2017

A novel approach to support evidence-based medicine: should sulfonylureas remain an acceptable therapy for diabetes?

BOSTON UNIVERSITY

SCHOOL OF PUBLIC HEALTH

Dissertation

A NOVEL APPROACH TO SUPPORT EVIDENCE-BASED MEDICINE: SHOULD SULFONYLUREAS REMAIN AN ACCEPTABLE THERAPY FOR DIABETES?

by

W. RYAN POWELL

B.A., Indiana University, 2004
M.A., Wake Forest University, 2008

Submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

2017

Approved by

First Reader

Cindy L. Christiansen, Ph.D.
Associate Professor of Health Law, Policy and Management

Second Reader

Lewis E. Kazis, Sc.D.
Professor of Health Law, Policy and Management
Boston University, School of Public Health

Research Assistant Professor of Medicine
Boston University, School of Medicine

Third Reader

Donald R. Miller, Sc.D.
Core Investigator
Center for Healthcare Organization & Implementation Research
Edith Nourse Rogers Memorial VA Hospital, Bedford, MA

Outside Reader

Manuel Cifuentes, M.D., M.P.H., Sc.D.
Associate Professor of Public Health
Regis College

Outside Reader

Varsha Vimalananda, M.D., M.P.H.
Assistant Professor of Medicine
Boston University, School of Medicine

ACKNOWLEDGMENTS

I am grateful to many people who made this work possible. First, thank you to my committee: Cindy L. Christiansen, Lewis E. Kazis, and Donald R. Miller, and my external readers: Manuel Cifuentes and Varsha Vimalananda. I am fortunate to have a great set of mentors who graciously offered their valuable expertise to strengthen this dissertation. I really appreciate the many conversations that encouraged me to think deeper about complex issues around health services research without losing focus on the broader perspective.

Thank you to my wife for standing by me throughout the entire process. Her steadfast support, reliable encouragement, and all-round positive thinking helped make this work possible. Thank you to my two-year-old daughter who taught me to constantly keep everything in perspective and for always insisting I step away from the work for a bit of laughter, foolishness, and fun. Finally, I acknowledge and thank my entire family for their unwavering support throughout my life.

A NOVEL APPROACH TO SUPPORT EVIDENCE-BASED MEDICINE: SHOULD SULFONYLUREAS REMAIN AN ACCEPTABLE THERAPY FOR DIABETES?

W. RYAN POWELL

Boston University School of Public Health, 2017

Major Professor: Cindy L. Christiansen, Ph.D., Associate Professor of Health Law, Policy and Management

ABSTRACT

A key element in evidence-based medicine approaches is the ability for clinicians to evaluate the scientific rigor and relevance of research evidence. In the treatment of diabetes, clinicians make increasingly difficult decisions about which drug regimens are best for their patients with limited evidence-based information.

While the consensus is that metformin should be the initial drug treatment when diet and exercise are not sufficient, clinicians disagree on whether sulfonylureas should remain a suitable therapy after metformin. While this would be improved with further research investigating the comparative safety of therapeutic options, there is also need for better ways to synthesize available information to guide evidence-based decision-making in health services research.

Study 1 summarizes the pre-existing evidence on the long-term safety risks associated with sulfonylurea therapy relative to other drug classes. Results from a series of meta-analyses provide some evidence that sulfonylureas are associated with elevated all-cause mortality and cardiovascular risks relative to several other medications, either as a monotherapy or in combination with metformin.

Study 2 analyzes the comparative safety of second-line treatment in diabetic patients in the Veterans Health Administration to address gaps in the literature. Results suggest that second-line use of sulfonylureas is associated with increased risks compared to thiazolidinediones. Results also suggest that changes to existing metformin therapy may lead to differential hazards.

Clinicians may disagree about the quality of the evidence as well as the relevancy to their own treatment population. Improvements in methods for evidence-based medicine that take this into account are needed. Study 3 applies an underutilized research method that allows for a more thoughtful synthesis of all available evidence. This framework allows clinicians to incorporate the scientific rigor and relevancy of previous study results when integrating new data into their current knowledge base. Results suggest an elevated risk in all models for sulfonylureas compared to thiazolidinediones and highlight the need to design more focused research to support clinical decision-making around medication safety. This novel application to evidence synthesis shows promise as applied to a health services research problem and has potential as a useful framework in other health services research areas.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AT ..... As treated

DPP-4 inhibitor ..... Dipeptidyl peptidase-4 inhibitor

EBM ..... Evidence-based medicine

GLP-1 ..... Glucagon-like peptide-1

IPTW ..... Inverse probability of treatment weights

ITT ..... Intent to treat

MEGL ..... Meglitinide

METF ..... Metformin

RCT ..... Randomized controlled trial

SGLT-2 ..... Sodium-glucose co-transporter 2

SU ..... Sulfonylurea

TZD ..... Thiazolidinediones

CHAPTER 1 : INTRODUCTION

The practice of evidence-based medicine (EBM) relies, in part, on the integration of empirical research evidence into treatment decisions. Yet, the ability for clinicians to evaluate the scientific rigor and relevance of research evidence is difficult especially when evidence is limited or inconsistent. To make matters more challenging, research methods that synthesize information in a way that is consistent with EBM are underutilized and innovative approaches are left unexplored. In the treatment of diabetes, clinicians must make increasingly difficult decisions about which drug regimens to manage patients with limited evidence-based information to guide appropriate therapy.

Among the most pressing clinical decisions is which drugs should be used after metformin is no longer sufficient, and in particular, whether sulfonylureas should remain as a suitable second-line treatment. Sulfonylureas are widely prescribed but several studies suggest increased long-term safety risks relative to other drug classes. Still, clinical opinion is fragmented on whether sulfonylureas should remain as a suitable therapy after diet/exercise and metformin therapies fail to control glucose levels. Other classes of drugs are available but with limited long-term evidence regarding their long-term safety.

There have been insufficient long-term comparisons evaluating sulfonylureas relative to other medications. Clinical opinions are disparate and there is variation in treatment. While this would be improved with further research investigating the comparative safety of therapeutic options, there is also need for better ways to synthesize available information to guide evidence-based decision-making.

In this chapter, the existing evidence on the comparative safety of sulfonylureas is discussed. In addition, an overview of EBM concepts and the diabetes treatment model are described in further detail and are used to guide the research reported in Chapters Two, Three, and Four. The first study summarizes the pre-existing evidence on the long-term safety risks associated with sulfonylurea therapy relative to other drug classes. Using observational cohort and randomized controlled trial studies with at least one year of follow-up in patients without serious conditions at baseline, a series of meta-analyses is conducted to provide pooled estimates of the risks of cardiovascular events and death.

The second study addresses several gaps in the existing literature by analyzing the comparative safety of second-line treatment in diabetic patients served by the Veterans Health Administration. Long-term cardiovascular and mortality risks between sulfonylureas and other drug classes are evaluated as well as how changes to existing metformin use coinciding with second-line therapy influences hazard rates.

While Study One summarizes prior evidence on sulfonylurea safety risks and Study Two conducts a new comparative safety analysis to address weaknesses in the current comparative safety literature, clinicians may disagree about the quality of the evidence as well as its relevancy to their own treatment population. EBM friendly research methods that take this into account are lacking, despite the need for clinicians to use the best external evidence to inform complex treatment decisions. Study Three attempts to address this need by focusing on developing a research method that allows clinicians to vary the uncertainty around pre-existing evidence in a way that reflects a clinician's appraisal of the quality of the evidence and relevancy to their own treatment

population. Taken altogether, this dissertation provides a comprehensive evaluation and understanding of the comparative safety of sulfonylurea long-term risks and applies a research method to aid EBM practice.

Background: Should sulfonylureas remain as an acceptable second-line oral agent therapy in patients with Type 2 diabetes?

Sulfonylurea medications have been used for glycemic control in patients with type 2 diabetes since the 1950s, but they are increasingly controversial and have attracted considerable debate in recent years (Abrahamson, 2015; Genuth, 2015). While they are an inexpensive, widely used, and very effective therapy for glycemic control (Bennett, Wilson, et al., 2011; UKPDS-33, 1998), they are also linked to adverse outcomes, including higher rates of cardiovascular events and death. The existing evidence on these outcomes is briefly summarized below.

Cardiovascular Disease Morbidity and Cardiovascular Disease-Related Death

Cardiovascular disease is the main cause of death for people with diabetes mellitus, yet evidence related to how particular drug therapies contribute to increased cardiovascular events is unclear and insufficient. Despite its importance, most clinical studies are instead designed to assess glycemic control — often short-term assessments that lack the power to examine rare events. This in turn, results in imprecise estimates of the risk for cardiovascular morbidity and mortality.

Cardiovascular Mortality

Clinical trial evidence on the risk of cardiovascular-related death associated with

sulfonylureas is limited. In an early randomized control trial (RCT) from the 1960s, The University Group Diabetes Program (UGDP) found an increased risk in cardiovascular death for first generation sulfonylurea monotherapy (tolbutamide) compared to those who received a placebo, but has not been confirmed in later trials (e.g., ADOPT and UKPDS) for sulfonylurea monotherapy (Kahn et al., 2006; Meinert, Knatterud, Prout, & Klimt, 1970; UKPDS-33, 1998). When the United Kingdom Prospective Diabetes Study (UKPDS) published their primary results in 1998, it seemed to exculpate second generation sulfonylureas by reporting that patients randomized to sulfonylurea therapy did not differ significantly in diabetes-related death (which included death from a myocardial infarction and stroke) compared to those in the diet only control group (UKPDS-33, 1998). Yet, a substudy investigating the risk for patients who added metformin to sulfonylurea therapy found other results (UKPDS-34, 1998). They found when sulfonylurea was used in combination with metformin it was related to a higher risk of diabetes-related death compared to those receiving only sulfonylurea monotherapy – a finding the authors concluded was due to chance only (type 1 error). This comparison has yet to be re-evaluated in newer RCTs, but remains a concern. More recent evidence from observational studies provide additional evidence for an elevated risk in cardiovascular-related death for sulfonylureas compared to metformin (Eurich, Majumdar, McAlister, Tsuyuki, & Johnson, 2005; Evans, Ogston, Emslie-Smith, & Morris, 2006; Johnson, Majumdar, Simpson, & Toth, 2002; Roumie et al., 2012).

Related to other therapy class comparisons, only one major RCT compared thiazolidinediones (TZDS) against sulfonylureas (ADOPT trial) but there was not

sufficient power to assess a statistically significant difference (Kahn et al., 2006). Also, there has been little research comparing sulfonylureas to newer drugs that may eventually replace sulfonylurea use such as GLP1-agonists, DPP-4 inhibitors, and SGLT-2 inhibitors. Of these classes, DPP-4 inhibitors have been studied the most. In a pooled analysis, Zhang et al. (2014) compared DDP-4 inhibitor use to sulfonylurea therapy using data from four RCTs lasting two years or less. They estimated that DDP-4 inhibitor use was associated with 47% fewer cardiovascular events when compared to sulfonylureas, but cautioned interpreting the results because of the relatively small number of studies used in the analysis and because of the limited study duration involved in each (Mantel-Haenszel OR=0.53, 95% CI [0.32, 0.87]).

Cardiovascular Morbidity

There is insufficient and unclear evidence for risk of cardiovascular morbidity with sulfonylurea use relative to other medications. The majority of comparisons have compared sulfonylureas to metformin. There is conflicting information in the literature with some studies supporting an elevated signal for sulfonylureas compared to metformin while others fail to confirm this finding. In the ADOPT trial, researchers found little difference in nonfatal MI and stroke compared to metformin (MI: 1.0% of sulfonylurea users vs. 1.4% of metformin users; Stroke: 1.2% for sulfonylureas and 1.3% for metformin) (Kahn et al., 2006). This is consistent with a large cohort study consisting of 91,521 patients with diabetes in the United Kingdom General Practice Research Database (i.e., a large primary care database that today is a part of the Clinical Practice Research Datalink) that found no evidence for an increased risk of non-fatal MI in the sulfonylurea

group (Tzoulaki et al., 2009).

