The feasibility of using serious adverse event rates as a measure of trial representativeness for pharmacological interventions: Using Randomised controlled trials of sodium-glucose co-transporter-2 inhibitors as an exemplar.

Khalid Alsallumi

BSc., MSc.

Submitted in fulfilment of the requirements for the degree of Doctor of Philosophy (PhD)

School of Health and Wellbeing

College of Medical, Veterinary and Life Sciences

University of Glasgow

September 2024

Abstract

Background

Randomised controlled trials (RCTs) are the gold standard for determining the efficacy and safety of medical interventions. However, their representativeness is often uncertain, as many trials tend to recruit healthier, younger, and less comorbid patients, potentially limiting the generalisability of findings to the real-world population. Assessing trial representativeness is complex, and currently, there is no gold standard measure to address this. Serious adverse events (SAEs) are clearly defined by the Food and Drug Administration (FDA), and regulatory bodies obligate trial sponsors to report all SAEs, regardless of causation. SAEs may reflect the underlying health status of participants, as they include events such as hospitalisations that may not be directly related to the intervention, thus serving as a potential marker of how sicker the participants are. Consequently, SAE reporting could provide insights into trial representativeness. In this thesis, I will explore whether SAE rates can be used to measure trial representativeness, using RCTs of sodium-glucose co-transporter-2 (SGLT-2) inhibitors as an exemplar.

Methods

Clinical trials of SGLT-2 inhibitors were identified from a recent systematic review and included in this thesis. I extracted data on SAE reporting, eligibility criteria, and multiple variables for trial and baseline characteristics. ClinicalTrials.gov, other trial registries, clinical study reports, and trial-relevant publications were used to extract the required data. To assess the feasibility of using SAE rate as a measure of trial representativeness, I initially explored whether SGLT-2 trials reported adequate information to calculate the SAE rates. I then compared SAE rates between trial arms to determine whether SAE rates varied depending on treatment and, by extension, whether each arm should be considered separately or together when assessing SAE rates. In the absence of a gold standard, I used different approaches to explore whether SAEs reflect trial

representativeness. First, I used the pragmatism metric (PRECIS-2 tool, PRagmatic Explanatory Continuum Indicator Summary) to examine its association with SAE rates. I operationalised the PRECIS-2 tool to align with the characteristics of SGLT-2 trials and assessed the trials retrospectively by scoring the nine PRECIS-2 domains. Second, I compared SAE rates with the PRECIS-2 score based on the differences in their associations with trial and baseline characteristics (included as fair umpires) to explore whether SAE is a better metric. Lastly, I examined the association between trial eligibility criteria and the SAE rates to further assess their potential as a measure of trial representativeness.

Results

A total of 146 RCTs for SGLT-2 inhibitors were identified and included in the analysis. In my literature review I found a lack of representativeness in clinical trials, which limited the generalisability of their findings. This lack of representativeness is driven by underrepresentation of older patients, racial/ethnic minorities, and the impact of strict eligibility criteria.

Trials of SGLT-2 inhibitors reported sufficient information, including the number of participants who experienced SAEs, the number of subjects at risk of SAEs, and the timeframe for these events, allowing for the calculation of SAE rates. Trials registered on ClinicalTrials.gov showed better SAE reporting than other trials. The reporting of major adverse cardiovascular events (MACE) was consistently included in the reporting of SAEs. There was no significant difference in the rates of SAE between the intervention and placebo arms (incidence rate ratio [IRR] 0.89, 95% confidence interval [CI] 0.73-1.07), and between the intervention and active comparator arms (IRR 0.90, 95% CI 0.69-1.17). Trials registered on ClinicalTrials.gov had higher SAE rates than trials registered on other registries (IRR 1.94, 95% CI 1.23-3.02). Multinational trials showed higher SAE rates than national trials (IRR 1.79, 95% CI 1.32-2.41). Additionally, SAE rates were higher in trials with hard outcomes (e.g., MACE) compared to those with soft outcomes (IRR 2.86, 95% CI 1.84-4.74).

A higher mean PRECIS-2 score was associated with higher SAE rates (β = 0.17, 95% CI 0.00–0.33). The score for the setting domain was also associated with higher SAE rates (β = 0.51, 95% CI 0.26–0.76). Furthermore, the SAE rate and mean PRECIS-2 score were positively associated with certain baseline/trial characteristics (umpires), such as diabetes duration and sponsorship. The trend for most umpires generally favoured SAE as a metric of trial representativeness; however, only age and type of blinding umpires strongly favoured SAE over PRECIS-2 (β = 0.97; 95% CI 0.26 to 1.70, β = 0.72; 95% CI 0.02 to 1.43, respectively). Trials with more restrictive eligibility criteria had lower SAE rates (IRR 0.75, 95% CI 0.73-0.77) than trials with more permissive criteria (IRR 1.33, 95% CI 1.30-1.37).

Conclusion

Clinical trials of SGLT-2 inhibitors reported sufficient information on SAE, allowing for the calculation of SAE rates. There was no difference in SAE rates between trial arms, enabling the combination of total SAEs for both arms. Pragmatic trials, which resemble real-world practice according to the mean PRECIS-2 score, showed higher SAE rates. The trend for most umpires favoured SAE more than PRECIS-2, and trials with restrictive eligibility criteria showed lower SAE rates compared to those with permissive criteria. Therefore, SAE rates may help assess the representativeness of clinical trials.

Table of Contents

ABSTRACT 2

TABLE OF CONTENTS 5

LIST OF TABLES 11

LIST OF FIGURES 13

ACKNOWLEDGEMENTS 14

AUTHOR'S DECLARATION 15

ABBREVIATIONS 16

CHAPTER 1 INTRODUCTION 21

1.1 Chapter overview 21

1.2 Randomised controlled trials (gold standard) 21

1.2.1 Eligibility criteria 21

1.2.2 Representativeness, applicability, pragmatism 22

1.3 Representativeness of randomised controlled trials 23

1.4 Aim and objectives of the thesis 24

CHAPTER 2 GENERALISABILITY OF CLINICAL TRIALS (LITERATURE REVIEW) 26

2.1 Chapter overview 26

2.2 Background 26

2.2.1 Rationale for choosing this review 26

2.2.2 What is generalisability? 27

2.2.3 Validity 27

2.3 Clinical trials generalisability review 28

2.3.1 Search strategy 28

2.3.2 Oncology trials 29

2.3.3 Cardiology trials 31
2.3.4 Psychiatry trials 31
2.3.5 Type 2 diabetes trials 33
2.3.6 Substance use disorder 33
2.3.7 HIV trials 34
2.3.8 Vitamin D trials 34
2.3.9 Nicotine dependence trials 35
2.3.10 Rheumatoid arthritis trials 35

2.4 Summary of the literature review 45

CHAPTER 3 RATIONALE AND OUTLINE OF THESIS 46

3.1 Chapter overview 46

3.2 Current approach used as a metric for trials representativeness 46

3.3 Serious adverse events (SAEs) 48
3.3.1 Overview 48
3.3.2 SAE as promising alternative measure of trial representativeness 49

3.4 Thesis research question and justification 50

3.5 Rationale for choosing T2DM and trials of SGLT-2 inhibitors as an example 51

3.6 Outline of thesis 52

CHAPTER 4 THE FEASIBILITY OF CAPTURING SAE RATES FROM RCTS OF SGLT-2 INHIBITORS: COVERAGE OF REPORTING 55

4.1 Chapter overview 55

4.2 Background 55

4.3 Aim and objectives 56

4.4 Methods 56
4.4.1 Study selection 56
4.4.2 Data sources 58
4.4.3 Data extraction and harmonisation 59
4.4.4 Statistical analysis 60

4.5 Results 60
4.5.1 Description and source of included studies 60

4.5.2 SAEs reporting in clinical trials of SGLT-2 inhibitors 62

4.6 Discussion 63

4.6.1 Summary of main findings 63

4.6.2 Interpretation 63

4.6.3 Comparison with other studies 64

4.6.4 Implications 65

4.6.5 Strengths and limitations 65

4.7 Conclusion 66

CHAPTER 5 DIFFERENCE IN SERIOUS ADVERSE EVENT RATES BETWEEN TRIAL ARMS OF SGLT-2 INHIBITORS 67

5.1 Chapter overview 67

5.2 Background 67

5.3 Aim and objectives 68

5.4 Methods 69

5.4.1 Data source 69

5.4.2 Data extraction 69

5.4.3 Statistical analysis 69

5.5 Results 71

5.5.1 Summary statistics 71

5.5.2 SAE rates between trial arms 71

5.5.3 Differences in SAE rates based on some factors 72

5.6 Discussion 72

5.6.1 Summary of main findings 72

5.6.2 Interpretation 73

5.6.3 Comparison with other studies 74

5.6.4 Implications 74

5.6.5 Strengths and limitations 75

5.7 Conclusion 76

CHAPTER 6 ASSESSMENT OF THE PRAGMATISM OF RCTS OF SGLT-2 INHIBITORS: USING THE PRAGMATISM METRIC PRECIS-2 TOOL 77

6.1 Chapter overview 77

6.2 Introduction 77
6.2.1 PRECIS-2 tool 77
6.2.2 Rationale for using PRECIS-2 tool 78
6.2.3 Domains of PRECIS-2 tool 78

6.3 Aim and objectives 83

6.4 Methods 83
6.4.1 Data source 83
6.4.2 Data extraction 84
6.4.3 PRECIS-2 ratings and protocolisation of assessment criteria 84
6.4.4 Statistical analysis 92

6.5 Results 93
6.5.1 PRECIS-2 scores 93
6.5.2 The association between SAE rates and PRECIS-2 score 96
6.5.3 The association between SAE rates and domains of PRECIS-2 96

6.6 Discussion 97
6.6.1 Summary of main findings 97
6.6.2 Interpretation 98
6.6.3 Challenge of using PRECIS-2 tool 99
6.6.4 Comparison with other studies 99
6.6.5 Strengths and limitations 100

6.7 Conclusion 100

CHAPTER 7 COMPARISON BETWEEN SAES AND PRECIS-2 TOOL BY EXPLORING THEIR ASSOCIATION WITH TRIAL AND BASELINE CHARACTERISTICS OF SGLT-2 INHIBITORS 102

7.1 Chapter overview 102

7.2 Introduction 102
7.2.1 Using SAEs to assess trial representativeness and lack of gold standard 102
7.2.2 Justification of using these umpires as marker of trial representativeness 103
7.2.3 Expected association of baseline and trial characteristics with SAEs 104

7.3 Aim and objectives 106

7.4 Methods 106
7.4.1 Data source 106
7.4.2 Data extraction 106
7.4.3 Statistical analysis 107

7.5 Results **108**
7.5.1 Descriptive statistics 108
7.5.2 The association of trial/baseline characteristics with SAEs and PRECIS-2 110

7.6 Discussion **114**
7.6.1 Summary of main findings 114
7.6.2 Interpretation and comparison with literature 114
7.6.3 Implications 116
7.6.4 Strengths and limitations 117

7.7 Conclusion **118**

CHAPTER 8 ASSOCIATION BETWEEN ELIGIBILITY CRITERIA AND SAE RATES IN CLINICAL TRIALS FOR SGLT-2 INHIBITORS 119

8.1 Chapter overview **119**

8.2 Background 119

8.3 Aim and objectives **120**

8.4 Methods **120**
8.4.1 Data source 120
8.4.2 Data extraction and harmonisation 120
8.4.3 Measure (restrictiveness of eligibility criteria) 124
8.4.4 Statistical analysis 130

8.5 Results **131**
8.5.1 Characteristics of eligibility criteria 131
8.5.2 Association between eligibility criteria and SAE rates 134
8.5.3 Association between eligibility criteria score and SAE rates 134

8.6 Discussion **135**
8.6.1 Summary of main findings 135
8.6.2 Interpretation 135
8.6.3 Comparison with other studies 137
8.6.4 Implications 138
8.6.5 Strengths and limitations 138

8.7 Conclusion **139**

CHAPTER 9 DISCUSSION 140

9.1 Chapter overview 140

9.2 Background of the thesis 140

9.3 Summary of findings 141

9.3.1 Representativeness of clinical trials and their generalisability 141

9.3.2 SAEs reporting in RCTs of SGLT-2 inhibitors 142

9.3.3 Comparison of SAE rates between trial arms 142

9.3.4 Association between pragmatism metric and SAE rates 143

9.3.5 Association of SAEs and PRECIS-2 with trial/baseline characteristics 143

9.3.6 Association between eligibility criteria and SAE rates 144

9.4 Strengths and limitations of this thesis 144

9.5 Contribution of the thesis 146

9.6 Implications and recommendations of research 147

9.6.1 Importance of SAE reporting in RCTs 148

9.6.2 Pragmatism of clinical trial and SAE rate 148

9.6.3 Importance of trial eligibility criteria 149

9.6.4 Further prospective research 149

9.7 Conclusion 151

REFERENCES 152

SUPPLEMENTARY 173

List of Tables

Table 2.1 Keywords used in the literature review search 28
Table 2.2 Summary of the main findings from the literature review 37
Table 4.1 Types of trial registers and data sources..... 61
Table 4.2 Sources of timeframe for the reported SAEs. 62
Table 4.3 Summary of trial registries and SAEs reporting on each registry. 62
Table 4.4 Reporting of MACE count within the reporting of SAEs 63
Table 5.1: Summary statistics for SAEs counts across arms of SGLT-2 trials. 71
Table 6.1 PRECIS-2 tool domains and variation between pragmatic and explanatory design..... 82
Table 6.2 The protocolisation of the PRECIS-2 for clinical trials of SGLT-2 inhibitors. 86
Table 6.3 Association between PRECIS-2 domains and SAE rates..... 97
Table 7.1 Expected associations of baseline/trials characteristics with the SAE rates 105
Table 7.2 Summary of baseline and trial characteristics (continuous variables) 109
Table 7.3: Summary of baseline and trial characteristics (binary variables) 110
Table 7.4 Comparison between SAE and PRECIS-2 112
Table 8.1 Extraction and harmonisation of eligibility criteria and their definitions 122
Table 8.2 Highly restrictive cutoffs and levels 125
Table 8.3 Restrictiveness score for each criterion based on specific cutoffs 126
Table 8.4 Summary statistics for eligibility criteria status 131
Table 8.5 Summary statistics of continuous criteria of SGLT-2 trials..... 132
Table 8.6: Summary statistics of categorical criteria of SGLT-2 trials..... 133
Table 8.7 Models result for the associations between SAE rates and eligibility criteria 135
Table S.1 Examples of pragmatic and explanatory trial for each domain..... 177
Table S.2 Summary of mean PRECIS-2 scores among domains and medications of SGLT-2 inhibitors 182
Table S.3 Sensitivity analysis for the association between SAE rates and PRECIS-2 domains (missing domains were assigned with the median domain score) 183

Table S.4 Sensitivity analysis for the association between SAE rates and PRECIS-2 domains (missing domains were imputed using MICE)..... 184
Table S.5 Missing data of trial and baseline characteristics before imputation. 184
Table S.6 Sensitivity analysis for the comparison between SAE and PRECIS-2 (using complete case for baseline and trial characteristics without imputation) 185

List of Figures

Figure 2.1 PRISMA diagram of the literature review..... 29
Figure 6.1 Heatmap of missing for all domains of PRECIS-2 tool..... 94
Figure 6.2 Histogram of the mean PRECIS-2 scores for all SGLT-2 trials..... 94
Figure 6.3 Wheel for the mean PRECIS-2 scores for all domains, this wheel is a visible figure to show the direction of score to demonstrate whether domain was pragmatic, explanatory, or equally pragmatic/explanatory..... 95
Figure 6.4 Frequency of PRECIS-2 score among all domains..... 96
Figure S.1 The frequency of SGLT-2 Medication..... 174
Figure S.2 Level of SAE reporting within trial registry..... 174
Figure S.3 The rates of serious adverse event..... 175
Figure S.4 Difference in SAE rates between intervention and control arms..... 176
Figure S.5 Frequency of PRECIS-2 score for all domains with missing information..... 180
Figure S.6 Wheel for mean PRECIS-2 scores for all domains with missing information..... 181
Figure S.7 Compare PRECIS-2 score between trial registries..... 183

Acknowledgements

First, I thank God Almighty for his kindness and easing everything for the preparation and writing of this thesis. I would like to express my great thanks to my supervisors, Professor David McAllister, Professor Jim Lewsey, Dr Claire Niedzwiedz, and Dr Peter Hanlon for their time, guidance, patience, and continued support throughout all stages of my PhD project. Special thanks to Dr Elaine Butterly for her assistance with eligibility criteria and her clinical input. I would also like to thank my mother for her encouragement and support throughout my studies. Huge thanks for my wife for the big role she made in facilitating my working from home. I will not forget to thank my colleagues Lili, Jamie, Saleh, Ryan, Zihao, it has always been encouraging to work with this team.

For my father' souls, who passed away during my PhD study.

"My Lord! Have mercy upon him as he cared for me since I was little". His continues support and weekly encouragement calls before he dead during the early years of my PhD have been a beacon of inspiration. His faith and love continue to guide me.

Author’s Declaration

I declare that the contents of this thesis and the work presented are my own work, and where the work of others has been used it has been stated by reference. This thesis has not been submitted previously for a degree or any other qualification at the University of Glasgow or any other institution.

Khalid Alsallumi

Abbreviations

AACT: Aggregate Analysis of ClinicalTrials.gov

ACR: Albert Cancer Registry

ACTG: AIDS clinical trial group

AE: Adverse events

ALND: Axillary lymph node dissection

AMI: Acute myocardial infarction

BB: Beta blocker

BMI: Body mass index

BPD: Borderline personality disorder

CHF: Chronic heart failure

ChiCTR: Chinese Clinical Trial Registry

CI: Confidence interval

CKD: Chronic kidney disease

COPD: Chronic obstructive pulmonary disease

CPCRA: Community Programs for Clinical Research on AIDS

CrCl: Creatinine Clearance

CSR: Clinical Study Reports

CSS: Cancer-specific survival

CV: Cardiovascular

DBP: Diastolic blood pressure

DCIN: Ductal carcinoma in situ

DIVA: D-Vitamin Intervention in Veteran Administration

DMARDs: Disease-modifying anti-rheumatic drugs

DPP-4: Dipeptidyl peptidase-4

eGFR: estimated Glomerular Filtration Rate

EMA: European medicines agency

FBS: Fasting blood glucose

FDA: Food and Drug Administration

GAD: Generalised anxiety disorder

GLM: generalized Linear Model

GLP-1: Glucagon-like peptide-1

HbA1c: Glycated Haemoglobin (A1c)

HR: Hazard ratio

IPD: Individual participants data

Ipi: Ipilimumab

IRR: Incidence Rate Ratio

ITT: Intention To Treat

Japic: Japan Pharmaceutical Information Center (Japic) Clinical Trials Information

KOA: Knee osteoarthritis

MACE: Major adverse cardiovascular events

MDD: Major depressive disorder

MICE: Multivariate Imputation by Chained Equations

MM: Multiple myeloma

MTT: Meal tolerance test

NCD: National Comorbidity Survey

NCI: National Cancer Institute

NESARC: National Epidemiologic Survey on Alcohol and Related Conditions

NHS: National Health Service

NIHR: National Institute for Health and Care Research

NKBC: Swedish National Breast Cancer Register

NYHA: New York Heart Association

PP: Per Protocol

PRECIS-2: PRagmatic Explanatory Continuum Indicator Summary

RAIN-DB: Rheumatology and Arthritis Investigational Network Database

RCTs: Randomised controlled trials

SAEs: Serious Adverse Events

SAIL: Secure Anonymised Information Linkage Databank

SBP: Systolic blood pressure

SEER: Surveillance Epidemiology and End Results

SGLT-2: Sodium-glucose co-transporter 2

SOCs: System organ classes

T2DM: Type 2 Diabetes Mellitus

TCTR: Thai Clinical Trials Registry

UHC: University Health Systems Consortium

UMIN: University Hospital Medical Information Network

US: United States

USD: Substance use disorder

VADT: Veterans Affairs Diabetes Trial

VARA: Veterans Affairs Rheumatoid Arthritis

Vem: Vemurafenib

WIHS: Women's Interagency HIV Study

Chapter 1 Introduction

1.1 Chapter overview

This chapter will provide a brief introduction about the importance of clinical trials. It will also cover the importance of trial representativeness and its consequences.

1.2 Randomised controlled trials (gold standard)

Randomised controlled trials (RCTs) are essential in the development and evaluation of pharmacological drugs (Houle, 2015). They are prospective studies that typically involve at least two groups of patients: an intervention group and a control group (Kendall, 2003). Trials are important for clinicians and decision-makers because they can determine whether a new treatment works, is effective compared to other treatments, or caused side effects (Selker et al., 2019). They employ randomisation to minimise bias and attribute any differences in outcomes to the treatment rather than to confounders. This randomisation ensures an equal distribution of patient's characteristics between treatment and control group, randomising known and unknown confounding factors that may bias results (Berger et al., 2021). This robust approach is not available with other study designs (e.g., observational studies), which mean that randomisation allows researchers to examine cause-effect relationships more rigorously (Concato, Shah and Horwitz, 2000). Trials are therefore considered the gold standard for examining the efficacy and safety of new treatments (Kabisch et al., 2011).

1.2.1 Eligibility criteria

Eligibility criteria define what individual-level characteristics are required for a person to be included in a trial and are crucial in conducting clinical trials (Weng et al., 2010). The success of RCTs and the reliability of their results rely heavily

on these criteria (Su, Cheng and Huang, 2023). They are essential for defining the patient population under investigation, specifying characteristics such as age, gender, comorbidities, severity of the condition, and medical history (Cragg et al., 2021). Eligibility criteria have several purposes: they specify the target population for the drug, ensure participants safety, facilitate trial conduct by excluding people who would struggle to participate, and ensure the trial reflects the intended use (Kim et al., 2017).

Appropriate selection of eligibility criteria is important to ensure that trial findings can be applied to routine care (Raymond et al., 2024). However, restrictive criteria can impact the representativeness of clinical trials and potentially compromise the generalisability of their results by limiting the diversity of trial participants (Li et al., 2020). This restrictiveness often excludes older people with comorbidities, who represent a high proportion of patients in the community. Consequently, these trials may not accurately reflect the broader patients that would use the intervention in routine care (Florisson et al., 2021). Given these potential challenges with generalising trial findings, it is important to be able to assess the representativeness of clinical trials to allow informed consideration of their implications (Qi et al., 2021).

1.2.2 Representativeness, applicability, pragmatism

As mentioned above, RCTs are the best design in evidence-based medicine but randomisation does not, of itself, promote the applicability of a trial's results to target population (Pibouleau et al., 2009). To understand the applicability and their context in clinical trials, some concepts and principles are defined below:

Trial representativeness: Refers to the extent to which the clinical trial samples mirror or represent broader samples (real-world population) to which the trials interventions are intended to be extrapolated. In other words, it concerns about how well the recruited participants reflect the diversity and characteristics of the broader population in daily practice (Kennedy-Martin et al., 2015).

Trial applicability: Atkins et al. (2011) defined the trial applicability as "the extent to which the effects observed in published studies are likely to reflect the expected results when a specific intervention is applied to the population of interest under 'real-world' conditions".

Trial pragmatism: RCTs were classified as either effectiveness or efficacy trial. In 1976, Schwartz and Lellouch characterised pragmatism as an "attitude to trial design rather than a characteristic of the trial itself". They introduced two terms for clinical trials: the pragmatic and explanatory approaches. The pragmatic trials are designed with the aim of enhancing the applicability of the examined intervention to real-world practice (Omerovic et al., 2024). Conversely, the design of explanatory trials aims to increase the intervention's ability to exhibit optimal effect under ideal conditions (Margolis et al., 2022). Although some authors seem to indicate that a trial is either pragmatic or explanatory, there is a continuum rather than a strict division between these trials (Treweek and Zwarenstein, 2009). Additional details and key comparisons between pragmatic and explanatory trials can be found in Chapter 6.

1.3 Representativeness of randomised controlled trials

Although clinical trials are the highest level of evidence and offer the most unbiased estimates of treatment efficacy, concern about their representativeness persist (Susukida et al., 2016). Often, the participants in these trials do not accurately represent the patients treated in routine care. This issue is particularly evident in type 2 diabetes mellitus (T2DM), a condition explored in this thesis. Indeed, compared to the population identified through record-linkage from the Scottish diabetes register, trial participants in T2DM trials were younger, had fewer comorbid diseases, and women were under-represented (Saunders et al., 2013).

The issue of representativeness is exacerbated by the frequent exclusion of individuals with polypharmacy and older people, often due to restrictive exclusion criteria or factors such as clinician preferences during the recruitment

process (Hanlon et al., 2021). When clinical trials are conducted using samples in which certain patient groups are under-represented or excluded, not only is the representativeness of the trial compromised, but also the impact and relevance of these studies affected (Domingo and Helman, 2022). This problem of under-representation may ultimately influence clinical guidelines by generating evidence that does not accurately meet the individual needs of patients in routine care (Lindsay et al., 2020). Consequently, a lack of representativeness may undermine trust in trials (Fortin et al., 2006).

Achieving representativeness in trials is crucial for enhancing the trust of both clinicians and the public in clinical trial findings (Clark et al., 2019). Improving the representativeness of trial populations is essential for maximising the applicability of trials to the target population, which in turn leads to better quality care for under-represented patients (Kennedy-Martin et al., 2015). Moreover, enhancing the representativeness of trial populations helps improve clinical guidelines that are often developed from evidence derived from studies involving unrepresentative populations (Mas-Llado et al., 2023). However, the absence of a practical method that can effectively assess trial representativeness could potentially compromise the usefulness of trials results and clinical guidelines. Therefore, it becomes necessary to find a method that can address this issue and appropriately assess the representativeness of clinical trials. This thesis aims to explore the feasibility of using serious adverse events (SAEs) as a metric for assessing the trial representativeness, particularly in the context of clinical trials for sodium glucose cotransporter-2 inhibitors (SGLT-2) as an exemplar.

1.4 Aim and objectives of the thesis

To determine whether SAEs reported in clinical trials for SGLT-2 inhibitors can be used as a metric of representativeness of those trials. To address this aim, the current thesis will conduct the following objectives:

1. Explore the feasibility of capturing SAE rates from clinical trials.

2. Compare the SAE rates between trial arms.

3. Apply a trial pragmatism metric (PRECIS-2 tool) to assess the pragmatism of RCTs of SGLT-2 inhibitors and compare this tool to SAE.

4. Examine the association of SAE rates and mean PRECIS-2 score with trial and baseline characteristics.

5. Explore the association between eligibility criteria and SAE rates.

The rationale for the thesis aims and objectives is presented in Chapter 3, which outlines the thesis structure.

Chapter 2 Generalisability of clinical trials (Literature review)

2.1 Chapter overview

This chapter provides a review of the issue of representativeness in clinical trials and its impact on the generalisability of their results across different index conditions, rather than focusing solely on a single condition. It summarises the factors contributing to the lack of representativeness.

2.2 Background

2.2.1 Rationale for choosing this review

The advantage of RCTs lies in their ability to control for confounding factors through randomisation, making them less prone to bias than observational studies (Ratain and Sargent, 2009). RCTs are therefore considered the gold standard for evaluating the efficacy of treatments. However, RCTs are limited in terms of generalisability (Gheorghe et al., 2015). Although these trials often provide precise estimates of intervention's effect for the patients recruited for the trial (referred as “internal validity”), they do not always provide applicable information about the effects on the boarder target population (referred as “external validity”) (Stuart et al., 2015). Some have raised concerns that trial findings are often more applicable to trial participants than the broader target population (Rothwell, 2005). This is because many trials have been shown not to be representative of the real-world populations, suggesting that their findings may not be applicable to all (Hanlon et al., 2022). Therefore, this review was conducted to explore trial representativeness, how the trial participants are representative to real-world populations, and how their findings are generalisable. Table 2.2 summarises the main findings of this review and the methods that have been used to assess trial generalisability.

2.2.2 What is generalisability?

Generalisability describes the extent to which the findings of a study with a specific sample can be applied to a broader target population (Kennedy-Martin et al., 2015). If the trial results can be extrapolated to a population different from the trial population, the trial has good generalisability (S. V. Wang et al., 2019). Conversely, poor generalisability indicates that a study's findings are applicable to a very limited population (Bornhöft et al., 2006).

2.2.3 Validity

The study's validity can be described as the degree to which the findings and interpretations derived from a study are reliable and likely to yield consistent results (Steg et al., 2007). This is can be achieved when the study methods and the nature of the population are comparable (Akobeng, 2008). Study validity can be classified into two types: internal and external validity. Internal validity is associated with the extent to which the design of the trial is performed to avoid or minimise biases (Jadad et al., 1996). The quality of evidence derived from the trial can be negatively affected if the trial has low internal validity (Stephenson and Babiker, 2000). Also, internal validity is a crucial prerequisite for achieving external validity, which is defined as the degree to which the trial results can be applied to real-world populations (Akobeng, 2008). If a trial lacks sufficient internal validity, it will also lack external validity (Steg et al., 2007). However, even if trial results are internally valid, they may have limited clinical utility if the external validity is low (Rothwell, 2005). Clinical trials should possess both internal and external validity. In practice, however, clinical practitioners often argue that the external validity of clinical trials is frequently poor (Reiss, 2019). Moreover, the key determinants of external validity are not always adequately reported (Bornhöft et al., 2006).

2.3 Clinical trials generalisability review

2.3.1 Search strategy

The relevant literature was searched using MEDLINE and EMBASE (via Ovid) databases. These databases were chosen due to their medical specialisation and relevance to the thesis topic. As illustrated in table 2.1, the literature search was conducted using different keywords related to the representativeness and generalisability of clinical trials. The included study must assess the representativeness and generalisability of clinical trials. Studies that were irrelevant to the topic were excluded; 768 out of 1159 articles were excluded after screening the titles, 237 were excluded after reviewing the abstract, and 126 were excluded after reviewing the full text. Eventually, 28 studies were included in the current review (Figure 2.1).

Table 2.1 Keywords used in the literature review search

Steps	Keywords used
1	randomised controlled trial or RCT or clinical experimentation or clinical experiment or clinical study or clinical studies or clinical trial or clinical research or clinical test or phase III trial or phase IV trial
2	reliability or reliable or validity or valid or external validity or internal validity
3	representativeness or representative or generalisability or generalizability or generalisation or generalization or applicability or applicable or generalizable or generalisable or real-world setting
4	1 and 2 and 3

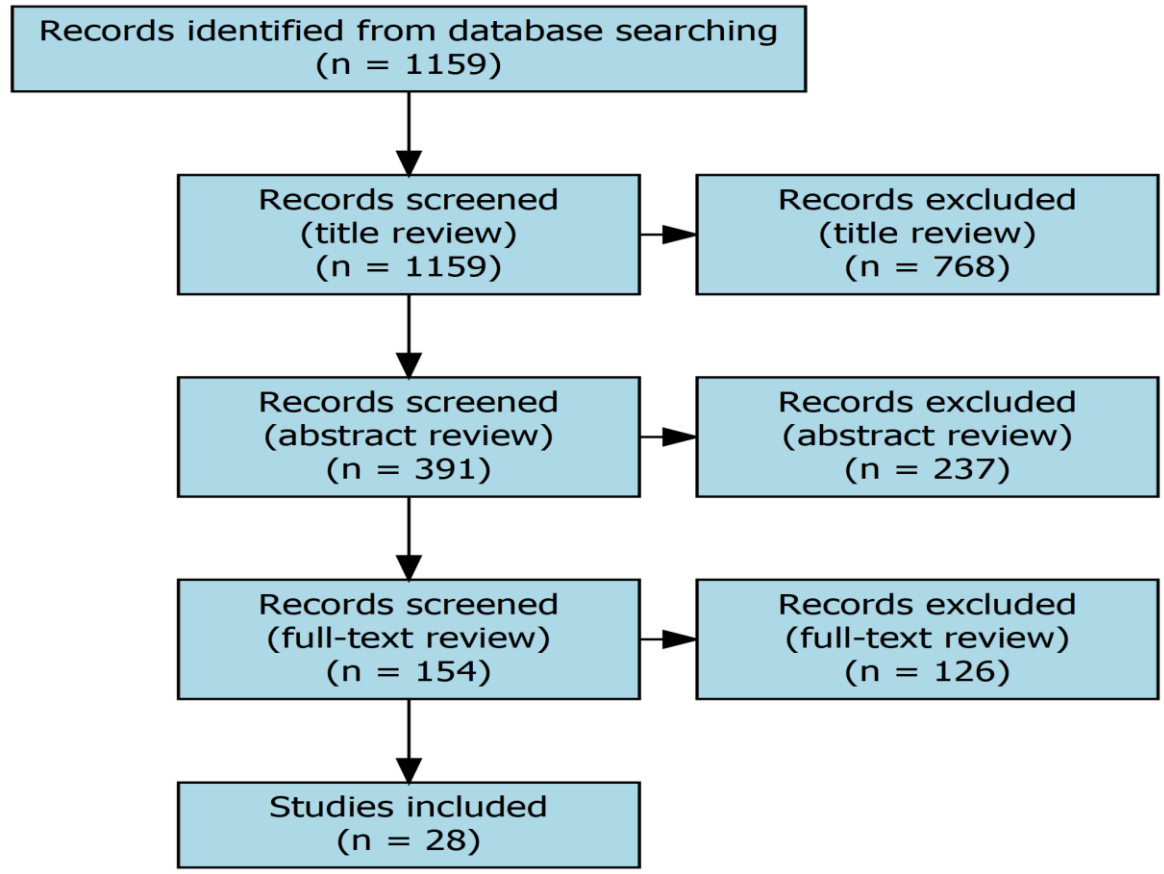

Figure 2.1 PRISMA diagram of the literature review

2.3.2 Oncology trials

The generalisability of cancer clinical trials is uncertain, given that less than 5% of patients participate (Chen, Skingley and Meyer, 2000). Clinical trials often exclude older patients, despite their significant representation among medication consumers, in favour of younger patients who typically exhibit better performance status and fewer comorbidities (Pitkala and Strandberg, 2022). Karim et al. (2019) examined the impact of eligibility criteria on the representativeness of oncology trials. They applied common exclusion criteria to a real-world population to determine the proportion of patients who would be eligible vs ineligible to participate in the RCTs. Their study found that 38% of cancer patients were ineligible for trials due to strict criteria of age (over 75 years), heart disease, and prior malignancy. Trial-eligible patients had higher cancer-specific survival (CSS) compared to trial-ineligible patients (HR 0.72, 95% CI 0.70-0.74). Similarly, 52% of patients with ductal carcinoma in situ (DCIS) were ineligible in the EORTC 10853 trial due to rigorous patient selection

process. Non-entered patients had higher recurrence rates than randomised patients, indicating that the trial results might not be applicable to all DCIS patients due to the unrepresentativeness of the selected participants (Bijker *et al.*, 2002). However, the study only focused on people receiving treatment in speciality cancer centres, which may not represent the broader healthcare setting. Moreover, trial-eligible patients for vemurafenib (Vem) and ipilimumab (Ipi) trials had improved survival compared to ineligible patients (HR 0.68, 95%CI 0.44–0.98) (Sam et al., 2018). Common reasons for ineligibility included recent primary cancer, comorbidities, and untreated brain metastases, which significantly impact the applicability of trial findings.

