. Pfeffer MA, Claggett B, Diaz R, Dickstein K, Gerstein HC, Køber L v., et al. Lixisenatide in Patients with Type 2 Diabetes and Acute Coronary Syndrome. New England Journal of Medicine. 2015 Dec 3;373(23):2247–57.

21.Schumi J, Wittes JT. Through the looking glass: Understanding non-inferiority. Trials [Internet]. 2011;12(1):106. Available from: http://www.trialsjournal.com/content/12/1/106

22. Carlos L. The EMPA-REG OUTCOME trial (empagliflozin). A critical appraisal. Drug and Therapeutics Bulletin of Navarre, Spain. 2006, Vol 24, nº3.

23. Mauri L, D'Agostino RB. Challenges in the design and interpretation of noninferiority trials. New England Journal of Medicine. 2017;377(14):1357–67.

24. Hahn S. Understanding noninferiority trials. Korean Journal of Pediatrics. 2012;55(11):403–7.

25. Chase Kruse B, Matt Vassar B. Unbreakable? An analysis of the fragility of randomized trials that support diabetes treatment guidelines. Diabetes Research and Clinical Practice [Internet]. 2017;134:91–105. Available from: https://doi.org/10.1016/j.diabres.2017.10.007

26. Tignanelli CJ, Napolitano LM. The Fragility Index in Randomized Clinical Trials as a Means of Optimizing Patient Care. JAMA Surgery. 2019;154(1):74–9.

27. RoB2: Learning live webinar series [Internet]. Available from: https://training.cochrane.org/rob-2-learning-live-webinar-series

28. Higgins J, Altman D, Sterne J. Chapter 8: Assessing risk of bias in included studies. Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 (updated March 2011). The Cochrane Collaboration, 2011. [Internet]. Vol. 8. 2011. Available from: http://www.handbook.cochrane.org

29. McMurray JJV, Solomon SD, Inzucchi SE, Køber L, Kosiborod MN, Martinez FA, et al. Dapagliflozin in Patients with Heart Failure and Reduced Ejection Fraction. New England Journal of Medicine. 2019 Nov 21;381(21):1995–2008.

30. McMurray JJV, Solomon SD, Inzucchi SE, Køber L, Kosiborod MN, Martinez FA, et al. Dapagliflozin in Patients with Heart Failure and Reduced Ejection Fraction. New England Journal of Medicine [Protocol]. 2019 Nov 21;381(21):1995–2008.

31. McMurray JJV, Solomon SD, Inzucchi SE, Køber L, Kosiborod MN, Martinez FA, et al. Dapagliflozin in Patients with Heart Failure and Reduced Ejection Fraction. [Supplementary Appendix] New England Journal of Medicine. 2019 Nov 21;381(21):1995–2008.

32. Perkovic V, Jardine MJ, Neal B, Bompoint S, Heerspink HJL, Charytan DM, et al. Canagliflozin and Renal Outcomes in Type 2 Diabetes and Nephropathy. New England Journal of Medicine. 2019 Jun 13;380(24):2295–306.

33. Perkovic V, Jardine MJ, Neal B, Bompoint S, Heerspink HJL, Charytan DM, et al. Canagliflozin and Renal Outcomes in Type 2 Diabetes and Nephropathy. [Protocol] New England Journal of Medicine. 2019 Jun 13;380(24):2295–306.

34. Perkovic V, Jardine MJ, Neal B, Bompoint S, Heerspink HJL, Charytan DM, et al. Canagliflozin and Renal Outcomes in Type 2 Diabetes and Nephropathy. [Supplementary Appendix] New England Journal of Medicine. 2019 Jun 13;380(24):2295–306.

35. Packer M, Anker SD, Butler J, Filippatos G, Pocock SJ, Carson P, et al. Cardiovascular and Renal Outcomes with Empagliflozin in Heart Failure. New England Journal of Medicine. 2020 Oct 8;383(15):1413–24.

36. Packer M, Anker SD, Butler J, Filippatos G, Pocock SJ, Carson P, et al. Cardiovascular and Renal Outcomes with Empagliflozin in Heart Failure. [Protocol] New England Journal of Medicine. 2020 Oct 8;383(15):1413–24.

37. Packer M, Anker SD, Butler J, Filippatos G, Pocock SJ, Carson P, et al. Cardiovascular and Renal Outcomes with Empagliflozin in Heart Failure. [Supplementary Appendix] New England Journal of Medicine. 2020 Oct 8;383(15):1413–24.

38. Heerspink HJL, Stefánsson B v., Correa-Rotter R, Chertow GM, Greene T, Hou F-F, et al. Dapagliflozin in Patients with Chronic Kidney Disease. New England Journal of Medicine. 2020 Oct 8;383(15):1436–46.

39. Heerspink HJL, Stefánsson B v., Correa-Rotter R, Chertow GM, Greene T, Hou F-F, et al. Dapagliflozin in Patients with Chronic Kidney Disease. [Protocol] New England Journal of Medicine. 2020 Oct 8;383(15):1436–46.

40. Heerspink HJL, Stefánsson B v., Correa-Rotter R, Chertow GM, Greene T, Hou F-F, et al. Dapagliflozin in Patients with Chronic Kidney Disease. [Supplementary Appendix] New England Journal of Medicine. 2020 Oct 8;383(15):1436–46.