However, several observational study findings favor metformin over sulfonylurea on cardiovascular events (Pantalone et al., 2009; Roumie et al., 2012; Tzoulaki et al., 2009). For example, in an analyses of 20,450 diabetics receiving care at the Cleveland Clinic, researchers found an increased risk for congestive heart failure for sulfonylureas compared to metformin after adjusting for baseline characteristics and risk factors (HR=0.76, 95% CI [0.64, 0.91]) (Pantalone et al., 2009). Research using other cardiovascular disease morbidity measures has also found elevated signals for sulfonylureas when compared against metformin. They include nonfatal cardiovascular hospitalizations, acute MI, coronary revascularization, and cardiovascular events composite outcomes (Johnson et al., 2002; McAfee, Koro, Landon, Ziyadeh, & Walker, 2007). For instance, in a retrospective cohort study of veterans, investigators compared the effects of sulfonylureas and metformin on the risk of cardiovascular events (composite outcome consisting of nonfatal acute MI and stroke) in new users of antidiabetic medications and found an elevated risk for patients who first were treated with sulfonylureas compared to those were given metformin (HR=1.13, 95% CI [1.03, 1.24]) (Roumie et al., 2012).

Worse outcomes have also been reported for sulfonylurea/metformin (SU+METF) combination therapy against metformin monotherapy (Hermann et al., 1994; McAfee et al., 2007). In a small six-month trial, researchers found that 14% (10 patients) of patients in the SU+METF combination therapy group had a cardiovascular event compared to a 5% (two patients) in the metformin only therapy group, though the difference was not

statistically significant (Hermann et al., 1994).

Compared to TZDs, evidence is inconsistent with some studies reporting elevated cardiovascular risks for sulfonylureas (Jain, Osei, Kupfer, Perez, & Zhang, 2006; McAfee et al., 2007), similar risks between TZDs and sulfonylureas (Kahn et al., 2006; Pantalone et al., 2009; St John Sutton et al., 2002), and lower risks for sulfonylureas (Brownstein et al., 2010; Hsiao et al., 2009; St John Sutton et al., 2002).

All-Cause Mortality

The extent to which medications used to manage type 2 diabetes affect all-cause mortality remains unclear. Insufficient evidence has been produced from RCTs. RCTs have not been designed to evaluate long-term safety. To date, the majority of RCTs have been less than one year (16–30 weeks) with relatively few or no deaths occurring during the study period (Chien et al., 2007; DeFronzo & Goodman, 1995; A. J. Garber, Donovan, Dandona, Bruce, & Park, 2003; Hemmingsen et al., 2013; Kahn et al., 2006). As a result, existing RCTs have not been powered to evaluate mortality. While RCTs have not been helpful in understanding mortality risk, several observational studies link sulfonylurea monotherapy to increased all-cause mortality when compared to other diabetic treatments, particularly metformin.

Observational studies supporting a higher risk of death in sulfonylurea monotherapy compared to metformin include: three observational studies using claims information from a cohort of patients in the Saskatchewan Health Registry (Eurich et al., 2005; Johnson et al., 2002; Simpson, Majumdar, Tsuyuki, Eurich, & Johnson, 2006), one study of patients receiving care from the United Kingdom National Health Service in

Scotland (Evans et al., 2006), one study of patients receiving care at the Cleveland Clinic (Pantalone et al., 2009), one study of patients in northwestern Italy without cancer at baseline (Bo et al., 2012), two studies of patients in the United Kingdom General Practice Research Database (Azoulay, Schneider-Lindner, Dell'Aniello, Schiffrin, & Suissa, 2010; Gulliford & Latinovic, 2004; Tzoulaki et al., 2009), and a study of patients in the Veterans Health Administration (Kahler et al., 2007). Average follow-up times for these studies were considerably larger than existing RCTs with most ranging 5–8 years of follow-up.

While there appears to be evidence for at least a slight risk of death compared to metformin, few studies have compared sulfonylurea to TZD, newer drug classes (e.g., DPP-4 inhibitors, GLP-1 agonists, SGLT-2 inhibitors) and most combination therapies. Only one study has compared risks of TZDs compared with other treatments. In that assessment, Cleveland Clinic patients had a higher risk when treated with sulfonylurea compared to TZDs (HR=1.69; 95% CI [1.23, 2.33]). Related to GLP-1 agonists, a six-month RCT in Japan was insufficient to assess mortality as a long-term outcome, with only one death in the GLP-1 group and no deaths in the sulfonylurea group (Seino, Min, Niemoeller, & Takami, 2012).

Outcome Summary

While many antidiabetic medications appear to have equal glucose-lowering efficacy alone and when combined with metformin (Morgan et al., 2012), further research is needed to determine whether they also provide greater long-term safety. RCTs are limited in their design to evaluate long-term outcomes, resulting in few events

and providing little evidence (Bennett, Wilson, et al., 2011; Bolen et al., 2007). The focus of many trials has been to make direct head-to-head comparisons to assess which medications work best at managing glucose-levels and were not designed to examine long-term risks (Holman, Paul, Bethel, Matthews, & Neil, 2008; Kahn et al., 2006; Lipska KJ & Krumholz HM, 2014; Nathan et al., 2013; Ryan et al., 2003; UKPDS-34, 1998). These trials have typically not been designed to evaluate long-term safety (e.g., small in size with relatively short follow-up periods), limiting the ability to obtain precise estimates of risk. Thus, much of the evidence is derived from observational studies, but the methodological rigor of such studies is challenged (e.g., internal threats to validity such as selection bias and unmeasured confounding are possible).

There are other shortcomings in existing comparative safety analyses. While sulfonylurea is commonly compared to metformin and TZD, there is insufficient comparative safety research on how newer classes of medications (e.g., DPP-4 inhibitor, GLP-1 agonists, SGLT-2 inhibitor) compared against sulfonylurea therapy. Also, there are even fewer comparative safety analyses that examine METF+SU combination therapy against other metformin combinations.

Significance: Clinical Opinion is Fragmented on Sulfonylurea Use.

Diabetic treatment involves a sequencing of therapies over time when glycemic control is not being obtained by the current therapy. These are commonly referred to as 'lines' of treatment. Clinicians have developed different opinions regarding the relative therapeutic benefits and harms of sulfonylureas, leading to the support of different guidelines recommendations. While the current consensus is that metformin should be

the initial first-line agent when diet and exercise are not sufficient, the debate is whether sulfonylurea should be used as a) a second-line agent after metformin, b) a third- or fourth-line treatments, or c) not at all. For example, The American College of Physicians (ACP) developed guidelines based on their evaluation of the existing comparative effectiveness of type 2 diabetes medications, which included outcomes related to glycemic control, mortality, cardiovascular events as well as safety outcomes and recommends sulfonylurea use as a suitable second-line therapy in the treatment of type 2 diabetics (Qaseem A, Barry MJ, Humphrey LL, Forciea M, & for the Clinical Guidelines Committee of the American College of Physicians, 2017; Qaseem, Humphrey, Sweet, Starkey, & Shekelle, 2012).

While this is a stance in agreement with the American Diabetes Association, the National Institute for Health and Care Excellence, the International Diabetes Federation, and the European Association for the Study of Diabetes, it differs from guidelines put forth by the American Association of Clinical Endocrinologists and American College of Endocrinology guidelines which cautions against the use of sulfonylurea as a first- or second-line agent in the treatment of type 2 diabetics (Garber et al., 2017; International Diabetes Federation Guideline Development Group, 2014; Inzucchi et al., 2012, 2015; Marathe, Gao, & Close, 2017; National Institute for Health and Clinical Excellence, 2014).

Significance: Diabetes is a Major Public Health Problem.

This research area is of public health importance given that approximately 9% (~29 million) of the adult population in the United States are diagnosed as diabetic, is the

seventh leading cause of death, and costs upwards of $245 billion or 1 in 5 healthcare dollars spent (as of 2012) (Centers for Disease Control and Prevention, 2014; Geiss LS, Wang J, Cheng YJ, & et al, 2014; Go et al., 2014). The prevalence is even higher amongst U.S. veterans (Reiber, Koepsell, Maynard, Haas, & Boyko, 2004) and it is particularly high among veterans receiving health care from the Department of Veteran Affairs (VA) (the study population in Study 2), with an estimated prevalence of about 25% (Veterans Health Administration, 2011).

Evidence based medicine model of clinical practice and diabetes medication treatment model

Evidence-based medicine conceptual model

As described above, different evaluations of the same evidence have contributed to the creation of different treatment guidelines. Evaluating the scientific rigor and relevance of research evidence is a key element for informing clinical decision-making in evidence-based medicine (EBM) approaches. EBM is commonly defined as "the integration of best research evidence with clinical expertise and patient values" (Rosenberg & Haynes with Sackett, Straus Sharon E. and Richardson W. Scott, 2000) and involves "the conscientious, explicit, and judicious use of clinically relevant research in making decisions about the care of individual patients" (Sackett, Rosenberg, Gray, Haynes, & Richardson, 1996). This is the basis for our conceptual model developed by Sacket et al. (2000) and is depicted in the diagram in Figure 1.1.

The integration of three main components are involved in most EBM conceptual models of decision-making: Best Research Evidence; Individual Clinical Expertise; and Patient Values and Preferences. First, the model acknowledges the importance of previous research in informing clinical practice. Best External Evidence involves evaluating evidence from systematic investigations that are clinically relevant. For diabetes research, this includes existing comparative effectiveness and safety evaluations. The gold standard in evidence is appropriately designed RCTs given their strong internal validity to infer causality. In EBM approaches physicians are "urged to assume that the clinical effectiveness results from RCTs could be applied to their own patients, unless there was a good reason not to make this default assumption" (Charles, Gafni, & Freeman, 2011; Straus, Richardson, Glasziou, & Haynes, 2005). However, RCT designs have their limitations. Evidence solely from RCTs is inadequate for the questions in this dissertation for several reasons. First, the relatively short follow-up and small sample sizes of RCTs make long-term evaluations difficult and underpowered (susceptible in making type II errors). Second, the generalizability is of concern. This is particularly true in populations that have multiple comorbidities (e.g., veteran and elderly populations) – a common criterion for study exclusion. That may not reflect common characteristics in clinical practice.

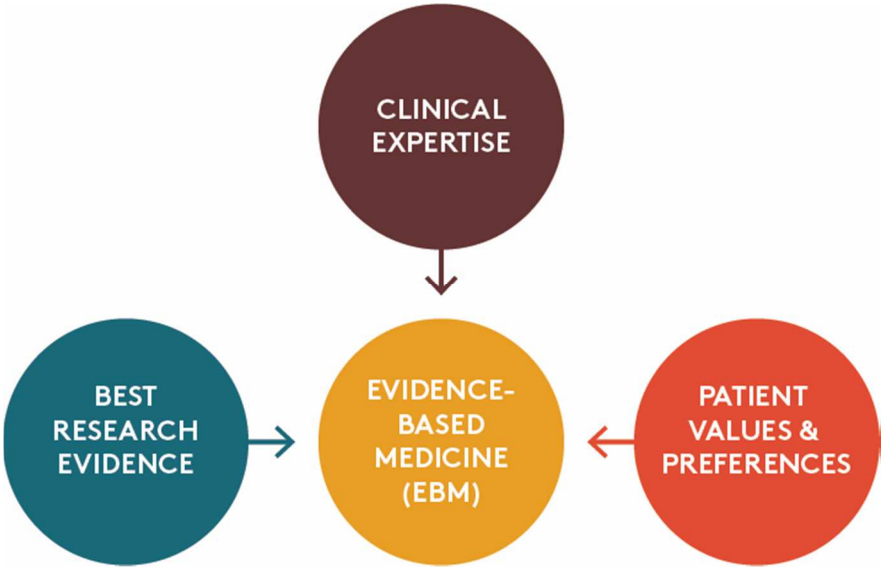


Figure 1.1. Evidence-based medicine conceptual model

Evidence from other study designs should be included to a certain degree. A key challenge for researchers is the extent in which to incorporate evidence from other study designs that are quasi-experimental. These observational studies may introduce bias given that patients are not randomized into particular treatment groups, but are typically much larger in size and have longer follow-up duration than RCTs. Also, they may provide useful information about real clinical populations typically excluded from RCTs and how they may benefit/be harmed by treatments.

A clinician's experience, education, knowledge, and skills are also needed in the decision making process. The Individual Clinical Expertise component refers to "the proficiency and judgment that individual clinicians acquire through clinical experience and clinical practice" (Sackett et al., 1996). Also, the extent to which external evidence is judged to apply to an individual patient may rely on clinical expertise.