Costa, Hari and Kumar (2016) conducted a retrospective cohort study on multiple myeloma trials to determine how the dissimilarities between trial participants and the target population affect external validity. They found that real-world populations were older and had more advanced disease compared to trial participants. The stringent eligibility criteria, designed to exclude patients with more advanced disease, underrepresent these patients. While this enhances internal validity, it potentially compromises the external validity of trial results. However, this study was limited to trials conducted in the US, which may not accurately reflect the global patients. Similarly, Elting et al. (2006) compared the characteristics of trial participants and non-participants, finding significant differences in marital status, gender, race, health status, and comorbidities. In contrast, there were no significant differences in tumour characteristics, age, gender, and breast surgery between participants in the SENOMIC trial and the target population, suggesting the trial was representative (Andersson et al., 2019).

Mishkin et al. (2016) assessed the representativeness of NCI-sponsored gynaecological cancer trials by comparing trial participants' demographics (age, race, ethnicity) and insurance status to the real-world population from the NCI's SEER program. They found under-representation of Black women in ovarian trials and Hispanic participants in uterine and ovarian trials. Older patients were under-represented across all trials, while privately insured patients were overrepresented in ovarian trials. These discrepancies between NCI participants

and the target population may limit the generalisability of the trial results. Moreover, Tam et al. (2009) assessed the trials' generalisability by comparing the chemotherapy toxicity rates of irinotecan (grades 3 and 4 diarrhoea) between trial participants and non-participants. No significant difference was found, indicating the representativeness of the included participants.

2.3.3 Cardiology trials

The exclusion of potentially eligible patients from RCTs may lead to selection bias that may influence the representativeness of the trial and its external validity, as demonstrated by several studies (Kahan, Rehal and Cro, 2015). Steg et al. (2007) found that hospital mortality was doubled in eligible non-enrolled acute myocardial infarction patients compared to trial participants (7.1% vs 3.6%) (2.07; 95% CI, 1.44-2.97), indicating that the outcomes observed in the trial may not fully represent broader population. Similarly, Fareed, Suri and Qureshi (2012) reported that ischemic stroke patients recruited in RCTs had different characteristics and clinical outcomes than those not recruited. African American or white race are more likely to be recruited in RCTs than others. The rate of intracranial haemorrhage (6% vs 2%, $p<0.05$) and progression of stroke (12% vs 3%, $p<0.05$) during hospitalisation were also higher in patients recruited in RCTs. Further, Lim et al., (2022) compared heart failure trials with registries and found that over half (56%) of registry patients met the trial eligibility criteria. While all-cause mortality rates were similar between trial and registry patients (0.97; 95%CI 0.92–1.03), cardiovascular mortality rates were higher in trial patients (1.19; 95%CI 1.12–1.27). However, this study was limited to only 5 trials and 2 registries, which may not fully capture all heart failure patients.

2.3.4 Psychiatry trials

Clinical trials for psychiatric disorders such as borderline personality disorder (BPD) and generalised anxiety disorder (GAD) often exclude patients with different psychiatric comorbidities due to strict eligibility criteria (Stoffers et al., 2010), even though a significant proportion of real-world patients have these

comorbidities such as mood and anxiety disorders and substance dependence (Sjastad et al., 2012). Thus, the representativeness of these trials is affected, compromising their generalisability. Few studies have focused on the generalisability of pharmacological and psychotherapy treatment trials for these disorders. Hoertel et al. (2015) and Hoertel et al. (2012) examined the representativeness of BPD and GAD trials by applying trial eligibility criteria to the target population in a disease registry (National Epidemiologic Survey on Alcohol and Related Conditions (NESARC)) to determine the percentage of patients who would have been excluded. They found that 76% of the 2,232 participants were excluded from BPD trials, and 75% of participants were excluded from both pharmacological and psychotherapy efficacy GAD trials. A history of bipolar disorder and recent history of alcohol abuse were major reasons for exclusion. Additionally, current depression was the single criterion excluding the largest proportion of patients in both groups in GAD trials (55.43% participants).

In addition, Blanco et al. (2017) and Hoertel et al. (2013) applied the same approach to assess trial representativeness for major depressive disorder (MDD) and bipolar disorder. Using data from the National Comorbidity Survey (NCD) and NESARC registries, they found that 61.9% and 58.2% of the target population with MDD and bipolar disorder, respectively, would have been excluded due to a single criterion: the risk of suicide. Similarly, 75.8% of patients with major depression and 63.8% of patients with post-traumatic stress disorder (PTSD) would have been excluded due to a single criterion (Blanco et al., 2008; Franco et al., 2016). Although these studies utilised large representative datasets and examined both pharmacological and psychotherapy trials, the analysis was limited to eligibility criteria and did not evaluate patient outcomes.

Overall, these findings indicate that restrictive criteria lead to under-representation of certain patient categories in clinical trials, potentially reducing the external validity of the trials. Therefore, there is a need for more justification and balance in exclusion criteria to ensure a representative population and generalisable results to real-world settings.

2.3.5 Type 2 diabetes trials

Dugard et al. (2024) explored the representativeness of type 2 diabetes trials by comparing the characteristics of general practice patients to the baseline characteristics of participants enrolled in 23 trials. They found that general practice patients were older (mean age 68.8 vs 59.9 years), had a higher BMI (31.5 vs 28.2 kg/m^2), and used more hypoglycaemics and antihypertensives drugs compared to patients randomised in clinical trials. These findings indicate that the trial participants may not fully represent the typical patients seen in general practice, potentially affecting the generalisability of trial results. However, this study relies on trials conducted between 1995 and 2012, which may not reflect current routine care practice.

2.3.6 Substance use disorder

Susukida et al. (2017) conducted a study to compare the treatment effects of 10 RCTs for substance use disorder (USD) (3,592 participants) with population effects (1,602,226 patients). The characteristics of trial participants were adjusted using propensity scores to resemble those of the target population. They then re-compute the effects from the included RCTs for three outcomes: retention, abstinence, and urine toxicology using statistical weighting. After weighting the samples, the statistically significant effect observed in the trials became non-significant. However, this study would only be able to adjust for measured confounding variables using the propensity scores, whereas the trial results would account for both measured and unmeasured confounding through randomisation. Therefore, the findings of this study could be, at least in part, due to unmeasured confounding, which might influence the study's conclusion about the generalisability of RCT results.

Furthermore, when trial exclusion criteria were applied on routine patients diagnosed with alcohol dependence, half (50.5%) were found to be ineligible (Blanco et al., 2008). Financial situation and lack of compliance and motivation were the main criteria excluding these individuals. Similarly, the results of

tightly controlled trials for cannabis dependence may not be useful to the community, as their results cannot be extrapolated (Okuda et al., 2010). These trials included a very selective sample because they used numerous exclusion criteria that excluded 80% of community samples with cannabis dependence.

2.3.7 HIV trials

Gandhi et al. (2005) conducted a study that aimed to assess the representativeness of HIV clinical trials. They employed the same exclusion criteria used in HIV trials within the AIDS clinical trial group (ACTG) and Community Programs for Clinical Research on AIDS (CPCRA) to the Women's Interagency HIV Study (WIHS), which is a large cohort study for HIV women in the US. The eligibility criteria of 32 HIV trials in the ACTG were applied to 1,717 WIHS participants. The results showed that approximately 67.6% WIHS participants would have been ineligible for the trials based on the ACTG eligibility criteria. This high exclusion percentage reflects the limited representativeness of HIV clinical trials, as it suggests that a substantial portion of WIHS participants would not meet the inclusion criteria. However, this study focused only on trials involving HIV-positive women and did not explore how these eligibility criteria would affect HIV- positive men.

2.3.8 Vitamin D trials

Eisenberg et al. (2015) compare excluded to included patients in the D-Vitamin Intervention in Veteran Administration (DIVA) trial and assessed trial design according to the effectiveness vs efficacy continuum using the PRECIS tool. A retrospective chart review was used to compare data on HbA1c, lipid profile (cholesterol, LDL, HDL, and triglycerides), AST, and ALT measures between excluded and included patients, and 15 members of endocrinology department completed the PRECIS assessment. After comparing 178 patients included in the DIVA trial to 178 patients excluded from the trial, no significant differences were found between the subjects for most parameters. Furthermore, the PRECIS scores indicated that the trial was generally pragmatic.

2.3.9 Nicotine dependence trials

In the context of smoking cessation trials, it has been found that the attempt to maximise the safety and efficacy of treatment is addressed by the restrictive exclusion criteria (Weisberg, Hayden and Pontes, 2009). However, this approach widens the gap between daily clinical practice and experimental research, diminishing the representativeness of clinical trials (Kennedy-Martin et al., 2015). Le Strat et al. (2011) investigated the representativeness of nicotine dependence trials by applying the eligibility criteria to 4,962 adults obtained from NESARC. They found that approximately 66% of adults were excluded because of one criterion, specifically smoking fewer than ten cigarettes per day. Similarly, over 47% of the daily-smoking population was excluded from varenicline trials when the common eligibility criteria were applied (Motschman et al., 2016). Key exclusion factors included lack of motivation to quit and comorbidities such as COPD or cardiovascular disorders. These findings show the lack of trial representativeness, which may challenge extrapolating trial results to a broader population. In contrast, Howard-Pitney et al. (2001) found no significant differences in characteristics of trial participants in chewing tobacco cessation trials and the target populations, indicating that the trial participants are representative of the general population.

2.3.10 Rheumatoid arthritis trials

The representativeness of rheumatoid arthritis trials might be uncertain as they often include only patients with active disease, whereas many patients in routine practice have less active disease (Kingsley et al., 2005). Vashisht et al. (2016) investigated this issue by applying the eligibility criteria of 30 RCTs for rheumatoid arthritis to two observational clinical cohorts: the Veterans Affairs Rheumatoid Arthritis (VARA) registry (1,523 patients) and the Rheumatology and Arthritis Investigational Network Database (RAIN-DB) (1,548 patients). They assessed patient eligibility based on disease activity, recorded through the joint activity score. The mean activity score in RCTs was 6.95, double that of VARA (3.87) and RAIN-DB (3.65). This difference raises questions about the clinical relevance of findings from these RCTs. The high activity scores observed in the

trials suggest that they may not be representative of the broader patient population, which typically exhibits lower disease activity. However, the study was limited to only two cohorts (RAIN-DB and VARA), which may not accurately reflect the global population of rheumatoid arthritis patients. In contrast, Koog, Lee and Wi (2015) conducted a systematic review of RCTs for knee osteoarthritis (KOA) to assess their external validity. They observed that for every three patients screened, two were randomised, indicating a good generalisability for KOA trials.

Table 2.2 Summary of the main findings from the literature review

Speciality	Study	Type of study	Assessment approach	Data sources	Main findings
Oncology	Karim et al. (2019)	Retrospective cohort study.	Assess the impact of eligibility criteria of an oncology trial on the real-world population. They applied common exclusion criteria on 125,316 patients from the Albert Cancer Registry (ACR) diagnosed with 11 different types of malignancy between 2004 and 2015.	Albert Cancer Registry (ACR).	Among the cohort of 125,316 cancer patients, 48,149 (38%) would have been ineligible for trials. The median overall survival was 47 months for trial-ineligible patients compared to 135 months for trial-eligible patients ($p < 0.001$). High number of real-world populations cannot participate in clinical trials because of strict exclusion criteria.
	Bijker et al. (2002)	Retrospective cohort study.	Assessed trial representativeness by examining the impact of patient selection in a EORTC trial on the applicability of its results to a larger population with DCIS and compared the outcomes between ineligible and randomised patients.	EORTC Trial.	Out of 910 patients treated for DCIS, 477 (52%) were ineligible, primarily due to lesion size, leaving 433 eligible patients. Of these, 278 were randomized into the trial. Main reasons for non-entry of eligible patients included physician preference and patient refusal. At a four-year follow-up, non-entered patients who received local excision and radiotherapy had higher local recurrence rates than randomized patients.

					There were differences in outcome, and the trial results may not be applicable to target population.
	Sam et al. (2018)	Retrospective cohort study.	Reviewed unresectable melanoma patients from 2011−2014 and analysed the eligibility criteria for Vemurafenib (Vem) and Ipilimumab (Ipi) trials to assess the applicability of their findings.	Vem and Ipi Trials.	High number of ineligible patients has been found. Of the 290 patients in the Vem cohort, 49 were trial-eligible, and 36 received treatment, showing improved overall survival (HR 0.68, 95%CI 0.44–0.98) compared to ineligible patients. In the Ipi cohort, 119 out of 212 patients were trial-eligible, with 43 receiving treatment, also showing improved survival.
	Costa, Hari and Kumar (2016)	Retrospective cohort study.	Compare the similarity and differences between trial participants and target populations. They included 128 published trials for multiple myeloma (MM) in the United States (US) with a total of 8,869 participants and compared them to samples from Surveillance Epidemiology and End Results (SEER).	SEER and multiple myeloma trials.	There was under-representation of older individuals in the trials and differences in age, trial participants were younger (median age 61) than target populations younger (median age 69). The trial included patients who seem to have lower risk disease compared to unselected patients.

	Elting et al. (2006)		Examined the generalisability of cancer trials by evaluating the comparability of trial participants and non-participants. They used a cohort of 62,562 patients newly diagnosed with cancer at the University of Texas M. D. Anderson Cancer Centre during the period 1990–1997. They compared the characteristics of trial participants with the characteristics in the SEER population.	SEER and University of Texas M. D. Anderson Cancer Centre.	There were significant differences among participants in terms of marital status, gender, and race, since there were fewer African American participants than white and Hispanic participants. Although participants had more progressive cancer, their health was better than that of non-participants. Moreover, participants had fewer comorbidities, were younger, and had better performance.
	Andersson et al. (2019)	Comparative analysis.	Compare SENOMIC trial participants (548 patients) with 1,070 cases reported to the Swedish National Breast Cancer Register (NKBC).	SENOMIC trial and NKBC.	No significant differences were found in tumour characteristics, age, and breast surgery. Participants of SENOMIC trial were representative of the target population in NKBC.
	Mishkin et al. (2016)	Retrospective cohort study.	Compare the demographics of RCTs sponsored by the National Cancer Institute (NCI) for gynaecological malignancies to a target US population. They reviewed cervical, uterine, and ovarian cancer trials	NCI-sponsored gynaecological cancer trials and NCI's SEER program.	There are discrepancies between NCI participants and the target population, which may reduce the applicability of NCI trials.

			conducted between 2003 and 2012. The study included 18,913 participants from 156 trials, with 56% for ovarian, 32% for uterine, and 12% for cervical cancer.		
	Tam et al. (2009)	Retrospective cohort study.	Compare adverse events rates observed in the patient charts with those reported in the largest phase III clinical trials.	Retrospective review of patient charts from the Juravinski Cancer Centre (JCC).	No significant difference in the toxicity rates among non-trial patients (21%) and trial participants (31%) (P = 0.10).
Cardiology	Steg et al. (2007)	Retrospective cohort study.	Compare eligible patients not enrolled in RCTs to ineligible patients and participating patients. The sample size was 8,469 divided as follows: 953 RCT participants, 4,669 eligible non-enrolled patients, and 2,847 ineligible patients for RCT. The main outcome of this study was the rate of hospital mortality.	Global Registry of Acute Coronary Events (GRACE).	Ineligible patients had higher mortality rate than eligible non-enrolled patients and trial participants (11.4%, 7.1%, and 3.6%, respectively) (P<001). All three groups had AMI but were entirely different in terms of their baseline characteristics, outcomes, and treatment.
	Fareed, Suri and Qureshi (2012)	Retrospective cohort study.	Compare the characteristics of ischemic stroke participants participated in RCTs with those not participated and their effects on outcomes and generalisability.	University Health Systems Consortium (UHC) benchmarking	The characteristics of patients recruited in RCTs are different from those not recruited which affect trial representativeness.

			Data was collected for 1,256 patients with a mean age of 67 years. Of these, 77 patients were recruited in RCTs and 1,179 were not recruited.	project conducted in 2005.	
	Lim et al. (2022)	Comparative analysis.	Compared individual patient data (IPD) of five HF RCTs and two HF registries.	CHECK-HF, the SwedeHF registries, and 5 HF RCTs.	A total of 26,104 (56%) registry patients fulfilled the eligibility criteria. Trial patients were younger (mean 63.6 years vs. 72.7 years) and less frequently women (22% vs. 31%). Clinical outcomes were similar between trial patients and registry patients.
Psychiatry	Hoertel et al. (2015)	Cross-sectional.	Examined trial representativeness by applying trial eligibility criteria to target population in disease registry to explore the number of patients who would be ineligible.	NESARC for the period 2004−2005.	Restrictive eligibility criteria were responsible for excluding most of the target population with borderline personality disorder.
	Hoertel et al. (2012)	Cross-sectional.	Applied same method as above on GAD trials.	NESARC for the period 2001−2002.	Restrictive eligibility criteria were responsible for excluding the majority of the target population with generalised anxiety disorder.
	Blanco et al. (2017)	Cross-sectional.	Applied same method as above on MDD trials.	NCD for the period 2001−2004.	There is lack of representativeness for both pharmacological and psychotherapy trials, as the trials

					tend to include ideal patients rather than typical patients seen in practice.
	Hoertel et al. (2013)	Cross-sectional.	Applied same method as above on bipolar disorder trials.	NESARC for the period 2001−2002.	Majority of patients would have been excluded because of one criterion, which may limit the applicability of clinical trials.
	Franco et al. (2016)	Cross-sectional.	Applied same method as above on PTSD trials.	NESARC.	Same findings as above.
	Blanco et al. (2008)	Cross-sectional.	Applied same method as above on major depression trials.	NESARC.	Same findings as above.
Substance use disorder	Susukida et al. (2017)	Comparative analysis.	Assessed trial generalisability by re-compute the effects from the included RCTs.	10 RCTs of USD and Treatment Episodes Data Set-Admissions (TEDS-A) between 2001 and 2009.	The finding indicates that the trial results cannot be easily generalised to the target population.
	Blanco et al. (2008)	Cross-sectional.	They examined the impact of eligibility criteria on trial representativeness by applying a set of eligibility criteria for RCTs on routine patients with alcohol use disorder.	NESARC.	Half of the routine patients were excluded due to a single criterion.

	Okuda et al. (2010)	Cross-sectional.	Applied same method as above on cannabis dependence trials.	NESARC.	Eligibility criteria excluded 80% of community samples with cannabis dependence from clinical trials.
HIV trials	Gandhi et al. (2005)	Comparative observational study.	Applied same method as above on HIV trials.	32 major HIV RCTs and the Women's Interagency HIV Study (WIHS).	Around 67.6% of participants in the WIHS would have been excluded from the trial if the ACTG eligibility criteria had been applied.
Vitamin D trials	Eisenberg et al. (2015)	Comparative analysis.	Compare excluded to included patients in the-Vitamin D trial and assess trial design using the PRECIS tool.	DIVA.	No significant difference was found between patients, and the trial was generally pragmatic.
Nicotine dependence trials	Le Strat et al. (2011)	Cross-sectional.	The eligibility criteria of nicotine dependence trials were applied to a sample of target population.	NESARC.	Approximately 65.89% of adults were excluded because of one criterion.
	Howard-Pitney et al. (2001)	Comparative analysis.	Compare characteristics of trial participants (n=401) to sample of the general population (n=155).	Chewing tobacco cessation trials.	No significant difference was found between both groups.
	Motschman et al. (2016)	Systematic review.	Eligibility criteria and participants' characteristics of 32 varenicline trials were compared to national representative data.	National Health Interview Survey (NHIS) and National Survey on Drug Use and Health (NSDUH).	The application of this type of eligibility criteria was responsible to exclude more than 47% of patients.

Rheumatoid arthritis trials	Vashisht et al. (2016)	Cross-sectional.	Applied the eligibility criteria of 30 RCTs for rheumatoid arthritis to target population in registry.	VARA and RAIN-DB.	The clinical trials for rheumatoid arthritis lack representation to the population, as most patients do not meet the trial participation criteria.
	Koog, Lee and Wi (2015)	Systematic review.	Systematically reviewed and analysed 352 RCTs for knee osteoarthritis (KOA) to examine their external validity.	KOA RCTs.	They found that for every three patients screened, two were randomised.

2.4 Summary of the literature review

This chapter reviewed several studies that investigated the representativeness of clinical trials and provided an overview of how this impacted their generalisability. The literature showed that strict exclusion criteria prevented a significant number of real-world patients from participating in trials, potentially limiting the generalisability of the results. Moreover, many studies have noted that the characteristics of patients recruited in RCTs often differ from those not recruited, although this is not universally observed across all studies reviewed. These differences might reduce the representativeness of the trial and raise questions about the generalisability of their findings. Trialists need to balance external and internal validity by loosening the restrictiveness of eligibility criteria to include a more diverse population, leading to more representative and applicable trials. In this thesis I will examine the usefulness of using SAEs as tool to assess trial representativeness. It will focus on clinical trials for type 2 diabetes mellitus, particularly RCTs for SGLT-2 inhibitors.

Chapter 3 Rationale and outline of thesis

3.1 Chapter overview

Previous chapters (Introduction and literature review) highlighted issues in clinical trials regarding representativeness and their generalisability. In this chapter, I will present the approaches that decision-makers currently use to assess trial representativeness. I will also introduce the thesis research question and explain how I propose to examine the SAEs as a metric. Accordingly, I will outline the chapters conducted in this thesis to address the research question.

3.2 Current approach used as a metric for trials representativeness

Representativeness is a key part of generalisability; however, assessment of trial representativeness is a complex and challenging task. Despite this, some studies have assessed representativeness by comparing the trial population to the real-world population. This method compares baseline characteristics of trial participants, such as age, sex, and race/ethnicity to baseline characteristics of the target population in registries or routine healthcare data. Although this approach assesses representativeness by identifying underrepresented subgroups, trial descriptions may not capture all relevant characteristics of trial participants (Hanlon et al., 2022). For example, clinical characteristics (e.g., multimorbidity, disease severity), concomitant medication use, and lifestyle factors (e.g., exercise, diet) influence trial representativeness and are rarely included in such descriptions, thus they may not be captured through this approach (Leinonen et al., 2015).

Other studies have assessed trial representativeness by applying trial exclusion criteria to population samples obtained from disease registries or routine health-care data (Karim et al., 2019). Trials with restrictive eligibility criteria may exclude specific populations due to medical comorbidities, age, and concomitant

medication use (Beaver et al., 2017). While high restrictive criteria contribute to high internal validity by creating a homogeneous study population and controlling for potential confounding factors, they are associated with less representative samples that may potentially limit the generalisability of trial findings to the real-world population. Applying trial exclusion criteria on routine care patients can determine the proportion of people who would be ineligible for trials. However, assessing trial representativeness using this method is also challenging because eligibility criteria may not always determine trial representativeness; eligible patients may sometime refuse to participate in a trial due to concerns about potential risks or may be withdrawn during the trial (Kim et al., 2022; Hillman et al., 2023). Moreover, health outcomes in the trial population compared to routine patients cannot be directly assessed through this approach (Averitt et al., 2020).

An additional method is the Pragmatic Explanatory Continuum Indicator Summary (PRECIS-2) tool, which is developed to assess the pragmatism of clinical trials (Additional details on trial pragmatism and its assessment can be found in Chapter 6). Pragmatism focuses on the extent to which the trial design reflects real-world practice (Smelt et al., 2010), while representativeness refers to how well the trial population mirrors the target population in routine care for whom the intervention is intended (He et al., 2020). Although these concepts are different, both relate to a trial's applicability to everyday clinical settings. Therefore, the PRECIS-2 tool may indirectly reflect representativeness and contribute effectively to standard care settings (Usman et al., 2022). The PRECIS-2 tool assesses trials across nine domains, assigning scores ranging from 1 (least pragmatic) to 5 (most pragmatic). These scores help determine the degree to which trials align with usual care conditions. Trials scoring high are more likely to have inclusive eligibility criteria, and their protocols for intervention delivery and adherence reflect real-world scenarios. The most pragmatic trials exhibit trial designs and outcomes that reflect routine practice and included participants who closely represent real-world demographics. In contrast, trials scored 1, indicating a more explanatory nature, are less likely to be representative of real-world practice (Loudon et al., 2015). This approach offers a systematic framework that permits the assessment of representativeness, enabling the identification of trials that better reflect the broader population

and allow for the extrapolation of trial findings to clinical practices. However, this tool has limitations when used retrospectively to assess published trials. Lack of information in published protocols for scoring some domains may lead to inaccurate or incomplete assessment. Moreover, the subjectivity introduced when scoring domains may lead to biased assessments.

Overall, while these methods can be employed to assess trial representativeness, their limitations underscore the need for alternative methods. Serious adverse events represent an alternative approach that may be considered as a novel metric to assess trial representativeness, which will be explored in this thesis.

3.3 Serious adverse events (SAEs)

3.3.1 Overview

During the investigation of a new intervention, it is expected that SAEs may occur. The occurrence of SAEs, by definition, is not necessarily caused by the intervention being studied, and trials are required to report all SAEs regardless of cause (whether or not they are suspected to be related to the treatment under investigation). The purpose of reporting SAE data in clinical trials is to ensure the safety of trial participants (Arnaud-Coffin et al., 2019). SAEs are any undesirable events that result in death, cause or increase hospitalisation, lead to persistent or significant disability, or cause birth defects. Among these, hospitalisation and death are commonly reported SAEs in clinical trials (FDA, 2023). Moreover, trial reports routinely include SAEs, primarily accounting for all-cause hospitalisations and deaths.

3.3.2 SAE as promising alternative measure of trial representativeness

SAEs can provide insight into the health status of trial participants, including aspects such as disease severity and comorbid conditions, which may contribute to understanding the broader applicability of trial results (Sheppard et al., 2020). From first principles, it can be expected that SAEs would naturally be higher in populations that are sicker or have more severe underlying health conditions because they include by definition all-cause hospitalisation and deaths (Panagioti et al., 2015). Because trials are often criticised for excluding older people and those with more comorbidities, more representative trial would likely include sicker individuals. Thus, if trials are representative, they would be expected to report a higher incidence of SAEs than trials with more restrictive criteria (Hanlon et al., 2021). This expectation arises because these populations are generally more susceptible to SAEs. Consequently, this supports the use of SAEs as a tool for assessing trial representativeness, particularly when other measures are lacking (Hanlon et al., 2019).

If the trial population is representative of the target population, it is expected that the SAEs in the trial would be similar to the SAEs within the target population. Hanlon et al. (2022) previously showed that observed SAEs in a trial population are often significantly lower than SAEs in the target population in routine care. This difference indicates that trial participants were healthier and younger than the target population, raising concerns about the representativeness of clinical trials. It also highlights the potential feasibility of using SAEs to capture the representativeness of trial participants, SAEs may therefore be explored as a promising alternative metric to assess trial representativeness.

3.4 Thesis research question and justification

The research question that will be answered in this thesis is:

Can we use serious adverse event rates in clinical trials as a measure of trial representativeness?

SAEs offer advantages over other methods; they are an objective measure as clearly defined by the FDA. Also, they are well-documented, as trial sponsors are obligated to report all SAEs for both treatment and control arms, regardless of whether they are related to the intervention or not. Therefore, this thesis aims to explore the feasibility of using SAEs as a metric to assess trial representativeness. However, there might be challenges in studying SAE rates as metric. First, it is unclear whether trials of SGLT-2 inhibitors report sufficient information about SAEs and their timeframes to allow for the calculation of SAE rates. Consequently, I will explore the feasibility of capturing SAE rates from clinical trials. Second, it is not known whether the treatment arm needs to be considered separately. This consideration is important because if the treatment (SGLT-2) influences the SAE rate, then the SAE rates would differ between arms, and it may not be valid to compare the overall trial rate to some (untreated) target population. However, selecting only a single arm reduces statistical power. Therefore, I will compare the difference in SAE rates between trial arms to determine whether to consider each arm separately or use the total for both arms in the analysis when examining the SAE rates. Lastly, it is not known whether SAEs are related to representativeness, and there is no gold standard metric to compare with. Thus, I will use triangulation approach to explore whether SAE rates reflect trial representativeness by i) using a trial pragmatism metric (PRECIS-2 tool) to compare with SAE, ii) comparing SAE rates and PRECIS-2 score to trial and baseline characteristics (as a fair umpire), and iii) comparing SAE rates to eligibility criteria. I will use T2DM condition and clinical trials of SGLT-2 inhibitors as an exemplar, with the rationale explained in detail in section 3.5.

3.5 Rationale for choosing T2DM and trials of SGLT-2 inhibitors as an example

The interventions in the included trials are generally safer compared to more toxic intervention (e.g., chemotherapy). These trials were selected from a recent systematic review of novel antidiabetics (Butterly et al., 2022). Type 2 diabetes is a global and chronic health condition and is the ninth leading cause of mortality (Galicia-Garcia et al., 2020; Whicher et al., 2020). Approximately 462 million individuals, accounting for around 6.28% of the world's population, are affected by T2DM (Abdul et al., 2020). Furthermore, diabetic patients are commonly seen in primary care settings, constituting 65.8% of primary care visits in the US, which contributes to substantial health and economic burdens (Pilla, Segal and Maruthur, 2019). Therefore, given this prevalence and burden of T2DM, clinical trials of antidiabetics need to be highly representative of routine care.

Additionally, four large cardiovascular outcome trials have found that SGLT-2 inhibitors reduce the risk of cardiovascular events and mortality rate, particularly in T2DM patients with cardiovascular disease at baseline (Zelniker et al., 2019). Accordingly, clinical guidelines and treatment strategies have evolved, and the European Society of Cardiology now recommends SGLT-2 inhibitors as a first-line drug in patients with T2DM and CV diseases (Cosentino et al., 2020). However, the strict enrolment criteria vary among trials, which may impact their representativeness and limit generalisability (Birkeland et al., 2019). This uncertainty may, in turn, affect the clinical guidelines implemented in routine care. Given that the care of patients with T2DM is an integral part of everyday clinical practice, and the relevance of SGLT-2 trials extends beyond diabetes management, T2DM and SGLT-2 trials have been chosen as an example to explore the feasibility of using SAE rate as a metric to assess trial representativeness.

3.6 Outline of thesis

Chapter 1: Introduction

This chapter provides a brief introduction to clinical trials and the selection process of participants and defines trial representativeness and trial applicability. The representativeness of clinical trials, their importance, and how they often compromised are also discussed. The chapter finally presents the aim and objectives of this thesis.

Chapter 2: Literature review. Generalisability of clinical trials across different conditions.

This chapter explores how the lack of representativeness impacts the generalisability of clinical trials across different conditions, including cardiovascular disease, oncology, and psychiatry. It summarises the factors contributing to the lack of representativeness, including the role of strict inclusion and exclusion criteria, the exclusion of older people, and the underrepresentation of sicker people.

Chapter 3: Rationale and outline of thesis

This chapter provides the research question and the rationale for conducting this thesis. It outlines the approaches that decisions-makers currently use to assess trial representativeness. It also presents an overview of SAEs as a promising alternative measure. Finally, it outlines the chapters for the current thesis.

Chapter 4: The feasibility of capturing SAE rates from RCTs of SGLT-2 inhibitors.

This chapter aims to determine whether clinical trials of SGLT-2 inhibitors report the necessary information to calculate SAE rates, including the number of trial participants who experienced SAEs, the number of participants at risk of SAEs, and the timeframe of SAEs (the specific period during which SAEs occurred). It also explores the consistency in the reporting of SAEs (i.e., reporting MACE within SAE reporting). Additionally, this chapter describes the process of extracting these data and the sources from which SAE data were obtained.

Chapter 5: Difference in SAE rates between intervention and control arms of SGLT-2 inhibitors.

This chapter examines the difference in SAE rates between intervention and control arms, explaining why this analysis was conducted. It determines whether to consider each arm separately or to use the total for both arms in the analysis for the chapters that follow.

Chapter 6: Assessment of the pragmatism of RCTs of SGLT-2 inhibitors: using the pragmatism metric PRECIS-2 tools

Because there is no gold standard for assessing trial representativeness, the PRECIS-2 tool was used in this chapter to compare the tool with the SAE rates. This tool assesses the pragmatism of the included trials and distinguishes pragmatic trials undertaken in usual care from explanatory trials conducted in an idealised setting. The chapter details the assessment criteria used for assessing the included trials and scoring the nine PRECIS-2 domains. It presents the distribution of PRECIS-2 score across the trials and examines the association of SAE rates with the mean PRECIS-2 score and each domain.