41. Liraglutide and cardiovascular outcomes in type 2 diabetes. Vol. 54, Drug and Therapeutics Bulletin. BMJ Publishing Group; 2016. p. 101.

42. Liraglutide and cardiovascular outcomes in type 2 diabetes. [Protocol] Vol. 54, Drug and Therapeutics Bulletin. BMJ Publishing Group; 2016. p. 101.

43. Liraglutide and cardiovascular outcomes in type 2 diabetes. [Supplementary Appendix] Vol. 54, Drug and Therapeutics Bulletin. BMJ Publishing Group; 2016. p. 101.

44. Hernandez AF, Green JB, Janmohamed S, D'Agostino RB, Granger CB, Jones NP, et al. Albiglutide and cardiovascular outcomes in patients with type 2 diabetes and cardiovascular disease (Harmony Outcomes): a double-blind, randomised placebo-controlled trial. The Lancet. 2018 Oct 27;392(10157):1519–29.

45. Gerstein HC, Colhoun HM, Dagenais GR, Diaz R, Lakshmanan M, Pais P, et al. Dulaglutide and cardiovascular outcomes in type 2 diabetes (REWIND): a double-blind, randomised placebo-controlled trial. The Lancet. 2019 Jul 13;394(10193):121–30.

46. Steiner S. Empagliflozin, cardiovascular outcomes, and mortality in type 2 diabetes. Vol. 13, Zeitschrift fur Gefassmedizin. Krause und Pachernegg GmbH; 2016. p. 17–8.

47. Neal B, Perkovic V, Mahaffey KW, de Zeeuw D, Fulcher G, Erondu N, et al. Canagliflozin and Cardiovascular and Renal Events in Type 2 Diabetes. New England Journal of Medicine. 2017 Aug 17;377(7):644–57.

48. Neal B, Perkovic V, Mahaffey KW, de Zeeuw D, Fulcher G, Erondu N, et al. Canagliflozin and Cardiovascular and Renal Events in Type 2 Diabetes. [Protocol] New England Journal of Medicine. 2017 Aug 17;377(7):644–57.

49. Neal B, Perkovic V, Mahaffey KW, de Zeeuw D, Fulcher G, Erondu N, et al. Canagliflozin and Cardiovascular and Renal Events in Type 2 Diabetes. [Supplementary Appendix] New England Journal of Medicine. 2017 Aug 17;377(7):644–57.

50. Marso SP, Bain SC, Consoli A, Eliaschewitz FG, Jódar E, Leiter LA, et al. Semaglutide and Cardiovascular Outcomes in Patients with Type 2 Diabetes. New England Journal of Medicine. 2016 Nov 10;375(19):1834–44.

51. Marso SP, Bain SC, Consoli A, Eliaschewitz FG, Jódar E, Leiter LA, et al. Semaglutide and Cardiovascular Outcomes in Patients with Type 2 Diabetes. [Protocol] New England Journal of Medicine. 2016 Nov 10;375(19):1834–44.

52. Marso SP, Bain SC, Consoli A, Eliaschewitz FG, Jódar E, Leiter LA, et al. Semaglutide and Cardiovascular Outcomes in Patients with Type 2 Diabetes. [Supplementary Appendix] New England Journal of Medicine. 2016 Nov 10;375(19):1834–44.

53. Holman RR, Bethel MA, Mentz RJ, Thompson VP, Lokhnygina Y, Buse JB, et al. Effects of Once-Weekly Exenatide on Cardiovascular Outcomes in Type 2 Diabetes. New England Journal of Medicine. 2017 Sep 28;377(13):1228–39.

54. Holman RR, Bethel MA, George J, Sourij H, Doran Z, Keenan J, et al. Rationale and design of the EXenatide Study of Cardiovascular Event Lowering (EXSCEL) trial. [Protocol] American Heart Journal. 2016 Apr 1;174:103–10.

55. Holman RR, Bethel MA, Mentz RJ, Thompson VP, Lokhnygina Y, Buse JB, et al. Effects of Once-Weekly Exenatide on Cardiovascular Outcomes in Type 2 Diabetes. [Supplementary Appendix] New England Journal of Medicine. 2017 Sep 28;377(13):1228–39.

56. Pfeffer MA, Claggett B, Diaz R, Dickstein K, Gerstein HC, Køber L v., et al. Lixisenatide in Patients with Type 2 Diabetes and Acute Coronary Syndrome. New England Journal of Medicine. 2015 Dec 3;373(23):2247–57.

57. Cannon CP, Pratley R, Dagogo-Jack S, Mancuso J, Huyck S, Masiukiewicz U, et al. Cardiovascular Outcomes with Ertugliflozin in Type 2 Diabetes. New England Journal of Medicine. 2020 Oct 8;383(15):1425–35.

58. Cannon CP, McGuire DK, Pratley R, Dagogo-Jack S, Mancuso J, Huyck S, et al. Design and baseline characteristics of the eValuation of ERTugliflozin effIcacy and Safety CardioVascular outcomes trial (VERTIS-CV). [Protocol] American Heart Journal. 2018 Dec 1;206:11–23.