EBM models also acknowledge that the patient plays an important role in the

decision-making process. The Patient Values and Preferences component includes the patient's "own personal and unique concerns, expectations, and values" (Sackett et al., 1996). It includes the types of patient information that should be considered in treatment decisions (i.e., either treatment decisions made on behalf of the patient or with active participation of the patient in the process).

The three components do not have to be equally relevant in a given decision and determining the appropriate balance may not be an easy task. That is, any one component or components may be stronger in any particular decision.

Model of Medication Decisions Involved in Antidiabetic Therapy

Many medications exist to manage glucose levels, resulting in an even greater number of sequencing possibilities to treat type 2 diabetes patients as the illness progresses. Figure 1.2 describes the medication decisions patients and providers have to make when determining how to manage diabetes. During each step of treatment, clinicians should monitor short and long-term outcomes beyond how well current treatment manages a patient's glucose levels.

The common treatment pathway is as follows. Patients begin with lifestyle changes such as diet and exercise. They likely also receive diabetes education. Changes are recommended to existing therapy when glucose levels are not controlled. When initial approaches fail to manage glucose levels, monotherapy of an oral antidiabetic agent is added (commonly referred to as first-line treatment). While there are seven classes of medications that are FDA approved, metformin (a biguanide) is recommended as the first-line agent by all major clinical guidelines. When monotherapy fails to control

glucose levels, patients are then typically placed on one or more agents from the remaining classes: sulfonylurea, thiazolidinediones, meglitinides, dipeptidyl peptidase 4 inhibitor, glucagon-like peptide-1 agonist, sodium/glucose cotransporter-2 inhibitor, and/or insulin (note: the order of the agents in the diagram has no meaning of hierarchy or preference). With each step, a decision to augment existing medications (combination therapies) or replace existing therapies is made, keeping in mind the short and long-term outcome risks. These outcomes are indicators for the treatment's effectiveness. Shorter-term outcomes are primarily focused on glycemic control (HbA1c), while long term risks include both micro and macrovascular complications, safety and adverse effects as well as death. Also, the extent to which a treatment is considered safe and effective may vary depending on the subgroup being treated, where any effects may be modified by a particular variable (e.g., in patients over 65, medication adherence, etc.). In the next chapter, attention is focused on the safety risks associated with sulfonylureas in attempt to understand the extent that sulfonylureas should be included in this treatment model.

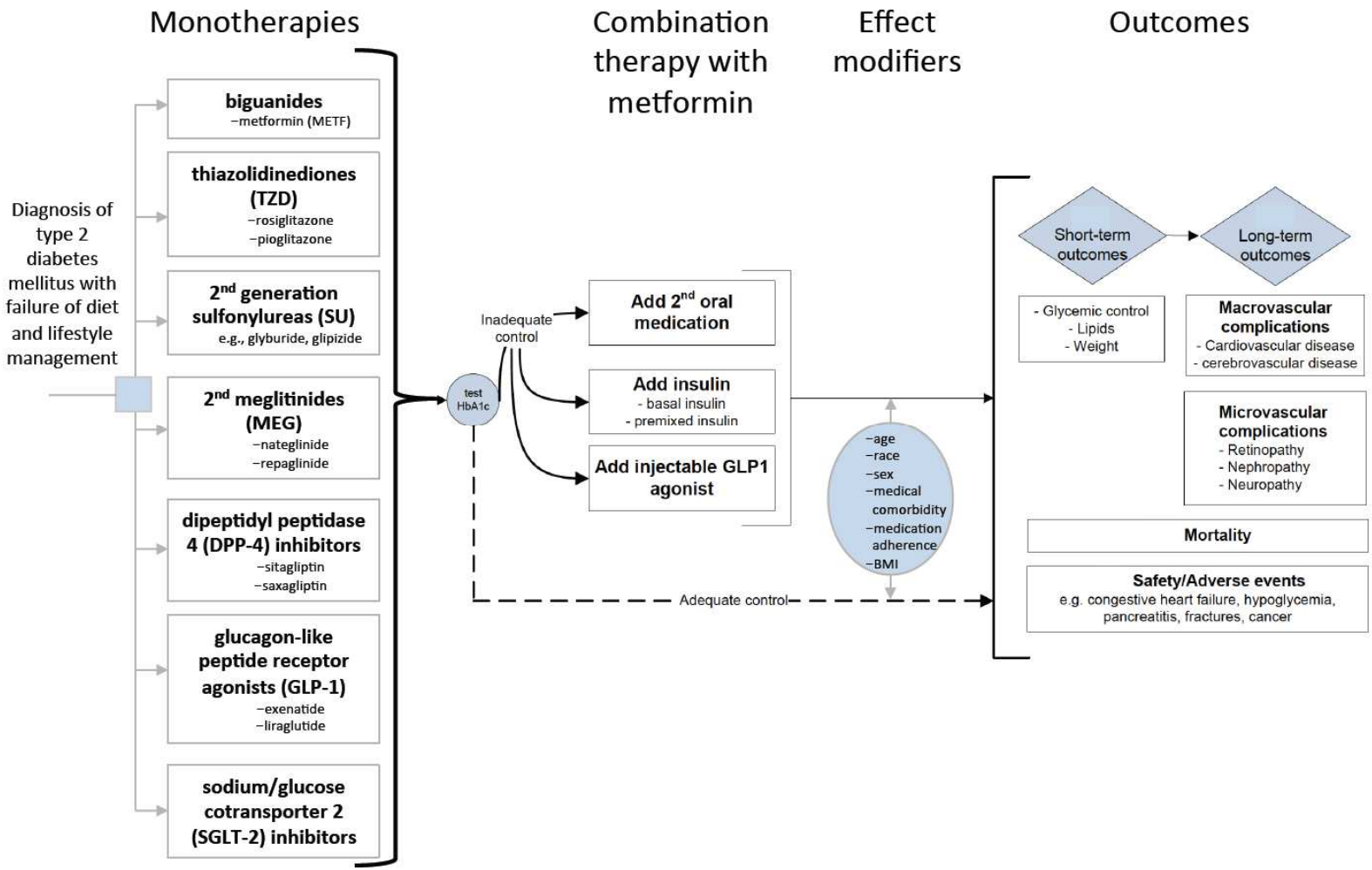


Figure 1.2. Medication treatment model in type 2 diabetes [adapted from (Bennett, Wilson, et al., 2011)]

CHAPTER 2
EVIDENCE-BASED SYNTHESIS OF SULFONYLUREA USE AND THE RISK OF MORTALITY AND CARDIOVASCULAR EVENTS COMPARED TO OTHER ORAL ANTIDABETIC TREATMENTS

Introduction

Type 2 diabetes is a common cause of death, linked to both micro and macrovascular complications, and costly to treat. A wide array of medications exists to manage glucose levels, resulting in an even greater number of sequencing possibilities to treat the illness as it progresses. Clinicians face increasingly complex decisions about which medications regimens to use to treat patients with limited evidence-based information to guide appropriate therapy.

The current clinical consensus is to treat patients with metformin when diet and exercise has failed to control glucose levels, but there is disagreement on whether sulfonylureas should be used as a second-line therapy. While sulfonylureas are a common, inexpensive, and effective way to manage glucose levels they have become increasingly controversial because of long-term safety concerns. Emerging evidence links sulfonylureas use with elevated risks for cardiovascular events and mortality compared to other antidiabetic drug therapies. Still, clinical opinion is fragmented on whether sulfonylureas should remain as a suitable therapy. This difference may be attributed, in part, to the fact that a number of studies reported elevated risks that are observational in nature and challenged on their methodological rigor. This, in combination with the lack of safety and efficacy RCTs that are designed to evaluate long-term outcomes, and to reflect actual clinical populations in such trials have likely contributed to the adoption of different clinical guidelines.

The purpose of this study is to use the existing evidence to summarize the risk of 1) cardiovascular events and 2) mortality (all-cause and cardiovascular) associated with sulfonylureas use relative to other therapies by conducting a series of meta-analyses using existing evidence for these outcomes.

Methods

Search strategy

The MEDLINE database (via PubMed) was searched for studies comparing the safety of sulfonylurea (monotherapy or in combination) relative to other diabetes medications in those with type 2 diabetes patients from 1965 through December 15th 2015.[1] Clinicaltrials.gov, a public database for clinical trials was also searched for unpublished data. In addition, reference lists of relevant articles were examined for studies not retrieved from the other search strategies. Finally, references from previous meta-analyses and Cochrane reviews were examined.

Figure 2.1 describes the selection process resulting from the MEDLINE and other search strategies. A total of 1,982 articles were extracted from MEDLINE. Two hundred sixty-four additional articles were culled from the other search strategies, resulting in a total of 2,246 articles extracted. Abstracts of each of these articles were reviewed for eligibility. Of these, 172 articles were reviewed in their entirety and a total of 52 met the eligibility requirements to be included in the meta-analyses.

Information regarding the effect size (e.g., hazard ratio, odds ratio, relative risk)

[1] Search terms included in Appendix A

or raw information to calculate it (e.g., number of major cardiovascular events, number of people who died), standard deviation (or 95% confidence interval), sample size (number of people in treatment group), and the study characteristics were extracted from each study if this information was provided. Adjusted estimates of the effect size were used if provided, otherwise unadjusted estimates were extracted. Authors were not contacted to obtain information if missing. For the purposes of this study, hazard ratios, odds ratios, and relative risk were treated as equivalent when pooling estimates.

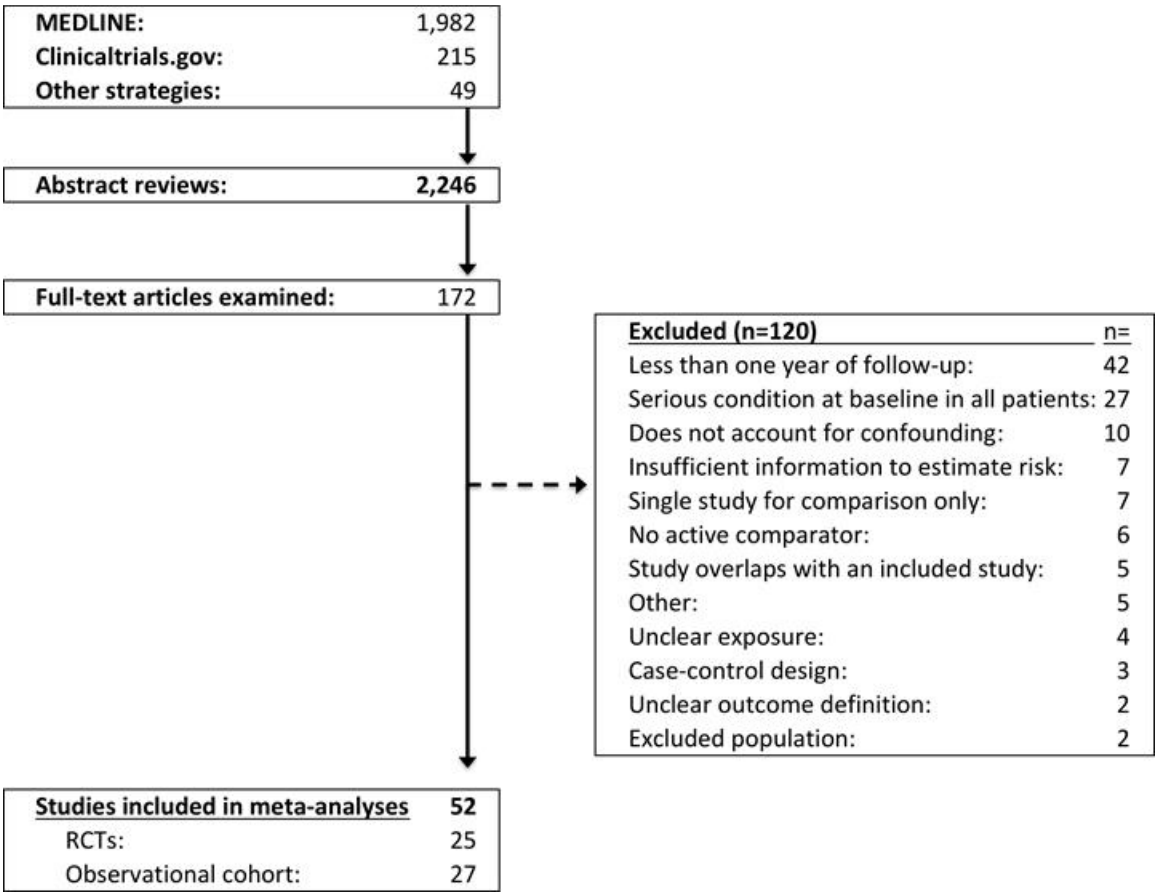


Figure 2.1. Study selection process

Study Selection Criteria

Randomized control trials and observational cohort studies were included in the study. All studies explicitly examining all-cause mortality, cardiovascular-related mortality, or major cardiovascular events were examined. Since the aim was to evaluate long-term cardiovascular and mortality risks, only studies with a year or more of follow-up from the date of the first prescription were included.