Chapter 7: Comparison between SAEs and PRECIS-2 tool by exploring the difference in their association with trial and baseline characteristics of SGLT-2 inhibitors.

This chapter compares the SAE and PRECIS-2 tool to trial and baseline characteristics. It presents the umpire test used in the absence of the gold standard and provides justification for each umpire used. It also presents the associations of SAE rates and PRECIS-2 score against trial and baseline characteristics.

Chapter 8: Association between eligibility criteria and SAE rates in clinical trials for SGLT-2 inhibitors.

This chapter further explores the feasibility of SAE rates as a metric by comparing them with the trial eligibility criteria. It explains the expected relationship between SAE rates and the type of eligibility criteria (restrictive vs permissive). The chapter describes the methods applied to measure the eligibility criteria and presents their distribution. It also explores and presents the associations between eligibility criteria and SAE rates.

Chapter 9: Discussion

This final chapter summarises the main findings and contributions of the thesis. It also discusses the limitations, strengths, implications, and future research. It finally presents the overall conclusions derived from the conducted thesis.

Chapter 4 The feasibility of capturing SAE rates from RCTs of SGLT-2 inhibitors: coverage of reporting

4.1 Chapter overview

Before examining the usefulness of SAE rates as a metric for trial representativeness, this chapter explores SAE reporting and determines whether clinical trials report sufficient information to calculate SAE rates within RCTs of SGLT-2 inhibitors. It also demonstrates the process and sources for extracting the required data to calculate the SAE rates, and it illustrates the source of the included trials. Additionally, it explores the consistency in reporting major adverse cardiovascular events (MACE) outcomes within SAE reporting.

4.2 Background

Assessing trial representativeness is essential for clinicians who rely on trial findings. Some literature has suggested using SAEs as a metric to assess trial representativeness (Hanlon et al., 2021). However, it is unknown whether clinical trials of SGLT-2 inhibitors adequately report SAEs. Trial sponsors are obligated to report all SAEs, but the specific requirements for reporting may vary depending on the regulatory authorities. For example, the FDA requires that SAE be reported within specific timeframes, whereas the European medicines agency (EMA) does not specify exact timelines (FDA, 2021). Adhering to the standard classification of SAEs helps maintain uniformity and accuracy in reporting these events, while also minimising the risk of misinterpretation (Gliklich et al., 2014).

To determine the usefulness of SAE rates as a metric of trial representativeness, it is crucial to prioritise the sufficient reporting of SAEs. The level of SAEs reporting would be considered adequate if the number of SAEs for each trial arm is reported. Failing to report this information adequately can hinder the calculation of SAE rates, thus, compromising the feasibility of using SAE rates as

a metric. In this chapter, I will explore the feasibility of capturing SAE rates using aggregated data from completed RCTs, focusing on whether these reports provide sufficient information to allow for the calculation of SAE rates.

4.3 Aim and objectives

In this chapter, in order to explore whether SAE rates can be used as a metric for trial representativeness, I will explore the feasibility of capturing SAE rates and determine whether clinical trials of SGLT-2 inhibitors report sufficient information about SAEs in each arm to calculate the SAE rates that will be used in this thesis. I will also compare the level of reporting between ClinicalTrials.gov and other trial registries. Additionally, I will explore the consistency of SAE reporting by determining whether MACE count was included within the reported SAE count.

4.4 Methods

4.4.1 Study selection

The selection of studies was obtained from a recent systematic review (Butterly et al., 2022). This systematic review encompasses various classes of novel antidiabetics, and specifically, 146 clinical trials of SGLT-2 inhibitors have been identified and selected for inclusion in this thesis.

Description of systematic review

The review identifies phase 3 and 4 RCTs of SGLT-2 inhibitors, GLP-1 RA, and DPP-4 inhibitors. It aims to compare the efficacy of these novel antidiabetics on Glycated Haemoglobin (HbA1c), body weight, and cardiovascular events in patients with type 2 diabetes mellitus by applying meta-analysis and calibrating to representative population from Scottish diabetes register.

Eligibility criteria (For the systematic review)

Inclusion criteria:

- Presence of type 2 diabetes mellitus.
- Phase III or IV RCTs.
- Age 18 years and older.
- Any country.
- Any subgroup population will be included e.g., patients with co-existing comorbidities.
- Any class of novel antidiabetics.

Exclusion criteria:

- Any other types of diabetes mellitus.
- Phase I and II RCTs.
- Diagnosis of pre-diabetes.
- People at risk of type 2 diabetes mellitus but not currently diagnosed.
- Other class of novel antidiabetics.

Interventions (for trials included in this thesis)

The interventional trials involved novel antidiabetic drugs of SGLT-2 inhibitors, including canagliflozin, dapagliflozin, empagliflozin, ertugliflozin, ipragliflozin, luseogliflozin, bexagliflozin, tofogliflozin, and sotagliflozin.

Comparators (for trials included in this thesis)

Comparators were either placebo alone, active medication alone (other antidiabetic medications), or both active and placebo.

4.4.2 Data sources

Data sources for the 146 trials varied depending on the type of trial registration and the availability of required data within trial registries. Data were obtained from ClinicalTrials.gov, University Hospital Medical Information Network (UMIN) Clinical Trial Registry, Japan Pharmaceutical Information Center (Japic) Clinical Trials Information, and Thai Clinical Trials Registry (TCTR), as well as from trial-relevant publications.

4.4.2.1 ClinicalTrials.gov registry

ClinicalTrials.gov served as the primary data source for this study. It is an extensive public database that provides information on both publicly and privately supported clinical trials across various diseases and conditions. This platform offers access to summary information about study design, aims, intervention details, eligibility criteria, and study results, including efficacy and safety data, particularly for all SAEs. The provided trial information is for completed and ongoing clinical trials. For trials registered to the US Clinical Trial registry at ClinicalTrials.gov, data was accessed via the Aggregate Analysis of ClinicalTrials.gov (AACT) database. This is a publicly available relational database that contains all information about studies registered in ClinicalTrials.gov. For some trials that did not post their results on ClinicalTrials.gov, required information was obtained from published research papers or study documents [e.g., Clinical Study Reports (CSR)]. The number of trials for which data were obtained from AACT or from published research papers is presented in the results (section 4.5.1).

4.4.2.2 Other trial registries and publications

For trials not registered on ClinicalTrials.gov, data were obtained from other trial registries such as UMIN, Japic, and TCTR. In case where some information was not available within these registries, the required data were collected from trial-relevant publications and CSR (detail in section 4.4.3).

4.4.3 Data extraction and harmonisation

Structured data extraction forms were used to collect information from the included studies. Data extraction for SAEs involved capturing the number of participants who experienced SAEs in each trial arm, the number of participants at risk of SAEs, and timeframe of these events. Additionally, the extracted data included trial registration identifier, type of trial registries, and arm names.

I conducted data extraction for reported SAEs in three steps. First, for trials registered on ClinicalTrials.gov, I extracted the SAE data for all trial arms from the total reported event tables in the AACT database. Second. for trials registered on registries other than ClinicalTrials.gov (e.g. UMIN, Japic, and TCTR), I manually extracted the SAE data from their registries using Excel spread sheets. Third, for trials that did not post their results for reported SAE data on ClinicalTrials.gov or other registries, I downloaded trial-relevant publications and manually extracted the SAE data from reported safety tables within the papers using Excel. Additionally, I explored the consistency in the reported SAEs by reviewing the number of MACE events to check if they were included within the SAE reports or not.

Furthermore, I initially extracted timeframes for the reported SAEs as text from the AACT database and then harmonised them into numeric values. I extracted the timeframe data for these SAEs from the reported event tables in the AACT database. However, some trials did not specify timeframes in their reported event tables. For these trials, I first extracted the timeframes from the result group table. If timeframes were also not specified in the result group table, I then extracted them from the primary outcome tables (sources for timeframes are presented in table 4.2 in results). For trials that did not post their results or were not registered on ClinicalTrials.gov, I manually extracted the timeframes from the published research papers. Furthermore, timeframes were reported in different formats such as days, weeks, months, or years across different data sources. I harmonised these timeframes into a standardised unit by converting them to years. The type of trial registry was classified as either registered on ClinicalTrials.gov or registered on other registries.

4.4.4 Statistical analysis

Descriptive statistics (counts and percentages) was conducted to summarise the reported SAEs across all the trials. This includes the number of trials reported SAEs, the number of participants for whom SAEs were reported in each trial, the number of participants at risk of SAEs, and the timeframe for which the SAEs were reported.

R software (R-4.1.2) was used to perform all data analysis and presentation for this thesis. Various packages within R software were employed to generate summary statistics, tables, charts, and regression analyses.

4.5 Results

4.5.1 Description and source of included studies

A total of 146 trials investigating SGLT-2 inhibitors were identified and included in this thesis, involving 116,207 participants. The distribution of trials was as follows: 51 trials for dapagliflozin (35%), 25 for empagliflozin (16%), 26 for canagliflozin (17%), 15 for ipragliflozin (10%), 9 for ertugliflozin (6%), 5 for tofogliflozin (3%), 3 for luseogliflozin (2%), 5 for bexagliflozin (3%), and 7 trials for sotagliflozin (5%) (Figure S.1). These trials were conducted in a variety of healthcare settings and countries. The included trials compared the safety and efficacy of SGLT-2 inhibitors either to placebo or active medication.

Out of the 146 trials, 120 (82%) were registered on ClinicalTrials.gov, while the remaining 26 trials (18%) were registered on other registries, including 17 on UMIN (11%), 4 on Japic (3%), 1 on TCTR (1%), and 4 trials were unregistered (3%) (Table 4.1). Of the 120 trials registered on ClinicalTrials.gov, 99 posted their results on ClinicalTrials.gov, and their data were extracted from the AACT database. The remaining 21 trials did not post their results on ClinicalTrials.gov; their data were manually extracted from publications. Additionally, for the 26

trials registered on other registries, all data were manually extracted from trial-relevant publications (Table 4.1).

Regarding timeframe reporting for trials registered on ClinicalTrials.gov, 87 trials reported the timeframes for SAE in the reported event tables within the AACT database, while 12 trials did not provide them in these tables. For these trials, timeframes were captured from the result group tables for 8 trials and from the primary outcome tables for 4 trials. For trials that did not post their results on ClinicalTrials.gov, and for trials registered on other registries, timeframes were captured from reported safety tables in corresponding journal articles (Table 4.2).

Table 4.1 Types of trial registers and data sources

Clinical trial registration	Number of trials	Data sources
ClinicalTrials.gov (Posted their results on ctgov)	99 (68%)	AACT
ClinicalTrials.gov (Not posted their results on ctgov)	21 (14%)	Publications and CSR
UMIN	17 (11%)	Trial registry and Publications
Japic	4 (3%)	Trial registry and Publications
TCTR	1 (1%)	Trial registry and Publications
Unregistered	4 (3%)	Publications

Table 4.2 Sources of timeframe for the reported SAEs.

Sources for timeframe	No. of Trials
Adverse events table	87 (59.6%)
Result Group table	8 (5.5%)
Primary outcome table	4 (2.7%)
Corresponding journal articles (for trials on other registries)	47 (32.2%)
Total	146 (100%)

4.5.2 SAEs reporting in clinical trials of SGLT-2 inhibitors

The feasibility of capturing SAE rates was determined by exploring the level of SAEs reporting in trials of SGLT-2 inhibitors, including a comparison of reporting between trials registered on ClinicalTrials.gov and those on other registries. SAEs were reported in all trials (100%) registered on ClinicalTrials.gov, indicating sufficient reporting that allows for the calculation of SAE rates. Among trials registered on other registry, only 5 trials (19%) did not reported SAEs (Table 4.3). As shown in table 4.4, all trials reporting MACE outcomes included MACE counts in the SAE reporting table (Table 4.4).

Table 4.3 Summary of trial registries and SAEs reporting on each registry.

SAE reporting	ClinicalTrials.gov	Other registry/Unregistered	Total
No. of trials reported SAEs	120 (100%)	21 (81%)	141 (97%)
No. of trials not reported SAEs	0	5 (19%)	5 (3%)
No. of participants experienced SAEs	21,026 (99%)	115 (1%)	21,141 (100%)
No. of persons at risk of SAEs	112,717 (97%)	3490 (3%)	116,207 (100%)

Table 4.4 Reporting of MACE count within the reporting of SAEs

Trial ID	MACE reported	Subject at risk for MACE	SAE reported	Subject at risk for SAE	Reported MACE within SAE	Results posted on ctgov
NCT01131676	772	7020	2777	7020	YES	YES
NCT01032629	1214	4330	1791	4327	YES	YES
NCT01989754	601	5812	1489	5807	YES	YES
NCT02065791	598	4401	1543	4397	YES	YES
NCT01730534	1559	17160	6623	17143	YES	YES
NCT02597049	5	423	13	423	YES	YES
NCT01986881	325	8238	3010	9790	YES	YES

4.6 Discussion

4.6.1 Summary of main findings

The analysis aimed to assess whether trials of SGLT-2 inhibitors report sufficient information to capture SAE rates, and the extent to which SAEs were reported in each registry category. A total of 141 (97%) trials provided the required information on reported SAEs, leaving 5 (3%) trials that did not report SAEs. All trials registered on ClinicalTrials.gov reported SAEs, and 81% of trials registered on other registries reported these events in their relevant publications.

4.6.2 Interpretation

The findings reveal that trials registered on ClinicalTrials.gov demonstrate a high level of reporting based on the number of SAEs reported for each arm, as all trials (100%) reported the SAEs. Conversely, some unregistered trials on ClinicalTrials.gov (19%) did not report SAEs. Trial-level factors, including

differences in reporting practices and potential underreporting of SAEs by trial sponsors, may contribute to this variation. This may also indicate that registration on ClinicalTrials.gov is a marker of sponsors adhering to the required regulations for FDA approval, which includes SAE reporting, thereby explaining the correlation between ClinicalTrials.gov registration and SAE reporting.

Additionally, only 60% of the trials reported the SAE timeframe in the reported event tables. For the other trials, timeframes were captured from either the result group tables or the primary outcome tables. However, this method is not perfect because the timeframe for SAEs may be longer than that of the result group or primary outcome. Nevertheless, it was the only method available to calculate the SAE rates.

Furthermore, the results provide reassurance regarding the consistency of SAEs reporting in trials with a hard serious outcome, particularly in the context of reporting cardiovascular events. Specifically, all trials reported hard outcomes (MACE), consistently included them in their reporting of SAEs, indicating a consistent and reliable reporting. Such consistency is crucial, as inconsistent reporting of MACE may lead to biased results by showing lower SAE rates, potentially impacts the implication of SAE rates as a metric of trial representativeness.

4.6.3 Comparison with other studies

Several studies examined the reporting of safety data in clinical trials across various diseases and interventions and found different level of reporting. A previous study examined SAE reporting for 156 trials studying 52 drugs registered on ClinicalTrials.gov and found that all trials (100%) reported SAE on ClinicalTrials.gov (Chen et al., 2022). Similarly, other study found that 199 (99%) of 202 trials registered on ClinicalTrials.gov reported SAE (Riveros et al., 2013). However, Yao et al. (2021) examined SAEs reporting in 160 publications on cancer trials and found that only 41 articles reported the SAEs. Additionally, higher SAEs reporting was observed in trials sponsored by pharmaceutical

companies compared to those sponsored by investigators. Reporting of SAEs in high-impact journals (e.g. New England Journal of Medicine) was higher than in other journals. Moreover, another study assessed the completeness of safety reporting and found that only 39% of 192 trials from 7 different medical areas reported the safety data adequately, indicating that reporting varied across medical areas and settings (Ioannidis and Lau, 2001).

4.6.4 Implications

The findings of this chapter have implications for clinicians, developers of reporting guidelines, and researchers. It highlights the importance of ClinicalTrials.gov registry in improving transparency in clinical research, as it adheres to stricter reporting standards, thus enhancing the accuracy and proper reporting for SAEs (Riveros et al., 2013). ClinicalTrials.gov conducts quality control reviews to ensure that the information submitted is complete and accurate (Tse et al., 2018). This contributes to a reliable database for researchers and healthcare professionals, facilitating evidence-based decision-making (Wieseler et al., 2012). On the other hand, other registries may be subject to varying reporting practices, potentially leading to underreporting or incomplete information regarding SAEs. SAE reporting is a regulatory requirement in trials, and investigators must inform sponsors and regulatory bodies of any events that meets the FDA criteria for SAE. Overall, the information on SAEs was reported for most trials of SGLT-2 inhibitors, making it feasible to calculate the SAE rates. Therefore, the use of SAE rates as a metric for trial representativeness can be investigated in this thesis. However, other trials with different interventions may fail to report it sufficiently.

4.6.5 Strengths and limitations

Strength of this analysis include exploring the reporting of SAEs across different trial registries and publications, providing a variety in SAEs reporting from these sources and demonstrating that most of these sources reported the SAEs. However, a limitation of the study is that the inconsistencies in SAE reporting

between trial registries and publications was not assessed. Moreover, another limitation is that, for some trials, the timeframe for reported events was captured from a different resource (result group or primary outcome tables) rather than the actual timeframe from the reported event tables. This may introduce bias in calculating the SAE rates, as the timeframes reported in these tables may not precisely match those in the event tables.

4.7 Conclusion

The included trials reported the sufficient information on SAEs, including the number of participants that experienced SAEs, the number of subjects at risk of SAEs, and, in most cases, the timeframe for these SAEs, which allows for the calculation of SAE rates. In the next chapter, I will calculate the SAE rates for both the intervention and control arms and compare the differences in rates between the two arms to determine which one to consider in the analyses for the subsequent chapters.

Chapter 5 Difference in serious adverse event rates between trial arms of SGLT-2 inhibitors

5.1 Chapter overview

This chapter addresses the initial examination of the feasibility of using SAE rates as a metric for trial representativeness. First, it examines whether there are significant differences in SAE rates between the trial arms of SGLT-2 inhibitors. This analysis will determine whether it is reasonable to consider SAE across all arms, or whether it is necessary to focus on an individual arm when analysing the SAE rate for this thesis. The chapter then explores how various trial characteristics, such as type of trial outcome and trial settings, influence the rates of SAE.

5.2 Background

According to the definition re-stated from chapter 3, SAEs include any untoward medical events that may result in death, are life-threatening, causes or prolong hospitalisation, result in persistent or significant disability, or cause a birth defect. If a trial were representative, it would be expected that the SAEs observed in the trial would be similar to those within the target population, suggesting that SAEs could serve as a suitable measure of trial representativeness. Generally, the incidence of SAEs is recorded regardless of treatment, including in the control arms. In some trials, particularly for less toxic treatments, it is likely that most SAEs are not directly related to the treatment, in which case it would be expected that SAE rate between treatment and control arms may be similar (Mahr et al., 2017). However, for some trials that involve potential toxic treatment like chemotherapy, the treatment arm is likely to have a higher SAE than the control arm. Thus, where there is a difference, the control arm would likely provide the suitable comparison with patients in routine care, especially when control arm reflects usual care by giving ‘standard’ treatment (Hanlon et al., 2022). In the context of SGLT-2

inhibitors (as in this thesis), it is not clear whether SAE rates vary depending on treatment arm in a trial, and it therefore remains uncertain which arm's SAE should be considered when determining the SAE rate as a metric of representativeness (whether the treatment, control, or the combined total of both). Therefore, in this chapter I will compare the differences in SAE rates between both arms to determine which SAE I will consider for analysis throughout this thesis.

In addition, SAE rates could be influenced by certain factors. For example, trials with hard serious outcomes, like cardiovascular events, may recruit older and sicker subjects, targeting patients with cardiovascular diseases who are inherently more vulnerable to SAEs (Wise et al., 2020). Moreover, ClinicalTrials.gov, as one of the largest registries, tends to have better reporting standards that might capture more SAEs than smaller registries. As a results, SAE rates in trials registered on ClinicalTrials.gov could be higher compared to those registered on other registries. Also, it is expected that SAE rates would be higher in multinational trials compared to national trials due to the broader and potentially more diverse participant base. Therefore, in this chapter, I will also explore how these variations in SAE rates across different trial outcomes, trial registries, and trial settings might reflect differences in trial representativeness as measured by SAEs. By examining these differences, I aim to assess whether SAE rates can serve as a reliable metric for trial representativeness.

5.3 Aim and objectives

In this chapter, in order to initially explore the feasibility of using SAE rates as a metric for trial representativeness, I will address the following objectives:

1. Examine the difference in SAE rates between treatment and control arm.

2. Examine the differences in SAE rates based on the following characteristics:

i. Type of trial outcome (hard vs soft outcome).

ii. Type of trial settings (single-centre sites vs multicentre, multinational vs one country).

iii. Type of trial registries (ClinicalTrials.gov vs other registries).

5.4 Methods

5.4.1 Data source

Trial selection and data sources are presented in detail in Chapter 4.

5.4.2 Data extraction

Data extracted included the number of reported SAEs, the number of subjects at risk of SAEs in each trial arm, and the timeframes for SAE occurrence. Additionally, the type of trial arm as intervention or control were extracted for each trial. Trials were classified based on the type of control arm, distinguishing between trials with placebo arms and trials with active comparator arms. This classification was determined by reviewing the designations of control arms and their prespecified aims. Furthermore, trial outcomes were categorised, distinguishing between MACE outcomes (considered a hard outcome) and surrogate endpoints like the change in HbA1c (considered a soft outcome). Other variables extracted included the type of trial registries, study centre (single or multi-centre), and the geographic scope (national or multinational). Additional variables of person-time and the rate for SAE were calculated for each trial.

5.4.3 Statistical analysis

Descriptive statistics were used to summarise the SAE counts across each trial arm. The SAE rate for each trial arm was calculated by dividing the number of SAEs by person-time. The person-time for each arm was computed using this formula:

(follow-up time [in years] × number of subjects at risk of SAEs − follow-up time [in years] × number of subjects affected by SAEs × 0.5).

First, I calculated the total time at risk by multiplying the follow-up time by the number of subjects at risk for SAEs. Then, I adjusted for subjects who experienced SAE by subtracting half of their follow-up time (using a factor of 0.5). The 0.5 factor assumes that participants who experienced SAE contributed, on average, only half of the follow-up period before the event occurred. This adjustment avoids overestimating person-time and provides a more accurate estimate when the exact timing of SAEs is unknown. This approach is a widely used approximation in epidemiology when the exact timing of events (e.g., SAEs) is not known within the follow-up period (Hanlon et al., 2021, 2022).

The first analysis was conducted to compare the differences in SAE rates between intervention and control arms across all trials using generalised linear models with Poisson likelihood regression. However, overdispersion was observed in the Poisson models. This was assessed by comparing the variance to the mean, where it was found that the variance was greater than the mean. Additionally, overdispersion was confirmed by calculating the ratio of the residual deviance to the residual degrees of freedom, which was significantly greater than 1, indicating overdispersion. Therefore, two separate negative binomial models were employed to more accurately compare the differences in SAE rates 1) between intervention and placebo arms across all trials, and 2) between intervention and active comparator arms across all trials. For both models, the SAE count was the outcome variable. To account for variations in person-time across trials, an offset was included in each model.

The second analysis employed separate negative binomial models to examine the differences in SAR rates between trials based on: types of trial outcome, types of trial registries, and types of trial settings. This analysis used combined SAE counts across all trial arms as the outcome variable. The choice of whether to use overall SAEs or individual arm SAEs was based on the first analysis. The incidence rate ratio (IRR) was compared for all analysis.

5.5 Results

5.5.1 Summary statistics

Reported SAEs were available at the arm level, except for five trials that did not report these events and were excluded from the analysis, leaving 141 trials with reported SAEs. Trials were classified according to the type of the comparator (control arm), specifically whether it was a placebo arm or an active comparator arm. The intervention arm received either one of the 9 SGLT-2 drugs (dapagliflozin, empagliflozin, canagliflozin, ertugliflozin, ipragliflozin, luseogliflozin, bexagliflozin, tofogliflozin and sotagliflozin). Out of the 141 trials analysed, 95 trials had a placebo as a control arm, while 46 trials had an active comparator. The total number of reported SAEs for trials with a placebo comparator (95 trials) and trials with an active comparator (46 trials) were 19,453 and 1688, respectively, across both arms (Table 5.1).

Table 5.1: Summary statistics for SAEs counts across arms of SGLT-2 trials.

Type of arms	Placebo comparator arms			Active comparator arms		
	Placebo	Intervention	Total	Active Comparator	Intervention	Total
No. of SAEs reported	8423	11,030	19,453	768	920	1688
Percentage (%)	43.3%	56.7%	100%	43.3%	56,7%	100%

5.5.2 SAE rates between trial arms

The difference in SAE rates between intervention and placebo arms across all trials of SGLT-2 inhibitors was examined. The results showed that SAE rates were 11% lower in the intervention arms compared to the placebo arms; however, the confidence interval included the null (IRR 0.89, 95% CI 0.73-1.07). Similarly, in the comparison between the intervention and active comparator arms across all trials of SGLT-2 inhibitors, the results showed that SAE rates were 10% lower in

the intervention arms than in the active comparator arms; however, the confidence interval included the null (IRR 0.90, 95% CI 0.69-1.17).

5.5.3 Differences in SAE rates based on some factors

For this analysis, the overall trial rate was used because there were no significant differences across trial arms. The analysis examined the differences in SAE rates among trials based on multiple factors, including trials registration on ClinicalTrials.gov, type of outcome, and trial settings. The rate was 94% higher in ClinicalTrials.gov registry compared to non ClinicalTrials.gov registries, and the confidence interval did not include the null (IRR 1.94, 95% CI 1.23-3.02). Trials with hard outcomes (MACE) had a higher rate than trials with soft outcomes (HbA1c) (IRR 2.86, 95% CI 1.84-4.74). Furthermore, trials that involved more than one country (multinational) showed 79% higher rates of SAE compared to trials that involved only one country (IRR 1.79, 95% CI 1.32-2.41). However, there was no difference in the rate between trials conducted at multicentre sites in comparison to those conducted at single-centre sites (IRR 1.07, 95% CI 0.55-2.11).

5.6 Discussion

5.6.1 Summary of main findings

The total SAEs across all trials was 21,141 events. No significant difference was found in SAE rates between the intervention and control arms across all SGLT-2 trials. This indicates that the intervention-related effect in the included trials was not the major driver of SAEs. The SAE counts for both intervention and control arms were combined to increase the statistical precision. Moreover, the rates of SAE were higher in trials registered on ClinicalTrials.gov, in trials with hard outcome, and in trials conducted in more than one country.

5.6.2 Interpretation

The findings show that the SAE rates did not significantly differ between the intervention and placebo arms nor between the intervention and active comparator arms in trials of SGLT-2 inhibitors. It is expected that there would be no significant difference between treatment and control arms, as in trial settings, most SAEs are not particularly related to the investigational product (Gebrie et al., 2021). Although SGLT-2 inhibitors are known to cause conditions such as diabetic ketoacidosis, amputation, and genital infections, the results suggest that these events are likely to be sufficiently rare. Therefore, for the purposes of modelling the overall SAE rate, the impact of these treatment-specific events is relatively small. The improvement in precision from using both arms together outweighs the potential bias that might be caused by including SAEs related to treatment. Furthermore, the similar rates between trial arms may suggest that SAE rates reflect the underlying health status of the trial participants as a population, which may reveal information that can be useful in assessing the representativeness of the trial.

Furthermore, trials registered on ClinicalTrials.gov showed higher SAE rates compared to those not registered, indicating better reporting in these trials. ClinicalTrials.gov has detailed reporting requirements and undergoes rigorous checks by staff members (ClinicalTrials.gov, 2023). These requirements involve the inclusion of information such as severity, types of events, and their relationship to the study intervention (Mathieu, Moher and Altman, 2009). These requirements may vary by registry, and other registries may not consistently document all SAEs, particularly if the event occurred at the end of the study or was unrelated to the investigated drug (Wortzel et al., 2020).

Additionally, the analysis revealed a difference in the rates of SAE based on the type of outcome, specifically between trials with hard and soft outcomes. The higher rate found in trials with a prespecified hard outcome, suggests that these trials may have broader eligibility criteria, possibly being more inclusive of older and sicker people who are ultimately more vulnerable to SAE. Moreover, it is possible that these trials deliberately recruited participants at higher risk of

events, such as heart attacks or strokes, especially since the studies are designed to detect MACE. Furthermore, the higher incidence of SAEs in multinational trials may be attributed to the heterogeneity of patient characteristics. Multinational trials cover a diverse population with varying disease progression and comorbidities compared to single-country trials. Additionally, healthcare standards and protocol adherence can differ across countries, influencing patient outcomes and the reporting of SAEs. Consequently, this factor may contribute to a higher incidence of SAEs.

5.6.3 Comparison with other studies

This comparison offers insights by showing no significant difference in SAE rates between the trial arms of SGLT-2 trials. While few studies have used SAEs to assess representativeness, this issue of treatment arms is an important aspect of such assessments. Hanlon et al. (2022) assessed trial representativeness across 21 index conditions and compared SAE rates between intervention and control arms, finding no differences between them. Another study examined the rates of SAE between treatment and control arms in RCTs for hypertension and similarly found no differences in SAE rates between both arms (IRR 0.81, 95% CI 0.59–1.08) (Hanlon et al., 2021). Furthermore, another study compared the difference in SAEs between the intervention and control arms in 26 naltrexone trials and found no significant difference in SAEs across both arms (Bolton et al., 2019). However, Wolff et al. (2022) found that SAEs were 1.3 times higher in the treatment arm than in the control arm of cancer trials. This difference could be related to the investigational agents, as cancer drugs are likely more toxic and potentially result in more SAEs than other interventions. However, it could also be attributed to the underlying disease, as cancer patients are more vulnerable to SAEs than others.

5.6.4 Implications

The comparison of SAE rates between trial arms has important implications for the analyses conducted throughout this thesis. It provides valuable information on which arms' SAE data should be considered when assessing SAE rates as a

metric of representativeness. This analysis suggests that combining SAEs across all arms is a reasonable approach when analysing SAE rates against the PRECIS-2 tool, trial/baseline characteristics, and eligibility criteria. Combining arms addresses the issue of lower event numbers in control arms and enhances statistical power. However, this approach may not be applicable to all trials, particularly those involving potentially toxic treatments (e.g. chemotherapy), where higher SAEs in the intervention arm are likely to be directly related to the treatment. Additionally, the comparison of SAE rates between trials with hard and soft outcomes offers insights into how the nature of trials' outcomes can influence SAEs reporting. Specifically, trials targeting hard serious outcomes, such as MACE, tend to report higher SAE rates, potentially due to their focus on outcomes that are inherently more serious. This suggests that the type of outcome a trial measures may impact the incidence of SAEs, offering a deeper understanding of the factors that contribute to SAE variability. Such insights are crucial for assessing the representativeness of clinical trials.

5.6.5 Strengths and limitations

The included trials were from different trial registries, offering a diverse selection of RCTs. This may help minimise the bias that could arise from the limitations of using trials from single registries. However, it is worth noting that this study has limitations. First, it is possible that there is a true difference between trial arms, but because the SAE are relatively rare, the sample may lack sufficient power to detect it. Second, the use of the 0.5 factor in calculating person-time assumes that SAEs occur, on average, midway through the follow-up period. This may not always reflect the true timing of events and could slightly overestimate or underestimate the actual person-time. Nonetheless, this approach is a widely accepted approximation in epidemiology when the exact timing of events is unavailable (Hanlon et al., 2021, 2022).

5.7 Conclusion

There was no difference in the rates of SAE between the intervention and control arms. Total SAEs from all trial arms for each trial will be combined in the subsequent analyses of this thesis. In the next three chapters, I will explore whether SAE rates reflect trial representativeness by examining their association with the pragmatism metric (PRECIS-2 tool), baseline and trial characteristics, and eligibility criteria.

Chapter 6 Assessment of the pragmatism of RCTs of SGLT-2 inhibitors: using the pragmatism metric PRECIS-2 tool

6.1 Chapter overview

I will first describe the Pragmatic Explanatory Continuum Indicator Summary (PRECIS-2) tool in detail and its importance, then present the findings from applying this tool on the RCTs of SGLT-2 inhibitors. I will also examine the association between SAE rates and PRECIS-2 score and each domain. The finding of comparing this pragmatism metric to SAEs against their association to baseline characteristics will be shown in Chapter 7.

6.2 Introduction

6.2.1 PRECIS-2 tool

The PRECIS-2 tool was employed to assess the pragmatism of clinical trials. It has undergone rigorous development to ensure its reliability and broad applicability across diverse clinical trial contexts. It was revised by Loudon et al. in 2015 as a modification of the original PRECIS tool that was developed by Thorpe et al. in 2009. This tool distinguishes pragmatic trials undertaken in usual care from explanatory trials conducted in an idealised setting. Its application assists trialists in determining whether the trial design matches the proposed aim (Loudon et al., 2015). The PRECIS-2 tool contains nine domains, including eligibility criteria, recruitment, setting, organisation, flexibility (delivery), flexibility (adherence), follow-up, primary outcome, and primary analysis. Each domain is assessed by assigning a score that reflects the pragmatic or explanatory nature of trials. The scoring system ranges from 1 (highly explanatory approach) to 5 (highly pragmatic approach), and labelled as follows: 1_ very explanatory, 2_ rather explanatory, 3_ equally pragmatic and explanatory, 4_ rather pragmatic, 5_ very pragmatic.