59. Cannon CP, Pratley R, Dagogo-Jack S, Mancuso J, Huyck S, Masiukiewicz U, et al. Cardiovascular Outcomes with Ertugliflozin in Type 2 Diabetes. [Supplementary Appendix] New England Journal of Medicine. 2020 Oct 8;383(15):1425–35.

60. Wiviott SD, Raz I, Bonaca MP, Mosenzon O, Kato ET, Cahn A, et al. Dapagliflozin and Cardiovascular Outcomes in Type 2 Diabetes. New England Journal of Medicine. 2019 Jan 24;380(4):347–57.

61. Wiviott SD, Raz I, Bonaca MP, Mosenzon O, Kato ET, Cahn A, et al. Dapagliflozin and Cardiovascular Outcomes in Type 2 Diabetes. [Protocol] New England Journal of Medicine. 2019 Jan 24;380(4):347–57.

62. Wiviott SD, Raz I, Bonaca MP, Mosenzon O, Kato ET, Cahn A, et al. Dapagliflozin and Cardiovascular Outcomes in Type 2 Diabetes. [Supplementary Appendix] New England Journal of Medicine. 2019 Jan 24;380(4):347–57.

63. Husain M, Birkenfeld AL, Donsmark M, Dungan K, Eliaschewitz FG, Franco DR, et al. Oral Semaglutide and Cardiovascular Outcomes in Patients with Type 2 Diabetes. New England Journal of Medicine. 2019 Aug 29;381(9):841–51.

64. Husain M, Birkenfeld AL, Donsmark M, Dungan K, Eliaschewitz FG, Franco DR, et al. Oral Semaglutide and Cardiovascular Outcomes in Patients with Type 2 Diabetes. [Protocol] New England Journal of Medicine. 2019 Aug 29;381(9):841–51.

65. Husain M, Birkenfeld AL, Donsmark M, Dungan K, Eliaschewitz FG, Franco DR, et al. Oral Semaglutide and Cardiovascular Outcomes in Patients with Type 2 Diabetes. [Supplementary Appendix] New England Journal of Medicine. 2019 Aug 29;381(9):841–51.

66. Rosenstock J, Perkovic V, Johansen OE, Cooper ME, Kahn SE, Marx N, et al. Effect of Linagliptin vs Placebo on Major Cardiovascular Events in Adults with Type 2 Diabetes and High Cardiovascular and Renal Risk: The CARMELINA Randomized Clinical Trial. JAMA - Journal of the American Medical Association. 2019 Jan 1;321(1):69–79.

67. Rosenstock J, Perkovic V, Alexander JH, Cooper ME, Marx N, Pencina MJ, et al. Rationale, design, and baseline characteristics of the CArdiovascular safety and Renal Microvascular outcomE study with LINAgliptin (CARMELINA®): A randomized, double-blind, placebo-controlled clinical trial in patients with type 2 diabetes and high cardio-renal risk. Cardiovascular Diabetology. 2018 Mar 14;17(1).

68. Rosenstock J, Perkovic V, Alexander JH, Cooper ME, Marx N, Pencina MJ, et al. Rationale, design, and baseline characteristics of the CArdiovascular safety and Renal Microvascular outcomE study with LINAgliptin (CARMELINA®): A randomized, double-blind, placebo-controlled clinical trial in patients with type 2 diabetes and high cardio-renal risk. [Supplementary Appendix] Cardiovascular Diabetology. 2018 Mar 14;17(1).

69. Rosenstock J, Kahn SE, Johansen OE, Zinman B, Espeland MA, Woerle HJ, et al. Effect of Linagliptin vs Glimepiride on Major Adverse Cardiovascular Outcomes in Patients with Type 2 Diabetes: The CAROLINA Randomized Clinical Trial. JAMA - Journal of the American Medical Association. 2019 Sep 24;322(12):1155–66.

70. Marx N, Rosenstock J, Kahn SE, Zinman B, Kastelein JJ, Lachin JM, et al. Design and baseline characteristics of the CARdiovascular Outcome Trial of LINAgliptin Versus Glimepiride in Type 2 Diabetes (CAROLINA®). Diabetes and Vascular Disease Research. 2015 May 9;12(3):164–74.

71. White WB, Cannon CP, Heller SR, Nissen SE, Bergenstal RM, Bakris GL, et al. Alogliptin after acute coronary syndrome in patients with type 2 diabetes. Vol. 7, Austrian Journal of Clinical Endocrinology and Metabolism. Krause und Pachernegg GmbH; 2014. p. 77.

72. White WB, Bakris GL, Bergenstal RM, Cannon CP, Cushman WC, Fleck P, et al. EXamination of CArdiovascular OutcoMes with AlogliptIN versus Standard of CarE in Patients with Type 2 Diabetes Mellitus and Acute Coronary Syndrome (EXAMINE): A cardiovascular safety study of the dipeptidyl peptidase 4 inhibitor alogliptin in patients with type 2 diabetes with acute coronary syndrome. American Heart Journal. 2011;162(4).