Studies were excluded if they met any of the following criteria: included only patients with serious conditions at baseline such as a history of major cardiovascular events or renal failure, consisted of only children (younger than 18 years of age), type 1 diabetes patients only, did not include an active comparator (e.g., diet/exercise, placebo), case-control design, research only on animals, and written in a language other than English. For studies with more than one publication, the study with the most complete data or involved the most recent follow-up was selected. For observational studies, an attempt to address confounding must have been implemented (matched in the design or model adjustment) by including basic demographic information (i.e., age, sex, and race) and relevant comorbidities at baseline (those adjusting for CVD risk at a minimum). This resulted in 25 RCTs and 27 observational cohort studies being included in this study.

Risk of Bias (Quality) Assessment

Details about potential biases in each randomized trial included in the meta-analysis was assessed using items from the Jadad scale, which asks about the study design and its appropriateness (randomization, double blind) as well as whether a description of the dropouts from the study is included (Jadad et al., 1996). The quality of

observational cohort studies were rated using the eight items from the Newcastle-Ottawa Scale (Wells et al., 2000). This scale assesses quality of three dimensions: sample selection, comparability of groups, and outcome assessment. An additional item for both study designs examined whether industry funding explicitly sponsored the study.

Details of the quality assessments are presented in the Appendix A-3 and A-4. Results from the Newcastle Ottawa Scale suggest that all studies met most of the quality assessments for each domain. Regarding RCTs, all studies were randomized, 21 of 25 were double blind and a description of the participant dropouts were described in 24 of the studies. However, industry funding was judged to be high in 65% of all studies (24 of 25 RCTs; 10 of 27 observational studies). With the exception of industry funded studies, most studies were assessed as low risk of bias on the domains assessed suggesting that the overall quality was fair to good in the selected studies (see Appendix A-3 and A-4 for study specific breakdown). Total scores from the quality assessments were not used as a way to exclude studies from the meta-analyses.

Analysis

Each outcome and comparison required two or more studies (Valentine, Pigott, & Rothstein, 2009). For RCT and observational designs, both fixed effects and random effect models were conducted and reported. In a fixed effects model, the assumption is that each study provides evidence towards one common effect size. That is, the model assumes the effect size should be the same and that the features of the study (e.g., study design, population,) should not impact the magnitude of the effect size. Therefore, the fixed effect model combined all study information together without taking into account

that studies can vary between each other as well as vary between different study designs. Weights given to each study are determined only by its within-study variance (study weight=1/within study variance). Since variance is a function of sample size, smaller studies will contribute less information to the weighted estimate than larger studies.

In the random effects model, the weights given to each study are determined not only by the within group variability (like for fixed effects) but also by the between-group variability. The implication is that relatively greater weight tends to be given to smaller studies than it would be in a fixed effect model approach since the weights for each study now account for between study design variability. In general, since random effects models also include between study variation, they will tend to have relatively wider confidence intervals compared to fixed effects models (Borenstein, Hedges, Higgins, & Rothstein, 2009). The inverse variance and the DerSimonian-Laird methods were used to estimate fixed and random effect estimates respectively using the METAN command in Stata 14.1 (StataCorp, 2015).

A particular challenge for researchers is how to synthesize results that are produced from two inherently different study designs: RCTs and quasi-experimental observational designs. Therefore, to address this methodological challenge, the approach uses a two-level hierarchical Bayesian design to synthesize result estimates across RCT and observational designs. This is a *random effects* model approach, and assumes that the effects derived from different study designs will be similar and also different to some extent. The combined effect is the weighted average of these two common effect sizes.

Overall pooled estimates were estimated using the 'bayesmh' command with

random effect of study design in Stata 14.1 (StataCorp, 2015). Thus the model accounts for heterogeneity from the different study design. This is similar to the approach by Peters et al. (2005) and involved Markov chain Monte Carlo (MCMC) estimation using a Metropolis-Hastings algorithm and Gibbs sampling with vague conjugate prior distributions specified on unknown parameters. Convergence diagnostics suggest fairly rapid convergence with no trend in trace plots, low autocorrelation, and acceptance rates for the Metropolis-Hastings algorithm around 75% (well above the 10% rule of thumb) and efficiencies above one percent for all analyses.

Heterogeneity across the studies was assessed via the I^2 statistic, with values greater than 50% benchmarked as indicating substantial heterogeneity (Higgins, Thompson, Deeks, & Altman, 2003). This statistic represents the percent of variance in the effect size attributable to heterogeneity with larger values indicating less overlap in confidence intervals across studies. A benefit of the statistic is that the number of studies involved in the meta-analysis has little influence on the I^2 statistic unlike other estimates.

In drug comparisons that included 10 or more studies, publication bias was assessed by testing for asymmetry in funnel plots (scatterplot for the log effect size by the log standard error) using Egger's tests (Egger, Davey Smith, Schneider, & Minder, 1997) via the METABIAS Stata command (Harbord, Harris, & Sterne, 2009). Tests for funnel plot asymmetry are not recommended in comparisons with less than 10 studies since power may be too low to detect moderate asymmetry (Higgins & Green, 2011).

Results

A total of 25 randomized clinical trials and 27 observational cohort studies were included in the series of meta-analyses. Meta-analytic summaries of the effect size (and 95% confidence or credible intervals) for each comparison and outcome are presented in Figures 2.2, 2.3, and 2.4. Further information, including both fixed and random effect models for each analysis is presented in Appendix A-2.

Pooled Effects By Design

Observational cohort design. Sixteen meta-analyses (from eight drug-to-drug comparisons) of only observational cohort studies suggest greater sulfonylurea risk compared to other therapies. Three of these comparisons involved sulfonylurea monotherapy against metformin (all-cause mortality: RR=1.38, 95% CI [1.35, 1.41], cardiovascular mortality: 1.21 95% CI [1.16, 1.27], cardiovascular composite: RR=1.18, 95% CI [1.15, 1.22]), TZD (all-cause mortality: RR=1.28, 95% CI [1.13, 1.45]), and combination METF+TZD (all-cause mortality: RR=1.76, 95% CI [1.41, 2.20], cardiovascular composite: RR=1.99, 95% CI [1.47, 2.70]).

There were also differential risks when sulfonylurea combination therapy was evaluated against sulfonylurea and metformin monotherapy. A lower risk was associated with METF+SU when compared to sulfonylurea monotherapy (all-cause mortality: RR=0.76, 95% CI [0.71, 0.80], cardiovascular mortality: RR=0.80, 95% CI [0.66, 0.97], cardiovascular composite: RR=0.84, 95% CI [0.77, 0.93]) and a higher risk was associated with sulfonylurea combination therapy compared against metformin monotherapy (all-cause mortality: RR=1.15, 95% CI [1.08, 1.22], cardiovascular

mortality: RR=1.47, 95% CI [1.18, 1.82]).

The remaining analyses found elevated effects for sulfonylurea combination therapy with metformin relative to other metformin combinations: METF+TZD (all-cause mortality: RR=1.20, 95% CI [1.08, 1.34], cardiovascular composite: RR=1.12, 95% CI [1.03, 1.23]), METF+DPP-4 (all-cause mortality: RR=1.45, 95% CI [1.32, 1.59], cardiovascular composite: RR=1.46, 95% CI [1.28, 1.68]), and METF+GLP-1 (all-cause mortality: RR=1.42, 95% CI [1.00, 2.01].

In addition, pooled results were statistically inconsistent in four analyses between the fixed inverse variance method and the DerSimonian and Laird random effect method, such that the added between study variance included in the random effects estimates produced wider confidence intervals for the pooled effect in all cases giving statistically non-significant estimates. Thus substantial heterogeneity existed within each of these analyses, with the I^2 statistic ranging from 74% to 93%. All of these analyses involved METF+SU combination therapy compared to monotherapies, and found a lower risk when compared to sulfonylurea alone (all-cause, cardiovascular composite) and found a higher risk when compared to METF monotherapy (on all cause mortality. cardiovascular death). With the exception of this last drug comparison, all of the inconsistent comparisons had similar magnitude and directions of the estimated pooled effects between random effects and fixed effects estimates (see Appendix A-2 for more details).

RCT. One significant elevated effect was found in the series of analyses using only RCTs. People randomized to receive combination METF+SU had an 86% increased risk of a cardiovascular composite event at any point in time compared to those assigned

METF+DPP-4 (Pooled RR=1.86, 95% CI [1.18, 2.93]). All other pooled estimates of RCT design studies failed to find a difference in risk between sulfonylurea therapy and other regimens for all outcomes. While most comparisons had the same direction in the effect as pooled observational cohort estimates, precision was often worse than its pooled observational cohort counterpart.

Overall combined across study design. None of the analyses suggested an elevated effect for sulfonylureas when results were combined across RCT and observational cohort study designs according to all two-level hierarchical Bayesian models. While the overall direction and magnitude of the effect estimates are similar to that of the pooled estimates from observational cohort designed studies, overall pooled estimates have considerably wider credible intervals. This is most likely a result of the added variation existing between study designs.

Publication Bias

Assessing publication bias was limited since most analyses were excluded if there were fewer than 10 studies included. There was no significant test suggesting publication bias according to Egger's test.