6.2.2 Rationale for using PRECIS-2 tool

The pragmatism of a trial refers to the extent to which its design reflects the setting of routine care (Smelt et al., 2010). On the other hand, trial representativeness refers to the extent to which a trial population reflects the target population in routine care (He et al., 2020). Although these concepts have different meanings, both are concerned with enhancing the applicability of clinical trials to routine clinical practice (Treweek and Zwarenstein, 2009). For example, pragmatic trials tend to recruit more representative patients by using inclusive eligibility criteria. The PRECIS-2 tool can assess trial pragmatism by scoring the eligibility criteria domain to assess its inclusiveness. The framework design of the tool includes various domain structures, focusing on multiple aspects of trial design. This structure offers a systematic and standardised evaluation with a multidimensional assessment of the degree of trial pragmatism (Loudon et al., 2015). Moreover, it is known that pragmatic trials investigate medications under usual conditions that resemble routine care, thereby enabling the extrapolation of trial findings to real-world practice. Hence, the PRECIS-2 tool, as reflected by its pragmatism score, potentially reflects how the trials represent routine practice, and whether the trial findings can be generalised to the population. Therefore, in the absence of a gold standard for assessing trial representativeness, comparing SAEs feasibility, as a metric, with the PRECIS-2 tool may serves as an approach to gauge the tool's feasibility (SAEs) in assessing trial representativeness.

6.2.3 Domains of PRECIS-2 tool

The different domains and their importance in trial design are presented below.

6.2.3.1 Eligibility criteria

A highly pragmatic trial is designed to include participants who are eligible for treatment in routine care. Thus, if the intervention under study is expected to be prescribed for diverse demographic groups in routine care, including older

people and patients with different comorbidities, it is essential to include these groups in the trials to yield pragmatic trials with applicable results (Kanzler et al., 2022). Conversely, if the eligibility criteria are restrictive, excluding certain groups, the trial becomes more explanatory, focusing on measuring intervention efficacy under ideal rather than real-world conditions typical in routine care. Trialists might limit trial pragmatism by implementing restrictive exclusion criteria (Loudon et al., 2015). However, for trials investigating the safety and efficacy of interventions, particularly in phase I and II trials, and due to ethical considerations, the trial requires to be as explanatory as possible (Le-Rademacher et al., 2023).

6.2.3.2 Recruitment

Recruitment is a preliminary step that precedes the trial's commencement, and trialists should carefully consider where and how participants will be recruited. Clarification of this process helps to differentiate between pragmatic and explanatory trials. A highly pragmatic approach would recruit participants from usual care settings, whereby individuals attending routine clinic appointments are recruited without additional efforts (Ford and Norrie, 2016). However, conducting additional assessments on individuals recruited from usual care, and employing methods of explanatory approach such as media advertising, sending letters, or emails to prospective patients, may lead to an explanatory trial. To ensure that the trial can provide more relevant results, participants should be recruited from multiple clinics rather than one (Dal-Ré, de Boer and James, 2020).

6.2.3.3 Setting

This domain is crucial in describing where the trial is conducted. While accepted guidelines for assessing the impact of setting on generalisability are lacking, there are several characteristics of setting that need to be considered. This includes geography, healthcare system, country, and ethnic mix of the included participants, all of which might affect the trial's generalisability (Dekkers et al.,

2010; St Sauver et al., 2012). Pragmatic trials are more likely to involve multiple centres (Sepehrvand et al., 2019). However, in some instances, trials conducted in a single centre may be still considered pragmatic if the intervention targets a highly specialised centre. Trialists should ensure that the trial setting mirrors the healthcare setting where their results are anticipated to be implemented.

6.2.3.4 Organisation

This domain focuses on the expertise and resources used to deliver the intervention, this aspect contributes to generating trials with broader applicability by encouraging the utilisation of the same resources and healthcare staff available in routine care settings. An explanatory trial often implements the intervention under different protocols compared to the usual care system; it might involve additional staff and provide extra training to healthcare professionals (Casey et al., 2022). However, these additional services may not be available in routine care practice.

6.2.3.5 Delivery

The method of delivering the intervention plays a role in assessing the pragmatism of clinical trials. A highly pragmatic design can be achieved by allowing participants to determine how the intervention will be delivered without imposing additional instructions. However, this doesn't imply that the trial must specify which other interventions are permitted or refrain from outlining specific delivery protocols, as these aspects are often part of usual practice. Essentially, the administration of the intervention should closely resemble the protocol followed in usual care (Loudon et al., 2015). If the trial team requires healthcare providers to execute additional intervention procedures, the trial would tend towards being more explanatory.

6.2.3.6 Adherence

A highly pragmatic trial would allow participants to engage with and adhere to the intervention without further instructions (Fitzpatrick et al., 2020). However, in regular practice, practitioners typically continue to persuade patients during each follow-up visit to comply with the intervention to the best of their ability. This encouragement doesn't significantly impact the pragmatism of the clinical trial since it aligns with routine care practices. Conversely, a trial protocol that uses adherence monitoring methods uncommon in routine care and excludes nonadherent patients would be considered highly explanatory.

6.2.3.7 Follow-up

Follow-up, in this context, refers to the domain that examines how closely participants are monitored and the procedures carried out during these follow-up visits. A highly pragmatic approach regarding follow-up would entail having follow-up visits no more frequent than those in usual care settings. It would involve collecting outcome data from administrative or clinical record systems without direct contact with the participant (Le-Rademacher et al., 2023). Furthermore, pragmatic trials should aim to collect as little additional data as possible while avoiding extensive data collection.

6.2.3.8 Primary outcome

This domain focuses on the trial’s outcome that is most relevant to the patients. The selection of the trial’s primary outcome is pivotal for trialists, who should prioritise outcomes that hold significance for patients rather than just practitioners (Kanzler et al., 2022). This approach aims to ensure the production of an applicable and pragmatic trial. Furthermore, the methods used to measure the outcome should be similar to the standard care protocol.

6.2.3.9 Primary analysis

This domain evaluates the extent to which data is incorporated into the analysis. If the intention-to-treat analysis (ITT) is used, all participants will be included in the data analysis, making the trial more pragmatic (Loudon et al., 2015).

Table 6.1 PRECIS-2 tool domains and variation between pragmatic and explanatory design

Domain	Explanatory approach	Pragmatic approach
Eligibility criteria *(what are the criteria in patient selection?)*	Highly selected patients; strict inclusion criteria, e.g., high restrictions on comorbidities, concomitant medication	Typical patients; minimal inclusion criteria, e.g., include any patients with condition of interest (T2DM)
Recruitment *(what are the ways to recruit participants?)*	Use methods and resources outside of, or in addition to, what is typical, e.g., recruit patients through using media advertising, and sending letters to participants	Recruited in usual healthcare setting; participants may include patients, providers, or health systems, e.g., use the regular appointment to recruit participants
Setting *(Where is the trial has been conducted?)*	Specialist practice or academic medical centre	Primary care clinic or setting where the trial results will be applied, e.g., usual care centre like primary care
Organisation *(What resources and expertise were used in the trial?)*	Changes the workflow, adds equipment, or need for extra staff training, or affects how care is typically delivered, e.g., use additional staff, and extensive resources	Changes to clinical delivery and resources are minimal, easy to implement in usual care after the trial, e.g., no more than usual healthcare staff and resources
Flexibility (delivery) *(How was the intervention delivered?)*	Highly specified, protocol-driven with timing of intervention tightly defined, e.g., restriction on drug administration and prevent use of other medication during the trial	Details of intervention delivery left to the care provider, e.g., no more direction than used in usual care
Flexibility (adherence) *(How were participants monitored for adherence?)*	Measures to monitor patient adherence and excludes patients judged not to be adherent, e.g., use some test and procedure to monitor and enhance adherence and compliance	No special measures to enforce intervention engagement or compliance, e.g., use similar protocol of usual care for adherence enhancement

Follow-up *(How were participants followed up?)*	Rigorous follow-up protocol distinct from usual care. e.g., monthly visits, or extensive data collection	Follow-up visits similar to usual care. e.g., clinic visits every 3-6 months
Primary outcome *(How the trial's primary outcome was relevant to participants?)*	Surrogate outcomes or measures distant from the key question, e.g., blood test	Outcomes of importance to patients, measured as they would be in usual care, e.g., MACE
Primary analysis *(how was the data included in the analysis?)*	Excludes noncompliant participants, or dropouts, e.g., per protocol analysis	Intention-to-treat analysis (ITT)

6.3 Aim and objectives

In this chapter, in order to determine whether SAEs reported in clinical trials for SGLT-2 inhibitors can be used as a marker of trial representativeness, I will use PRECIS-2 tool to assess the trial pragmatism and compare it with SAEs. I will also examine the association between SAE rates and individual domains of the PRECIS-2 tool.

6.4 Methods

6.4.1 Data source

Trial selection and data sources are presented in Chapter 4. Data required for assessing trial pragmatism and scoring all domains were obtained from trial-relevant publications. Additionally, trial protocols and CSRs were reviewed to collect data that were not reported in the publications.

6.4.2 Data extraction

For the pragmatism assessment, the extracted data included details of published trials (publication title, authors, trial ID, PMID, publication year) and key information required for assessing pragmatism and scoring PRECIS-2 domains. This information covers eligibility criteria, type of trial settings, recruitment method, intervention delivery and follow-up protocols, type of primary outcome, and type of population analysis. These data were manually extracted and organised in a table to be used as the rationale for scoring each domain. Data sources and extraction methods for SAEs and their timeframes are presented in Chapter 4.

6.4.3 PRECIS-2 ratings and protocolisation of assessment criteria

The assessment process involved a careful review of trial protocols and methods from the trial-relevant publications or supplementary materials retrieved from trial registries or sponsors. The original criteria of the PRECIS-2 tool underwent an extensive protocolisation process to create a scoring template that aligned with the specific characteristics of clinical trials for SGLT-2 inhibitors within the context of T2DM. Domains requiring clinical insight were reviewed by a diabetologist. For the eligibility criteria domain, the participants selection criteria were reviewed and tailored to SGLT-2 inhibitor trials in type 2 diabetes. Criteria such as age, comorbidities, and concomitant medications were carefully considered with the diabetologist before scoring the domain and to gauge the inclusiveness of trial populations. Furthermore, a multi-disciplinary team of researchers in diabetology and epidemiology was involved in the protocolisation process to mitigate the subjectivity in the assessment.

As illustrated in table 6.2, the scoring system for each domain ranged from 5 (pragmatic scale) towards 1 (explanatory scale) based on the protocolised criteria. The eligibility criteria domain was scored 5 if trial criteria were inclusive. The score reduced toward 1 as the criteria became more restrictive, excluding participants based on their comorbidities, adherence, and

concomitant medications. The recruitment domain was scored 5 if the trial recruited participants from a usual care setting where the interventions are likely to be prescribed. The score moved toward the explanatory scale if the trial used incentives (e.g., vouchers) or advertisements to recruit participants. The setting domain was scored 5 if the trial was conducted in multinational, multicentre and usual care clinics, and the score reduced toward 1 if the trial was conducted in a single centre or site other than the usual care clinic, such as an academic or specialist centre. The primary outcome domain scored 5 if the endpoint was important for both healthcare providers and the patients, such as adverse events. The score moved toward the explanatory scale if the outcome was a surrogate outcome, required central adjudication, or was measured earlier than in usual care. Furthermore, the adherence domain was scored 5 if the trial allowed complete flexibility for participants to adhere to the intervention without special measures or additional effort. If the trial included pre-screening stages to monitor patient adherence or a lead-in period to exclude non-adherent participants, it scored toward explanatory scale. Examples of pragmatic and explanatory trials are provided in Supplementary Table S.1. Details of protocolised assessment criteria for all PRECIS-2 domains are illustrated in table 6.2.

To ensure consistency in the trial assessment, a revision was conducted following a random sample of 10 trials by another PhD student, Saleh Almazam. Before commencing the assessment for all the trials, multiple discussion meetings have been conducted to resolve disagreements and reach a consensus. After scoring all domains and completing the assessment, all scores were reviewed by a diabetologist to ensure the reliability of the assessments. Each domain was scored based on the available information found in the trial protocol. If the required information was not reported there, other pertinent publications related to the assessed trial and the original protocol from ClinicalTrials.gov were reviewed. If the required information remained missing despite these efforts, then the domain was left blank, and during analysis, the missing domain was assigned a score of 3 (equally pragmatic/explanatory).

Table 6.2 The protocolisation of the PRECIS-2 for clinical trials of SGLT-2 inhibitors.

Domain	Score	Condition
Eligibility criteria	1	If the trial applies highly restrictive criteria, such as specific age groups or comorbidities (e.g., including diabetic patients with only chronic kidney disease).
	1	If the trial applies 4 or more of the below exclusion criteria that reduce the score by 1 point.
	2	If the trial applies 3 of the below exclusion criteria that reduce the score by 1 point.
	3	If the trial applies 2 of the below exclusion criteria that reduce the score by 1 point.
	4	If the trial includes participants who are eligible for treatment in routine care but only excludes participants who are not known to be highly adherent.
	4	If the trial includes participants who are eligible for treatment in routine care but only excludes participants with comorbid conditions (e.g. heart failure).
	4	If the trial includes participants who are eligible for treatment in routine care but only excludes participants using other medications (e.g. antihyperlipidemic drugs).
	4	If the trial includes participants who are eligible for treatment in routine care but only excludes participants due to the delivery challenges unrelated to the intervention (e.g. geographical location).
	4	If the trial includes participants who are eligible for treatment in routine care but only uses measures or tests to exclude participants (e.g. Glycaemic clamps or C-peptide).
	4	If the trial includes participants who are eligible for treatment in routine care but only excludes participants relies on carers.
	5	If the trial includes participants who are eligible for treatment in routine care but only excludes severely comorbid participants who are not likely to receive the intervention in routine care.
	5	If the trial includes participants who are eligible for treatment in routine care.
	Missing	If no information reported about eligibility criteria.

Recruitment	1	If the trial was single centres and used other usual care (e.g., hospitals, speciality clinics and research centres) to recruit participants for an intervention that is likely to be prescribed in primary care.
	1	If the trial recruited participants through advertisements (e.g. radio and television), or telephone calls, or used incentives and rewards (e.g. vouchers and cash payments).
	2	If the trial was national, multicentre and used other usual care (e.g., hospitals, speciality clinics and research centres) to recruit participants for an intervention that is likely to be prescribed in primary care.
	2	If the trial sent invitation letters or emails to the eligible participant identified through searching the medical records
	3	If the trial was multinational, multicentre and used other usual care (e.g., hospitals, speciality clinics and research centres) to recruit participants for an intervention that is likely to be prescribed in primary care.
	3	If the trial was national, single centre and used usual care clinics to recruit participants.
	4	If the trial was national, multicentre and used usual care clinics to recruit participants.
	5	If the trial was multinational, multicentre and used usual care clinics to recruit participants without any additional effort.
	Missing	If no information reported about recruitment.
Setting	1	If the trial was conducted in a single centre and site other than the usual care (e.g. hospitals, speciality clinics and research centres) for an intervention that is likely to be prescribed in usual care.
	2	If the trial was conducted in national, multicentre and sites other than usual care (e.g. hospitals, speciality clinics and research centres) for an intervention that is likely to be prescribed in usual care.
	3	If trial was conducted in a single centre and usual care clinic.
	3	If the trial was conducted in multinational, multicentre and sites other than usual care (e.g.

		hospitals, speciality clinics and research centres) for an intervention that is likely to be prescribed in usual care.
	4	If the trial was conducted in national, multicentre, and usual care clinics.
	5	If the trial was conducted in setting that mimics usual care setting.
	5	If trial was conducted in multinational, multicentre and usual care clinics.
	Missing	If no information reported about trial setting.
Organisation	1	If the trial applies 3 of the below criteria that reduce the score by 2 points.
	2	If the trial applies 2 of the below criteria that reduce the score by 2 points.
	3	If the trial provides training or education not required in usual care.
	3	If the trial requires a certification or an experience not required in usual care (e.g. nurses certified in diabetes management).
	3	If the trial uses additional resources or additional diagnostic procedures that are not used in usual care (e.g. MTT).
	3	If the trial uses an additional staff that is not used in usual care (e.g. pathologists).
	4	If the trial increases the resources used to deliver the intervention beyond what is typical in usual care (e.g. increasing the number of intervention providers or using additional facilities)
	5	If the trial was conducted without using additional resources, staff or requires certification that is not required in usual care.
	Missing	If no information reported about organisation.
	1	If the trial applies a highly specified protocol.

Flexibility (delivery)	1	If the trial applies 4 of the below restrictions or protocols that reduce the score by 1 point.
	2	If the trial applies 3 of the below restrictions or protocols that reduce the score by 1 point.
	3	If the trial applies 2 of the below restrictions or protocols that reduce the score by 1 point.
	4	If the trial applies restrictions on the number or the type of cointerventions.
	4	If the trial provides specific direction on managing side effects or complications of the intervention.
	4	If the trial provides specific direction to enhance the delivery of the intervention.
	4	If the trial tightly defines the timing of the intervention.
	4	If the trial undertakes additional interventions unavailable in usual care.
	5	If the trial does not specify permitted cointerventions or the delivery procedure mimicking the usual care.
	5	If the trial leaves the details of the delivery procedure to the health care provider.
	Missing	If no information reported about delivery procedure.
Flexibility (adherence)	1	If the trial withdrew non-adherent participants.
	2	If the trial applies a pre-screening stage (run-in or lead-in period) for adherence and excludes non-adherents.
	3	If the trial measures and monitors the adherence of subjects to the intervention.
	4	If the trial does not report any measurement of adherence or exclusion of non-adherents.

	5	If the trial does not apply measures or enforce the adherence of subjects to the intervention.
	5	If the trial permits full flexibility on how and when subjects take the intervention.
Follow-up	1	If the trial applies 2 of the below criteria that reduce the score by 3 points.
	2	If the trial has longer visits for follow-up or more extensive data collection than the usual care.
	2	If unscheduled follow-up visits were triggered by a primary outcome event.
	5	The score will be 5 for the trial that has follow-up visits and intervals no more than usual care.
	5	The score will be 5 for the trial that has no follow-up contact with participants to obtain data.
	____	The trial that does not report the follow-up frequencies or intervals will be assessed based on the frequency and the intervals of the primary outcome.
	____	The trial that has follow-up visits and intervals more than usual care will be assessed based on the intervals and the frequency of the visits as below:
	1	If the follow-up visits were every 2 weeks or less.
	2	If the follow-up visits were every 3 weeks or less.
	3	If the follow-up visits were every 4-6 weeks.
	4	If the follow-up visits were every 7-12 weeks.
	5	If the follow-up visits were every 8 weeks or more.

Primary outcome	1	If the trial applies 2 of the below criteria that reduce the score by 3 points.
	2	If the trial has a surrogate outcome, on which the intervention is expected to have a direct effect on.
	2	If the trial has an outcome that requires central adjudication or special training to measure it.
	2	If the trial has an outcome that is not of importance to the health care provider and the patients.
	3	If the trial has an outcome of importance only to the health care provider.
	4	If the trial measures the outcome at a time earlier than the usual care (less than 4 weeks).
	5	If the trial has hard composite outcome (e.g. MACE).
	5	If the trial has an outcome of importance to the health care provider and the patients (e.g. adverse events).
Primary analysis	1	If the trial uses per-protocol or as-treated analysis.
	2	If the trial includes data from all patients who were randomised and received a trial product with efficacy data.

	3	If the trial reports the use of modified intention-to-treat analysis without details of the modification.
	4	If the trial analyses its primary outcome based on an intention-to-treat analysis using all available data for subjects who received at least one dose of the study drug.
	5	If the trial analyses its primary outcome based on an intention-to-treat analysis using all available data for all randomised participants.
Interpretation of the scores: 1) very explanatory, 2) rather explanatory, 3) equally pragmatic/explanatory, 4) rather pragmatic, 5) very pragmatic.		

6.4.4 Statistical analysis

For each trial, the mean score of all PRECIS-2 domains was calculated. Also, the mean score for each domain across all trials was calculated. The PRECIS-2 score across all the trials were summarised using descriptive statistics (means and standard deviation). Histogram was used to visualise the distributions of the mean PRECIS-2 scores for all trials. PRECIS-2 wheel was used to show the mean scores for all PRECIS-2 domains, reflecting the overall pragmatism of SGLT-2 trials.

To allow for fair comparisons, the SAE rates, mean PRECIS-2 score, and the 9 domains need to be normally distributed and on the same scale. Therefore, they were transformed using the "orderNorm" function from the R package "bestNormalize". Linear regression was used to examine the association between SAE rates and mean PRECIS-2 score. Additionally, the association of SAE rates with individual domains of PRECIS-2 was estimated using simple and multiple linear regression. Simple linear regression was fitted for unadjusted analysis to

capture the association between SEAs and single domains independently, whereas multiple linear regression was fitted for adjusted analysis to estimate the association while accounting for the combined influence of other domains (as outlined in Table 6.3). This analysis will provide insight about how the design of trials, as reflected by assessing each PRECIS-2 domains, is associated with the SAE rates.

6.5 Results

6.5.1 PRECIS-2 scores

The required information for scoring some domains was not reported within some trials. Recruitment was the domain with most missing information (77%) (Figure 6.1). The mean (SD) PRECIS-2 score across 146 trials was 3.34 (0.48), indicating a medium level of pragmatism (Figure 6.2). Moreover, a sensitivity analysis for the mean (SD) PRECIS-2 score using complete case (i.e. where the mean score was calculated based only on the domains with non-missing scores) showed a similar medium level of pragmatism, with a mean (SD) score of 3.37 (0.55). The mean score of most and least pragmatic studies in the sample were 4.22 and 1.78, respectively. The frequency of scores for each domain within the PRECIS-2 tool are illustrated in figure 6.4. The mean (SD) PRECIS-2 score for trials categorised by SGLT-2 inhibitor medications are presented in Supplementary Table S.2. Moreover, the frequency of PRECIS-2 scores for all domains with missing information is shown in Supplementary Figure S.5.

Additionally, in terms of individual domains of PRECIS-2, setting and organisation had the highest scores, 4.18 (1.26), 4.08 (0.94) respectively, indicating a pragmatic level. Conversely, primary outcome and eligibility criteria had the lowest score, 2.16 (0.66), 2.45 (0.76) respectively, indicating an explanatory level (Figure 6.3). These mean scores were calculated after replacing missing domain scores with a value of 3. The mean PRECIS-2 scores for all domains, based only on the available non-missing scores, are displayed in Supplementary Figure S.6 and show similar results.

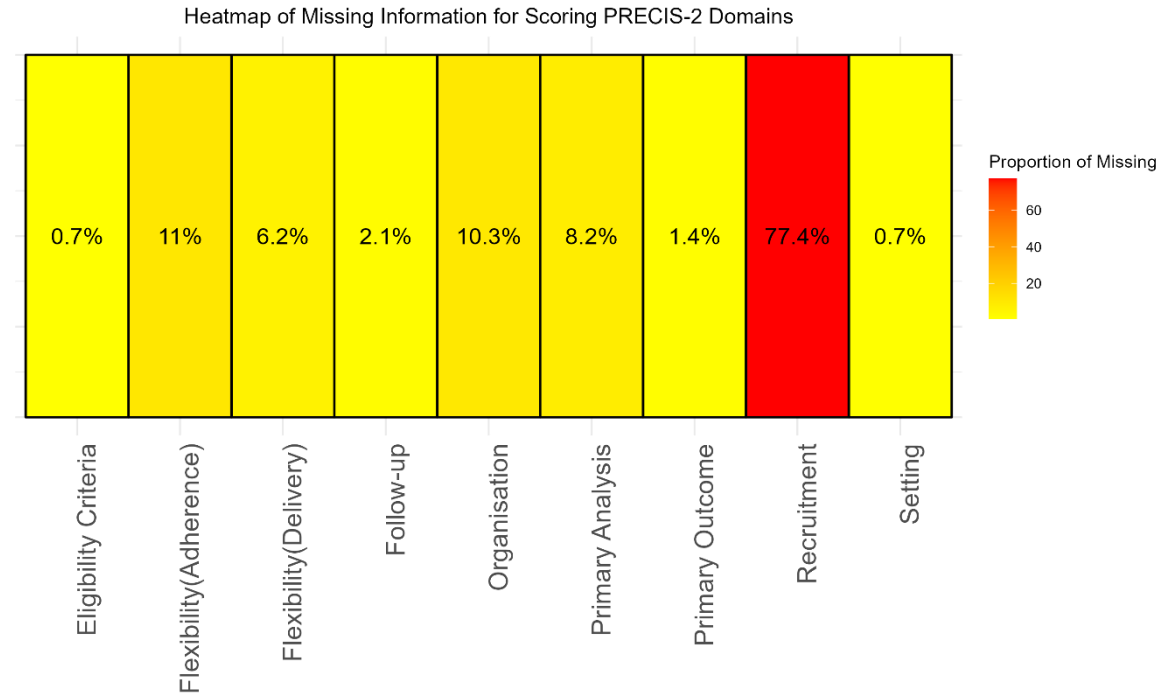

Figure 6.1 Heatmap of missing for all domains of PRECIS-2 tool

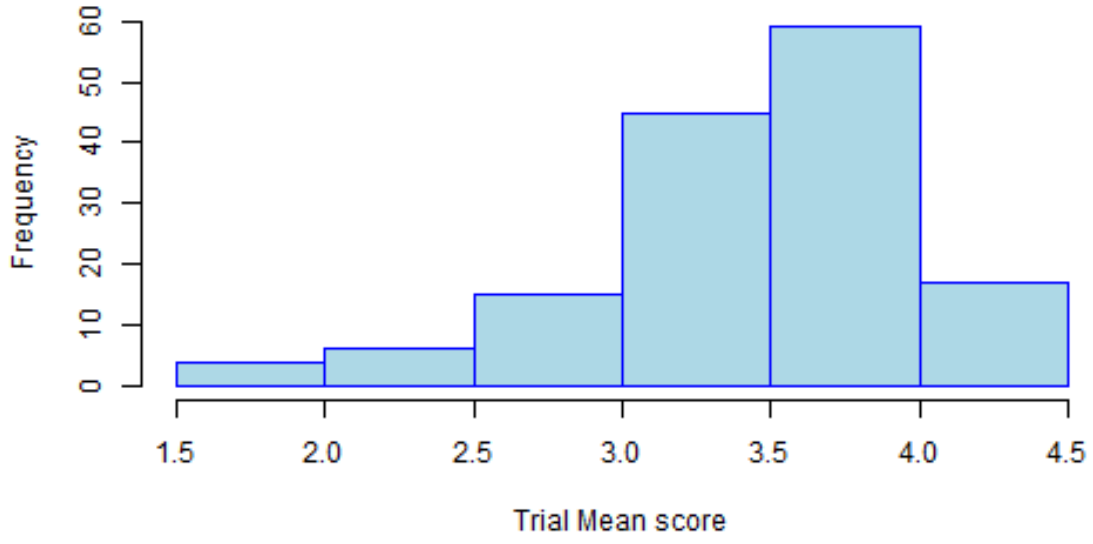

Figure 6.2 Histogram of the mean PRECIS-2 scores for all SGLT-2 trials

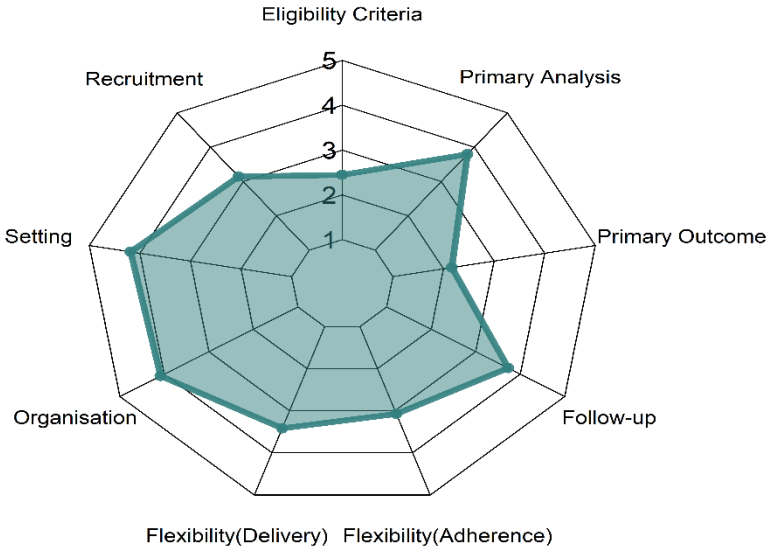

Figure 6.3 Wheel for the mean PRECIS-2 scores for all domains, this wheel is a visible figure to show the direction of score to demonstrate whether domain was pragmatic, explanatory, or equally pragmatic/explanatory

Figure 6.4 Frequency of PRECIS-2 score among all domains

6.5.2 The association between SAE rates and PRECIS-2 score

The mean PRECIS-2 score was positively associated with the SAE rates (β = 0.17, 95% CI 0.00–0.33). This positive coefficient indicates that an increase in the PRECIS-2 score corresponds to a 0.17 increase in the SAE rates. Moreover, a sensitivity analysis using complete case (i.e. where the mean PRECIS-2 score was calculated based only on the remaining domains with non-missing scores) showed a similar positive association with SAE rates (β = 0.19, 95% CI 0.02–0.35).

6.5.3 The association between SAE rates and domains of PRECIS-2

The setting domain was the only domain associated with SAE rates in both unadjusted and adjusted models, the confidence interval did not include the null (β = 0.51, 95% CI 0.26–0.76). Other domains did not show associations with SAE rates in either the unadjusted or unadjusted models (Table 6.3).

Table 6.3 Association between PRECIS-2 domains and SAE rates

Serious Adverse Event Rates						
PRECIS-2 Domains	Unadjusted			Adjusted		
	β	CI	P	β	CI	P
Eligibility criteria	0.14	-0.05 to 0.33	0.143	0.05	-0.14 to 0.25	0.596
Recruitment	0.11	-0.09 to 0.32	0.280	0.05	-0.16 to 0.24	0.652
Setting	0.49	0.29 to 0.69	<0.001	0.51	0.26 to 0.76	0.001
Organisation	0.09	-0.10 to 0.29	0.340	-0.11	-0.34 to 0.12	0.339
Flexibility delivery	0.09	-0.09 to 0.27	0.338	0.01	-0.22 to 0.24	0.951
Flexibility adherence	0.03	-0.15 to 0.21	0.774	0.03	-0.19 to 0.25	0.808
Follow up	-0.01	-0.22 to 0.19	0.903	-0.01	-0.22 to 0.20	0.915
Primary outcome	-0.09	-0.35 to 0.16	0.479	0.04	-0.22 to 0.31	0.754
Primary analysis	0.10	-0.09 – 0.28	0.300	0.03	-0.17 to 0.22	0.799

6.6 Discussion

6.6.1 Summary of main findings

This chapter utilised the pragmatism metric tool (PRECIS-2) to compare it with SAE rates. Applying this tool assessed the trial pragmatism and illustrated where they lie on a spectrum from pragmatic to explanatory. Based on the mean PRECIS-2 score, the 146 trials were, on average, in the middle of the efficacy and effectiveness continuum. Regarding individual domains, the eligibility criteria and primary outcome domains were more explanatory compared to the setting and organisation domains, which represented a more pragmatic approach. The domains of recruitment, delivery, and adherence demonstrated a medium level of pragmatism.

The SAE rates and the mean PRECIS-2 score were found to be associated with each other. Also, the SAE rate was significantly associated with only one PRECIS-2 domain (setting).

6.6.2 Interpretation

The medium level of pragmatism (a score of 3.34 out of 5) for the included trials might be explained by the individual trials' aim of addressing the efficacy and effectiveness of the intervention. This analysis highlights the importance of clinical trials design by assessing each domain of PRECIS-2 tool. Particularly, the restrictive eligibility criteria, which excluded many patients with comorbidities, contributed to the low mean PRECIS-2 score in the eligibility criteria domain. Notably, patients with uncontrolled hyperglycaemia and/or concurrent use of other antidiabetic or anti-obesity medications were frequently excluded. These restrictions may maximise the internal validity of the trials and consequently enhance the efficacy of the intervention being studied. This could be related to the study's aim, as some trials were designed to explore how the intervention works under controlled conditions, thus including selective patients (characterised as explanatory trial). However, this approach might limit trial representativeness, impacting decision-making in real-world practice (Lu et al., 2017). Additionally, the lower mean PRECIS-2 score for the primary outcome reflects that, in most trials, a surrogate endpoint was used, which is considered explanatory. Conversely, the higher mean PRECIS-2 score for the setting domain is due to many trials being multinational and multicentre.

In addition, the pragmatic trials are designed to reflect real-world practice and enhance representativeness. Therefore, the observed association between the mean PRECIS-2 score, indicative of pragmatism, and the SAE rates may suggest that SAE rates could reflect trial representativeness. This association may also indicate the possibility of more SAEs being observed in pragmatic trials. Similarly, the pragmatic setting domain was associated with higher SAE rates, which is consistent with findings in chapter 5 that showed higher SAE rates in multinational trials. This suggests that the pragmatic trial design, as reflected by this PRECIS-2 domain, could influence SAE rates. Thus, explanatory trials, which

are less likely to be representative of real-world practice, might be associated with lower SAE rates.

6.6.3 Challenge of using PRECIS-2 tool

The PRECIS-2 tool was a suitable approach to compare with SAE as a metric of trial representativeness, especially in the absence of gold standard. However, this tool presents some challenges for retrospective assessment. Information required for scoring certain PRECIS-2 domains, particularly the recruitment domain, was often missing. Most published RCTs did not provide details in their protocols or other relevant publications about how they recruited participants. Moreover, it is time-consuming as it requires a thorough review of RCT reports to assess each of the nine domains. Specifically, each trial required approximately 60 minutes to extract the required information and an additional 40 to 60 minutes to assess each trial.