73. Scirica BM, Bhatt DL, Braunwald E, Steg PG, Davidson J, Hirshberg B, et al. The design and rationale of the Saxagliptin Assessment of Vascular Outcomes Recorded in patients with diabetes mellitus-Thrombolysis in Myocardial Infarction (SAVOR-TIMI) 53 Study. American Heart Journal. 2011;162(5).

74. Scirica BM, Bhatt DL, Braunwald E, Steg PG, Davidson J, Hirshberg B, et al. Saxagliptin and Cardiovascular Outcomes in Patients with Type 2 Diabetes Mellitus. New England Journal of Medicine [Internet]. 2013 Oct 3;369(14):1317–26. Available from: http://www.nejm.org/doi/10.1056/NEJMoa1307684

75. Green JB, Bethel MA, Armstrong PW, Buse JB, Engel SS, Garg J, et al. Effect of Sitagliptin on Cardiovascular Outcomes in Type 2 Diabetes. New England Journal of Medicine. 2015 Jul 16;373(3):232–42.

76. Walsh M, Srinathan SK, McAuley DF, Mrkobrada M, Levine O, Ribic C, et al. The statistical significance of randomized controlled trial results is frequently fragile: A case for a Fragility Index. Journal of Clinical Epidemiology. 2014;67(6):622–8.

77. Carter RE, McKie PM, Storlie CB. The fragility index: A P-value in sheep's clothing? European Heart Journal. 2017;38(5):346–8.

78.Cefalu WT, Kaul S, Gerstein HC, Holman RR, Zinman B, Skyler JS, et al. Cardiovascular outcomes trials in type 2 diabetes: Where Do We Go From Here? Ref lections From a Diabetes Care Editors' Expert Forum. Diabetes Care. 2018;41(1):14–31.

14. ANNEX

RoB Tool assessment, detailed.

Box 2. The RoB 2 tool (part 1): Preliminary considerations

Study design

- ☐ Individually-randomized parallel-group trial
- ☐ Cluster-randomized parallel-group trial
- ☐ Individually randomized cross-over (or other matched) trial

For the purposes of this assessment, the interventions being compared are defined as

: Experimental: _____ Comparator: _____

Specify which outcome is being assessed for risk of bias _____

Specify the numerical result being assessed. In case of multiple alternative analyses being presented, specify the numeric result (e.g. RR = 1.52 (95% CI 0.83 to 2.77) and/or a reference (e.g. to a table, figure or paragraph) that uniquely defines the result being assessed. _____

Is the review team's aim for this result…?

- ☐ to assess the effect of *assignment to intervention* (the 'intention-to-treat' effect)
- ☐ to assess the effect of *adhering to intervention* (the 'per-protocol' effect)

If the aim is to assess the effect of *adhering to intervention*, select the deviations from intended intervention that should be addressed (at least one must be checked):

- ☐ occurrence of non-protocol interventions
- ☐ failures in implementing the intervention that could have affected the outcome
- ☐ non-adherence to their assigned intervention by trial participants

Which of the following sources were <u>obtained</u> to help inform the risk-of-bias assessment? (tick as many as apply)

- ☐ Journal article(s)
- ☐ Trial protocol
- ☐ Statistical analysis plan (SAP)
- ☐ Non-commercial trial registry record (e.g. ClinicalTrials.gov record)
- ☐ Company-owned trial registry record (e.g. GSK Clinical Study Register record)
- ☐ 'Grey literature' (e.g. unpublished thesis)
- ☐ Conference abstract(s) about the trial
- ☐ Regulatory document (e.g. Clinical Study Report, Drug Approval Package)
- ☐ Research ethics application
- ☐ Grant database summary (e.g. NIH RePORTER or Research Councils UK Gateway to Research)
- ☐ Personal communication with trialist
- ☐ Personal communication with the sponsor

Box 4. The RoB 2 tool (part 2): Risk of bias arising from the randomization process