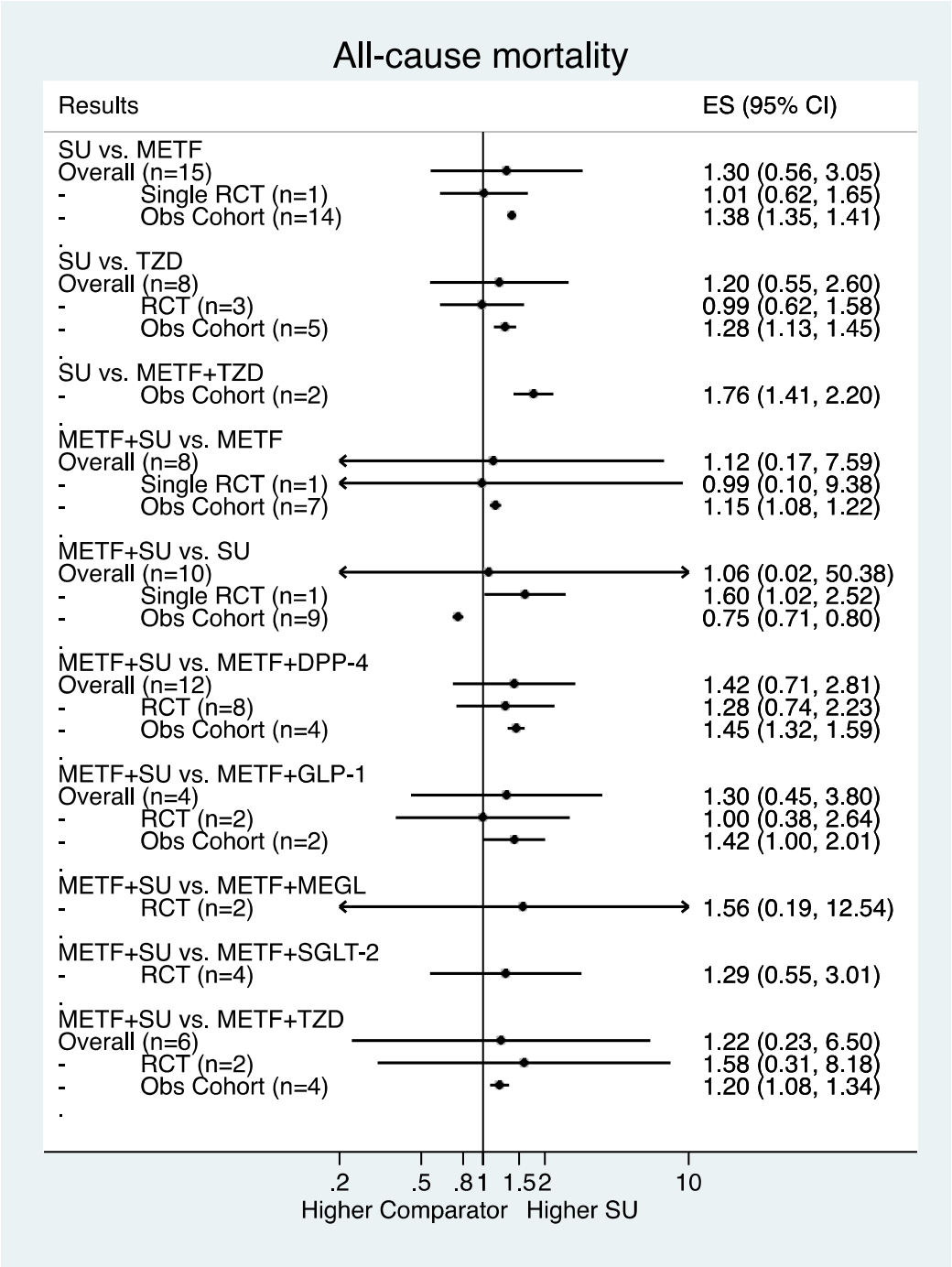


Figure 2.2. Pooled relative risks for all-cause mortality. Inverse variance fixed effect estimates are shown for pooled estimates by study design. Two-level hierarchical Bayesian estimates shown for overall pooled estimates. Relative risk and 95% confidence interval presented for results by study design. Relative risk and 95% credible intervals for overall pooled estimates. Note: ES=Effect Size

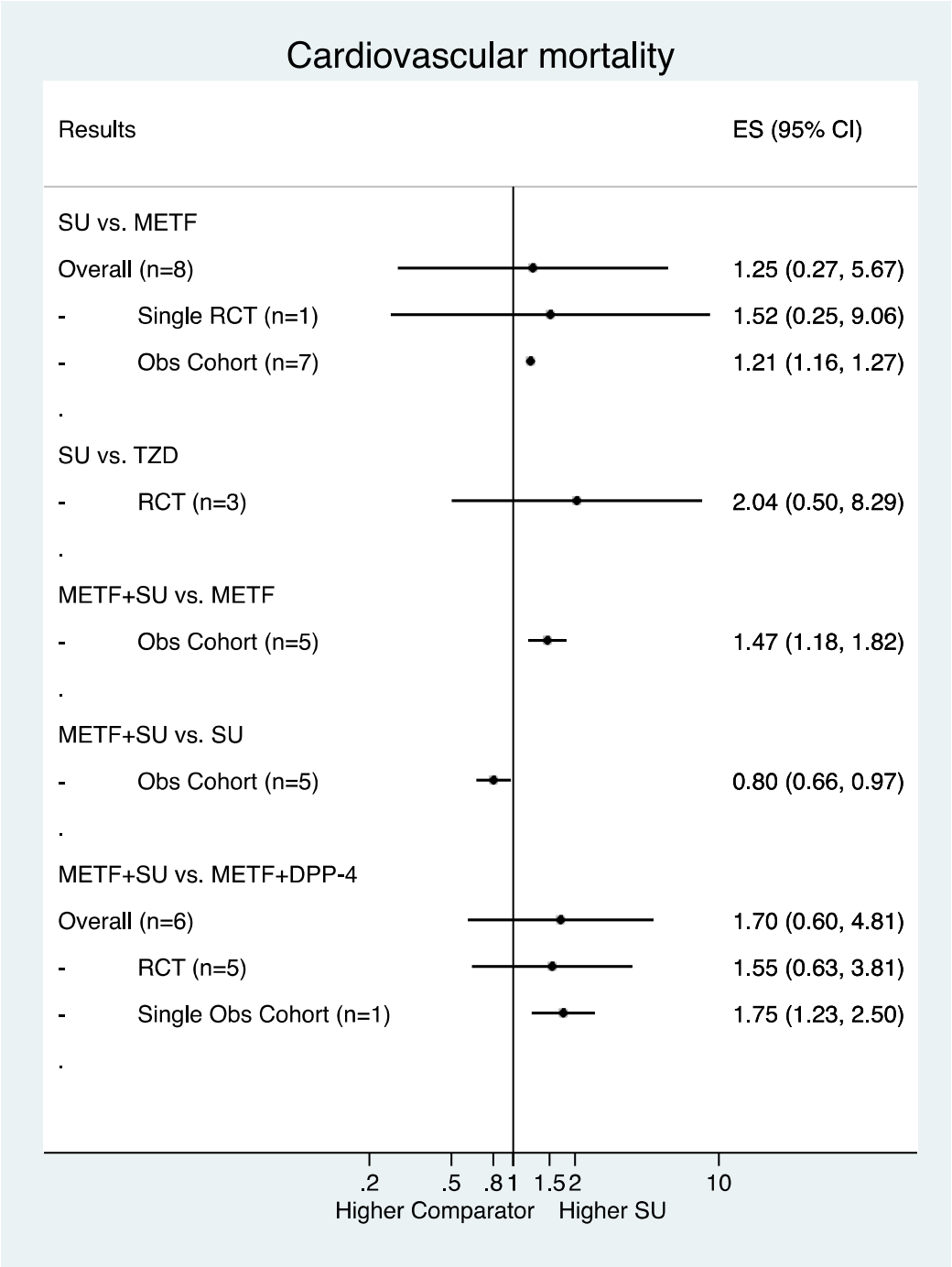


Figure 2.3. Pooled relative risks for cardiovascular mortality. Inverse variance fixed effect estimates are shown for pooled estimates by study design. Two-level hierarchical Bayesian estimates shown for overall pooled estimates. Relative risk and 95% confidence interval presented for results by study design. Relative risk and 95% credible intervals for overall pooled estimates. Note: ES=Effect Size

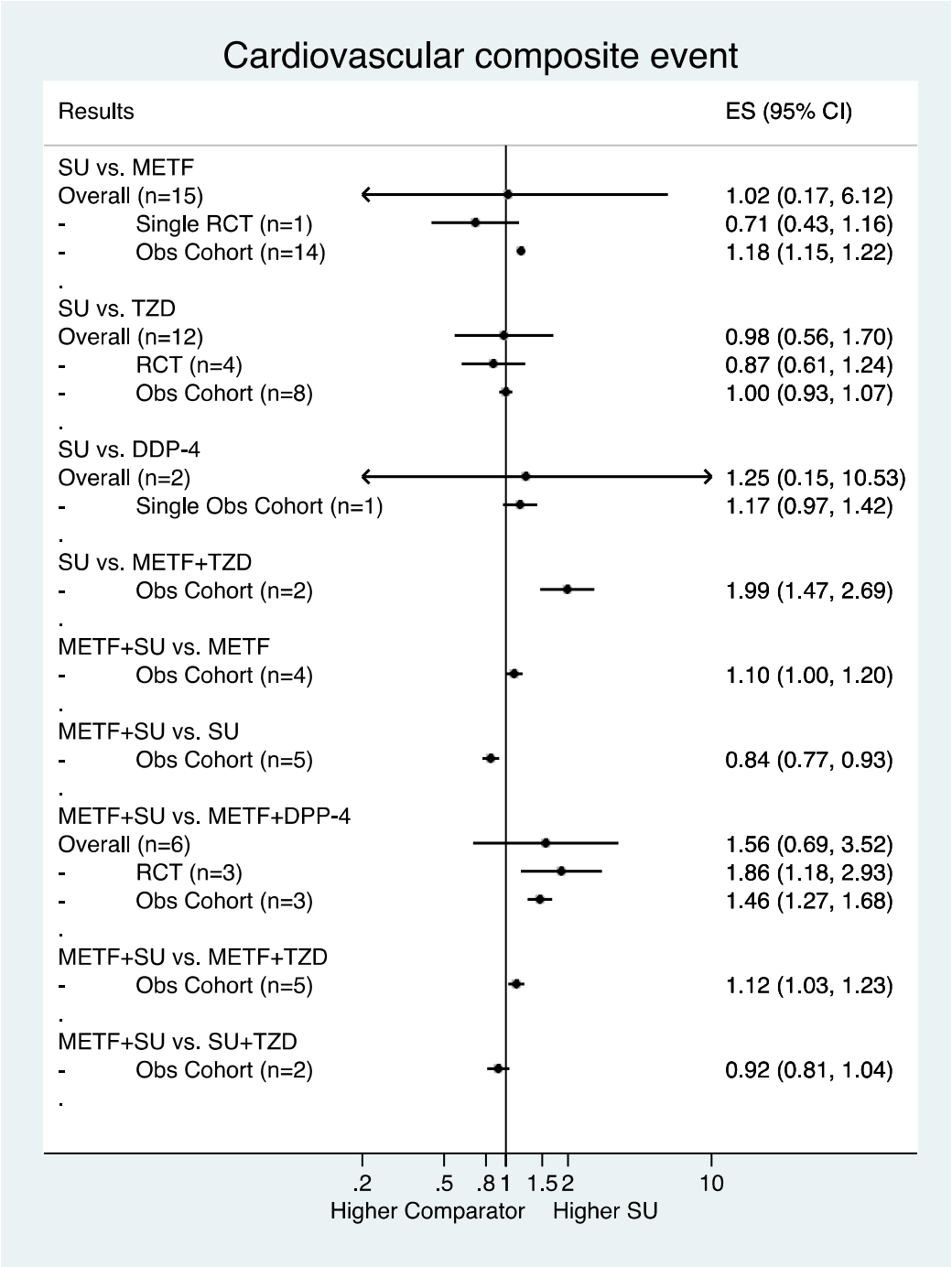


Figure 2.4. Pooled relative risks for cardiovascular composite events. Inverse variance fixed effect estimates are shown for pooled estimates by study design. Two-level hierarchical Bayesian estimates shown for overall pooled estimates. Relative risk and 95% confidence interval presented for results by study design. Relative risk and 95% credible intervals for overall pooled estimates. Note: ES=Effect Size

Discussion

Cardiovascular disease is the main cause of death for people with diabetes mellitus, yet evidence on whether particular drug therapies contribute to an increase in cardiovascular events and mortality has been unclear and insufficient. Despite its importance, most clinical studies are instead designed to assess glycemic control — often short-term assessments that lack the power to examine rare events. This in turn, results in imprecise estimates of the risk for cardiovascular morbidity and mortality.

Based on this series of meta-analyses, there is some evidence that sulfonylurea therapy is associated with elevated risk relative to other drug classes (against metformin, TZD, and DPP-4s) when compared alone (as a monotherapy) and when used in combination with metformin.

These significant findings are almost entirely derived from observational data (with one exception). In monotherapy, higher pooled relative risk is reported for sulfonylurea monotherapy when compared to metformin on all three safety outcomes and TZD on all-cause mortality.

Evidence from observational cohort studies also found sulfonylureas to have higher long-term risks when compared to three other potential second-line drugs for one or more outcomes. Against METF+DDP-4, results suggest that combination METF+SU has an increased risk of cardiovascular composite events (in agreement with RCT pooled results) as well as all-cause mortality. Against METF+GLP-1 there was an elevated risk for all-cause mortality. In both sulfonylurea alone and METF+SU combination therapy,

the results suggest higher risk than METF+TZD for all-cause mortality and cardiovascular composite events.

While most RCT derived estimates were in the same direction and had similar magnitude to their observational cohort counterpart, the uncertainty surrounding the effect were much larger. Therefore when evidence is pooled using both types of studies design, there is high variability around effect estimates (wide credible intervals) as a result from the imprecise estimates reported from prior RCT studies.