6.6.4 Comparison with other studies

The original PRECIS tool was developed to prospectively assess pragmatism when writing a trial protocol. However, this study utilises the developed PRECIS-2 tool to retrospectively assess pragmatism, as many investigators have used it for retrospective assessment for trials that have already been completed (Forbes et al., 2017). Saesen et al. (2023) conducted a study to assess the degree of pragmatism of clinical trials for antineoplastic treatments. Using the PRECIS-2 tool, they assessed 42 trials that were tagged as pragmatic in their titles. In terms of the PRECIS-2 score, their findings were consistent with this analysis; with a mean PRECIS-2 score of 3.13. Moreover, Choi et al. (2019) retrospectively assessed the pragmatism of 96 RCTs in rheumatoid arthritis and similarly found that the eligibility criteria domain was very explanatory, reflecting strict inclusion criteria, and the setting domain was pragmatic, because many trials were multinational and conducted in multiple centres.

Furthermore, previous literature that applied PRECIS-2 for retrospective assessment used the mean score of all PRECIS-2 domains to describe the pragmatism of each trial, which is similar to the approach used in this study (Palese et al., 2014; Choi et al., 2019; Sepehrvand et al., 2019; Fitzpatrick et al., 2020). Additionally, like this study, previous studies addressed domains with missing information for scoring by assigning a score of 3 (equally pragmatic/explanatory) (Devos et al., 2019; Fitzpatrick et al., 2020).

The developers of the PRECIS-2 tools assume that every clinical trial lies on a continuum, meaning they are not purely pragmatic or purely explanatory (Loudon et al., 2015). This assumption is consistent with the findings of the study on the Pragmatic Nature of Cardiovascular Clinical Trials (Sepehrvand et al., 2019), which found no trials that were completely pragmatic or explanatory in their samples.

6.6.5 Strengths and limitations

The strength of this study lies in the application of the PRECIS-2 tool on several trials (n=146), whereas many other studies used the tool on a smaller scale. This is the first study to assess the level of pragmatism for SGLT-2 inhibitors using the PRECIS-2 tool. Furthermore, subjectivity is possible when assessing the trials; also, some domains necessitate clinical insight for proper scoring. However, this issue was mitigated by consulting two diabetologists, who also reviewed the operationalised assessment criteria. These criteria were developed to create a scoring template tailored for use in SGLT-2 trials based on original PRECIS-2 criteria. The required information for scoring some domains was not available in some trials, which is considered a limitation of retrospective PRECIS-2 assessments.

6.7 Conclusion

This chapter assesses the pragmatism of the included trials using the PRECIS-2 tool and shows a medium level of trial pragmatism. PRECIS-2 tools are viable

resources that trial units can use to determine whether the trial design aligns with their intended purpose. Trials with more pragmatic designs showed higher SAE rates, suggesting that SAE rates could reflect trial representativeness. In the next chapter, I will further explore whether SAE rates can reflect trial representativeness by comparing SAE rates and PRECIS-2 scores to baseline and trial characteristics.

Chapter 7 Comparison between SAEs and PRECIS-2 tool by exploring their association with trial and baseline characteristics of SGLT-2 inhibitors

7.1 Chapter overview

This chapter further explores the usefulness of SAE rates as a measure for trial representativeness. First, it explores the trend and strength of the association of the SAE rates with trial and baseline characteristics. It also explores the association of the mean PRECIS-2 score with these characteristics. Then, it compares the two metrics based on the differences in the estimates of their associations with these characteristics.

7.2 Introduction

7.2.1 Using SAEs to assess trial representativeness and lack of gold standard

SAEs might be a feasible approach that can be used retrospectively to assess the representativeness of clinical trials. This approach can be utilised by capturing the observed SAE rates and comparing them to the expected SAE rates seen in routine healthcare. However, this approach requires access to real-world data from daily practice, often through record-linkage, where trial data are linked with external sources such as electronic health records or national registries. Exploring the association of SAE rates with trial and baseline characteristics could help assess the feasibility of SAE as a metric because these characteristics may predict SAEs (see section 7.2.3). To investigate this feasibility further, SAEs can be compared with other metrics such as PRECIS-2 tool to determine if SAE is a better measure. However, one difficulty of examining the feasibility of using SAEs as a metric for trial representativeness is the lack of real gold standard. One approach to address this issue is to use ‘fair umpire’ test, which incorporates baseline and trial characteristics as umpires to compare between SAEs and the current metric (PRECIS-2 tool). The concept of a ‘fair umpire’ test

is based on a test that reasonably discriminates for outcome of interest and is mechanistically unrelated to the comparator tests (Glasziou, Irwig and Deeks, 2008). Hence, in this study, the baseline and trial characteristics included serve as impartial referees (fair umpire), not favouring one metric (PRECIS-2) over the other (SAEs), as they are neutral and not inherently biased towards both comparator metrics. This approach for comparisons ensures a fair assessment and offers an informative decision regarding the feasibility of using SAEs as a metric for trial representativeness.

7.2.2 Justification of using these umpires as marker of trial representativeness

The selected baseline characteristics (umpires) may reflect trial representativeness. Previous research has indicated that trials often exclude older people, those with longer disease durations, and those with higher levels of HbA1c (Neven et al., 2022). Characteristics such as age, disease duration, gender, and HbA1c are likely to be affected by the restrictiveness of the eligibility criteria, which in turn may reflect the representativeness of the trial. Trials with baseline characteristics that closely resemble population in real-world practice are likely to be more representative (Kirkman et al., 2012; Kennedy-Martin et al., 2015). Moreover, trials with larger sample size and longer follow-up durations may enhance the external validity of the trial findings by providing more precise estimates and a better understanding of long-term efficacy (Faber and Fonseca, 2014; Sepehrvand et al., 2019). While these aspects do not directly reflect representativeness, they do help in judging the applicability of the findings to broader populations.

From another perspective, trials characteristics could serve as an effective marker for trial representativeness. For instance, phase 4 trials may be more representative as they target a wider community and have less restrictive criteria than phase 3 trials (Sen et al., 2018). Trials with double or triple-blinding may serve as good markers for representativeness by minimising potential reporting bias compared to single-blinded or open-label trials (Pitre et al., 2023). Moreover, industry-sponsored trials are often larger and more inclusive

than trials sponsored by other entities (Yao et al., 2013). Trial with longer duration could reflect real-world practice because they are likely to monitor patients over several years (Sen et al., 2018).

Therefore, these markers (trial and baseline characteristics) were chosen as umpires because they reflect trial representativeness and are not biased towards both measures. Other characteristics were not included as umpires because they may bias the comparison, as they were part of the PRECIS-2 assessment. For example, the type of trial settings (multicentre, multinational) was considered in scoring the setting domain of the PRECIS-2 tool. The type of trial outcome (hard or soft) and type of population analysis (intention to treat or per protocol) were also considered in scoring the primary outcome and primary analysis domains of the PRECIS-2 tool.

7.2.3 Expected association of baseline and trial characteristics with SAEs

The occurrence of SAEs may be influenced by trial characteristics and patient demographics (Table 7.1). Participants with higher ages, longer durations of diabetes, and higher HbA1c levels are generally sicker and more likely to have comorbidities, potentially increasing the risk of SAEs. Larger trials are expected to detect more events compared to smaller trials (Lessing et al., 2010). Similarly, longer trials may allow for the observation of events that might not be observed in shorter trials (Dworkin et al., 2021). Thus, these trials may result in more SAEs than shorter trials with small sample sizes. Moreover, phase 4 trials are often conducted after drug approval for the community and usually aim to provide more evidence of treatment safety. These trials often examine longer treatment durations, employ different outcome measures, and focus on specific populations such as older people (Dworkin et al., 2021). As a result, phase 4 trials are expected to exhibit more events compared to other phases (FDA, 2018). Female patients with diabetes are at a higher risk of cardiovascular disease compared to males (Y. Wang *et al.*, 2019), potentially leading to a higher incidence of SAEs in this group. Thus, a higher proportion of male participants may result in lower SAEs. Consequently, both trial and baseline characteristics

may play an essential role in influencing the incidence of SAEs in clinical trials, underscoring their importance in assessing trial representativeness.

Based on these exceptions, the feasibility of using SAE rates as a metric can be investigated by examining the association of these characteristics with SAE rates and the PRECIS-2 score, and then capturing the difference in the estimates of the associations between these two metrics will further indicate whether these characteristics (umpires) strongly favour SAEs as a metric, potentially reflecting trial representativeness (as outlined in Table 7.4). If these umpires strongly favoured SAE rates, it suggests that SAE rates may reflect trial representativeness.

Table 7.1 Expected associations of baseline/trials characteristics with the SAE rates

Variables	**Expected direction of association (if SAEs are a metric of trial representativeness)**
Baseline sample size; increase	Increase SAEs
Age (years); increase	Increase SAEs
Male, %; increase	Decrease SAEs
Duration of diabetes; increase	Increase SAEs
HbA1c, %; increase	Increase SAEs
Trial year difference (the difference between when a clinical trial was first registered and when the intervention being studied was first trialled); increase	Increase SAEs
Trial duration; increase	Increase SAEs
Sponsorship; (being Industrial sponsored trials)	Increase SAEs
Trial phase; (being Phase IV)	Increase SAEs
Blinding (being Double or more blinding)	Increase SAEs

7.3 Aim and objectives

To determine whether SAE rates in clinical trials for SGLT-2 inhibitors can be used as a marker of representativeness of those trials, I will address the following objectives:

1. Estimate the association between SAE rates and trial/baseline characteristics.

2. Estimate the association between the trial pragmatism metric (PRECIS-2 tool), as a mean score, with baseline and trial characteristics.

3. Compare SAE with the PRECIS-2 tool as metrics of trial representativeness based on the differences in the estimates of their associations with these characteristics (fair umpires).

7.4 Methods

7.4.1 Data source

Trial selection and data sources are presented in Chapter 4. For the PRECIS-2 metric, data was utilised from Chapter 6, where they were synthesised by assessing the pragmatism of the included trials using this tool.

7.4.2 Data extraction

For this analysis, baseline characteristics were extracted for each trial, including mean age, proportion of males, duration of diabetes, sample size (number of participants), and the levels of HbA1c, Additionally, trial characteristics were extracted, including types of blinding, duration of trial, trial year difference (the difference between when a clinical trial was first registered and when the intervention being studied was first trialled), types of sponsorship, and trial phases. Types of blinding were classified into single and double blinding, trial

phases were classified into phase 3 and 4, and types of sponsorship were classified to industrial and non-industrial-sponsored trials. Data extraction for SAEs is presented in Chapter 4 (section 4.4.3).

Baseline characteristics were extracted from baseline tables within trial publications using TableTidier software (https://tabletidier.org/). This software was utilised to extract and harmonise the data, as it is designed to convert structured tables with non-standard terminologies into standard formats. Trial characteristics were manually extracted from trial registries and published research papers.

The metrics of interest, indicative of trial representativeness, were the SAE rates and the mean PRECIS-2 scores. SAE data were collected from selected trials as a count of the number of participants who experienced SAEs. The rate for these events was then calculated by dividing the number of SAEs by person-time. Additionally, the analysis used SAEs at the trial level rather than individual arms; the total number of SAEs for each trial arm (treatment and control arms) was aggregated.

7.4.3 Statistical analysis

Descriptive statistics (means and standard deviation for continuous variables and counts and percentages for categorical variables) were used to summarise trial and baseline characteristics.

To address the study objectives, a bivariate Bayesian regression model was fitted using the *brms* package. This model simultaneously estimated the associations of SAE rates and PRECIS-2 scores with various trial/baseline characteristics, allowing for the borrowing of strength between outcomes and accounting for potential correlation across the two metrics (Xiaochen Yuan, 2019). This approach effectively handles complex relationships and dependencies, offering greater efficiency and precision than separate univariate models. Therefore, it is the most suitable method to compare between SAEs and PRECIS-2 based on the

differences in their associations with these characteristics. The model used default non-informative priors as set by the brms package to ensure that the conclusions drawn were primarily driven by the data rather than prior assumptions. This choice allowed the data to exert a more significant role in determining the posterior distributions, especially in the absence of strong prior knowledge specific to the relationships being modelled. For comparing SAE rates and PRECIS-2 scores, the output from the Bayesian model was used. For each covariate (mean age, duration of diabetes, sample size, proportion of males, levels of HbA1c, trial year difference, duration of trial, types of sponsorship, trial phases, and types of blinding), 20,000 samples for the parameter representing the association with SAE rates and PRECIS-2 scores were obtained. For each sample, I subtracted the SAE estimate from the PRECIS-2 estimate. I then summarised these differences via the mean and 2.5^{th} and 97.5^{th} centiles. This represents the differences in association, providing a basis for comparing PRECIS-2 and SAE. The outcomes of interest in this model were the SAE rate and the mean PRECIS-2 score. Continuous variables in the model included mean age, proportion of males, duration of diabetes, sample size, levels of HbA1c, trial year difference, and duration of trial. Binary variables were types of sponsorship, trial phases, and types of blinding.

7.5 Results

7.5.1 Descriptive statistics

The mean PRECIS-2 score across 146 trials was 3.34. A total of 21,141 SAEs were reported, involving 116,207 participants. The mean (SD) age of the included participants was 58.7 years (4.46). The average duration of type 2 diabetes, and the average HbA1c level were 9.1 (3.65) 8.1 % (0.46), respectively. The trial and baseline characteristics are summarised in table 7.2 and table 7.3.

Table 7.2 Summary of baseline and trial characteristics (continuous variables)

Variables	Minimum	Maximum	Mean (SD)
Number of participants	25	17160	808.9
Mean age (years)	51.35	77.2	58.7 (4.46)
Male, %	42	85	59 (1)
Duration of diabetes (years)	1.11	19.6	9.1 (3.65)
HbA1c, %	6.60	10.20	8.1 (0.46)
Year difference (years)	0	12.53	3.87 (3.4)
Trial duration (years)	0.08	7.38	0.88 (0.93)

Table 7.3: Summary of baseline and trial characteristics (binary variables)

Variables	Count (%)
Total	146 (100%)
Sponsorship	
Industrial sponsored trials	119 (81%)
Non-industrial sponsored trials	27 (19%)
Trial phase	
Phase III	107 (73%)
Phase IV	39 (27%)
Blinding	
Single or open-label	27 (19%)
Double or more blinding	119 (81%)

7.5.2 The association of trial/baseline characteristics with SAEs and PRECIS-2

The SAE rates and PRECIS-2 score were positively associated with some trial and baseline characteristics. Duration of diabetes was strongly associated with both the SAEs rate and the mean PRECIS-2 score, and the confidence interval did not include the null (β = 0.81; 95% CI 0.23 to 1.39, β = 0.89; 95% CI 0.28 to 1.5, respectively). This means that trials with long disease duration are linked to increased SAE rates and higher PRECIS-2 score. Moreover, no observed difference in the estimate was found in the association between the two metrics (β = -0.08; 95% CI -0.89 to 0.75), suggesting that diabetes duration has similar association with both measures (Table 7.4).

In addition, baseline sample size was positively associated with the PRECIS-2 score (β = 0.27; 95% CI 0.12 to 0.43), while no association was observed with the SAE rates (β = 0.03; 95% CI -0.12 to 0.18). Also, the difference in the estimate between both metrics was observed (β = -0.24; 95% CI -0.45 to -0.03), favouring the PRECIS-2 score, and the confidence interval did not include the null (Table 7.4). This implies that sample size is more strongly associated with PRECIS-2 than with SAEs, indicating that as the number of trial participants increases, the mean PRECIS-2 score tends to increase. Conversely, age had notable negative

association with PRECIS-2 (β = -0.58; 95% CI -1.11 to -0.05). Moreover, it showed positive association with the SAE rates, however the confidence interval included the null (β = 0.39; 95% CI -0.09 to 0.90). The large difference in the estimate suggests that age strongly favour SAE as a metric of trial representativeness more than PRECIS-2, and the confidence interval did not include the null (β = 0.97; 95% CI 0.26 to 1.70).

Industrial-sponsored trials had a strong positive association with SAE rates (β = 0.78; 95% CI 0.10 to 1.46). However, the difference in the estimate did not strongly favour SAE (β = 0.17; 95% CI -0.83 to 1.16). Moreover, trial year difference was positively associated with SAE rates (β = 0.55; 95% CI 0.06 to 1.03). The difference in the estimate favoured SAE; however, the confidence interval included the null (β = 0.69; 95% CI -0.02 to 1.39). Type of blinding was positively associated with SAE rates while negatively associated with PRECIS-2. The difference strongly favoured SAE, and the confidence interval did not include the null (β = 0.72; 95% CI 0.02 to 1.43). Furthermore, no notable associations were observed between trial duration, trial phase, and level of HbA1c with both SAEs and PRECIS-2, and no large differences in estimates were found; all confidence intervals included the null.

Table 7.4 Comparison between SAE and PRECIS-2

Metrics of Representativeness								
Umpire variables	**SAE**		**PRECIS-2**		**Difference in estimates**		**Significant favouring *1**	**Direction of favouring *2**
	β	**95% CI**	**β**	**95% CI**	**β**	**CI**		
Type of blinding; Double or more	0.44	-0.06 to 0.93	-0.28	-0.79 to 0.22	0.72	0.02 to 1.43	SAE	SAE
Trial phases; Phase IV	-0.23	-0.79 to 0.33	-0.06	-0.66 to 0.53	-0.16	-0.97 to 0.65	Neutral	PRECIS-2
Sponsorship; industrial-sponsored trial	0.78	0.10 to 1.46	0.62	-0.11 to 1.33	0.17	-0.83 to 1.16	Neutral	SAE
Trial duration/years	0.02	-0.18 to 0.21	-0.12	-0.33 to 0.08	0.14	-0.14 to 0.42	Neutral	SAE
Year difference/decade; (the difference between when a clinical trial was first registered and when the intervention being studied was first trialled)	0.55	0.06 to 1.03	-0.14	-0.66 to 0.37	0.69	-0.02 to 1.39	Neutral	SAE
Baseline sample size	0.03	-0.12 to 0.18	0.27	0.12 to 0.43	-0.24	-0.45 to -0.03	PRECIS-2	PRECIS-2
Age/decade	0.39	-0.09 to 0.90	-0.58	-1.11 to -0.05	0.92	0.26 to 1.70	SAE	SAE

Proportion of male	-1.14	-2.61 to 0.33	-0.73	-2.31 to 0.84	-0.41	-2.61 to 1.81	Neutral	SAE
Duration of diabetes/decade	0.81	0.23 to 1.39	0.89	0.28 to 1.50	-0.08	-0.89 to 0.75	Neutral	Neutral
HbA1c	-0.09	-0.49 to 0.31	-0.26	-0.71 to 0.18	0.18	-0.42 to 0.78	Neutral	SAE

*[1] **Criteria for significant favouring (CI does not include the null):** SAE: if the difference in association significantly favours SAEs. PRECIS-2: if the difference in association significantly favours the PRECIS-2. Neutral: if there is no difference in associations between metrics.

*[2] **Criteria for direction of favouring (direction of point estimate):** SAE: if the difference estimate in association is in the direction of SAEs. PRECIS-2: if the difference estimate in association is in the direction of the PRECIS-2. Neutral: if the difference in associations between metrics was close to zero (null).

7.6 Discussion

7.6.1 Summary of main findings

Some baseline/trial characteristics showed a degree of association with the two trial representativeness metrics (SAEs and PRECIS-2). These associations were consistent and followed the same direction, although the magnitude of the associations varied. Therefore, the findings of this analysis, to some extent, align with the initial expectations that these characteristics might be associated with an increase in the SAE rate (Table 7.1).

Regarding the differences in associations of umpires with both metrics, only age and type of blinding umpires strongly favoured SAE more than PRECIS-2. For other umpires, including trial year difference, sponsorship, trial duration, proportion of males, and HbA1c, the direction of these differences generally favoured SAEs, although the confidence intervals included the null. On other hand, two umpires of baseline sample size and trial phases favoured PRECIS-2; however, for trial phases the confidence intervals included the null. Moreover, only one umpire, the duration of diabetes, showed no difference between the metrics.

7.6.2 Interpretation and comparison with literature

Longer duration of diabetes for the included participants revealed increases in the rates of SAE as well as the pragmatism score (PRECIS-2). Severity and complications of diabetes tend to increase over time (Ghouse et al., 2020). Patients with longer duration of diabetes are more likely to have higher risk characteristics, clinical impairment, and more comorbidity (Badano et al., 2003), eventually expected to experience more SAEs than other. Similarly, the pragmatic trials that resemble routine care are designed to include patients with long disease duration to reflect real-world practice (Loudon et al., 2015). This observed finding was consistent with the prior expectation that disease duration could contribute to trial representativeness. Trials that included participants

with longer disease duration are expected to have higher SAEs than trials with shorter disease duration. Although no difference was observed in the estimate between SAEs and PRECIS-2, this indicates that diabetes duration factor may influence both metrics, potentially reflecting the trial representativeness.

The sample size partially meets the expectation as it positively associated with the PRECIS-2 but not associated with SAEs. The difference found in the estimate indicates that an increase in sample size likely results in a higher PRECIS-2 score without a notable increase in SAE rates. This finding is consistent with a study that investigated the nature of cardiovascular trials using the PRECIS-2 tool, which found that pragmatic trials, as reflected by their high scores, had larger sample sizes (Sepehrvand et al., 2019).

In addition, the positive association observed between SAEs and age might be attributed to the recruitment for older participants who have a greater risk of SAEs (Luo et al., 2016; Datta et al., 2018). Excluding older people from clinical trials may limit the generalisability. Thus, patient characteristics like age may consider an important factor that can help in predicting SAEs. Increasing age is associated with higher SAEs rates, which could help assess trial representativeness. Hence, age meets expectations and could serve as a predictive marker for representativeness, particularly in relation to SAEs. On the other hand, age was unexpectedly associated with PRECIS-2, indicating that recruiting older participants may lead to lower score. This finding contradicts existing literature with regard to the PRECIS-2 tool, which claims that an increase in age contributes to obtaining pragmatic trials (Loudon et al., 2015). Nevertheless, age is more relevant to patient characteristics, as reflected by SAEs, than to trial design, which is captured by PRECIS-2. Indeed, the difference in the estimate suggests that age strongly favours SAE rates as a metric more than PRECIS-2.

Additionally, umpire of year difference was associated with higher SAE rates. The direction of the difference in association favoured SAE. This might be attributed to the fact that recent trials tend to be more inclusive as they aim to examine the effectiveness of already-marketed drugs, compared to earlier trials that are

still in the licensing process which require strict protocols to maximise the internal validity. Moreover, industrial-sponsored trials were strongly associated with higher SAE rates, which may be due to these trials typically involving larger enrolments and a more diverse group of participants (Versavel et al., 2023). However, the difference in association did not strongly favour SAEs. The type of blinding strongly favoured SAE, which meets prior expectations that double-blinded trials, when compared to single-blinded or open-label trials, could be more representative by keeping participants and investigators unaware of the treatment assignment (Monaghan et al., 2021).

Although some studies have used trial eligibility criteria and patient characteristics to assess trial representativeness, no study has directly compared SAE rates to the PRECIS-2 score. Nevertheless, two previous studies assessed trial representativeness by comparing SAE rates in trial populations to SAE rates in the target population and found that trial populations relatively have lower SAE rates than population in routine care (Hanlon et al., 2021, 2022). Another study conducted on psoriasis compared SAE rates between trial-eligible and trial-ineligible patients and found lower SAE rates among eligible patients (Garcia-Doval et al., 2012). Additionally, another study investigated the pragmatism of cardiovascular trials and whether it changed over time, comparing trial characteristics such as sample size and duration of follow-up with the mean PRECIS-2 score. They found that the level of pragmatism increased over time. Trials with larger sample sizes and longer follow-up durations also had higher PRECIS-2 scores (Sepehrvand et al., 2019).

7.6.3 Implications

Assessing the representativeness of clinical trials is challenging, especially when there is no gold standard metric. The differences in estimates of association of SAE and PRECIS-2 with some baseline/trial characteristics (umpires) favoured SAE over PRECIS-2. This may indicate the possibility of SAE as a proxy for assessing trial representativeness. However, not all umpires favour SAEs, so, exploring the plausible association could be further investigated by including other characteristics, such as socioeconomic status, that might reflect trial

representativeness. The strong association observed between diabetes duration and SAEs is reassuring, offering valuable insight into the influence of disease duration on capturing SAEs and potentially reflecting real-world disease duration in routine care. The association with industry sponsorship should be considered in predicting SAEs and assessing trial representativeness. Specifically, future research should investigate whether industrial-sponsored trials report higher SAEs compared to those sponsored by other entities and examine how these differences might impact the representativeness. While the current analysis involves a range of trial and baseline characteristics, incorporating additional variables and a higher number of trials could potentially enhance the associations observed.

7.6.4 Strengths and limitations

This analysis has several strengths. First, it could explore the SAE rates from different aspects by including multiple covariates of both trial and baseline characteristics, rather than a single characteristic. Moreover, the clinical trials were selected through a systematic review and were obtained from different trial registries with various sponsors. Additionally, the trials were conducted in different settings across multiple countries, which contributes to efficient comparison. However, there were some limitations. The failure to reach significant associations between all trial/baseline characteristics and the SAE rates might be attributed to the relatively small sample size (i.e. the small number of clinical trials included). Moreover, some of the selected umpires may not directly assess trial representativeness. For example, while larger sample size and longer follow up duration are useful for providing more precise estimates and understanding of long-terms efficacy, they may not be perfect markers of representativeness. Additionally, factors such as multimorbidity and frailty, which may relate to trial representativeness and could influence SAE rates, were not included in this analysis because such data are rarely reported in trials.

7.7 Conclusion

Some trial and baseline characteristics were associated with both SAEs and PRECIS-2. Although the direction of the difference in associations for most umpires generally favoured SAE rates as a metric of trial representativeness, age and type of blinding were the only umpires that strongly favoured SAE rates over PRECIS-2. Therefore, these findings suggest that it may be premature to conclusively determine that the SAE rate is a reliable metric of trial representativeness. In the next chapter, I will further explore whether SAE rates reflect trial representativeness by examining the associations between SAE rates and eligibility criteria.

Chapter 8 Association between eligibility criteria and SAE rates in clinical trials for SGLT-2 inhibitors

8.1 Chapter overview

In this chapter, further examination will be conducted to analyse SAE rates against the elements of eligibility criteria to determine whether SAEs are associated with trials eligibility criteria, potentially reflecting the representativeness of clinical trials.

8.2 Background

Clinical trials are one of the most reliable sources of evidence on treatment effectiveness. However, there is uncertainty regarding the applicability of trial findings due to the lack of representativeness. Eligibility criteria are an essential component in conducting RCTs. Trials with broader eligibility criteria are expected to have more SAEs and to be more representative (Kim et al., 2017). Inappropriate eligibility criteria may enrol participants who do not accurately represent the community (Li et al., 2020). Where this results in trial participants who are healthier, on average, than people treated in the community, the SAE rate is also likely to be lower. It has been found that the SAEs observed in clinical trials were lower than those in routine care (Hanlon et al., 2022). Underrepresented populations, such as older people, comorbid patients, and those with a long duration of T2DM, heart failure, and renal dysfunction, represents a high proportion of patients in the community. These patients are expected to experience more SAEs than other populations, and excluding them through restrictive criteria may not accurately capture the true incidence of SAEs (Florisson et al., 2021). Therefore, adopting trials with permissive eligibility criteria may better reflect the true SAEs expected in routine practice, potentially enhancing the trials representativeness. Thus, the association between SAEs rates and eligibility criteria could address this issue and suggest the feasibility of using SAEs as a metric for trial representativeness. In this

chapter, the focus will be on how the eligibility criteria, whether they were permissive or restrictive, are associated with SAE rate.

8.3 Aim and objectives

In this chapter, in order to determine whether SAEs reported in clinical trials for SGLT-2 inhibitors can be used as a marker of trial representativeness, I will examine the association between eligibility criteria and SAE rate using two different approaches to quantify eligibility criteria:-

1. Using a decision tree to assign each trial to an eligibility status.

2. Using a single measure where each criterion contributes to an overall score.

8.4 Methods

8.4.1 Data source

The elements of eligibility criteria for 146 clinical trials involving nine medications of SGLT-2 inhibitors were manually extracted from ClinicalTrials.gov. For trials on other registries, the criteria were manually extracted from published research papers or study documents (e.g., Clinical Study Reports). Data sources and extraction methods for SAEs and their timeframes are presented in Chapter 4.

8.4.2 Data extraction and harmonisation

Initially, all eligibility criteria across the trials were identified. Subsequently, a diabetologist and an epidemiologist were consulted to review the identified criteria and advise on which were relevant to the context and needed to be extracted. Accordingly, 22 common exclusion criteria were extracted and

harmonised. After several meetings, criteria that conveyed similar information were dropped, leaving 16 key exclusion criteria to be included in the analysis. Moreover, the cutoff for each criterion was determined based on clinical practice and input from the clinicians, as illustrated in table 8.2 (detail in section 8.4.3).

The common eligibility criteria extracted included both continuous and categorical variables. Continuous variables included body mass index (BMI), age, systolic blood pressure (SBP), duration of diabetes, HbA1c, estimated Glomerular Filtration Rate (eGFR), and creatinine clearance (CrCl). These criteria were extracted as minimum and maximum values of exclusion level, as presented in each eligibility criterion (Table 8.1). For example, if a trial's criteria specified that subjects with a BMI of 18-45 kg/m^2 are eligible, I extracted 18 and 45 as the minimum and maximum values, respectively, for exclusion. Moreover, some trials reported the criteria for HbA1c in mmol/l, which I standardised to a percentage (%) using this formula: (HbA1c (mmol/mol) / 10.929) + 2.15. Also, criteria for the duration of diabetes were reported in days, weeks, or months, which I harmonised to weeks.

Furthermore, categorical variables included heart failure classes, glycaemic control medicines, stability of diabetes treatment, non-cardiovascular comorbidities, other medications, alcohol/drug abuse, investigator discretion, previous surgery, and participation in other study. These criteria were extracted and harmonised based on their exclusion status. For example, non-cardiovascular comorbidities criteria were categorised as follows: comprehensive – excludes if presence of all/comprehensive comorbidities, any other – excludes if presence of other comorbidities, cautions/contraindications – excludes only based on reasonable exclusions related to comorbidities appearing as cautions or contraindications to the intervention, and none – does not exclude based on other comorbidities. Heart failure criteria were extracted according to the New York Heart Association (NYHA) class. The criteria for glycaemic control medicines were extracted as mono, dual, or triple, based on the number of antidiabetics allowed. Similarly, criteria for exclusion based on medications other than antidiabetics were extracted and harmonised into three levels: steroids, other,

or none (Table 8.1). Criteria for the stability of diabetes treatment were extracted as stable or unspecified. Other criteria of alcohol/drug abuse, investigator discretion, previous surgery, and participation in other study, were extracted as yes or no. For details regarding the extraction of eligibility criteria, as well as definitions for all variables and their highly restrictive cutoffs, please refer to table 8.1 and table 8.2.

Table 8.1 Extraction and harmonisation of eligibility criteria and their definitions

Eligibility criteria elements		
Criterion	**Extracted value**	**Interpretation**
Non-cardiovascular comorbidities	Comprehensive	Excludes if presence of all/comprehensive comorbidities.
	Any other	Excludes if presence other comorbidities.
	Cautions/Contraindications	Excludes only on reasonable exclusions based on comorbidities which appear cautions or contraindications to the intervention.
	None	Does not exclude based on other comorbidities.
Other medications	Steroids	Excludes only if high dose steroids, which would generally require either sulfonylurea or Insulin treatment.
	Other	Excludes based on other non-glucose lowering drugs.
	None	Does not exclude based on other non-glucose lowering drugs.
Glycaemic control medicines	Triple	Allow inclusion of participants with three antidiabetics.
	Dual	Allow inclusion of participants with two antidiabetics.

	Mono	Allow inclusion of participants with only one antidiabetic.
Stability of treatment	Stable	Requires a stable dose of diabetes treatment.
	Unspecified	Dose not specify a requirement for a stable dose of diabetes treatment.
Alcohol and drugs	Yes	Exclude participants with history of alcohol or drug abuse.
	No	Does not require to exclude participants with history of alcohol or drug abuse.
Investigator discretion	Yes	Exclude participants based on unspecified investigator discretion.
	No	Does not exclude participants based on unspecified investigator discretion.
Investigations	Yes	Exclude participants according to investigation not part of routine care like C-peptide.
	None	Does not exclude participants according to investigation not part of routine care.
Previous surgery	Yes	Excludes if a participant had previous surgery.
	No	Does not exclude if a participant had previous surgery.
Participation in other study	Yes	Excludes if a participant has participated in a previous study
	No	Does not exclude if a participant has participated in a previous study
Age, BMI, SBP, HbA1c, diabetes duration, eGFR, and CrCl	Continuous value	The Maximum value for exclusion (e.g., BMI > 45 kg/m^2 were excluded).

	Continuous value	The Minimum value for exclusion (e.g., BMI < 18 kg/m^2 were excluded).