Signalling questions	Elaboration	Response options
1.1 Was the allocation sequence random?	Answer 'Yes' if a random component was used in the sequence generation process. Examples include computer-generated random numbers; reference to a random number table; coin tossing; shuffling cards or envelopes; throwing dice; or drawing lots. Minimization is generally implemented with a random element (at least when the scores are equal), so an allocation sequence that is generated using minimization should generally be considered to be random. Answer 'No' if no random element was used in generating the allocation sequence or the sequence is predictable. Examples include alternation; methods based on dates (of birth or admission); patient record numbers; allocation decisions made by clinicians or participants; allocation based on the availability of the intervention; or any other systematic or haphazard method. Answer 'No information' if the only information about randomization methods is a statement that the study is randomized. In some situations a judgement may be made to answer 'Probably no' or 'Probably yes'. For example, , in the context of a large trial run by an experienced clinical trials unit, absence of specific information about generation of the randomization sequence, in a paper published in a journal with rigorously enforced word count limits, is likely to result in a response of 'Probably yes' rather than 'No information'. Alternatively, if other (contemporary) trials by the same investigator team have clearly used non-random sequences, it might be reasonable to assume that the current study was done using similar methods.	Y/PY/PN/N/NI
1.2 Was the allocation sequence concealed until participants were enrolled and assigned to interventions?	Answer 'Yes' if the trial used any form of remote or centrally administered method to allocate interventions to participants, where the process of allocation is controlled by an external unit or organization, independent of the enrolment personnel (e.g. independent central pharmacy, telephone or internet-based randomization service providers). Answer 'Yes' if envelopes or drug containers were used appropriately. Envelopes should be opaque, sequentially numbered, sealed with a tamper-proof seal and opened only after the envelope has been irreversibly assigned to the participant. Drug containers should be sequentially numbered and of identical appearance, and dispensed or administered only after they have been irreversibly assigned to the participant. This level of detail is rarely provided in reports, and a judgement may be required to justify an answer of 'Probably yes' or 'Probably no'. Answer 'No' if there is reason to suspect that the enrolling investigator or the participant had knowledge of the forthcoming allocation.	Y/PY/PN/N/NI
1.3 Did baseline differences between intervention groups suggest a problem with the randomization process?	*Note that differences that are compatible with chance do not lead to a risk of bias. A small number of differences identified as 'statistically significant' at the conventional 0.05 threshold should usually be considered to be compatible with chance.* Answer 'No' if no imbalances are apparent or if any observed imbalances are compatible with chance. Answer 'Yes' if there are imbalances that indicate problems with the randomization process, including: (1) substantial differences between intervention group sizes, compared with the intended allocation ratio; or (2) a substantial excess in statistically significant differences in baseline characteristics between intervention groups, beyond that expected by chance; or	Y/PY/PN/N/NI
	(3) imbalance in one or more key prognostic factors, or baseline measures of outcome variables, that is very unlikely to be due to chance and for which the between-group difference is big enough to result in bias in the intervention effect estimate. Also answer 'Yes' if there are other reasons to suspect that the randomization process was problematic: (4) excessive similarity in baseline characteristics that is not compatible with chance. Answer 'No information' when there is no *useful* baseline information available (e.g. abstracts, or studies that reported only baseline characteristics of participants in the final analysis). The answer to this question should not influence answers to questions 1.1 or 1.2. For example, if the trial has large baseline imbalances, but authors report adequate randomization methods, questions 1.1 and 1.2 should still be answered on the basis of the reported adequate methods, and any concerns about the imbalance should be raised in the answer to the question 1.3 and reflected in the domain-level risk-of-bias judgement. Trialists may undertake analyses that attempt to deal with flawed randomization by controlling for imbalances in prognostic factors at baseline. To remove the risk of bias caused by problems in the randomization process, it would be necessary to know, and measure, all the prognostic factors that were imbalanced at baseline. It is unlikely that all important prognostic factors are known and measured, so such analyses will at best reduce the risk of bias. If review authors wish to assess the risk of bias in a trial that controlled for baseline imbalances in order to mitigate failures of randomization, the study should be assessed using the ROBINS-I tool.	
Risk-of-bias judgement	See Table 3, Table 4 and Figure 1.	Low / High / Some concerns
Optional: What is the predicted direction of bias arising from the randomization process?	If the likely direction of bias can be predicted, it is helpful to state this. The direction might be characterized either as being towards (or away from) the null, or as being in favour of one of the interventions.	NA / Favours experimental / Favours comparator / Towards null /Away from null / Unpredictable

Box 6. The RoB 2 tool (part 3): Risk of bias due to deviations from the intended interventions (*effect of assignment to intervention*)