Insufficient evidence has been produced from RCTs. This is reflected in the imprecise estimates reported in this study (relative to pooled observational estimates). In addition, 34 RCT designed studies were excluded solely on their duration (less than one year). Despite these limitations, one analysis did suggest a significant elevated effect for combination METF+SU against METF+DPP-4 for cardiovascular events. Across all RCTs in this study, the majority of RCTs evaluating long-term safety outcomes had small sample sizes with relatively few or no events in a given drug group occurring during the study period. As a result, existing RCTs were not powered sufficiently to evaluate long-term safety outcomes.

While several pooled analyses of observational cohorts provide evidence and suggest different risks, the extent to which observational data should be used as evidence is unresolved. Despite some evidence from observational studies, when combined with RCTs, pooled estimates did not suggest an increased risk for long-term adverse events. A methodological challenge for researchers is how/whether evidence from observational cohort and RCTs can be combined. In this study, a two-level Bayesian model was used

to explore how results can be synthesized across study designs. Since there were fewer RCTs relative to observational studies this tended to give more weight to RCTs than it otherwise would if simply combined without consideration of study design. However, it also included additional variance in the form of between study design variance as well as variability from highly uncertain (i.e., uniformed) prior distributions in the Bayesian models.

Future studies should explore whether there are other suitable methods to account for uncertainty and pooling estimates across study designs for the purpose of advancing empirical knowledge and informing evidence-based medicine practice. In particular, Bayesian multilevel models that use informed prior distributions that are formally specified to reflect the relative strength of RCT designs compared to observational designs would be most beneficial. This would assign less weight to study design types that are more susceptible to bias (e.g., observational designs) relative to RCT designs. Empirically, these weights might be developed via meta-regression examining how effect estimates vary by study design as has been suggested previously (Goodman, 2013). Additionally, expert judgments may be elicited via survey or using a Delphi or group consensus approach, where this information may be quantified in the form of a prior probability distribution.

Finally, it is important to note that there are several shortcomings in existing comparative safety analyses that need to be explored in future research. While sulfonylurea therapy is commonly compared to metformin and TZD, there is limited comparative safety research on how newer classes of medications compare against

sulfonylurea therapy. Also, there are even fewer comparative safety analyses that parse out the different sequencing possibilities involving sulfonylurea combination therapy such as whether existing therapy (e.g., often METF monotherapy) is discontinued or augmented when a second-line therapy is introduced.

Also, there were few comparisons that included 10 or more studies to examine publication bias so this cannot be ruled out. In addition, other biases could influence study effect sizes. In future work, meta-regression is one way to explore the influence that various study characteristics as well as other effect modifying factors had on these estimates.

Conclusion

While previous studies suggest other antidiabetic medications appear to have equal glucose-lowering efficacy alone and when combined with metformin (Morgan et al., 2012), further research is needed to determine whether they also provide greater long-term safety. While meta-analyses using only observational cohort evidence suggest elevated sulfonylurea risk, RCTs to date have been poorly designed to evaluate long-term outcomes, resulting in few events and providing little evidence. The focus of many trials has been to make direct head-to-head comparisons to assess which medications work best at managing glucose-levels and were not designed to examine long-term risks (Holman et al., 2008; Kahn et al., 2006; Lipska KJ & Krumholz HM, 2014; Nathan et al., 2013; Ryan et al., 2003; UKPDS-33, 1998). These trials have typically been small in size with relatively short follow-up periods, limiting the ability to obtain precise estimates of risk.

While much of the evidence is derived and will continue to come from observational studies, the methodological rigor of such studies is questionable (e.g., internal threats to validity such as selection bias and unmeasured confounding are possible). Well-designed, rigorous observational studies may provide additional evidence, but more RCTs examining long-term risk are needed.

CHAPTER 3
A COMPARATIVE SAFETY ANALYSIS OF SECOND-LINE DIABETES TREATMENTS IN A VETERAN POPULATION

Background

Evaluating the scientific rigor and relevance of research evidence is a key element for informing clinical decision-making in evidence-based medicine (EBM) approaches. In the treatment of type 2 diabetes, care involves a sequencing of therapies over time as the illness progresses. EBM is commonly defined as "the integration of best research evidence with clinical expertise and patient values" (Rosenberg & Haynes with Sackett, Straus Sharon E. and Richardson W. Scott, 2000) and involves "the conscientious, explicit, and judicious use of clinically relevant research in making decisions about the care of individual patients" (Sackett et al., 1996). With physicians making difficult decisions about which medications regimens to use to treat patients with limited or uncertain evidence-based information to guide appropriate therapy more comparative safety evidence is needed.

While the current clinical consensus is that metformin should be the initial first-line agent when diet and exercise are not sufficient, disagreement exists amongst clinicians about whether sulfonylureas are a suitable second-line of treatment after metformin fails to control glucose levels in type 2 diabetes patients. While they have been used for glycemic control in patients with type 2 diabetes since the 1950s, they are increasingly controversial and have attracted considerable debate in recent years (Abrahamson, 2015; Genuth, 2015). Knowledge of the benefits and risks relative to other therapies in the short-term are fairly well known, while the longer-term risks are poorly

understood. In the short-term, several pooled analyses of RCT and observational cohort studies suggest that other classes of diabetes medications yield similar reductions in hemoglobin A1c levels relative to sulfonylureas and are associated with lower rates of hypoglycemic events compared to sulfonylureas (Bennett, Maruthur, et al., 2011; Bolen et al., 2007; DeFronzo & Goodman, 1995; A. Garber et al., 2009; Hemmingsen et al., 2014; Morgan et al., 2012; Scott, Wu, Sanchez, & Stein, 2007; Seaquist et al., 2013; UKPDS-33, 1998).

The results from the series of meta-analyses conducted presented in Chapter Two of this dissertation, suggest that there is some evidence for increased risks for sulfonylureas compared to other oral antidiabetic treatments from observational studies. RCTs have been, for the most part, poorly designed to evaluate long-term outcomes, often resulting in fewer events over the longer term and also provide little evidence for long term outcomes (Bennett, Wilson, et al., 2011; Bolen et al., 2007). This has led to imprecise estimates of pooled risk across RCT and observational studies.

Results from Study One also highlight that there are critical gaps in the literature that deserve attention. Specifically, few studies to date have investigated second-line therapies after metformin is no longer sufficient. Since all major guidelines support metformin as the initial pharmacologic therapy to maintain glycemic control, studies should be designed to evaluate the safety when a second-line therapy is needed. Related, information regarding the impact that discontinuing or using metformin in combination with second-line medications is also lacking.

With limited evidence-based information on long-term outcomes to guide

appropriate therapy, second-line treatment considerations rely upon scientific evidence from studies of patients with no prior oral medication use or that ignore the sequencing of medication. These cohorts may be inherently different from actual treatment populations who are considering second-line treatment. Second-line considerations also rely upon short-term safety (severe hypoglycemic events) and efficacy (influence on glucose levels). Well-designed observational studies with large sample sizes and longer follow-up can provide additional evidence on the comparative safety of sulfonylureas relative to other second-line drugs on long-term outcomes. The purpose of this study is to provide new evidence on the cardiovascular and mortality risks for diabetes agents relative to sulfonylureas by conducting a new comparative safety study in a veteran diabetic population on second-line glycemic control treatment both as add-on therapies and second-line monotherapies (i.e., where metformin is discontinued) with longer term follow-up than previously reported.

Methods

Data Source

A national analysis of Veterans Health Administration (VA) medical and administrative observational data was conducted. Patients in the Diabetes Epidemiology Cohorts (DEpiC), a diabetes registry developed and maintained for over 10 years (Miller & Pogach, 2008), were used to reliably identify patients with type 2 diabetes and to link VA data over time. Information gathered included face-to-face outpatient visits and inpatient stays, which contain codes for diagnoses and medical procedures performed, as

well as VA prescriptions (e.g., drug agent, number of days supplied, and dispense date) and laboratory test results. Since many VA patients also receive care outside the VA, we used non-VA data sources including Medicare data. For mortality data, the VA Vital Status File and National Death Index were used (Centers for Disease Control and Prevention, 2017).

Design and Analysis.

A comparative safety analysis was conducted on the risk of adverse outcomes in a retrospective cohort study of VA veteran patients with diabetes who received a prescription for metformin and subsequently augmented or switched to other diabetic oral treatments from fiscal year 1998 to 2012. Specifically, patients were included whose first diabetes medication was metformin as a first-line therapy, remained adherent to metformin until they subsequently augmented or switched to sulfonylurea or another second-line oral glycemic control medication. Similar to a run-in period in clinical trials, patients must have remained on that second-line regime, defined as at least two prescriptions and more than 90 days supplied within 180 days of initial second-line prescription to ensure that the cohorts consist of patients who tolerated and remained on the second-line therapy. Other inclusion criteria are: type 2 diabetic (according to DEpiC registry) in at least one fiscal year; received regular care in VA (at least two face-to-face outpatient visits in the 731 days before the initial second-line prescription); known gender and age; first VA service use (outpatient visit or inpatient admission) at least 365 days before initial second-line prescription date; no insulin use prior to the start of follow-up;

and did not die before the start of follow-up. A flowchart for the sample derivation is presented in Figure 3.1.

Study Design, Cohort Entry and Exposure

To compare medications for patients at similar stages in diabetes progression, a variant of a new user design in which incident use of a second-line medication was implemented after prior metformin prescribing. Cohort entry was defined on the basis of the initial second-line prescription, with exposure beginning on the day of the initial second-line prescription and ending 180 days after. As previously stated, the second line regimen had to be tolerated and prescribed relatively long-term, which is defined as consisting of two or more prescriptions totaling more than 90 days supplied in the 180-day exposure window. Follow-up began after the criteria for cohort entry and exposure status had been established and therefore started 181 days after the initial second-line prescription.

All second-line oral agents were considered and classified according to their drug class. However, prescribing for newer medications like DPP-4, GLP-1 analogues, SGL-2, and alpha glucosidase inhibitors was low and cohorts for these groups lacked sufficient power for analysis. As a result, second-line sulfonylurea and thiazolidinedione (TZD) therapies are only reported in this study.

In a second series of analyses, cohorts were subdivided into whether second-line regimens were added to existing metformin use or if patients were switched off metformin when the second-line monotherapy was initiated (i.e., metformin prescribing discontinued).

In addition, all models were also stratified by variables determined *a priori* to potentially modify any estimated effect. These variables included dual Medicare/VA beneficiaries, prior cardiovascular disease, and being 65 years of age or older.

Two analysis approaches were employed: intent-to-treat and as-treated. The two analytic approaches treat changes in the exposure to drug treatment differently. Intent-to-treat ignores subsequent changes to drug exposure and analyzes the initial exposure cohort, censoring only at the end of study period. In the as-treated analysis, there will be fewer events since follow-up is censored when the second-line therapy is discontinued (defined as a prescription gap of cohort medicine lasting more than 90 days or more), when another class of oral glucose lowering medication for long-term use was prescribed, when insulin was prescribed, or at the end of the study period. Ninety days was chosen to define discontinuation as it has been shown to represent true non-persistence in a study of sulfonylurea and metformin therapies in an earlier study (Greevy et al., 2011).

Outcomes

Time to all-cause death and time to a cardiovascular composite event were the outcomes of interest. Death was ascertained from National Death Index and VA vital Status data. Events included in the cardiovascular composite event measure are listed in Table 3.1 and include new hospitalized acute myocardial infarctions, stroke hospitalizations, congestive heart failure, cardiac surgery, and all-cause mortality. Non-death related cardiovascular events included in the composite were identified by *International Classification of Disease, Ninth Revision [ICD-9] and Current Procedural Terminology* codes from the medical encounter data.