8.4.3 Measure (restrictiveness of eligibility criteria)

To explore the association between SAE rates and eligibility criteria, two analysis approaches were applied. These approaches aimed to simplify the complexity of eligibility criteria elements into a single variable. This step was important for reducing the dimensionality and facilitating the creation of plots and regression models. Therefore, redundant variables that reflected similar information of other criteria were dropped. For example, fasting blood glucose (FBG) criterion was dropped because it conveyed similar information of Hb1Ac criterion. For first approach, highly restrictive cutoffs were defined for each element of eligibility criteria to assign the restrictiveness of each trial. Each criterion was categorised into two levels (restrictive-unrestrictive) by defining a threshold that represents the highly restrictive cutoff. For example, a criterion such as BMI, which excludes participants with a BMI over 45 kg/m^2, was defined as highly restrictive (cutoffs for each criterion are presented in table 8.2). These thresholds were defined based on clinician judgment (from a diabetologist), which reflects common practice in routine care. Using these cutoffs, each trial was labelled as either restrictive or permissive. For example, if a trial excluded patients older than 60 years, it was labelled as restrictive; if not, the process proceeded to the next criterion, which is BMI. If it excluded participants with a BMI over 45 kg/m^2, the trial was labelled as restrictive; if not, the process proceeded to the next criterion. If no other criteria excluded participants based on highly restrictive cutoffs (as illustrated in table 8.2), then the trial was labelled as permissive. Subsequently, I examined the association between the trial eligibility status and SAE rates.

For the second approach, an algorithmic approach was applied to generate a continuous restrictiveness score for eligibility criteria that ranges from least to most restrictive, which reflects the restrictiveness of the eligibility criteria for each trial (Table 8.3). These scores were calculated by summing the score for

each criterion and analysing it against SAE rates. The combination of these analysis approaches provides a comprehensive framework for addressing the research objectives and assessing the association between eligibility criteria and the rates of SAE.

Table 8.2 Highly restrictive cutoffs and levels

Eligibility criteria	**Highly restrictive criteria if value extracted for exclusion was:**
Age	Max value is ≤ 60 years old.
BMI	Min value is ≥ 45, and/or Max value is ≤ 30 kg/m^2.
HbA1c	Min value is ≥ 11% (97mmol/mol), and/or Max value is ≤ 8% (64mmol/mol).
T2DM Duration	Min duration of disease is ≥ 120 months, and Max duration is ≤ 12 months.
Heart failure	Class I-IV (exclude all heart failure).
SBP	Trial Max SBP is ≤ 140 mmHg.
eGFR	Trial Min eGFR is ≥ 60 ml/min/1.73m2.
CrCl	Trial Min CrCl is ≥ 100 ml/min.
Medicines	Other.
Non-CV Comorbidities	Comprehensive.
Previous surgery	Yes.
Participation in other study	Yes.
Stability of treatment	Stable.
Alcohol and drugs	Yes.
Investigations	Yes.
Glycaemic control medicines	Mono.
Investigator discretion	Yes.

Table 8.3 Restrictiveness score for each criterion based on specific cutoffs

Criterion	Condition	Restrictiveness score given
BMI; Minimum value for exclusion	If the minimum value for exclusion was ≥40 kg/m^2 (i.e. at least morbidly obese).	4 (most restrictive)
	If the minimum value for exclusion was ≥30 but ≤40 kg/m^2 (i.e. at least obese).	3 (middle restrictive)
	If the minimum value for exclusion was ≥25 kg/m^2 but ≤30 kg/m^2 (i.e. at least overweight).	2 (middle restrictive)
	If no minimum value for exclusion or was ≥18.5 kg/m^2 but <25 kg/m^2 (i.e. normal weight).	1 (least restrictive)
BMI; Maximum value for exclusion	If the maximum value for exclusion was ≤35 kg/m^2.	3 (most restrictive)
	If the maximum value for exclusion was ≥35 but ≤40 kg/m^2.	2 (middle restrictive)
	If no maximum value for exclusion or was ≥40 kg/m^2.	1 (least restrictive)
Age; Minimum value for exclusion	If the minimum value for exclusion was ≥60 years.	5 (most restrictive)
	If the minimum value for exclusion was ≥40 years but ≤60 years.	4 (middle restrictive)
	If the minimum value for exclusion was ≥30 years but ≤40 years.	3 (middle restrictive)
	If the minimum value for exclusion was ≥21 years but ≤30 years.	2 (middle restrictive)
	If no minimum value for exclusion or was ≥18 years but ≤21 years.	1 (least restrictive)
Age; Maximum value for exclusion	If the maximum value for exclusion was ≤60 years.	5 (most restrictive)

	If the maximum value for exclusion was ≥60 but ≤65 years.	4 (middle restrictive)
	If the maximum value for exclusion was ≥65 but ≤75 years.	3 (middle restrictive)
	If the maximum value for exclusion was ≥75 but ≤85 years.	2 (middle restrictive)
	If no maximum value for exclusion or was ≥85 years.	1 (least restrictive)
HbA1c; Minimum value for exclusion	If the minimum value for exclusion was ≥80 mmol/mol (9.5%).	4 (most restrictive)
	If the minimum value for exclusion was ≥69.5 mmol/mol (8.5%) and ≤80 mmol/mol (9.5%).	3 (middle restrictive)
	If the minimum value for exclusion was ≥58.5 mmol/mol (7.5%) and ≤69.5 mmol/mol (8.5%).	2 (middle restrictive)
	If no minimum value for exclusion or was ≥53 mmol/mol (7%) and ≤58.5 mmol/mol (7.5%).	1 (least restrictive)
HbA1c; Maximum value for exclusion	If the maximum value for exclusion was ≤ 75 mmol/mol (9%).	4 (most restrictive)
	If the maximum value for exclusion was ≥75 mmol/mol (9%) but ≤86 mmol/mol (10%).	3 (middle restrictive)
	If the maximum value for exclusion was ≤97 mmol/mol (11%) and ≥ 86 mmol/mol (10%).	2 (middle restrictive)
	If no maximum value for exclusion or was ≥97 mmol/mol (11%).	1 (least restrictive)
eGFR; Minimum value for exclusion	If the minimum value for exclusion was ≥60 ml/min/1.73 m^2.	3 (most restrictive)
	If the minimum value for exclusion was ≥45 ml/min/1.73 m^2.	2 (middle restrictive)

	If no minimum value for exclusion or was ≥30 ml/min/1.73 m^2.	1 (least restrictive)
CrCl; Minimum value for exclusion	If the minimum value for exclusion was ≥100 ml/min.	3 (most restrictive)
	If the minimum value for exclusion was ≥60 ml/min.	2 (middle restrictive)
	If no minimum value for exclusion or was ≤60 ml/min.	1 (least restrictive)
SBP; Minimum value for exclusion	If the minimum value for exclusion was ≥120 mmHg.	3 (most restrictive)
	If the minimum value for exclusion was ≥100 mmHg and ≤120 mmHg.	2 (middle restrictive)
	If no minimum value for exclusion or was ≤100 mmHg.	1 (least restrictive)
SBP; Maximum value for exclusion	If the maximum value for exclusion was ≤140 mmHg.	3 (most restrictive)
	If the maximum value for exclusion was ≥140 and ≤180 mmHg.	2 (middle restrictive)
	If no maximum value for exclusion or was ≥180 mmHg.	1 (least restrictive)
Heart failure	Excludes Class I-IV.	4 (most restrictive)
	Excludes Class II-IV or III-IV.	3 (middle restrictive)
	Excludes Class IV.	2 (least restrictive)
	No exclusion.	1 (least restrictive)
Alcohol and drugs	Exclude participants with history of alcohol or drug abuse.	2 (most restrictive)
	Does not require to exclude participants with history of alcohol or drug abuse.	1 (least restrictive)
Stability of treatment	Requires a stable dose of diabetes treatment.	2 (most restrictive)

	Dose not specify a requirement for treatment stability.	1 (least restrictive)
Glycaemic control medicines	Allow inclusion of participants with only one antidiabetic.	3 (most restrictive)
	Allow inclusion of participants with two antidiabetics.	2 (middle restrictive)
	Allow inclusion of participants with three or more antidiabetics.	1 (least restrictive)
Other medicines	Excludes participants based on other non-glucose lowering drugs (anti-obesity or antihypertensive drugs).	2 (most restrictive)
	Does not exclude participants based on other non-glucose lowering drugs.	1 (least restrictive)
Non-cardiovascular comorbidities	Excludes if presence of all/comprehensive comorbidities.	3 (most restrictive)
	Excludes if presence other comorbidities (not cautions or contraindications)	2 (middle restrictive)
	Does not exclude based on other comorbidities or excludes only on reasonable exclusions based on comorbidities which appear in BNF as cautions or contraindications to the intervention.	1 (least restrictive)
Investigator discretion	Allows unspecified investigator discretion to exclude participants	2 (most restrictive)
	Does not allow unspecified investigator discretion to exclude participants	1 (least restrictive)
Previous surgery	Excludes if a participant had a previous surgery.	2 (most restrictive)

	Does not exclude if a participant had a previous surgery.	1 (least restrictive)
Participation in other study	Excludes if a participant participated in a previous study	2 (most restrictive)
	Does not exclude if a participant participated in a previous study	1 (least restrictive)

8.4.4 Statistical analysis

The eligibility criteria across all the trials were summarised using descriptive statistics (means and standard deviation for continues criteria and counts and percentages for categorical criteria).

For both analysis approaches, Poisson regression models were fitted to examine the association between eligibility criteria and the SAE rates. However, overdispersion was observed in the Poisson models, which was assessed by comparing the variance to the mean, where it was found that the variance was greater than the mean. Furthermore, overdispersion was confirmed by calculating the ratio of the residual deviance to the residual degrees of freedom, which was significantly greater than 1, indicating overdispersion. Therefore, negative binomial models were fitted for both analysis approaches. The first analysis used negative binomial model to examine the association between highly restrictive criteria and SAE. The outcome of interest was the SAE count, and the variable representing highly restrictive criteria was coded as binary, categorising each trial as either restrictive or permissive, coded as (1/0). The second analysis employed also negative binomial model to examine the association between restrictiveness scores of eligibility criteria and the SAE rates. Another negative binomial model was fitted to examine the association between individual restrictive criteria and SAE rates. To account for variations in person-time across trials, an offset was included in each model.

8.5 Results

8.5.1 Characteristics of eligibility criteria

The included eligibility criteria were 16 variables that included medical history, patient characteristics, laboratory values, comorbidities, and medication history. The trials were classified as either restrictive or permissive trial based on the defined restrictive cutoffs. A summary of SAE counts and the number of participants in each trial is presented in table 8.4.

The highest exclusion value for the age criteria was 130 years. Additionally, the maximum exclusion values for HbA1c, SBP, and BMI were 12.5%, 181 mmHg, 50 kg/m^2, respectively. Summaries of the categorical and continuous criteria are presented in table 8.5 and table 8.6.

Table 8.4 Summary statistics for eligibility criteria status

Outcomes	Restrictive RCTs	Unrestrictive RCTs	Total
No. of Trials	122 (84%)	24 (16%)	146
No. of Participants	71,202 (61%)	45,005 (39%)	116,207
No. of SAEs Reported	8946 (42%)	12,195 (58%)	21,141

Table 8.5 Summary statistics of continuous criteria of SGLT-2 trials

Criteria	Minimum	Maximum	Mean (SD)
Minimum age; years	18	65	20.9 (7.01)
Maximum age; years	65	130	81.8 (16.56)
Minimum BMI; kg/m^2	18	29	20.6 (2.47)
Maximum BMI; kg/m^2	35	50	44.4 (2.24)
Maximum SBP; mmHg	140	181	169.9 (10.94)
Minimum HbA1c; %	6	8	7 (0.38)
Maximum HbA1c; %	8	12.5	10.3 (0.72)
Minimum Diabetes duration; weeks	8	48	15.8 (12.67)
Minimum eGFR; ml/min/1.73m2	15	60	48.7 (12.58)
Minimum CrCl; ml/min	50	60	58 (4.14)

Table 8.6: Summary statistics of categorical criteria of SGLT-2 trials

Criteria	Count (%)
Non-cardiovascular comorbidities	146
Comprehensive	39 (27%)
Any other	54 (37%)
None	53 (36%)
Heart failure	
Class III-IV	44 (30%)
Class II-IV	2 (1%)
Class IV	28 (19%)
Do not exclude even if they have HF	72 (50%)
Other medications	
Other	46 (32%)
None	100 (68%)
Glycaemic control medicines	
Triple	70 (48%)
Dual	61 (42%)
Mono	15 (10%)
Stability of treatment	
Stable	68 (47%)
Unspecified	78 (53%)
Alcohol and drugs	
Yes	41 (28%)
No	105 (72%)
Investigator discretion	
Yes	56 (38%)
No	90 (62%)
Previous surgery	
Yes	32 (22%)
No	114 (78%)
Participation in other study	
Yes	39 (27%)
No	107 (73%)

8.5.2 Association between eligibility criteria and SAE rates

The analysis using Poisson model showed association between restrictive eligibility criteria and a reduction in SAE rates within clinical trials of SGLT-2 inhibitors. Trials with restrictive criteria had decreased SAE rates by 25%, the confidence interval did not include the null (IRR 0.75, 95% CI 0.73-0.77). Furthermore, as shown in table 8.7, the results of negative binomial model showed a similar association between restrictive criteria and SAE rates. Moreover, the analysis aimed to explore which individual eligibility criteria had a higher association with SAE rates compared to other criteria. The exclusion of participants with impaired renal function, as defined by a highly restrictive cutoff for eGFR (>60 ml/min/1.73m2), was significantly associated with a reduction in SAE rates (IRR 0.64, 95% CI 0.60-0.68). Similarly, trials that excluded patients based on the stability of diabetes treatment were associated with decreased SAE rates (IRR 0.64, 95% CI 0.62-0.66). The most restrictive criterion associated with decreased SAE rates was the criterion for glycaemic control medicines (IRR 0.38, 95% CI 0.33-0.45).

8.5.3 Association between eligibility criteria score and SAE rates

The restrictiveness score of eligibility criteria was analysed against SAE rates. It showed that an increase in the total sum of restrictiveness scores was associated with a 41% decrease in the SAE rates, the confidence interval did not include the null (IRR 0.59, 95% CI 0.55-0.63). Additionally, the negative binomial model showed similar association as observed in the Poisson model (Table 8.7).

Table 8.7 Models result for the associations between SAE rates and eligibility criteria

Models	IRR	95% CI	P-value
Highly restrictive criteria and SAE rates			
Poisson	0.75	0.73-0.77	0.001
Negative Binomial	0.71	0.53-0.93	0.017
Restrictiveness score of eligibility criteria and SAE rates			
Poisson	0.59	0.55-0.63	0.001
Negative Binomial	0.49	0.33-0.72	0.001

8.6 Discussion

8.6.1 Summary of main findings

There is uncertainty regarding the representativeness of clinical trials, which affects extrapolation of trial findings to a broader population. This chapter explores the feasibility of using SAEs as a metric for trial representativeness by comparing the eligibility criteria to the SAE rates using two approaches. The eligibility criteria of RCTs for SGLT-2 inhibitors demonstrated a decrease in SAEs with highly restrictive criteria (IRR 0.75) compared to unrestrictive criteria (IRR 1.33). The restrictiveness score of eligibility criteria also showed a decrease (41%) in the SAE rates (IRR 0.59).

8.6.2 Interpretation

The findings for the association suggest that the trial eligibility criteria may predict the incidence of SAEs. The decrease in SAEs observed with restrictive criteria compared to permissible criteria could be attributed to trial designs that aim to exclude individuals with certain medical conditions who are more susceptible to SAEs. For example, criteria such as stability of diabetes treatment, which require the patients to be on a stable dose of antidiabetics, might result in lower SAE rates compared to those with unstable doses commonly

seen in practice. By excluding these individuals and enrolling healthier participants, trials may capture a relatively lower rates of SAE. If such criteria were applied to real-world practice, different SAE rates might be observed compared to what is typically expected, potentially resulting in relatively lower SAEs, as observed in some RCTs.

Criterion for glycaemic control medicine, as defined by including only patients on a monotherapy regimen of antidiabetics, was the most restrictive criterion associated with a greater decrease in SAE rates compared to other criteria. This criterion is potentially intended to include patients with adequately controlled glycaemia, who, as a result, may experience fewer SAEs compared to those on dual or triple therapy of antidiabetics with poor glycaemic control. While this criterion may appear restrictive, it often has a reasonable rationale, as trials, at least in some cases, may implement such criteria to define different clinical scenarios. For example, evaluating whether SGLT-2 inhibitors work as initial first line-therapy (versus metformin) poses a different clinical question than assessing their usefulness as an add-on therapy in people not well-controlled on metformin. Thus, restrictive criteria are not necessarily wrong and may not all be about excluding certain people; rather, they are about being precise in what is being tested. Nevertheless, the broader question of whether these findings then reflect real-world practice remain an important consideration, particularly where patients often have more complex treatment regimens.

Applying stringent criteria to clinical practice may render the majority of routine patients ineligible to participate in clinical trials, thus potentially limiting the representativeness of the trial. For instance, using specific eligibility criteria, like renal disease, and implementing strict standards that exclude participants with eGFR higher than 60 ml/min/1.73m^2, may result in the exclusion of individuals commonly encountered in routine care.

Conducting clinical trials that closely resemble routine clinical practice by implementing permissive and inclusive eligibility criteria, which include underrepresented and high-risk people like older adults and patients with comorbidities, may show higher SAE rates.

8.6.3 Comparison with other studies

Based on the association between eligibility criteria and SAEs, the included participants may not highly represent the real-world population because lower SAEs were observed with restrictive eligibility criteria. It is also expected that the SAE rates in routine care populations would be higher than in trial populations. Study conducted by Hanlon et al. (2022) compared the rates of SAE in trial participants to the SAE rates in routine care patients. They found that the SAE rates were higher in routine care than in clinical trials and concluded that most trials substantially have lower SAE rates and were unrepresentative.

Additionally, the estimates of SAE incidence in the community would be lower when stringent criteria similar to those used in clinical trials are applied to real-world practice, compared to estimates derived without applying such stringent criteria. This was consistent with a cohort study conducted on psoriasis that compared the SAE rates between trial-eligible and trial-ineligible patients. The study found that eligible patients for trial participation had lower SAE rates compared to ineligible patients (Garcia-Doval et al., 2012).

Furthermore, a cohort study aimed to evaluate the impact of trial eligibility criteria on cancer treatment and to characterise the potential eligibility of real-world population for oncology clinical trials. They applied common cancer trial eligibility criteria to populations in routine care to determine the proportion of patient who would be ineligible for clinical trials. The study found that a significant proportion, approximately 38% of patients, were unable to participate in clinical trials because the restrictive criteria and were considered trial ineligible. Age and heart disease criteria were the most common reasons for ineligibility (24% and 16% respectively) (Karim et al., 2019). Moreover, He et al. (2020) conducted a systematic review to estimate the percentage of population with different chronic health conditions who would be excluded from RCTs of treatments for those conditions. They found that a high rate of population would be ineligible for most trials, including hypertension (83%), T2DM (81%), COPD (84%), and asthma (96%).

8.6.4 Implications

The observed association between eligibility criteria and a decreased SAE rate, indicates that the restrictive trial may select individuals at a lower risk of SAEs. In this context, the characteristics of eligibility criteria, whether considered individually or in combination, may predict the incidence of SAEs. Trialists should employ eligibility criteria that include a representative patient cohort reflecting the expected SAEs in real-world populations. Therefore, assessing the association between eligibility criteria and SAE rates might be a useful approach to explore the feasibility of SAEs as a metric for assessing trial representativeness. However, it is important to be cautious when extrapolating these findings regarding the association to other clinical trials because the associations could vary across different therapeutic and disease areas. Confirming the usefulness of SAEs as a metric based on this association needs further investigations by conducting comparative analyses for other RCTs in different contexts like examining this association in cancer trials involving toxic medications. Moreover, examining the association of eligibility criteria with trial SAEs, alongside the association of characteristics of routine care patients with routine SAEs, and then comparing these two associations, may contribute to more informed decisions about its feasibility.

8.6.5 Strengths and limitations

The included trials were obtained from different registries and conducted in various countries and type of settings, which provides an efficient estimation for the association. Additionally, the analysis focused on common key exclusion criteria that are consistently measured and defined, such as age, duration of diabetes, and laboratory measurements like HbA1c, eGFR, and BMI. Moreover, the restrictiveness cutoffs were operationalised for the included trials with input from diabetologists to reflect the real-world settings. However, one limitation is that the associations found in this study are primarily applicable to trials involving SGLT-2 inhibitors. Differences in pharmacological drugs and medical condition like chemotherapy and cancer may not capture the same association observed between eligibility criteria and SAEs rate, thus generalising these

findings to other clinical trials might be limited. Moreover, reporting of eligibility criteria may vary between sources which I did not assessed this; I only looked at other sources when information was not available with trial registry.

8.7 Conclusion

Restrictive eligibility criteria were associated with a decrease in SAE rates. Trials with restrictive criteria may exclude participants at a higher risk of SAEs. Eligibility criteria related to the regimen of glycaemic control medicines were the most stringent criteria. The occurrence of SAEs can be attributed either to the intervention being examined or the characteristics of the included participants. As eligibility criteria reflect participant characteristics, their association with SAEs suggests the feasibility of using SAEs as a potential metric for assessing trial representativeness.

Chapter 9 Discussion

9.1 Chapter overview

This chapter presents a discussion of the findings from the previous chapters. Firstly, it will present the background of the thesis, and a summary of the findings related to the research question and each objective, followed by strengths and limitations of the research in this thesis. Subsequently, the contributions, implications and recommendations of research will be discussed.

9.2 Background of the thesis

RCTs provide the most reliable evidence on treatment effectiveness. However, their generalisability has been a longstanding concern due to the lack of representativeness in such clinical trials. Assessing trial representativeness is challenging; current approaches often rely on participants characteristics. However, these methods are complex and may be unfeasible for widespread use because not all trials report these characteristics consistently. SAEs reporting provides useful safety information for each trial and could be used to address the issue of trial representativeness. This thesis aimed to explore the feasibility of using SAE rates as a possible metric to assess trial representativeness, using RCTs of SGLT-2 inhibitors as an exemplar. However, it was not known if trials of SGLT-2 inhibitors reported adequate information on SAEs to calculate SAE rates. It was also unclear whether to consider trial arms separately or use the total SAE count across all trial arms. Furthermore, it was unclear whether SAE rates reflected trial representativeness, and there was no gold standard measure for trial representativeness. Therefore, to address the thesis's aim, I conducted the following objectives:

1. Examined the feasibility of calculating SAE rates and assessing whether RCTs of SGLT-2 inhibitors reported sufficient SAE data.

2. Compared the SAE rates between trial arms to determine which arm to consider in the analysis; whether the treatment, the control, or the combined total of both.

3. Examined the differences in the SAE rates based on some characteristics (e.g., type of trial outcome, type of trial settings, and type of trial registries).

4. Given the absence of a gold standard for measuring trial representativeness, I employed triangulation approach to explore whether SAE rates could serve as an indicative metric by:

I. Examined the association between SAE rates and the PRECIS-2 score.

II. Compared SAE rates with PRECIS-2 score based on the differences in their associations with trial and baseline characteristics that included as fair umpires.

III. Examined the association between eligibility criteria and the SAE rates.

9.3 Summary of findings

A summary of the findings for the above research objectives is presented below.

9.3.1 Representativeness of clinical trials and their generalisability

The literature review covered the generalisability of RCTs across various index conditions. It found that, in most trials, a high proportion of patients would be ineligible to participate in trials due to the restrictiveness of those trials' eligibility criteria. The generalisability of the trials was limited due to their lack of representativeness driven by these eligibility criteria.

9.3.2 SAEs reporting in RCTs of SGLT-2 inhibitors

Chapter 4 explored whether trials of SGLT-2 inhibitors reported sufficient information to capture SAE rates and explored the extent to which SAEs were reported across different registry categories. Out of the 146 trials, 141 (97%) trials provided the required information on reported SAEs. In terms of comparison between trial registries, all trials registered on ClinicalTrials.gov (100%) reported all SAEs, and 81% of trials registered on other registries reported them. The high level of SAE reporting on ClinicalTrials.gov is reassuring and provides a valuable resource and a strong basis for using SAEs to examine their usefulness as a metric for trial representativeness. Moreover, all trials with MACE outcomes reported their MACE events within the safety reporting for SAEs.

9.3.3 Comparison of SAE rates between trial arms

The differences in SAE rates between treatment and control arms, whether placebo or active comparator, were examined in chapter 5. This comparison was conducted to determine whether SAE rates varied depending on treatment and to decide which SAE to consider in this thesis, whether from individual arms or the total for both arms when analysing SAE rates. There was no significant difference in the rates of SAE between the intervention and placebo arms, nor between the intervention and active comparator arms. This indicates that the SAEs may be attributed to the underlying participants' characteristics rather than the intervention. Therefore, the total SAEs for both interventions and control arms were combined to increase statistical precision. Furthermore, the differences in SAE rates based on some characteristics, including type of trial outcome, type of trial settings, and type of trial registries, were examined. Trials with prespecified hard outcomes (e.g., MACE) had a significantly higher SAE rate compared to trials with soft outcomes (e.g., HbA1c). Trials registered on ClinicalTrials.gov showed higher SAE rates than trials not registered on ClinicalTrials.gov. Trials conducted in more than one country (multinational) showed higher rates of SAE compared to national trials.

9.3.4 Association between pragmatism metric and SAE rates

The PRECIS-2 tool was used retrospectively to assess the pragmatism of clinical trials. The mean PRECIS-2 score of all trials was 3.34 out of 5, indicating a medium level of pragmatism. The mean PRECIS-2 score was associated with higher SAE rates, suggesting that more pragmatic trials, which reflect real-world practice and could be more representative, may result in higher SAE rates compared to explanatory trials, which may lack representativeness. Among the nine PRECIS-2 domains, only the setting domain was associated with SAE rates. The setting domain may play a role in the incidence of SAEs.

9.3.5 Association of SAEs and PRECIS-2 with trial/baseline characteristics

Some baseline and trial characteristics showed expected association with both SAE rates and the mean PRECIS-2 score, although the magnitudes of these associations varied. Longer duration of diabetes was strongly associated with both the mean PRECIS-2 score and SAE rates. Baseline sample size was strongly associated with the PRECIS-2 score but not with SAE rates. Moreover, baseline age had a negative association with the PRECIS-2 score and a positive association with the SAE rates. Industrial-sponsored trials were strongly associated with SAE rates. Other characteristics, such as trial duration, trial phase, and level of HbA1c, showed no notable association with SAE rates. The lack of substantial associations suggests that these characteristics may not serve as direct measures of trial representativeness. Moreover, SAE metric may capture representativeness from different dimensions rather than being associated with certain baseline and trial characteristics.

In terms of comparison between metrics, the difference in the estimates for trial blinding and age strongly favoured SAE as a metric of trial representativeness more than PRECIS-2. However, baseline sample size strongly favoured PRECIS-2 over SAE. Other umpires, including trial year difference, proportion of males, HbA1c, sponsorship, and trial duration, showed no significant favouring of either metric, but the general trend of these umpires leaned towards favouring SAE

rates as metric. Therefore, these findings did not strongly support the claim of using SAE rates as a metric of trial representativeness. While these umpires are expected to be associated with higher SAE rates, they may not fully capture trial representativeness. This could be attributed to different reasons; for example, umpires of trial phase and baseline sample size may reflect the trial design more than the representativeness of the trial population.

9.3.6 Association between eligibility criteria and SAE rates

SAE rates were associated with the status of eligibility criteria. Among 146 trials, 122 trials with restrictive criteria showed a reduction in SAE rates compared to 24 trials with permissive criteria, which showed an increase in the rates of SAE. Trials employing restrictive criteria enrolled participants with fewer comorbidities, shorter durations of diabetes, lower HbA1c levels, and younger ages compared to trials using permissive criteria. Trials that implemented restrictive criteria may have intended to enrol participants at lower risks for SAE, while trials with permissive criteria aimed to enrol participants that more closely represent real-world settings without such restrictions.

9.4 Strengths and limitations of this thesis

The current thesis has several strengths. Firstly, the literature review on trial representativeness was conducted across various index conditions, rather than focusing on a single condition. This approach provides a broad overview of the representativeness and generalisability of clinical trials. Moreover, data were obtained from different sources, including trial registries, trial protocols, CSR, and trial-relevant publications, especially when data were not available within trial registries. This approach effectively minimised data missingness, potentially enhancing the reliability of the findings in this thesis. Additionally, the assessment and scoring of PRECIS-2 domains were based on operationalisation assessment criteria that were developed with the assistance of a clinical diabetologist to avoid subjective assessment and add a clinical perspective. Finally, the analysis explored SAE as a metric from different dimensions, utilising

a triangulation approach that incorporates multiple aspects of clinical trials. This included comparing SAE rates with trial and baseline characteristics, drivers of trial representativeness (eligibility criteria), and a commonly used pragmatism metric (PRECIS-2). This methodology enhances the reliability of the findings and strengthens the conclusion of the study.

However, there were several limitations. Firstly, different data sources were used only when information was not available within a single source. Although I did check consistency for random samples from 5 trials and found consistent reporting between ClinicalTrials.gov and publications, this sample size may not be sufficient to rule out potential inconsistencies across all trials. Broader inconsistencies may still exist. Moreover, the research focuses only on SGLT-2 inhibitors, and the observations and associations derived from this study are valuable within the context of SGLT-2 trials. Therefore, it is important to note that the findings may not necessarily extrapolate to other clinical trials involving different drug classes. Moreover, two baseline characteristics (smoking status and race/ethnicity) were excluded from this analysis due to the significant amount of missing data within these variables. Additionally, some required information for scoring the PRECIS-2 domains was not reported in either the clinical study report or the trial-relevant publications. This highlights the need for improved reporting of trial protocols. The subjectivity in the assessment of some PRECIS-2 domains was observed, which may introduce bias in scoring these domains. However, this has been minimised by the protocolised assessment criteria that specific to RCTs of SGLT-2 inhibitors. Furthermore, a few trials from other registries did not report SAEs; however, the number of such unreported SAEs was small, including only five trials. In addition, the study examined SAEs as a metric on a single index condition, specifically T2DM. Examining on multiple index conditions, rather than just one, could have potentially contributed to more informative findings.

9.5 Contribution of the thesis

The current thesis introduces the following contributions:

1. This thesis explored the coverage of SAEs reporting on ClincalTrial.gov and other registries for RCTs of SGLT-2 inhibitors, and it found that they provided sufficient information on SAEs to enable the calculation of SAE rates. This suggests that SAE rates can be captured for future research to investigate their feasibility as a metric of trial representativeness. Furthermore, trials registered on ClincalTrial.gov had better SAE reporting than unregistered trials, indicating that ClincalTrial.gov is a reliable source for obtaining such data.

2. This thesis contributed to the literature by showing that the SAEs rates did not differ between treatment and control arms in trials of SGLT-2 inhibitors. This finding suggests that future analyses, which compare SAE rates in SGLT-2 trials to rates of SAE among people eligible for treatment in routine care, can reasonably include the entire trial population when calculating the rates. This approach will yield more precise estimates when assessing SAE rates as a metric of trial representativeness. However, this may not be applicable to trials involving potential toxic treatments like chemotherapy, where SAE rates might differ between treatment and control arms.

3. This thesis contributed to T2DM trials by demonstrating the consistency of MACE reporting within SAEs reporting. All trials with prespecified hard serious outcomes consistently reported MACE in their SAE reports. This finding provides a basis for future research analysing SAE to assess trial representativeness, particularly the consist reporting offers reassurance about the reliability of these reports.

4. This thesis retrospectively assessed the pragmatism of RCTs for SGLT-2 inhibitors using the PRECIS-2 tool. This assessment could help to understand how the trial applicability may relate to trial design. Moreover, it examined the association between pragmatic trial design, as

measured by the PRECIS-2 score, and SAE rates. The finding indicates that trials with higher PRECIS-2 scores, which are more reflective of real-world conditions, tend to report higher SAE. This association allows researchers and clinicians to better understand how pragmatic trials may influence SAE.

5. The assessment was conducted after extensively protocolising the original criteria of the PRECIS-2 tool to suit the unique context and characteristics of T2DM trials. It aimed to ensure that the assessment was objective, practical, and robust, contributing to the credibility of the findings. Moreover, this protocolisation for the criteria provides a framework that can facilitate the application of this tool in T2DM trials for future research.

6. This thesis examined the association between the rates of SAE and: 1) baseline characteristics, 2) trial characteristics, and 3) compared it with the mean PRECIS-2 score. This exploration contributes to the literature by illustrating how SAE rates vary based on patients' characteristics and assess whether SAE reflect trial representativeness.

7. This study is the first to examine the association between the rates of SAE and trial eligibility criteria within clinical trials of SGLT-2 inhibitors. It contributes to clinical practice by showing that restricting the eligibility criteria and excluding representative participants is associated with lower SAE rates. This finding may suggest that future studies could use SAE rates as a measure to assess the restrictiveness of eligibility criteria.

9.6 Implications and recommendations of research

The current thesis contributes to a better understanding of the clinical trial generalisability and the issue of trial representativeness, exploring the feasibility of using SAEs to address this issue. The findings in these thesis chapters can help in clinical practice, trial applicability, and future research.

9.6.1 Importance of SAE reporting in RCTs

The high level of SAE reporting on ClinicalTrials.gov is reassuring and encourages the exploration of SAEs as a metric for trial representativeness. It also suggests that other trial registries should improve their quality of reporting and implement reporting standards similar to those of ClinicalTrials.gov. The reporting of SAEs can be incorporated by researchers for evaluation and enhancing the generalisability of clinical trials. The consistent reporting of MACE and SAEs reporting within ClinicalTrials.gov indicates reliable and transparent documentation of SAEs within this registry. Moreover, the absence of a significant difference in SAE rates between treatment and control arms suggests that combining SAE rates from both arms is feasible. This approach would overcome the issue of lower event and enhance statistical power when assessing trial representativeness. Additionally, the higher SAE rates observed in trials with hard outcomes, multinational trials, and trials registered on ClinicalTrials.gov suggest a potential role of outcome types, trial registration and trial setting in influencing SAE. While these factors do not confirm SAE as a metric, they highlight their importance for future investigation to understand how they might reflect the nature of the trials.