Signalling questions	Elaboration	Response options
2.1. Were participants aware of their assigned intervention during the trial?	If participants are aware of their assigned intervention it is more likely that health-related behaviours will differ between the intervention groups. Blinding participants, most commonly through use of a placebo or sham intervention, may prevent such differences. If participants experienced side effects or toxicities that they knew to be specific to one of the interventions, answer this question 'Yes' or 'Probably yes'.	Y/PY/PN/N/NI
2.2. Were carers and people delivering the interventions aware of participants' assigned intervention during the trial?	If carers or people delivering the interventions are aware of the assigned intervention then its implementation, or administration of non-protocol interventions, may differ between the intervention groups. Blinding may prevent such differences. If participants experienced side effects or toxicities that carers or people delivering the interventions knew to be specific to one of the interventions, answer question 'Yes' or 'Probably yes'. If randomized allocation was not concealed, then it is likely that carers and people delivering the interventions were aware of participants' assigned intervention during the trial.	Y/PY/PN/N/NI
2.3. If Y/PY/NI to 2.1 or 2.2: Were there deviations from the intended intervention that arose because of the trial context?	For the effect of assignment to intervention, this domain assesses problems that arise when changes from assigned intervention that are inconsistent with the trial protocol arose because of the trial context. We use the term trial context to refer to effects of recruitment and engagement activities on trial participants and when trial personnel (carers or people delivering the interventions) undermine the implementation of the trial protocol in ways that would not happen outside the trial. For example, the process of securing informed consent may lead participants subsequently assigned to the comparator group to feel unlucky and therefore seek the experimental intervention, or other interventions that improve their prognosis. Answer 'Yes' or 'Probably yes' only if there is evidence, or strong reason to believe, that the trial context led to failure to implement the protocol interventions or to implementation of interventions not allowed by the protocol. Answer 'No' or 'Probably no' if there were changes from assigned intervention that are inconsistent with the trial protocol, such as non-adherence to intervention, but these are consistent with what could occur outside the trial context. Answer 'No' or 'Probably no' for changes to intervention that are consistent with the trial protocol, for example cessation of a drug intervention because of acute toxicity or use of additional interventions whose aim is to treat consequences of one of the intended interventions. If blinding is compromised because participants report side effects or toxicities that are specific to one of the interventions, answer 'Yes' or 'Probably yes' only if there were changes from assigned intervention that are inconsistent with the trial protocol and arose because of the trial context. The answer 'No information' may be appropriate, because trialists do not always report whether deviations arose because of the trial context.	NA/Y/PY/PN/N/NI
2.4 If Y/PY to 2.3: Were these deviations likely to have affected the outcome?	Changes from assigned intervention that are inconsistent with the trial protocol and arose because of the trial context will impact on the intervention effect estimate if they affect the outcome, but not otherwise.	NA/Y/PY/PN/N/NI
2.5. If Y/PY/NI to 2.4: Were these deviations from intended intervention balanced between groups?	Changes from assigned intervention that are inconsistent with the trial protocol and arose because of the trial context are more likely to impact on intervention effect estimate if they are not balanced between the intervention groups.	NA/Y/PY/PN/N/NI
2.6 Was an appropriate analysis used to estimate the effect of assignment to intervention?	Both intention-to-treat (ITT) analyses and modified intention-to-treat (mITT) analyses excluding participants with missing outcome data should be considered appropriate. Both naïve 'per-protocol' analyses (excluding trial participants who did not receive their assigned intervention) and 'as treated' analyses (in which trial participants are grouped according to the intervention that they received, rather than according to their assigned intervention) should be considered inappropriate. Analyses excluding eligible trial participants post-randomization should also be considered inappropriate, but post-randomization exclusions of ineligible participants (when eligibility was not confirmed until after randomization, and could not have been influenced by intervention group assignment) can be considered appropriate.	Y/PY/PN/N/NI
2.7 If N/PN/NI to 2.6: Was there potential for a substantial impact (on the result) of the failure to analyse participants in the group to which they were randomized?	This question addresses whether the number of participants who were analysed in the wrong intervention group, or excluded from the analysis, was sufficient that there could have been a substantial impact on the result. It is not possible to specify a precise rule: there may be potential for substantial impact even if fewer than 5% of participants were analysed in the wrong group or excluded, if the outcome is rare or if exclusions are strongly related to prognostic factors.	NA/Y/PY/PN/N/NI
Risk-of-bias judgement	See Table 5, Table 6 and Figure 2.	Low / High / Some concerns
Optional: What is the predicted direction of bias due to deviations from intended interventions?	If the likely direction of bias can be predicted, it is helpful to state this. The direction might be characterized either as being towards (or away from) the null, or as being in favour of one of the interventions.	NA / Favours experimental / Favours comparator / Towards null /Away from null / Unpredictable

Box 8. The RoB 2 tool (part 5): Risk of bias due to missing outcome data

Signalling questions	Elaboration	Response options
3.1 Were data for this outcome available for all, or nearly all, participants randomized?	The appropriate study population for an analysis of the intention to treat effect is all randomized participants. "Nearly all" should be interpreted as that the number of participants with missing outcome data is sufficiently small that their outcomes, whatever they were, could have made no important difference to the estimated effect of intervention. For continuous outcomes, availability of data from 95% of the participants will often be sufficient. For dichotomous outcomes, the proportion required is directly linked to the risk of the event. If the observed number of events is much greater than the number of participants with missing outcome data, the bias would necessarily be small. Only answer 'No information' if the trial report provides no information about the extent of missing outcome data. This situation will usually lead to a judgement that there is a high risk of bias due to missing outcome data. Note that imputed data should be regarded as missing data, and not considered as 'outcome data' in the context of this question.	Y/PY/PN/N/NI
3.2 If N/PN/NI to 3.1: Is there evidence that the result was not biased by missing outcome data?	Evidence that the result was not biased by missing outcome data may come from: (1) analysis methods that correct for bias; or (2) sensitivity analyses showing that results are little changed under a range of plausible assumptions about the relationship between missingness in the outcome and its true value. However, imputing the outcome variable, either through methods such as 'last-observation-carried-forward' or via multiple imputation based only on intervention group, should not be assumed to correct for bias due to missing outcome data.	NA/Y/PY/PN/N
3.3 If N/PN to 3.2: Could missingness in the outcome depend on its true value?	If loss to follow up, or withdrawal from the study, could be related to participants' health status, then it is possible that missingness in the outcome was influenced by its true value. However, if all missing outcome data occurred for documented reasons that are unrelated to the outcome then the risk of bias due to missing outcome data will be low (for example, failure of a measuring device or interruptions to routine data collection). In time-to-event analyses, participants censored during trial follow-up, for example because they withdrew from the study, should be regarded as having missing outcome data, even though some of their follow up is included in the analysis. Note that such participants may be shown as included in analyses in CONSORT flow diagrams.	NA/Y/PY/PN/N/NI
3.4 If Y/PY/NI to 3.3: Is it likely that missingness in the outcome depended on its true value?	This question distinguishes between situations in which (i) missingness in the outcome could depend on its true value (assessed as 'Some concerns') from those in which (ii) it is likely that missingness in the outcome depended on its true value (assessed as 'High risk of bias'). Five reasons for answering 'Yes' are: 1. Differences between intervention groups in the proportions of missing outcome data. If there is a difference between the effects of the experimental and comparator interventions on the outcome, and the missingness in the outcome is influenced by its true value, then the proportions of missing outcome data are likely to differ between intervention groups. Such a difference suggests a risk of bias due to missing outcome data, because the trial result will be sensitive to missingness in the outcome being related to its true value. For time-to-event-data, the analogue is that rates of censoring (loss to follow-up) differ between the intervention groups. 2. Reported reasons for missing outcome data provide evidence that missingness in the outcome depends on its true value;	NA/Y/PY/PN/N/NI
	3. Reported reasons for missing outcome data differ between the intervention groups; 4. The circumstances of the trial make it likely that missingness in the outcome depends on its true value. For example, in trials of interventions to treat schizophrenia it is widely understood that continuing symptoms make drop out more likely. 5. In time-to-event analyses, participants' follow up is censored when they stop or change their assigned intervention, for example because of drug toxicity or, in cancer trials, when participants switch to second-line chemotherapy. Answer 'No' if the analysis accounted for participant characteristics that are likely to explain the relationship between missingness in the outcome and its true value.	
Risk-of-bias judgement	See Table 9, Table 10 and Figure 4.	Low / High / Some concerns
Optional: What is the predicted direction of bias due to missing outcome data?	If the likely direction of bias can be predicted, it is helpful to state this. The direction might be characterized either as being towards (or away from) the null, or as being in favour of one of the interventions.	NA / Favours experimental / Favours comparator / Towards null /Away from null / Unpredictable