Table 3.1. Definition of cardiovascular composite event end-point

Cardiovascular condition	Specification
Myocardial infarction	**ICD-9-CM diagnosis code [inpatient only]**: 410, 410.xx, except 410.x2
Acute coronary syndrome	**ICD-9-CM diagnosis code:** 411.1, 411.81, 411.89
Stroke	**ICD9-CM code: single code or specified combination:** (inpatient codes for 430-432.xx, 434, 434.xx, 436, 436.xx) or ((inpatient codes: 342, 342.xx, 433, 433.xx, 435, 435.xx, 438, 438.xx) + (inpatient or outpatient) V57) or or ((inpatient codes: 433, 433.xx, 435, 435.xx) + ((inpatient or outpatient) 342, 342.xx))
Cardiac surgery	**ICD-9 Procedure code:** 36.01, 36.02, 36.05, 36.10-36.16, 36.19 **or CPT-4 procedure code:** 33510-33519, 33521-33523, 33533-33536, 33572, 92973-92975, 92977, 92980-92982, 92984, 92986, 92995, 92996
Congestive heart failure	ICD-9-CM code: 428,428.xx, 402.01, 402.11, 402.91, 404.01, 404.11, 404.91
All-cause death	National Death Index, VA Vital Status file

Covariates

The following types of information were used to minimize bias and were assessed before cohort entry: demographics/sociodemographics (e.g., age, sex, race, income), comorbidities in the prior two years (e.g., cardiovascular, cerebrovascular, peripheral vascular disease, retinopathy, neuropathy), health care utilization in the prior six months (hospitalizations, outpatient visits, Medicare use), physiologic measures in the past twelve months (A1c, creatinine, blood pressure), prescriptions (e.g., antihypertensive agents, lipid lowering medications), and access to care (VHA health insurance benefits).

See Appendix B-1 for selected covariates and their baseline frequencies and or means by drug exposure. In instances where patients had covariates with missing data, multiple imputations using Markov-chain Monte-Carlo methods were used to impute missing values (consisting of 20 imputation models using Jeffrey's prior) via PROC MI in SAS 9.3. Model estimates were subsequently pooled using PROC MIANALYZE and reported.

Statistical Analyses

To mitigate against the influence of selection bias, propensity score methods were used (Rosenbaum & Rubin, 1983). When applied in observational studies, propensity scoring attempts to approximate a randomized study as much as possible by creating balance in the distribution of observed baseline covariates across the sulfonylurea treatment group and its comparator (Austin, 2011; Rubin, 2007). Thus if propensity scores are properly specified, then they should make the treatment and comparison group similar on key characteristics. If the sample is balanced, then estimates of outcome are unbiased and model dependence is decreased (i.e., the way in which the statistical model is specified leads to very little change is the estimated effect).

Descriptive statistics at baseline were calculated for all covariates before and after propensity score weighting. The effect of sulfonylureas relative to TZDs is reported in two ways. The absolute effect of sulfonylurea treatment relative to its comparator was estimated by graphing the unadjusted Kaplan-Meier curves. Additionally, Cox proportional hazards models were used to estimate the relative effect via hazard ratios for each of the outcomes. TZD second-line therapy was used as the reference group for the

primary analyses and metformin augmented by sulfonylurea was used as the reference group for the secondary analysis. A series of unadjusted and weighted adjusted hazard ratios are reported. To check the assumption of proportional hazards, plots of the log (-log) Kaplan-Meier survival by the log of time were examined. Patients were weighted by the inverse probability of treatment using the propensity score (Austin, 2013; Austin & Stuart, 2015). Using inverse probability of treatment weights (IPTW) gives greater weight to a) patients in non-sulfonylurea groups with high propensity scores and b) sulfonylurea patients with low propensity scores.

Weights were determined by the following steps. First, the propensity score was estimated by modeling the predicted probability of receiving a sulfonylurea as a second-line medication compared to other antidiabetic drug agents conditioned on observed baseline characteristics. For each patient, the estimated predicted probabilities were obtained from a logistic regression predicting second-line sulfonylurea prescribing from all observed covariates plausibly related to both treatment and the outcomes *a priori*. Covariates included were based on subject matter expertise and a review of existing literature as has been proposed previously (Austin, 2014). IPTW was calculated as the inverse of the propensity score, defined as $(1/\rho)$ for the sulfonylurea group and $1/(1-\rho)$ for the comparison group, with ρ denoting the propensity score (Austin, 2013; Harder, Stuart, & Anthony, 2010). To reduce the influence of very large weights that may lead to imprecise (or biased) estimates, IPTWs were stabilized by multiplying each IPTW by the mean propensity score (inverse mean for comparator). In addition, very large weights were truncated to the 95th percentile of the distribution as previously recommended (Cole

& Hernán, 2008; Xiao, Moodie, & Abrahamowicz, 2013). IPTWs were then normalized to preserve the original sample size (Garrido et al., 2014).

The performance of the propensity score was evaluated in several ways. First, to assess how well the propensity score was specified, the overlap in the distribution of propensity scores between the treatment and comparison was examined, typically referred to as common support. Propensity score distributions between groups should be similar. If there are areas with no overlap it suggests that the propensity score may not be properly specified and/or that it may not be possible to reduce confounding through propensity score techniques (Harder et al., 2010).

If the propensity score is properly specified, it should serve as a balancing score: the distribution of each covariate should be the same across cohorts (Brookhart, Wyss, Layton, & Stürmer, 2013). Imbalance is the source of model dependence. When there is high model dependence, the way a model is specified will lead to different effect estimates and potentially different interpretations of the results. Small changes in the model specification may lead to changes in the effect estimates. This leaves results up to the researcher's discretion, which then ultimately leads to bias. Small changes in specifications produce big changes in the subset of results. Therefore assessing how well weighting diminishes imbalance (the relationship between the control variable and a key causal variable) is important. If balanced, it suggests that it is one adequate way of specifying the propensity model that limits confounding.

The extent weighting balanced covariates was assessed by comparing the standardized mean difference (SMD) before and after weighting for each baseline

covariate. SMD is similar to the effect size, and calculated by taking the difference in means of each covariate across the two therapy groups and dividing by the standard deviation in the sulfonylurea group (Stuart, Lee, & Leacy, 2013). As a rule of thumb, it is suggested that a standardized mean difference lower than 25% should be considered a relatively balanced covariate (Harder et al., 2010). See Appendix B-2 for more information.

Research assessing the performance of various propensity methods for estimating effects (in particular hazard ratios) suggests weighting may minimize bias more than other existing methods (Austin, 2013). While propensity adjustment (Rosenbaum & Rubin, 1983), stratifying (Rosenbaum & Rubin, 1984), and matching are other propensity score techniques that all attempt to remove confounding, weighting was chosen for several other reasons. Propensity score matching involves excluding patients from the sample who are unmatched (pruning), thus removing information and lowering the sample size. Removing patients/observations may introduce bias. In addition, propensity score matching, despite being a widely popular technique in observational studies, is increasingly controversial with research finding that it leads to biased estimates and performs suboptimally compared to other matching techniques (King & Nielsen, 2016). Propensity score matching throws away information and makes decisions on one dimension. Stratification approaches (Rosenbaum & Rubin, 1984) group patients with similar propensity scores, but the interpretation may be difficult when clinically meaningful distinctions exist between strata, (Curtis, Hammill, Eisenstein, Kramer, & Anstrom, 2007) and may result in the greatest bias compared to the other available

propensity score methods (Austin & Schuster, 2014).

As an additional step to reduce residual bias in the face of moderate levels of imbalance in some of the covariates, models were adjusted for covariates from five domains at baseline: demographics, utilization, laboratory values, comorbidities, and non-diabetes medications.

For stratified analyses, the hazard ratios obtained in each subgroup were compared to each other to test whether there is evidence supporting different effects (Altman & Bland, 2003).

Results

Figure 3.1 describes the sample derivation to obtain the analytic cohorts. There were 148,404 patients in the analysis cohort who met all the inclusion criteria: 138,097 received sulfonylurea second-line treatment (either as a monotherapy or in combination with metformin) and 7,153 were treated with TZDs. Of those treated with a second-line medication, 103,181 of sulfonylurea users and 4,897 of TZDs users were used in combination with metformin.

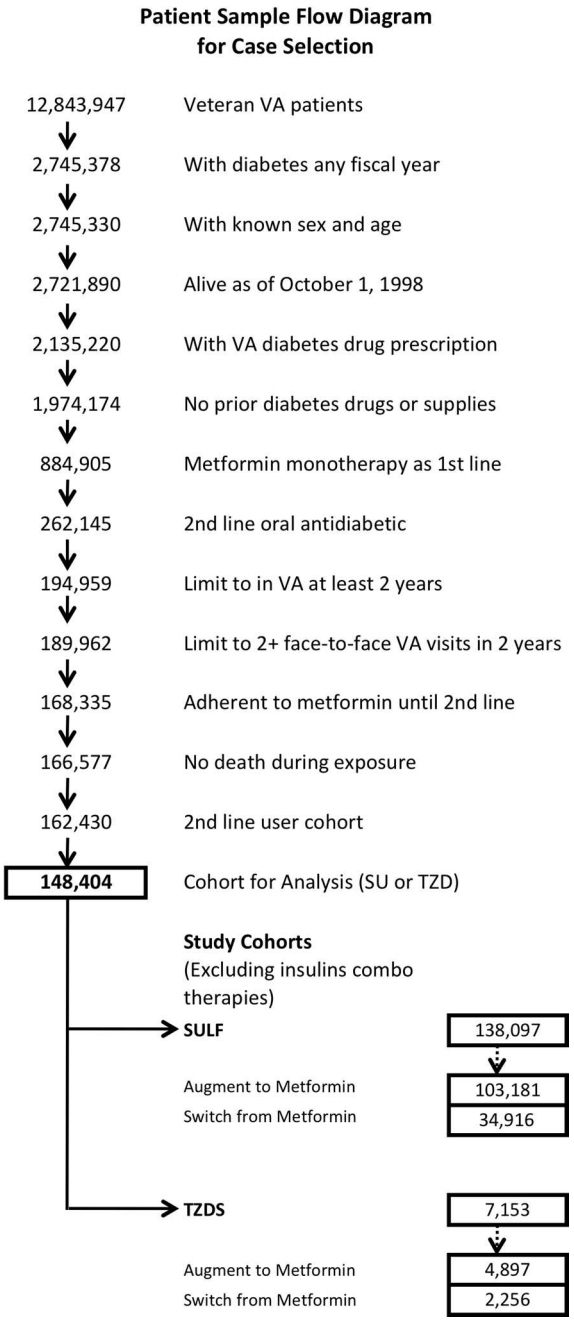


Figure 3.1. Sample derivation for patient selection into analytic sample

The distributions of estimated propensity scores for sulfonylurea and TZD users show an overlap in their range and density (i.e., common support) suggesting the two samples are comparable and propensity score methods can be applied (Figure 3.2).

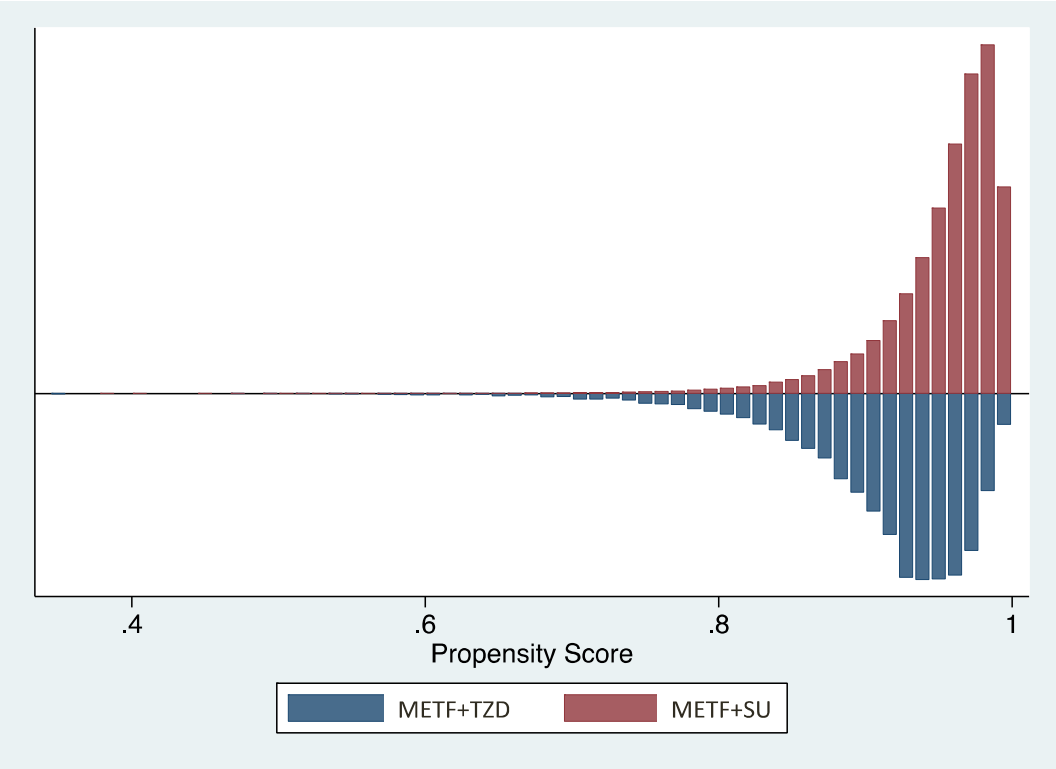


Figure 3.2. Overlap in propensity scores for METF+SU vs. METF+TZD

Figure 3.3 summarizes the absolute effect of sulfonylureas compared to TZDs for each outcome by comparing unadjusted Kaplan-Meier estimates of the survival function, and suggests that the two survival functions are significantly different with sulfonylurea having a higher risk of all-cause death and the cardiovascular composite end-point (log-rank chi-square=41.15, $p<0.0001$ and chi-square=57.344, $p=0.0001$ respectively).