9.6.2 Pragmatism of clinical trial and SAE rate

The positive association between the mean PRECIS-2 score and SAE rates offer a better understanding of the interplay between trial design pragmatism and the incidence of SAEs. As trials become more pragmatic, the SAE rates tend to increase, suggesting that such trials are likely to recruit sicker, more representative participants. This provides foundation for future research on the importance of considering trial pragmatism when estimating the SAEs and underscores the potential of using SAE rates to assess trial representativeness. However, future research is required to establish reliable SAE benchmarks for various trial designs. Moreover, domain-specific analysis revealed that setting plays a role in SAE, emphasising the need for careful consideration of trial settings in the planning and conduct of clinical trials.

9.6.3 Importance of trial eligibility criteria

Lack of representativeness can be across various characteristics such as age and comorbidity, which quite directly reflect health status; trial eligibility criteria also reflect the characteristics of included participants. Thus, the association between restrictive criteria and a reduction in SAE rates suggests that SAEs can be a reasonable marker to capture these characteristics and reflect trial representativeness. Future research may also employ the SAE rates to assess the restrictiveness of trials and distinguish between restrictive and permissive criteria.

9.6.4 Further prospective research

Future studies should investigate trial recruitment processes, focusing on how underrepresented eligible subjects (e.g. vulnerable patients) in routine care are excluded from trials. This investigation would aim to identify the selectivity in patient selection and uncover the obstacles that sometimes prevent these cohorts from participating in trials. Future studies could involve looking at system organ classes (SOCs) within SAEs. It may be that SAEs within specific organ systems (e.g., cardiovascular) may better reflect trial representativeness either with respect to the severity of the index condition (e.g., cardiometabolic diseases) or comorbid conditions (e.g., other systems).

In addition, future research could focus on SAE rates among diabetes patients in routine practice, particularly by comparing the observed SAE rates in RCTs of SGLT-2 inhibitors to the SAE rates in real-world practice. To achieve this, the use of record-linkage (e.g., the Secure Anonymised Information Linkage (SAIL) Databank) could enhance the comparison of SAE rates between trial populations and real-world populations. This approach would offer a clearer understanding of whether trials adequately reflect the health status and comorbidity burden of the target population. Such analysis would help determine the feasibility of using SAEs as a metric for trial representativeness. Additionally, future studies could further explore this feasibility by comparing the SAE rates in new trials to

the SAE rates in landmark representative trials. Furthermore, future work may utilise individual participant data (IPD), specifically to compare the difference in SAE rates between subgroups in clinical trials. Furthermore, future work may utilise individual participant data (IPD) to compare the difference in SAE rates between subgroups (e.g., age, sex, comorbidities) in clinical trials. This approach may provide insight into whether certain subgroups are underrepresented in trials. Additionally, it can help identify if older people, patients with multimorbidity, or those with specific comorbidities experience higher SAE.

9.7 Conclusion

This thesis examined the feasibility of using SAE rates as a metric for assessing trial representativeness. Clinical trials of SGLT-2 inhibitors reported all relevant information for SAEs, indicating the feasibility of calculating the SAE rates. SAE rates were similar between intervention and control arms, allowing for the combination of total SAEs for both arms. The similar rates also suggest that SAE rates in RCTs may reflect patient characteristics and could assess trial representativeness. Furthermore, the pragmatic design of RCTs that reflects real-world practice showed higher SAE rates.

The direction of the difference in association for most umpires generally favoured SAE rates. However, only age and type of blinding umpires that strongly favoured SAE as a metric more than PRECIS-2. The findings from these associations and comparison do not strongly support the hypothesis of using SAE rates as a metric for trial representativeness. However, the eligibility criteria analysis showed a significant reduction in SAE rates in trials with more restrictive criteria compared to trial with more permissive criteria.

Therefore, based on the association of SAE rates with eligibility criteria, the PRECIS-2 score, and given that SAEs are commonly reported in RCTs with sufficient information, SAE rates might be utilised as a practical method to assess the representativeness of clinical trials. However, for SAE rates to be informative for clinicians and trialists, it is essential to understand the expected rates of SAE and explore at what levels they may indicate a lack of representativeness. Future studies should explore these dimensions across different index conditions to fully assess the reliability of SAEs as a metric of trial representativeness and to implement this tool effectively.

References

Abdul, M. *et al.* (2020) 'Epidemiology of Type 2 Diabetes – Global Burden of Disease and Forecasted Trends', *Journal of Epidemiology and Global Health*, 10, pp. 107–111.

Akobeng, A. K. (2008) 'Assessing the validity of clinical trials', *Journal of Pediatric Gastroenterology and Nutrition*, 47(3), pp. 277–282. doi: 10.1097/MPG.0b013e31816c749f.

Andersson, Y. *et al.* (2019) 'Do clinical trials truly mirror their target population? An external validity analysis of national register versus trial data from the Swedish prospective SENOMIC trial on sentinel node micrometastases in breast cancer', *Breast Cancer Research and Treatment*. Springer US, 177(2), pp. 469–475. doi: 10.1007/s10549-019-05328-3.

Arnaud-Coffin, P. *et al.* (2019) 'A systematic review of adverse events in randomized trials assessing immune checkpoint inhibitors', *International Journal of Cancer*, 145(3), pp. 639–648. doi: 10.1002/ijc.32132.

Atkins, D. *et al.* (2011) 'Assessing applicability when comparing medical interventions: AHRQ and the Effective Health Care Program', *Journal of Clinical Epidemiology*. Elsevier Inc, 64(11), pp. 1198–1207. doi: 10.1016/j.jclinepi.2010.11.021.

Averitt, A. J. *et al.* (2020) 'Translating evidence into practice: eligibility criteria fail to eliminate clinically significant differences between real-world and study populations', *npj Digital Medicine*. Springer US, 3(1). doi: 10.1038/s41746-020-0277-8.

Badano, L. P. *et al.* (2003) 'Patients with chronic heart failure encountered in daily clinical practice are different from the "typical" patient enrolled in

therapeutic trials', *Italian Heart Journal*, 4(2), pp. 84–91.

Berger, V. W. *et al.* (2021) 'A roadmap to using randomization in clinical trials', *BMC Medical Research Methodology*, 21(1), pp. 1–24. doi: 10.1186/s12874-021-01303-z.

Bibbins-Domingo, K. and Helman, A. (2022) *Improving Representation in Clinical Trials and Research: BUILDING RESEARCH EQUITY FOR WOMEN AND UNDERREPRESENTED GROUPS*, *Improving Representation in Clinical Trials and Research: Building Research Equity for Women and Underrepresented Groups*. doi: 10.17226/26479.

Bijker, N. *et al.* (2002) 'Effects of patient selection on the applicability of results from a randomised clinical trial (EORTC 10853) investigating breast-conserving therapy for DCIS', *British Journal of Cancer*, 87(6), pp. 615–620. doi: 10.1038/sj.bjc.6600514.

Birkeland, K. I. *et al.* (2019) 'How representative of a general type 2 diabetes population are patients included in cardiovascular outcome trials with SGLT2 inhibitors? A large European observational study', *Diabetes, Obesity and Metabolism*, 21(4), pp. 968–974. doi: 10.1111/dom.13612.

Blanco, C., Olfson, M., Goodwin, R. D., *et al.* (2008) 'Generalizability of clinical trial results for major depression to community samples: Results from the National Epidemiologic Survey on Alcohol and Related Conditions.', *The Journal of Clinical Psychiatry*. Blanco, Carlos: New York State Psychiatric Institute, 1051 Riverside Dr., Box 69, New York, NY, US, 10032, cb255@columbia.edu: Physicians Postgraduate Press, pp. 1276–1280. doi: 10.4088/JCP.v69n0810.

Blanco, C., Olfson, M., Okuda, M., *et al.* (2008) 'Generalizability of clinical trials for alcohol dependence to community samples', *Drug and Alcohol Dependence*, 98(1–2), pp. 123–128. doi: 10.1016/j.drugalcdep.2008.05.002.

Blanco, C. *et al.* (2017) 'Generalizability of clinical trial results for adolescent major depressive disorder', *Pediatrics*, 140(6). doi: 10.1542/peds.2016-1701.

Bolton, M. *et al.* (2019) 'Serious adverse events reported in placebo randomised controlled trials of oral naltrexone: A systematic review and meta-analysis', *BMC Medicine*. BMC Medicine, 17(1), pp. 1–13. doi: 10.1186/s12916-018-1242-0.

Bornhöft, G. *et al.* (2006) 'Checklist for the qualitative evaluation of clinical studies with particular focus on external validity and model validity', *BMC Medical Research Methodology*, 6, pp. 1–13. doi: 10.1186/1471-2288-6-56.

Butterly, E. *et al.* (2022) 'Calibrating a network meta-analysis of diabetes trials of sodium glucose cotransporter 2 inhibitors, glucagon-like peptide-1 receptor analogues and dipeptidyl peptidase-4 inhibitors to a representative routine population: a systematic review protocol', *BMJ Open*, 12(10). doi: 10.1136/bmjopen-2022-066491.

Chen, C. I., Skingley, P. and Meyer, R. M. (2000) 'A comparison of elderly patients with aggressive histology lymphoma who were entered or not entered on to a randomized phase II trial', *Leukemia and Lymphoma*, 38(3–4), pp. 327–334. doi: 10.3109/10428190009087023.

Chen, K. Y. *et al.* (2022) 'Reporting of clinical trial safety results in ClinicalTrials.gov for FDA-approved drugs: a cross-sectional analysis', 19(4), pp. 442–451. doi: 10.1177/17407745221093567.Reporting.

Choi, M. Y. *et al.* (2019) 'Pragmaticism of Randomized Controlled Trials of Biologic Treatment With Methotrexate in Rheumatoid Arthritis : A Systematic Review', 71(5), pp. 620–628. doi: 10.1002/acr.23620.

Clark, L. T. *et al.* (2019) 'Increasing Diversity in Clinical Trials: Overcoming Critical Barriers', *Current Problems in Cardiology*, 44(5), pp. 148–172. doi:

10.1016/j.cpcardiol.2018.11.002.

Concato, J., Shah, N. and Horwitz, R. I. (2000) 'Randomized controlled trials observational studies', *N Engl J Med*, 342(25), pp. 1187–1892.

Cosentino, F. *et al.* (2020) '2019 ESC Guidelines on diabetes, pre-diabetes, and cardiovascular diseases developed in collaboration with the EASD', *European Heart Journal*, 41(2), pp. 255–323. doi: 10.1093/eurheartj/ehz486.

Costa, L. J., Hari, P. N. and Kumar, S. K. (2016) 'Differences between unselected patients and participants in multiple myeloma clinical trials in US: a threat to external validity', *Leukemia and Lymphoma*, 57(12), pp. 2827–2832. doi: 10.3109/10428194.2016.1170828.

Cragg, W. J. *et al.* (2021) 'Clinical trial recruiters' experiences working with trial eligibility criteria: results of an exploratory, cross-sectional, online survey in the UK', *Trials*. Trials, 22(1), pp. 1–12. doi: 10.1186/s13063-021-05779-4.

Dal-Ré, R., de Boer, A. and James, S. K. (2020) 'The design can limit PRECIS-2 retrospective assessment of the clinical trial explanatory/pragmatic features', *Journal of Clinical Epidemiology*. Elsevier Inc, 126, pp. 193–201. doi: 10.1016/j.jclinepi.2020.03.027.

Datta, R. *et al.* (2018) 'Serious adverse events of older adults in nursing home and community intervention trials', *Contemporary Clinical Trials Communications*. Elsevier, 9(December 2017), pp. 77–80. doi: 10.1016/j.conctc.2017.12.004.

Dekkers, O. M. *et al.* (2010) 'How to assess the external validity of therapeutic trials: A conceptual approach', *International Journal of Epidemiology*, 39(1), pp. 89–94. doi: 10.1093/ije/dyp174.

Devos, F. *et al.* (2019) 'Study characteristics impacted the pragmatism of randomized controlled trial published in nursing: a meta-epidemiological study', *Journal of Clinical Epidemiology*, 116, pp. 18–25. doi: 10.1016/j.jclinepi.2019.07.017.

Dugard, A., Giraudeau, B. and Dibao-Dina, C. (2024) 'Do patients with type 2 diabetes mellitus included in randomised clinical trials differ from general-practice patients? A cross-sectional comparative study', *BMJ Open*, 14(4), pp. 1–5. doi: 10.1136/bmjopen-2023-077582.

Dworkin, R. H. *et al.* (2021) 'Essential statistical principles of clinical trials of pain treatments', *Pain Reports*, 6(1), pp. 1–25. doi: 10.1097/PR9.0000000000000863.

Eisenberg, Y. *et al.* (2015) 'Similarities and differences between patients included and excluded from a randomized clinical trial of vitamin d supplementation for improving glucose tolerance in prediabetes: Interpreting broader applicability', *Trials*. Trials, 16(1), pp. 1–8. doi: 10.1186/s13063-015-0812-0.

Elizabeth A. Stuart, Catherine P. Bradshaw, P. J. L. (2015) 'Assessing the Generalizability of Randomized Trial Results to Target', *Prevention Science*, 176(1), pp. 475–485. doi: 10.1007/s11121-014-0513-z.Assessing.

Elting, L. S. *et al.* (2006) 'Generalizability of cancer clinical trial results: Prognostic differences between participants and nonparticipants', *Cancer*, 106(11), pp. 2452–2458. doi: 10.1002/cncr.21907.

Faber, J. and Fonseca, L. M. (2014) 'How sample size influences research outcomes', *Dental Press Journal of Orthodontics*, 19(4), pp. 27–29. doi: 10.1590/2176-9451.19.4.027-029.ebo.

Fareed, M., Suri, K. and Qureshi, A. I. (2012) 'Recruitment of Ischemic Stroke Patients in Clinical trials in General Practice and Implications for Generalizability of Results.', *Journal of vascular and interventional neurology*, 5(1), pp. 27–32.

Fitzpatrick, C. *et al.* (2020) 'Effect of pragmatic versus explanatory interventions on medication adherence in people with cardiometabolic conditions: A systematic review and meta-analysis', *BMJ Open*, 10(7), pp. 1–13. doi: 10.1136/bmjopen-2019-036575.

Florisson, S. *et al.* (2021) 'Are older adults insufficiently included in clinical trials?—An umbrella review', *Basic and Clinical Pharmacology and Toxicology*, 128(2), pp. 213–223. doi: 10.1111/bcpt.13536.

Forbes, G. *et al.* (2017) 'Understanding the applicability of results from primary care trials: lessons learned from applying PRECIS-2', *Journal of Clinical Epidemiology*. Elsevier Inc, 90, pp. 119–126. doi: 10.1016/j.jclinepi.2017.06.007.

Ford, I. and Norrie, J. (2016) 'Pragmatic Trials', *New England Journal of Medicine*, 375(5), pp. 454–463. doi: 10.1056/nejmra1510059.

Fortin, M. *et al.* (2006) 'Randomized controlled trials: do they have external validity for patients with multiple comorbidities?', *Annals of family medicine*, 4(2), pp. 104–8. doi: 10.1370/afm.516.INTRODUCTION.

Franco, S. *et al.* (2016) 'Generalizability of pharmacologic and psychotherapy clinical trial results for posttraumatic stress disorder to community samples', *Journal of Clinical Psychiatry*, 77(8), pp. e975–e981. doi: 10.4088/JCP.15m10060.

Galicia-Garcia, U. *et al.* (2020) 'Pathophysiology of type 2 diabetes mellitus', *International Journal of Molecular Sciences*, 21(17), pp. 1–34. doi:

10.3390/ijms21176275.

Gandhi, M. *et al.* (2005) 'Eligibility criteria for HIV clinical trials and generalizability of results: The gap between published reports and study protocols', *Aids*, 19(16), pp. 1885–1896. doi: 10.1097/01.aids.0000189866.67182.f7.

Gebrie, D., Getnet, D. and Manyazewal, T. (2021) 'Cardiovascular safety and efficacy of metformin-SGLT2i versus metformin-sulfonylureas in type 2 diabetes: systematic review and meta-analysis of randomized controlled trials', *Scientific Reports*. Nature Publishing Group UK, 11(1), pp. 1–18. doi: 10.1038/s41598-020-80603-8.

Gheorghe, A. *et al.* (2015) 'Evaluating the Generalisability of Trial Results: Introducing a Centre- and Trial-Level Generalisability Index', *PharmacoEconomics*. Springer International Publishing, 33(11), pp. 1195–1214. doi: 10.1007/s40273-015-0298-3.

Ghouse, J. *et al.* (2020) 'Effect of diabetes duration on the relationship between glycaemic control and risk of death in older adults with type 2 diabetes', *Diabetes, Obesity and Metabolism*, 22(2), pp. 231–242. doi: 10.1111/dom.13891.

Glasziou, P., Irwig, L. and Deeks, J. J. (2008) 'When should a new test become the current reference standard?', *Annals of Internal Medicine*, 149(11), pp. 816–821. doi: 10.7326/0003-4819-149-11-200812020-00009.

Gliklich RE, Leavy MB and Dreyer NA (2014) 'Registries for Evaluating Patient Outcomes: A User's Guide Fourth Edition Registries for Evaluating Patient Outcomes: A User's Guide', *AHRQ Publication*, 19(20)-EH, p. 360.

Hanlon, P. *et al.* (2019) 'Representation of people with comorbidity and multimorbidity in clinical trials of novel drug therapies: An individual-level

participant data analysis', *BMC Medicine*. BMC Medicine, 17(1), pp. 1–12. doi: 10.1186/s12916-019-1427-1.

Hanlon, P. *et al.* (2021) 'Observed and expected serious adverse event rates in randomised clinical trials for hypertension: an observational study comparing trials that do and do not focus on older people', *The Lancet Healthy Longevity*. The Author(s). Published by Elsevier Ltd. This is an Open Access article under the CC BY 4.0 license, 2(7), pp. e398–e406. doi: 10.1016/s2666-7568(21)00092-1.

Hanlon, P. *et al.* (2022) 'Assessing trial representativeness using serious adverse events: an observational analysis using aggregate and individual-level data from clinical trials and routine healthcare data', *BMC Medicine*. BioMed Central, 20(1), pp. 1–15. doi: 10.1186/s12916-022-02594-9.

He, Z. *et al.* (2020) 'Clinical Trial Generalizability Assessment in the Big Data Era: A Review', *Clinical and Translational Science*, 13(4), pp. 675–684. doi: 10.1111/cts.12764.

Hege Nordem Sjastad, Rolf W. Grawe, J. E. (2012) 'Affective Disorders among Patients with Borderline Personality Disorder', *PloS one*, 7(4), pp. 1–10. doi: 10.1371/Citation.

Hillman, S. L. *et al.* (2023) 'Rates of and Factors Associated with Patient Withdrawal of Consent in Cancer Clinical Trials', *JAMA Oncology*, 9(8), pp. 1041–1047. doi: 10.1001/jamaoncol.2023.1648.

Hoertel, N. *et al.* (2012) 'Generalizability of clinical trial results for generalized anxiety disorder to community samples', *Depression and Anxiety*, 29(7), pp. 614–620. doi: 10.1002/da.21937.

Hoertel, N. *et al.* (2013) 'Generalizability of clinical trial results for bipolar disorder to community samples: Findings from the national epidemiologic survey

on alcohol and related conditions', *Journal of Clinical Psychiatry*, 74(3), pp. 265–270. doi: 10.4088/JCP.12m07935.

Hoertel, N. *et al.* (2015) 'Generalizability of pharmacological and psychotherapy clinical trial results for borderline personality disorder to community samples.', *Personality Disorders: Theory, Research, and Treatment*. Hoertel, Nicolas: Department of Psychiatry, Corentin Celton Hospital, Paris Descartes University, 4 parvis Corentin Celton, Issy-les-Moulineaux, France, 92130, nico.hoertel@yahoo.fr: Educational Publishing Foundation, pp. 81–87. doi: 10.1037/per0000091.

Houle, S. (2015) 'An introduction to the fundamentals of randomized controlled trials in pharmacy research', *Canadian Journal of Hospital Pharmacy*, 68(1), pp. 28–32. doi: 10.4212/cjhp.v68i1.1422.

Howard-Pitney, B., Fortmann, S. P. and Killen, J. D. (2001) 'Generalizability of findings from a chewing tobacco cessation clinical trial', *Nicotine and Tobacco Research*, 3(4), pp. 347–352. doi: 10.1080/14622200110050439.

Ignacio Garcia-Doval, Gregorio Carretero, Francisco Vanaclocha, Carlos Ferrandiz, Esteban Dauden, Jose-Luis Sanchez-Carazo, Merce Alsina, Enrique Herrera-Ceballos, Francisco-Jose Gomez-Garcıa, Marta Ferran, J.-L. L.-E. and Jose-Manuel Hernanz, Isabel Belinchon-Romero, Jaime Vilar-Alejo, Raquel Rivera, Jose-Manuel Carrascosa, C. C. (2012) 'Risk of Serious Adverse Events Associated With Biologic and Nonbiologic Psoriasis Systemic Therapy: Patients Ineligible vs Eligible for Randomized Controlled Trials.', *Archives of Dermatology*, 148(4), p. 463. doi: 10.1001/archdermatol.2011.2768.

Ioannidis, J. P. A. and Lau, J. (2001) 'Completeness of safety reporting in randomized trials: An evaluation of 7 medical areas', *Jama*, 285(4), pp. 437–443. doi: 10.1001/jama.285.4.437.

Jadad, A. R. *et al.* (1996) 'Assessing the quality of reports of randomized clinical

trials: Is blinding necessary?', *Controlled Clinical Trials*, 17(1), pp. 1–12. doi: 10.1016/0197-2456(95)00134-4.

Jinzhang He, D. R. M. and B. G. (2020) 'Exclusion rates in randomized controlled trials of treatments for physical conditions: a systematic review', *Trials*. Trials, 22(1), pp. 1–11. doi: 10.1186/s13063-021-05019-9.

Jonathan D Casey, Laura M Beskow, Jeremy Brown, Samuel M Brown, Étienne Gayat, Michelle Ng Gong, Michael O Harhay, Samir Jaber, Jacob C Jentzer, Pierre-François Laterre, John C Marshall, Michael A Matthay, Todd W Rice, Yves Rosenberg, Alison E Turnbull, L, S. P. C. (2022) 'Use of pragmatic and explanatory trial designs in acute care research: lessons from COVID-19', *thelancet*, (January), pp. 19–21.

Julia A. Beaver, M.D., Gwynn Ison, M.D., and Richard Pazdur, M. D. (2017) 'Reevaluating Eligibility Criteria — Balancing Patient Protection and Participation in Oncology Trials', *New England Journal of Medicine*, 376(16), pp. 1502–1504. doi: 10.1056/nejmp1614972.

Kabisch, M. *et al.* (2011) 'Randomisierte kontrollierte studien: Teil 17 der serie zur bewertung wissenschaftlicher publikationen', *Deutsches Arzteblatt*, 108(39), pp. 663–668. doi: 10.3238/arztebl.2011.0663.

Kahan, B. C., Rehal, S. and Cro, S. (2015) 'Risk of selection bias in randomised trials', *Trials*. Trials, 16(1), pp. 1–7. doi: 10.1186/s13063-015-0920-x.

Kanzler, K. E. *et al.* (2022) 'Conducting a Pragmatic Trial in Integrated Primary Care: Key Decision Points and Considerations', *Journal of Clinical Psychology in Medical Settings*. Springer US, 29(1), pp. 185–194. doi: 10.1007/s10880-021-09790-4.

Karim, S. *et al.* (2019a) 'Generalisability of Common Oncology Clinical Trial

Eligibility Criteria in the Real World', *Clinical Oncology*, 31(9), pp. e160–e166. doi: 10.1016/j.clon.2019.05.003.

Karim, S. *et al.* (2019b) 'Generalisability of Common Oncology Clinical Trial Eligibility Criteria in the Real World', *Clinical Oncology*, 31(9), pp. e160–e166. doi: 10.1016/j.clon.2019.05.003.

Kendall, J. M. (2003) 'Designing a research project: Randomised controlled trials and their principles', *Emergency Medicine Journal*, 20(2), pp. 164–168. doi: 10.1136/emj.20.2.164.

Kennedy-Martin, T. *et al.* (2015) 'A literature review on the representativeness of randomized controlled trial samples and implications for the external validity of trial results', *Trials*. Trials, 16(1), pp. 1–14. doi: 10.1186/s13063-015-1023-4.

Kim, E. S. *et al.* (2017) 'Broadening eligibility criteria to make clinical trials more representative: American society of clinical oncology and friends of cancer research joint research statement', *Journal of Clinical Oncology*, 35(33), pp. 3737–3744. doi: 10.1200/JCO.2017.73.7916.

Kim, J., Kim, M. G. and Lim, K. M. (2022) 'Participation in and withdrawal from cancer clinical trials: A survey of clinical research coordinators', *Asia-Pacific Journal of Oncology Nursing*. The Authors, 9(4), pp. 197–201. doi: 10.1016/j.apjon.2021.12.015.

Kingsley, G. H. *et al.* (2005) 'Are clinical trials in rheumatoid arthritis generalizable to routine practice? A re-evaluation of trial entry criteria', *Rheumatology*, 44(5), pp. 629–632. doi: 10.1093/rheumatology/keh565.

Kirkman, M. S. *et al.* (2012) 'Diabetes in older adults', *Diabetes Care*, 35(12), pp. 2650–2664. doi: 10.2337/dc12-1801.

Koog, Y. H., Lee, J. S. and Wi, H. (2015) ‘External validity of knee osteoarthritis clinical trials: A systematic review’, *Archives of Rheumatology*, 30(1), pp. 83–90. doi: 10.5606/ArchRheumatol.2015.5054.

Le-Rademacher, J. *et al.* (2023) ‘Clinical Trials Overview: From Explanatory to Pragmatic Clinical Trials’, *Mayo Clinic Proceedings*. Mayo Foundation for Medical Education and Research, 98(8), pp. 1241–1253. doi: 10.1016/j.mayocp.2023.04.013.

Leinonen, A., Koponen, M. and Hartikainen, S. (2015) ‘Systematic review: Representativeness of participants in RCTs of acetylcholinesterase inhibitors’, *PLoS ONE*, 10(5), pp. 1–14. doi: 10.1371/journal.pone.0124500.

Lessing, C. *et al.* (2010) ‘Impact of sample size on variation of adverse events and preventable adverse events: Systematic review on epidemiology and contributing factors’, *Quality and Safety in Health Care*, 19(6), pp. 10–14. doi: 10.1136/qshc.2008.031435.

Li, Q. *et al.* (2020) ‘Using Real-World Data to Rationalize Clinical Trials Eligibility Criteria Design: A Case Study of Alzheimer’s Disease Trials’, *AMIA ... Annual Symposium proceedings. AMIA Symposium*, 2020(2), pp. 717–726.

Lim, Y. M. F. *et al.* (2022) ‘Generalizability of randomized controlled trials in heart failure with reduced ejection fraction’, *European Heart Journal - Quality of Care and Clinical Outcomes*, 8(7), pp. 761–769. doi: 10.1093/ehjqcco/qcab070.

Lindsay, W. A. *et al.* (2020) ‘Age, sex, race and ethnicity representativeness of randomised controlled trials in peri-operative medicine’, *Anaesthesia*, 75(6), pp. 809–815. doi: 10.1111/anae.14967.

Loudon, K. *et al.* (2015) ‘The PRECIS-2 tool: Designing trials that are fit for

purpose', *BMJ (Online)*, 350. doi: 10.1136/bmj.h2147.

Lu, L. *et al.* (2017) 'The application of PRECIS-2 ratings in randomized controlled trials of Chinese herbal medicine', 8(63), pp. 107002–107010.

Luo, J. *et al.* (2016) 'Population analysis of adverse events in different age groups using big clinical trials data', *JMIR Medical Informatics*, 4(4), pp. 1–11. doi: 10.2196/medinform.6437.

Mahr, A. *et al.* (2017) 'Types, frequencies, and burden of nonspecific adverse events of drugs: analysis of randomized placebo-controlled clinical trials', *Pharmacoepidemiology and Drug Safety*, 26(7), pp. 731–741. doi: 10.1002/pds.4169.

Margolis, K. L. *et al.* (2022) 'Comparison of explanatory and pragmatic design choices in a cluster-randomized hypertension trial: effects on enrollment, participant characteristics, and adherence', *Trials*, 23(1), pp. 1–14. doi: 10.1186/s13063-022-06611-3.

Mas-Llado, C. *et al.* (2023) 'Representativeness in randomised clinical trials supporting acute coronary syndrome guidelines', *European Heart Journal - Quality of Care and Clinical Outcomes*, 9(8), pp. 796–805. doi: 10.1093/ehjqcco/qcad007.

Mathieu, S., Moher, D. and Altman, D. G. (2009) 'and Published Primary Outcomes in Randomized Controlled Trials', *General Medicine*, 302(9), pp. 977–984.

Mishkin, G. *et al.* (2016) 'The generalizability of NCI-sponsored clinical trials accrual among women with gynecologic malignancies', *Gynecologic Oncology*. Elsevier Inc., 143(3), pp. 611–616. doi: 10.1016/j.ygyno.2016.09.026.

Monaghan, T. F. *et al.* (2021) 'Blinding in clinical trials: Seeing the big picture', *Medicina (Lithuania)*, 57(7), pp. 1–13. doi: 10.3390/medicina57070647.

Motschman, C. A. *et al.* (2016) 'Selection criteria limit generalizability of smoking pharmacotherapy studies differentially across clinical trials and laboratory studies: A systematic review on varenicline', *Drug and Alcohol Dependence*. Elsevier Ireland Ltd, 169, pp. 180–189. doi: 10.1016/j.drugalcdep.2016.10.018.

Neven, A. *et al.* (2022) 'Representativeness of trial participants: linking the EORTC boost-no boost trial to the Netherlands cancer registry', *Journal of Clinical Epidemiology*. The Author(s), 148, pp. 54–64. doi: 10.1016/j.jclinepi.2022.04.014.

Okuda, M. *et al.* (2010) 'Generalizability of clinical trials for cannabis dependence to community samples', *Drug and Alcohol Dependence*. Elsevier Ireland Ltd, 111(1–2), pp. 177–181. doi: 10.1016/j.drugalcdep.2010.04.009.

Omerovic, E. *et al.* (2024) 'Pragmatic randomized controlled trials: strengthening the concept through a robust international collaborative network: PRIME-9—Pragmatic Research and Innovation through Multinational Experimentation', *Trials*. BioMed Central, 25(1), pp. 1–13. doi: 10.1186/s13063-024-07935-y.

Palese, A., Bevilacqua, M. G. and Dante, A. (2014) 'Do randomized controlled nursing trials have a pragmatic or explanatory attitude? Findings from the Pragmatic-Explanatory Continuum Indicator Summary (PRECIS) tool exercise', *Journal of Nursing Research*, 22(3), pp. 216–220. doi: 10.1097/jnr.0000000000000045.

Panagioti, M. *et al.* (2015) 'Multimorbidity and patient safety incidents in primary care: A systematic review and meta-analysis', *PLoS ONE*, 10(8), pp. 1–30. doi: 10.1371/journal.pone.0135947.

Pibouleau, L. *et al*. (2009) 'Applicability and generalisability of published results of randomised controlled trials and non-randomised studies evaluating four orthopaedic procedures: Methodological systematic review', *BMJ (Online)*, 339(7732), p. 1239. doi: 10.1136/bmj.b4538.

Pilla, S. J., Segal, J. B. and Maruthur, N. M. (2019) 'Primary Care Provides the Majority of Outpatient Care for Patients with Diabetes in the US: NAMCS 2009–2015', *Journal of General Internal Medicine*. Journal of General Internal Medicine, 34(7), pp. 1089–1091. doi: 10.1007/s11606-019-04843-9.

Pitkala, K. H. and Strandberg, T. E. (2022) 'Clinical trials in older people', *Age and Ageing*, 51(5), pp. 1–9. doi: 10.1093/ageing/afab282.

Pitre, T. *et al*. (2023) 'The impact of blinding on trial results: A systematic review and meta-analysis', *Cochrane Evidence Synthesis and Methods*, 1(4). doi: 10.1002/cesm.12015.

Qi, M. *et al*. (2021) 'Quantifying representativeness in randomized clinical trials using machine learning fairness metrics', *JAMIA Open*, 4(3), pp. 1–11. doi: 10.1093/jamiaopen/ooab077.

Ratain, M. J. and Sargent, D. J. (2009) 'Optimising the design of phase II oncology trials: The importance of randomisation', *European Journal of Cancer*. Elsevier Ltd, 45(2), pp. 275–280. doi: 10.1016/j.ejca.2008.10.029.

Raymond, J. *et al*. (2024) 'Understanding why restrictive trial eligibility criteria are inappropriate', *Neurochirurgie*. Elsevier Masson SAS, 70(6), p. 101589. doi: 10.1016/j.neuchi.2024.101589.

Reiss, J. (2019) 'Against external validity', *Synthese*. Springer Netherlands, 196(8), pp. 3103–3121. doi: 10.1007/s11229-018-1796-6.

Riveros, C. *et al.* (2013) 'Timing and Completeness of Trial Results Posted at ClinicalTrials.gov and Published in Journals', *PLoS Medicine*, 10(12), pp. 1–12. doi: 10.1371/journal.pmed.1001566.

Rothwell, P. M. (2005) 'External validity of randomised controlled trials: "to whom do the results of this trial apply?".', *Lancet (London, England)*. England, 365(9453), pp. 82–93. doi: 10.1016/S0140-6736(04)17670-8.