Box 10. The RoB 2 tool (part 6): Risk of bias in measurement of the outcome

Signalling questions	Elaboration	Response options
4.1 Was the method of measuring the outcome inappropriate?	This question aims to identify methods of outcome measurement (data collection) that are unsuitable for the outcome they are intended to evaluate. The question *does not* aim to assess whether the choice of outcome being evaluated was sensible (e.g. because it is a surrogate or proxy for the main outcome of interest). In most circumstances, for pre-specified outcomes, the answer to this question will be 'No' or 'Probably no'. Answer 'Yes' or 'Probably yes' if the method of measuring the outcome is inappropriate, for example because: (1) it is unlikely to be sensitive to plausible intervention effects (e.g. important ranges of outcome values fall outside levels that are detectable using the measurement method); or (2) the measurement instrument has been demonstrated to have poor validity.	Y/PY/PN/N/NI
4.2 Could measurement or ascertainment of the outcome have differed between intervention groups?	Comparable methods of outcome measurement (data collection) involve the same measurement methods and thresholds, used at comparable time points. Differences between intervention groups may arise because of 'diagnostic detection bias' in the context of passive collection of outcome data, or if an intervention involves additional visits to a healthcare provider, leading to additional opportunities for outcome events to be identified.	Y/PY/PN/N/NI
4.3 If N/PN/NI to 4.1 and 4.2: Were outcome assessors aware of the intervention received by study participants?	Answer 'No' if outcome assessors were blinded to intervention status. For participant-reported outcomes, the outcome assessor is the study participant.	NA/Y/PY/PN/N/NI
4.4 If Y/PY/NI to 4.3: Could assessment of the outcome have been influenced by knowledge of intervention received?	Knowledge of the assigned intervention could influence participant-reported outcomes (such as level of pain), observer-reported outcomes involving some judgement, and intervention provider decision outcomes. They are unlikely to influence observer-reported outcomes that do not involve judgement, for example all-cause mortality.	NA/Y/PY/PN/N/NI
4.5 If Y/PY/NI to 4.4: Is it likely that assessment of the outcome was influenced by knowledge of intervention received?	This question distinguishes between situations in which (i) knowledge of intervention status could have influenced outcome assessment but there is no reason to believe that it did (assessed as 'Some concerns') from those in which (ii) knowledge of intervention status was likely to influence outcome assessment (assessed as 'High'). When there are strong levels of belief in either beneficial or harmful effects of the intervention, it is more likely that the outcome was influenced by knowledge of the intervention received. Examples may include patient-reported symptoms in trials of homeopathy, or assessments of recovery of function by a physiotherapist who delivered the intervention.	NA/Y/PY/PN/N/NI
Risk-of-bias judgement	See Table 11, Table 12 and Figure 5.	Low / High / Some concerns

Box 11. The RoB 2 tool (part 7): Risk of bias in selection of the reported result

Signalling questions	Elaboration	Response options
5.1 Were the data that produced this result analysed in accordance with a pre-specified analysis plan that was finalized before unblinded outcome data were available for analysis?	If the researchers' pre-specified intentions are available in sufficient detail, then planned outcome measurements and analyses can be compared with those presented in the published report(s). To avoid the possibility of selection of the reported result, finalization of the analysis intentions must precede availability of unblinded outcome data to the trial investigators. Changes to analysis plans that were made before unblinded outcome data were available, or that were clearly unrelated to the results (e.g. due to a broken machine making data collection impossible) do not raise concerns about bias in selection of the reported result.	Y/PY/PN/N/NI
Is the numerical result being assessed likely to have been selected, on the basis of the results, from...		
5.2. ... multiple eligible outcome measurements (e.g. scales, definitions, time points) within the outcome domain?	A particular outcome domain (i.e. a true state or endpoint of interest) may be **measured** in multiple ways. For example, the domain pain may be measured using multiple scales (e.g. a visual analogue scale and the McGill Pain Questionnaire), each at multiple time points (e.g. 3, 6 and 12 weeks post-treatment). If multiple measurements were made, but only one or a subset is reported on the basis of the results (e.g. statistical significance), there is a high risk of bias in the fully reported result. Attention should be restricted to outcome measurements that are eligible for consideration by the RoB 2 tool user. For example, if only a result using a specific measurement scale is eligible for inclusion in a meta-analysis (e.g. Hamilton Depression Rating Scale), and this is reported by the trial, then there would not be an issue of selection even if this result was reported (on the basis of the results) in preference to the result from a different measurement scale (e.g. Beck Depression Inventory). Answer 'Yes' or 'Probably yes' if: There is clear evidence (usually through examination of a trial protocol or statistical analysis plan) that a domain was measured in multiple eligible ways, but data for only one or a subset of measures is fully reported (without justification), and the fully reported result is likely to have been selected on the basis of the results. Selection on the basis of the results can arise from a desire for findings to be newsworthy, sufficiently noteworthy to merit publication, or to confirm a prior hypothesis. For example, trialists who have a preconception, or vested interest in showing, that an experimental intervention is beneficial may be inclined to report outcome measurements selectively that are favourable to the experimental intervention. Answer 'No' or 'Probably no' if: There is clear evidence (usually through examination of a trial protocol or statistical analysis plan) that all eligible reported results for the outcome domain correspond to all intended outcome measurements.	Y/PY/PN/N/NI

	or There is only one possible way in which the outcome domain can be measured (hence there is no opportunity to select from multiple measures). or Outcome measurements are inconsistent across different reports on the same trial, but the trialists have provided the reason for the inconsistency and it is not related to the nature of the results. Answer 'No information' if: Analysis intentions are not available, or the analysis intentions are not reported in sufficient detail to enable an assessment, and there is more than one way in which the outcome domain could have been measured.	
5.3 ... multiple eligible analyses of the data?	A particular outcome measurement may be analysed in multiple ways. Examples include: unadjusted and adjusted models; final value vs change from baseline vs analysis of covariance; transformations of variables; different definitions of composite outcomes (e.g. 'major adverse event'); conversion of continuously scaled outcome to categorical data with different cut-points; different sets of covariates for adjustment; and different strategies for dealing with missing data. Application of multiple methods generates multiple effect estimates for a specific outcome measurement. If multiple estimates are generated but only one or a subset is reported on the basis of the results (e.g. statistical significance), there is a high risk of bias in the fully reported result. Attention should be restricted to analyses that are eligible for consideration by the RoB 2 tool user. For example, if only the result from an analysis of post-intervention values is eligible for inclusion in a meta-analysis (e.g. at 12 weeks after randomization), and this is reported by the trial, then there would not be an issue of selection even if this result was reported (on the basis of the results) in preference to the result from an analysis of changes from baseline. Answer 'Yes' or 'Probably yes' if: There is clear evidence (usually through examination of a trial protocol or statistical analysis plan) that a measurement was analysed in multiple eligible ways, but data for only one or a subset of analyses is fully reported (without justification), and the fully reported result is likely to have been selected on the basis of the results. Selection on the basis of the results arises from a desire for findings to be newsworthy, sufficiently noteworthy to merit publication, or to confirm a prior hypothesis. For example, trialists who have a preconception or vested interest in showing that an experimental intervention is beneficial may be inclined to selectively report analyses that are favourable to the experimental intervention. Answer 'No' or 'Probably no' if: There is clear evidence (usually through examination of a trial protocol or statistical analysis plan) that all eligible reported results for the outcome measurement correspond to all intended analyses. or	Y/PY/PN/N/NI
	Answer 'No' or 'Probably no' if: There is clear evidence (usually through examination of a trial protocol or statistical analysis plan) that all eligible reported results for the outcome measurement correspond to all intended analyses. or	

	There is only one possible way in which the outcome measurement can be analysed (hence there is no opportunity to select from multiple analyses). or Analyses are inconsistent across different reports on the same trial, but the trialists have provided the reason for the inconsistency and it is not related to the nature of the results. Answer 'No information' if: Analysis intentions are not available, or the analysis intentions are not reported in sufficient detail to enable an assessment, and there is more than one way in which the outcome measurement could have been analysed.	
Risk-of-bias judgement	See Table 13, Table 14 and Figure 7.	Low / High / Some concerns
Optional: What is the predicted direction of bias due to selection of the reported result?	If the likely direction of bias can be predicted, it is helpful to state this. The direction might be characterized either as being towards (or away from) the null, or as being in favour of one of the interventions.	NA / Favours experimental / Favours comparator / Towards null /Away from null / Unpredictable