Saesen, R. *et al.* (2023) 'Analysis of the characteristics and the degree of pragmatism exhibited by pragmatic-labelled trials of antineoplastic treatments', *BMC Medical Research Methodology*. BioMed Central, 23(1), pp. 1–17. doi: 10.1186/s12874-023-01975-9.

Sam, D. *et al.* (2018) 'Generalizability of clinical trials of advanced melanoma in the real- world , population-based setting', *Medical Oncology*. Springer US, 35(7), pp. 1–8. doi: 10.1007/s12032-018-1167-7.

Saunders, C. *et al.* (2013) 'External validity of randomized controlled trials of glycaemic control and vascular disease: How representative are participants?', *Diabetic Medicine*, 30(3), pp. 300–308. doi: 10.1111/dme.12047.

Schwartz, D. and Lellouch, J. (1967) 'Explanatory and pragmatic attitudes in therapeutical trials', *Journal of Chronic Diseases*, 20(8), pp. 637–648. doi: 10.1016/0021-9681(67)90041-0.

Selker, H. P. *et al.* (2019) 'Efficacy and Effectiveness Too Trials: Clinical Trial Designs to Generate Evidence on Efficacy and on Effectiveness in Wide Practice', *Clinical Pharmacology and Therapeutics*, 105(4), pp. 857–866. doi: 10.1002/cpt.1347.

Sen, A. *et al.* (2018) 'The representativeness of eligible patients in type 2 diabetes trials: A case study using GIST 2.0', *Journal of the American Medical*

Informatics Association, 25(3), pp. 239–247. doi: 10.1093/jamia/ocx091.

Sepehrvand, N. *et al.* (2019) 'Trends in the Explanatory or Pragmatic Nature of Cardiovascular Clinical Trials over 2 Decades', *JAMA Cardiology*, 4(11), pp. 1122–1128. doi: 10.1001/jamacardio.2019.3604.

Sheppard, J. P. *et al.* (2020) 'Generalizability of Blood Pressure Lowering Trials to Older Patients: Cross-Sectional Analysis', *Journal of the American Geriatrics Society*, 68(11), pp. 2508–2515. doi: 10.1111/jgs.16749.

Smelt, A. F. H. *et al.* (2010) 'How usual is usual care in pragmatic intervention studies in primary care? An overview of recent trials', *British Journal of General Practice*, 60(576), pp. 305–318. doi: 10.3399/bjgp10X514819.

St Sauver, J. L. *et al.* (2012) 'Generalizability of epidemiological findings and public health decisions: An illustration from the Rochester Epidemiology Project', *Mayo Clinic Proceedings*. Elsevier Inc., 87(2), pp. 151–160. doi: 10.1016/j.mayocp.2011.11.009.

Steg, P. G. *et al.* (2007) 'External Validity of Clinical Trials in Acute Myocardial Infarction', *Archives of Internal Medicine*, 167(1), pp. 68–73. doi: 10.1001/archinte.167.1.68.

Stephenson, J. M. and Babiker, A. (2000) 'methods', pp. 244–247.

Stoffers, J. *et al.* (2010) 'Pharmacological interventions for borderline personality disorder', *Cochrane Database of Systematic Reviews*. doi: 10.1002/14651858.cd005653.pub2.

le Strat, Y., Rehm, J. and le Foll, B. (2011) 'How generalisable to community samples are clinical trial results for treatment of nicotine dependence: A comparison of common eligibility criteria with respondents of a large

representative general population survey', *Tobacco Control*, 20(5), pp. 338–343. doi: 10.1136/tc.2010.038703.

Su, Q., Cheng, G. and Huang, J. (2023) 'A review of research on eligibility criteria for clinical trials', *Clinical and Experimental Medicine*. Springer International Publishing, (0123456789). doi: 10.1007/s10238-022-00975-1.

Susukida, R. *et al.* (2016) 'Assessing sample representativeness in randomized controlled trials: application to the National Institute of Drug Abuse Clinical Trials Network', *Addiction (Abingdon, England)*, 111(7), pp. 1226–1234. doi: 10.1111/add.13327.

Susukida, R. *et al.* (2017) 'Generalizability of findings from randomized controlled trials: application to the National Institute of Drug Abuse Clinical Trials Network', *Addiction*, 112(7), pp. 1210–1219. doi: 10.1111/add.13789.

Tam, V. C. *et al.* (2009) 'Generalizability of toxicity data from oncology clinical trials to clinical practice: Toxicity of irinotecan-based regimens in patients with metastatic colorectal cancer', *Current Oncology*, 16(6), pp. 13–20. doi: 10.3747/co.v16i6.426.

Thorpe, K. E. *et al.* (2009) 'A pragmatic-explanatory continuum indicator summary (PRECIS): a tool to help trial designers', *Journal of Clinical Epidemiology*. Elsevier Inc, 62(5), pp. 464–475. doi: 10.1016/j.jclinepi.2008.12.011.

Treweek, S. and Zwarenstein, M. (2009) 'Making trials matter: Pragmatic and explanatory trials and the problem of applicability', *Trials*, 10, pp. 1–9. doi: 10.1186/1745-6215-10-37.

Tse, T., Fain, K. M. and Zarin, D. A. (2018) 'How to avoid common problems when using clinicaltrials.gov in research: 10 issues to consider', *BMJ (Online)*,

361, pp. 1–8. doi: 10.1136/bmj.k1452.

Usman, M. S. *et al.* (2022) ‘The need for increased pragmatism in cardiovascular clinical trials’, *Nature Reviews Cardiology*. Springer US, 19(11), pp. 737–750. doi: 10.1038/s41569-022-00705-w.

Vashisht, P. *et al.* (2016) ‘Generalizability of Patients With Rheumatoid Arthritis in Biologic Agent Clinical Trials’, *Arthritis Care and Research*, 68(10), pp. 1478–1488. doi: 10.1002/acr.22860.

Versavel, S. *et al.* (2023) ‘Diversity, equity, and inclusion in clinical trials: A practical guide from the perspective of a trial sponsor’, *Contemporary Clinical Trials*. Elsevier Inc., 126(September 2022), p. 107092. doi: 10.1016/j.cct.2023.107092.

Wang, S. V. *et al.* (2019) ‘Using Real-World Data to Extrapolate Evidence From Randomized Controlled Trials’, *Clinical Pharmacology and Therapeutics*, 105(5), pp. 1156–1163. doi: 10.1002/cpt.1210.

Wang, Y. *et al.* (2019) ‘Sex differences in the association between diabetes and risk of cardiovascular disease , cancer , and all-cause and cause-specific mortality.’, *BMC Cardiovascular Disorders*. BMC Medicine, 17(136), pp. 1–18.

Weisberg, H. I., Hayden, V. C. and Pontes, V. P. (2009) ‘Selection criteria and generalizability within the counterfactual framework: Explaining the paradox of antidepressant-induced suicidality?’, *Clinical Trials*, 6(2), pp. 109–118. doi: 10.1177/1740774509102563.

Weng, C. *et al.* (2010) ‘Formal representation of eligibility criteria: A literature review’, *Journal of Biomedical Informatics*, 43(3), pp. 451–467. doi: 10.1016/j.jbi.2009.12.004.

Whicher, C. A., O'Neill, S. and Holt, R. I. G. (2020) 'Diabetes in the UK: 2019', *Diabetic Medicine*, 37(2), pp. 242–247. doi: 10.1111/dme.14225.

Wieseler, B. *et al.* (2012) 'Impact of document type on reporting quality of clinical drug trials: A comparison of registry reports, clinical study reports, and journal publications', *BMJ (Online)*, 344(7838), pp. 1–11. doi: 10.1136/bmj.d8141.

Wise, R. A. *et al.* (2020) 'Adjudication of cardiovascular events in patients with chronic obstructive pulmonary disease: SUMMIT trial', *Clinical Trials*, 17(4), pp. 430–436. doi: 10.1177/1740774520920897.

Wolff, J. E. *et al.* (2022) 'Adverse Events in Placebo Arms of Cancer Trials', *Anticancer Research*, 42(6), pp. 2805–2812. doi: 10.21873/anticanres.15761.

Wortzel, J. R. *et al.* (2020) 'Trends in mental health clinical research: Characterizing the ClinicalTrials.gov registry from 2007–2018', *PLoS ONE*, 15(6), pp. 1–28. doi: 10.1371/journal.pone.0233996.

Xiaochen Yuan, S. B. (2019) 'Bivariate Logistic Bayesian LASSO for Detecting Rare Haplotype Association with Two Correlated Phenotypes', *Physiology & behavior*, 176(5), pp. 139–148. doi: 10.1002/gepi.22258.Bivariate.

Yao, B. *et al.* (2013) 'Safety monitoring in clinical trials', *Pharmaceutics*, 5(1), pp. 94–106. doi: 10.3390/pharmaceutics5010094.

Yao, Y. *et al.* (2021) 'Serious Adverse Events Reporting in Phase III Randomized Clinical Trials of Colorectal Cancer Treatments: A Systematic Analysis', *Frontiers in Pharmacology*, 12(November), pp. 1–16. doi: 10.3389/fphar.2021.754858.

Zelniker, T. A. *et al.* (2019) 'SGLT2 inhibitors for primary and secondary prevention of cardiovascular and renal outcomes in type 2 diabetes: a systematic

review and meta-analysis of cardiovascular outcome trials', *The Lancet*. Elsevier Ltd, 393(10166), pp. 31–39. doi: 10.1016/S0140-6736(18)32590-X.

Supplementary

Data management

The final dataset was stored in a CSV file. Four separate datasets were used in this thesis. The first dataset contained the scores of the 9 domains for the PRECIS-2 tool (10 variables). This dataset was analysed in chapter 6, which focused on the PRECIS-2 assessment. The second dataset contained the SAEs data (6 variables), which was analysed in chapter 4 and 5. The third dataset included trial and baseline characteristics (10 variables) which was analysed in the chapter 7. The fourth dataset contained elements of eligibility criteria (16 variables) and was employed in chapter 8, which explored the association between SAEs and eligibility criteria.

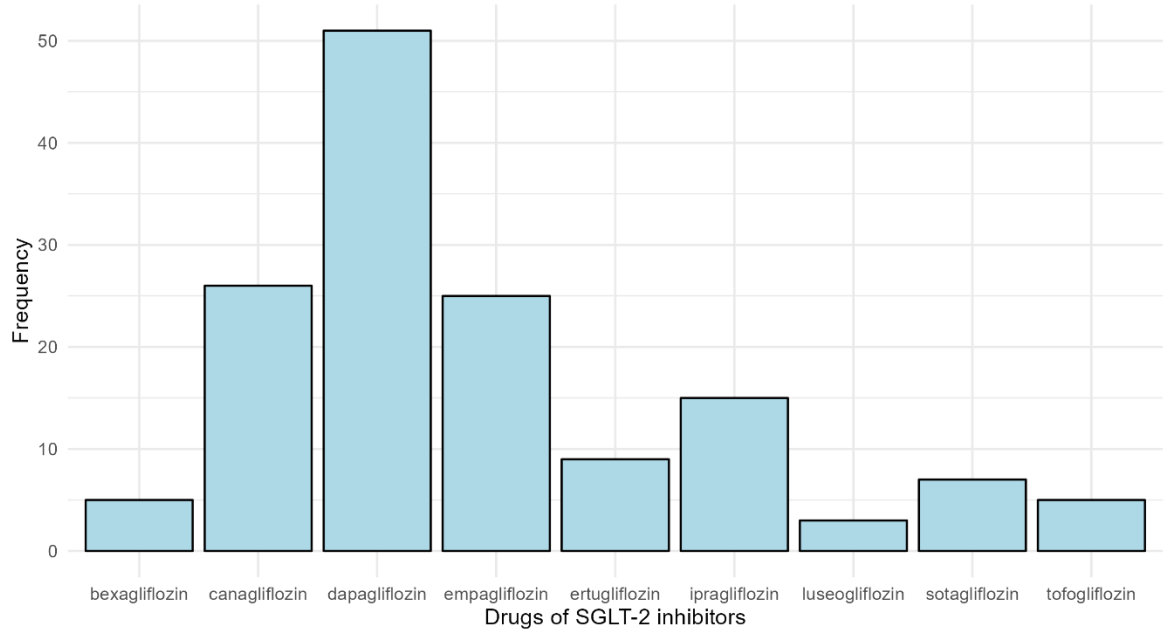

Figure S.1 The frequency of SGLT-2 Medication

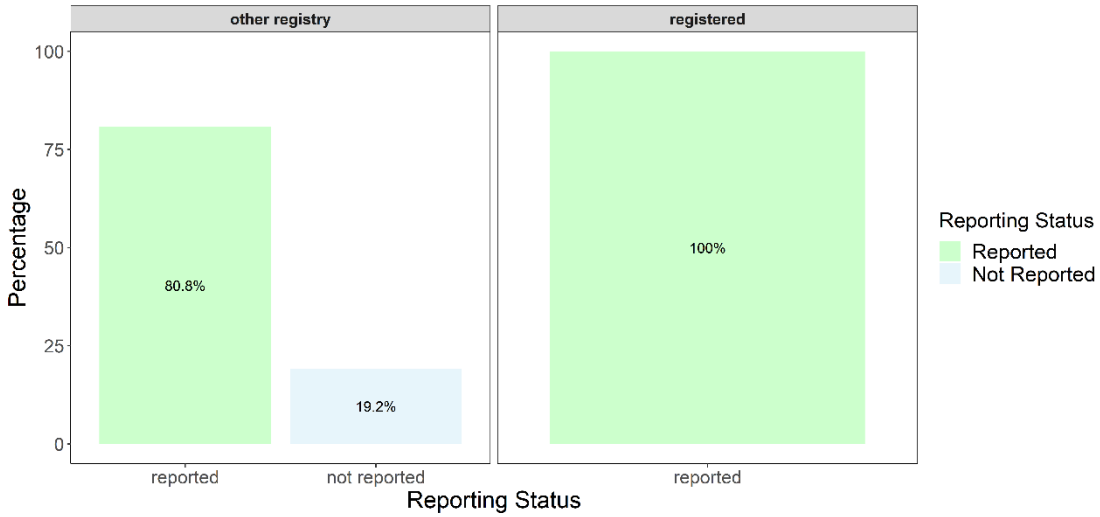

Figure 0.2 Level of SAE reporting within trial registry

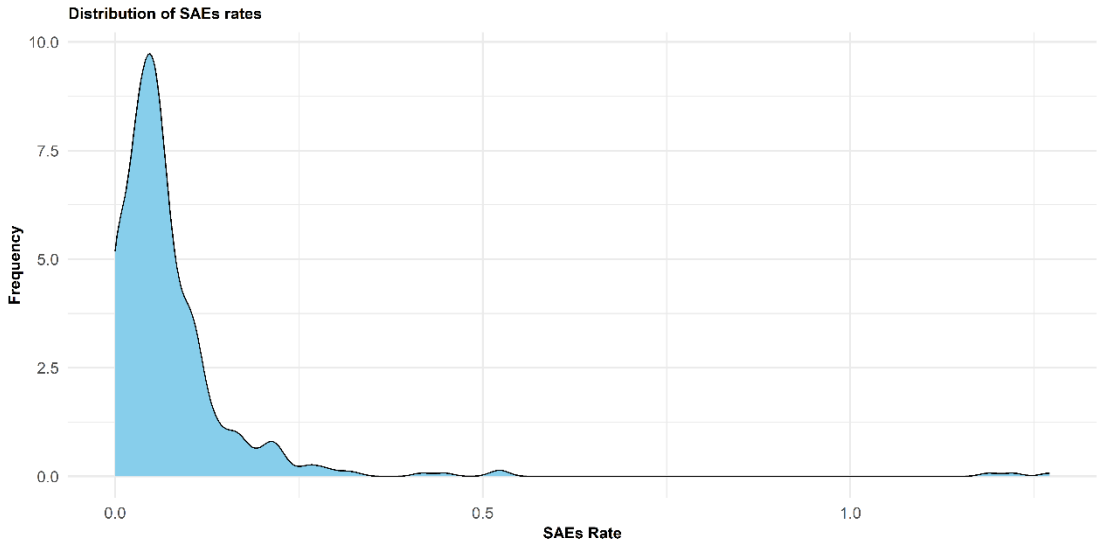

Figure 0.3 The rates of serious adverse event

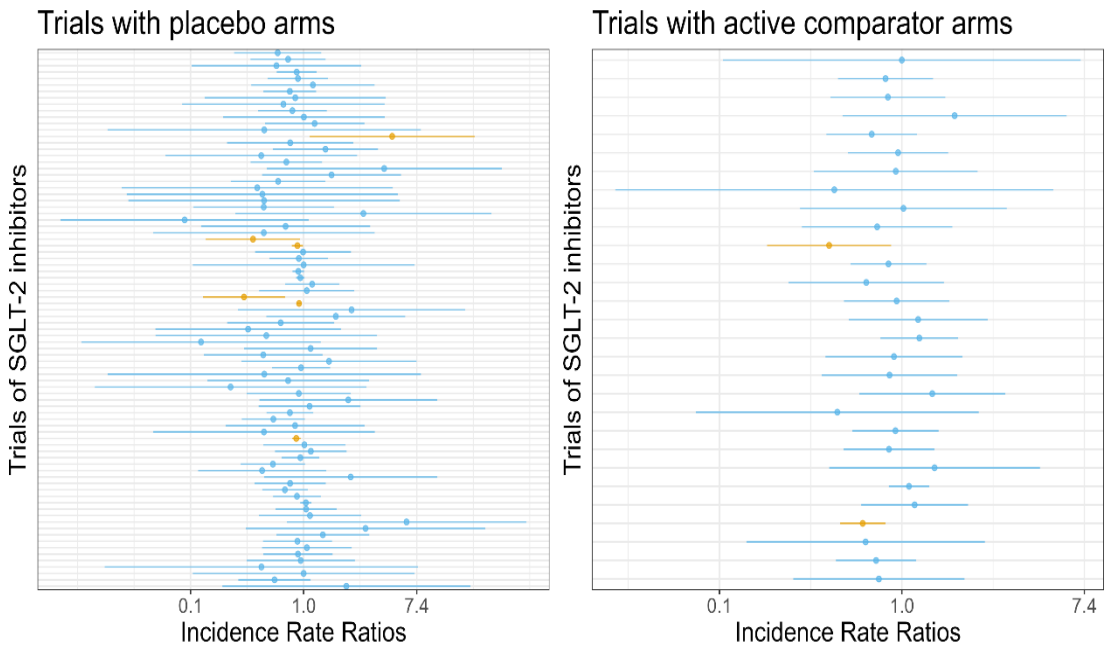

Figure 0.4 Difference in SAE rates between intervention and control arms

Table S.1 Examples of pragmatic and explanatory trial for each domain

Domain	Example of pragmatic trial	Example of explanatory trial
Eligibility criteria	Inclusion criteria included: patients with T2DM aged ≥20years, with a body mass index (BMI) of ≤45.0 kg/m^2 and HbA1c concentrations ≥7.0and ≤10.0% at screening, despite diet and exercise regimens and monotherapy with an SU, biguanide, TZD, AGI, DPP-4inhibitor or glinide, were eligible for the study. Key exclusion criteria included: uncontrolled hyperglycaemia (>240mg/dl) after an overnight fast, (eGFR) during screening or run-in of<60 ml/min/1.73m2, treatment with anti-obesity drugs <12 weeks before consent.	Eligible patients were adults with type 2diabetes and an HbA1c of 7.0-10.5%. The trial excluded the following: History of pancreatitis (acute or chronic), Any of the following: myocardial infarction, stroke or hospitalization for unstable angina, or transient ischemic attack within the past 180 days prior to the day of screening, renal impairment defined as estimated Glomerular Filtration Rate <60 ml/min/1.73m2, treatment with any medication for the indication of diabetes or obesity and history of diabetic ketoacidosis.
Recruitment	The recruitment of the initial cohort was done at 386 centres in 24 countries, commencing in December 2009 and completing in March 2011	Patients were recruited through advertisements in newspapers, radio and television.
Setting	This was a phase 3, randomized, double blind, parallel-group study conducted from August 2011 to September 2013 in 197 centres in 22 countries.	The present randomized, 24-week, open-label, parallel-group comparative clinical trial enrolled patients treated at Shiga University of Medical Science Hospital, Shiga, Japan.
Organisation	Patients received diet and exercise counselling based on local recommendations.	After assessing eligibility criteria by preliminary telephone interview, the volunteers underwent detailed clinical evaluation, anthropometry, carotid ultrasound,

		echocardiography, ophthalmologic evaluation and laboratory evaluation. Clinical and laboratory measurements were performed at the Clinical Research Centre (CRC) and Atherosclerosis and Vascular Biology Laboratory (AtheroLab) at UNICAMP.
Flexibility (delivery)	After a 2-week placebo run-in period, subjects were randomized (1:1:1:1:1) to receive empagliflozin 25 mg/linagliptin 5 mg as a fixed-dose combination (FDC) tablet, empagliflozin 10 mg/linagliptin 5 mg FDC tablet, empagliflozin 25 mg, empagliflozin 10 mg, or linagliptin 5 mg for 52 weeks as add-on to metformin at an unchanged dose. FDC tablets, empagliflozin tablets, and linagliptin tablets were taken once daily in the morning. No restriction, or specific direction.	During the study, patients were prohibited from using antihyperglycemic drugs other than the allocated study drug. Patients were also prohibited from using continuous systemic corticosteroid treatment for 2 weeks, appetite suppressants and other investigational products. However, other drugs were allowed, and their use was to be recorded. Compliance with study drug administration was assessed by the investigator who recorded the number of days taken, and the first and last day taken, using the case report forms. Regarding diet/exercise therapy, the investigator recorded the type of therapy, duration of therapy, details (including prescribed energy intake) throughout the study period and the reason for initiation of the therapy. Patients were instructed to continue their diet and exercise therapies unchanged from before the study until the end of the follow-up period.
Flexibility (adherence)	The investigators were not given specific guidelines for	The purpose of the run-in period was to evaluate participants'

	encouraging adherence to diet/exercise therapy, only a guideline that the prior therapy should be continued.	willingness to adhere to the long-term treatment and follow-up planned in the trial.
Follow-up	Post randomization follow-up in the first 12 months is scheduled at 3, 13, 26, 39, and 52 weeks after randomization and thereafter at 3-month intervals	All patients will be followed up for 4 weeks after the last dose of study drug. Furthermore, the study design includes a 4-week follow-up period to assess, using a mixed MTT.
Primary outcome	The primary outcome of the study is time to first occurrence of CV death, non-fatal myocardial infarction (MI, excluding silent MI), or non-fatal stroke i.e., 3-point major adverse cardiovascular events (3P-MACE). Doctor relevant outcome. Composite outcome.	The primary endpoint is the changes in UACR (urine albumin-to-creatinine ratio) from the baseline after a 2-year observation.
Primary analysis	The impact of the intervention on the primary endpoint will be assessed by an intention-to-treat analysis.	we analysed the primary efficacy endpoint in the per-protocol analysis set.

Check assumptions of linear regression model

The assumptions of linear regression were tested before conducting the analysis using the Global Validation of Linear Models Assumptions (gvlma) package in R. This package checks essential assumptions, including linearity, normality of residuals, and homoscedasticity (constant variance of residuals). All key assumptions were met, indicating the reliability of the model.

Figure 0.5 Frequency of PRECIS-2 score for all domains with missing information

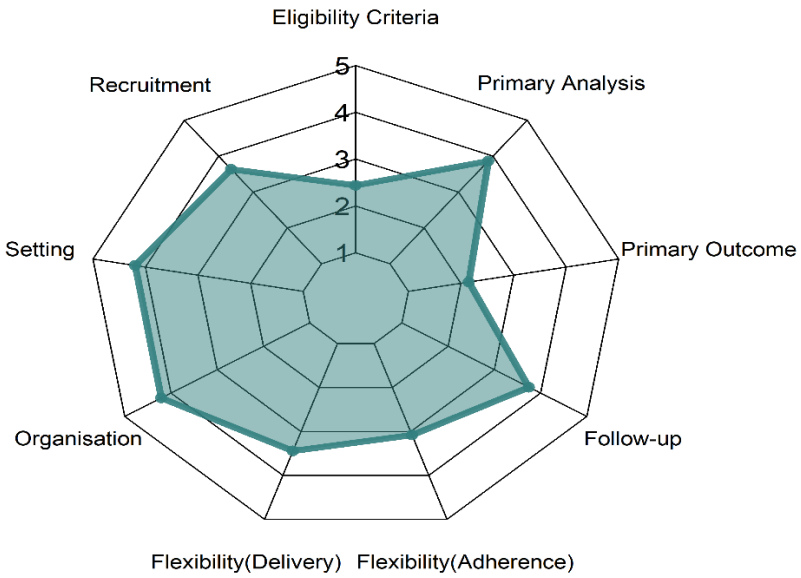

Figure 0.6 Wheel for mean PRECIS-2 scores for all domains with missing information

Table S.2 Summary of mean PRECIS-2 scores among domains and medications of SGLT-2 inhibitors

PRECIS- 2 domains	Mean score (SD)	Medication	Frequency	Mean score (SD)
Eligibility Criteria	2.45 (0.76)	Dapagliflozin	51	3.32 (0.58)
Recruitment	3.14 (0.72)	Empagliflozin	25	3.31 (0.47)
Setting	4.18 (1.26)	Canagliflozin	26	3.54 (0.4)
Organisation	4.08 (0.94)	Ipragliflozin	15	3.13 (0.37)
Flexibility (delivery)	3.42 (1.26)	Ertugliflozin	9	3.31 (0.46)
Flexibility (adherence)	3.08 (1.4)	Tofogliflozin	5	3.73 (0.23)
Follow-up	3.73 (0.75)	Luseogliflozin	3	2.93 (0.06)
Primary outcome	2.16 (0.66)	Bexagliflozin	5	3.47 (0.27)
Primary analysis	3.79 (0.89)	Sotagliflozin	7	3.16 (0.18)
Trial's Mean Score			3.34 (0.48)	
Trial's Median Score			3.33	

Comparison of PRECIS-2 scores based on types of their registries

Trials registered on ClinicalTrial.gov had higher PRECIS-2 scores compared with other trials (3.57 [0.49] vs 3.07 [0.66]), although there was considerable overlap. Nonetheless, of the 17 trials which were most pragmatic (mean PRECIS-2 score > 4) 16 were registered on ClincialTrials.gov (Figure S.7).

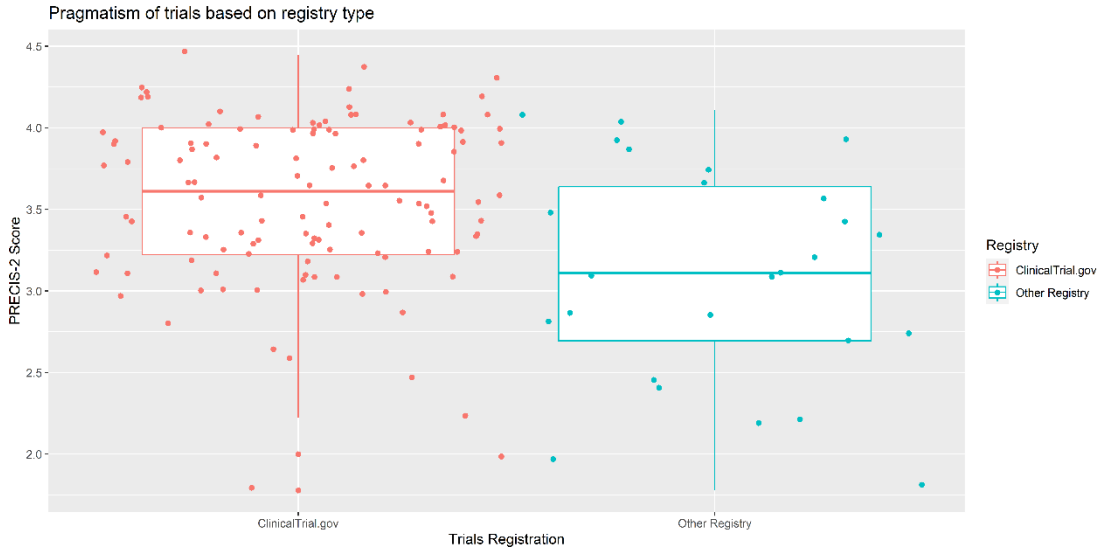

Figure 0.7 Compare PRECIS-2 score between trial registries

Table S.3 Sensitivity analysis for the association between SAE rates and PRECIS-2 domains (missing domains were assigned with the median domain score)

Serious Adverse Event Rates						
PRECIS-2 Domains	**Unadjusted**			**Adjusted**		
	β	**CI**	**P**	**β**	**CI**	**P**
Eligibility criteria	0.14	-0.05 – 0.33	0.146	0.07	-0.13 – 0.27	0.499
Recruitment	0.39	0.16 – 0.62	0.001	0.24	-0.01 – 0.50	0.062
Setting	0.48	0.28 – 0.68	<0.001	0.42	0.17 – 0.67	0.001
Organisation	0.07	-0.13 – 0.27	0.469	-0.11	-0.34 – 0.11	0.309
Flexibility delivery	0.10	-0.08 – 0.28	0.275	0.01	-0.22 – 0.24	0.925
Flexibility adherence	0.02	-0.16 – 0.21	0.804	0.05	-0.17 – 0.26	0.659
Follow up	-0.04	-0.25 – 0.17	0.688	-0.01	-0.22 – 0.19	0.893
Primary outcome	-0.09	-0.35 – 0.16	0.472	0.06	-0.21 – 0.32	0.673
Primary analysis	0.06	-0.13 – 0.25	0.519	0.02	-0.17 – 0.22	0.817

Table S.4 Sensitivity analysis for the association between SAE rates and PRECIS-2 domains (missing domains were imputed using MICE)

Serious Adverse Event Rates						
PRECIS-2 Domains	Unadjusted			Adjusted		
	β	CI	P	β	CI	P
Eligibility criteria	0.14	-0.05 – 0.33	0.143	0.10	-0.10 – 0.31	0.334
Recruitment	0.46	0.24 – 0.67	<0.001	0.23	-0.17 – 0.62	0.260
Setting	0.48	0.28 – 0.68	<0.001	0.36	-0.01 – 0.72	0.057
Organisation	0.05	-0.15 – 0.24	0.633	-0.18	-0.41 – 0.05	0.115
Flexibility delivery	0.10	-0.08 – 0.28	0.281	0.03	-0.20 – 0.27	0.794
Flexibility adherence	0.03	-0.16 – 0.21	0.786	0.01	-0.21 – 0.23	0.938
Follow up	0.00	-0.21 – 0.21	0.997	-0.01	-0.22 – 0.19	0.887
Primary outcome	-0.10	-0.36 – 0.17	0.468	0.06	-0.23 – 0.34	0.705
Primary analysis	0.11	-0.07 – 0.30	0.231	0.03	-0.17 – 0.22	0.784

Table S.5 Missing data of trial and baseline characteristics before imputation

Variables	Missing count	Missing percentages
Baseline sample size	5	3.4
Mean age (years)	7	4.8
Male, %	12	8.2
Duration of diabetes (years)	38	26
HbA1c, %	22	15
Year difference (years)	0	0
Trial duration (years)	0	0
Sponsorship	0	0
Trial phase	18	12.3
Blinding	0	0

Table S.6 Sensitivity analysis for the comparison between SAE and PRECIS-2 (using complete case for baseline and trial characteristics without imputation)

Metrics of Representativeness								
Umpire variables	**SAE**		**PRECIS-2**		**Difference in estimates**		**Significant favouring *1**	**Direction of favouring *2**
	β	**95% CI**	**β**	**95% CI**	**β**	**CI**		
Type of blinding; Double or more	0.22	-0.47 to 0.90	-0.78	-1.50 to -0.04	0.99	0.00 to 1.99	SAE	SAE
Trial phases; Phase IV	0.07	-0.59 to 0.74	-0.36	-1.08 to 0.36	0.43	-0.53 to 1.38	Neutral	SAE
Sponsorship; industrial-sponsored trial	0.79	-0.22 to 1.82	-0.76	-1.90 to 0.40	1.55	0.06 to 3.07	SAE	SAE
Trial duration/years	-0.06	-0.29 to 0.17	-0.14	-0.39 to 0.12	0.08	-0.26 to 0.43	Neutral	Neutral
Year difference/decade; (the difference between when a clinical trial was first registered and when the intervention being studied was first trialled)	0.49	-0.07 to 1.09	-0.09	-0.72 to 0.55	0.58	-0.30 to 1.48	Neutral	SAE
Baseline sample size	0.04	-0.14 to 0.23	0.44	0.24 to 0.65	-0.39	-0.68 to -0.12	PRECIS-2	PRECIS-2

Age/decade	0.91	0.20 to 1.64	-0.61	-1.41 to 0.23	1.51	0.40 to 2.59	SAE	SAE
Proportion of male	-1.37	-3.17 to 0.52	-0.44	-2.53 to 1.67	-0.92	-3.66 to 1.86	Neutral	SAE
Duration of diabetes/decade	0.49	-0.30 to 1.29	0.90	0.05 to 1.77	-0.41	-1.60 to 0.74	Neutral	PRECIS-2
HbA1c	0.01	-0.52 to 0.53	-0.13	-0.71 to 0.46	0.15	-0.63 to 0.92	Neutral	SAE

*[1] **Criteria for significant favouring (CI does not include the null):** SAE: if the difference in association significantly favours SAEs. PRECIS-2: if the difference in association significantly favours the PRECIS-2. Neutral: if there is no difference in associations between metrics.

*[2] **Criteria for direction of favouring (direction of point estimate):** SAE: if the difference estimate in association is in the direction of SAEs. PRECIS-2: if the difference estimate in association is in the direction of the PRECIS-2. Neutral: if the difference in associations between metrics was close to zero (null).