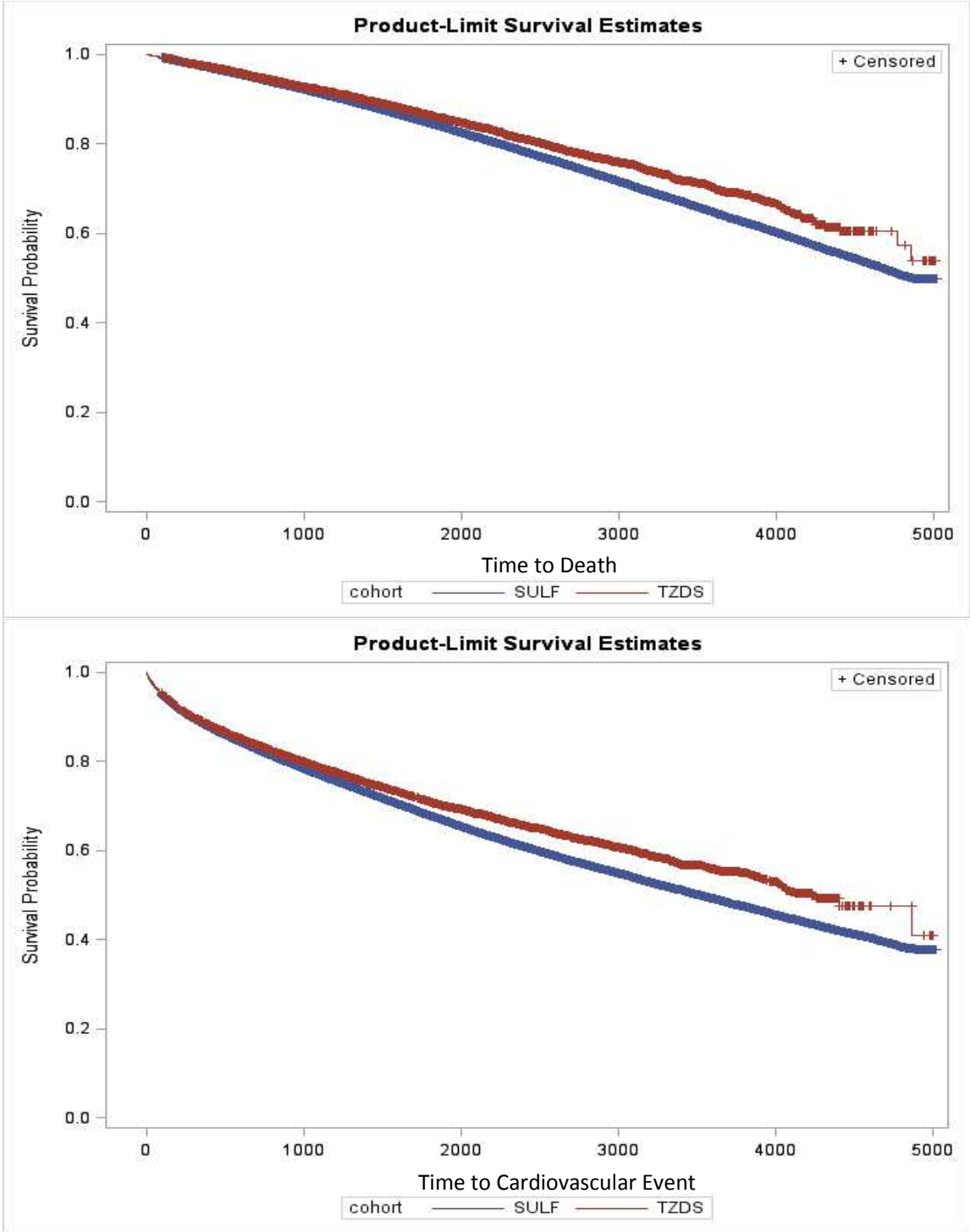


Figure 3.3. Unadjusted Kaplan-Meier curves for death (top) and cardiovascular event (bottom) for sulfonylureas (SULF) and TZD second-line treatments

Primary Intention-to-Treat Analysis.

Results for the relative estimated effect of sulfonylurea vs. TZD on each outcome are also reported. Table 3.2 presents the crude, weighted, and adjusted hazard ratios (HRs) separately for all-cause mortality and cardiovascular event. All models suggest an elevated risk of all-cause mortality and a cardiovascular event. In the final model (weighted, fully adjusted), second-line sulfonylurea users had a 15% increased risk of all-cause mortality and a 12% increased risk of a cardiovascular event when compared to TZD second-line treatment, (HR=1.15, 95% CI [1.08, 1.22]; HR=1.12, 95% CI [1.07, 1.17]). Thus, at any point in time, the risk of dying or having a cardiovascular event while receiving sulfonylureas treatment was 15% and 12% higher than the hazard of TZD treatment, respectively. When models were stratified, results were similar for all analyses (all *p*-values of comparisons were greater than 0.05), providing no evidence to suggest effects differ by age, Medicare status, or prior cardiovascular disease (see Appendix B).

Table 3.2. Primary intention-to-treat analysis: Crude, propensity adjusted and weighted, and weighted adjusted hazard ratios for all-cause mortality and cardiovascular events comparing sulfonylurea to TZD treatment

Treatment cohort	Patients	Events	Person-years	Incidence rate	Crude HR (95% CI)	PS adjusted (95% CI)	Weighted, unadjusted (95% CI)*	Weighted, adjusted (95% CI)*†
2nd line treatment								
All-cause mortality								
Sulfonylurea	138,097	28,386	719,850	3.94	1.19 (1.13-1.25)	1.19 (1.13-1.25)	1.22 (1.15-1.3)	1.15 (1.08-1.22)
TZD	7,153	1,504	45,647	3.29	1.00 (reference)	1.00 (reference)	1.00 (reference)	1.00 (reference)
Cardiovascular composite event								
Sulfonylurea	138,097	46,308	580,860	7.97	1.17 (1.12-1.22)	1.17 (1.12-1.22)	1.19 (1.13-1.24)	1.12 (1.07-1.17)
TZD	7,153	2,514	38,234	6.58	1.00 (reference)	1.00 (reference)	1.00 (reference)	1.00 (reference)

CI , confidence interval; HR, hazard ratio; *PS* , propensity score

Note: Incidence rate per 10,000 person-years

*Weighted by the inverse probability of treatment

†Adjusted for demographic, utilization, laboratory values, medications, and comorbidities at baseline

Table 3.3 summarizes the relative results of the secondary analysis in which medications are further classified into whether a new second-line treatment is added to metformin (i.e., second-line combination therapy) or switched off metformin (second-line monotherapy). In the crude model, the risk of death at any point was lower for the METF+SU combination group compared to switching off metformin to either TZDs or sulfonylureas. Suggestions of a lower risk for METF+SU combination therapy was supported in subsequent weighted and adjusted models vs. sulfonylurea monotherapy (weighted, fully adjusted model HR=0.88, 95% CI [0.85, 0.90]), but not vs. TZD monotherapy when confounding is minimized (weighted, fully adjusted model HR=1.06 95% CI [0.96, 1.18]). Additionally, the METF+SU combination was associated with an increased risk of death compared to the METF+TZD combination group in the unadjusted model and remained elevated in subsequent models controlling for measured confounding (weighted fully adjusted model HR=1.20, 95% CI [1.14, 1.28]).

For weighted adjusted estimates comparing sulfonylurea monotherapy to TZD mono and combination therapies, sulfonylurea monotherapy had a higher risk of death against TZD monotherapy (HR=1.21, 95% CI [1.09, 1.89]) but not compared to METF+TZD combination therapy (HR=1.37, 95% CI [0.98,1.93]). Final models find a difference in risk between METF+TZD and TZD only therapies (HR=0.88, 95% CI [0.56, 1.38]; reference=TZD monotherapy)

Similar conclusions for each comparison can be made regarding the risk of a cardiovascular event. In the final model, combination METF+SU had a lower risk of cardiovascular composite events compared to sulfonylurea monotherapy (HR=0.90, 95%

CI [0.88, 0.93]), a similar risk to TZD monotherapy (HR=1.06, 95% CI [99, 1.15]), and a higher risk than METF+TZD combination therapy (HR=1.15, 95% CI [1.10, 1.20]).

For weighted adjusted estimates comparing sulfonylurea monotherapy to TZD mono- and combination therapies, sulfonylurea monotherapy had an elevated risk of cardiovascular composite events relative to TZD monotherapy (HR=1.18, 95% CI [1.10, 1.72]) but not against METF+TZD combination therapy (HR=1.28, 95% CI [0.95, 1.72]). Final models fail to find a difference in risk between METF+TZD and TZD therapies (HR=0.93, 95% CI [0.63, 1.35]; reference=TZD monotherapy).

Table 3.3. Secondary intention-to-treat analysis: Crude, propensity adjusted and weighted, and weighted adjusted hazard ratios for all-cause mortality and cardiovascular events comparing second-line treatments as metformin combinations or monotherapy

Treatment cohort	Patients	Events	Person-years	Incidence rate	Crude HR (95% CI)	PS adjusted (95% CI)	Weighted, unadjusted (95% CI)*	Weighted, adjusted (95% CI)*†
2nd line treatment								
All-cause mortality								
Metformin+Sulfonylurea	103,181	20,229	560,043	3.61	1.00 (reference)	1.00 (reference)	1.00 (reference)	1.00 (reference)
Sulfonylurea only	34,916	8,157	159,808	5.10	1.51 (1.47-1.55)	1.08 (1.05-1.11)	1.08 (1.05-1.11)	1.14 (1.11-1.17)
TZD only	2,256	643	13,545	4.75	1.36 (1.26-1.48)	0.91 (0.84-0.99)	0.92 (0.84-1.02)	0.94 (0.85-1.04)
Metformin+TZD	4,897	861	32,103	2.68	0.76 (0.71-0.82)	0.68 (0.64-0.73)	0.60 (0.57-0.64)	0.83 (0.78-0.88)
*Cardiovascular composite event*ξ								
Metformin+Sulfonylurea	103,181	33,381	453,551	7.36	1.00 (reference)	1.00 (reference)	1.00 (reference)	1.00 (reference)
Sulfonylurea only	34,916	12,927	127,309	10.15	1.35 (1.32-1.37)	1.05 (1.03-1.08)	1.05 (1.03-1.08)	1.11 (1.08-1.13)
TZD only	2,256	969	10,865	8.92	1.24 (1.16-1.32)	0.93 (0.87-0.99)	0.92 (0.85-1.00)	0.94 (0.87-1.01)
Metformin+TZD	4,897	1,545	27,369	5.65	0.80 (0.76-0.84)	0.74 (0.70-0.78)	0.70 (0.67-0.73)	0.87 (0.83-0.91)

CI , confidence interval; HR, hazard ratio; *PS* , propensity score; TZDS, thiazolidinedione

Note: Incidence rate per 10,000 person-years

*Weighted by the inverse probability of treatment

†Adjusted for demographic, utilization, laboratory values, medications, and comorbidities at baseline

ξ Defined as composite of myocardial infarction, acute coronary syndrome, stroke, cardiac surgery, congestive heart failure, and all-cause death

Figure 3.4 reports the absolute effect (unadjusted) for each second-line combination and monotherapy in each outcome.

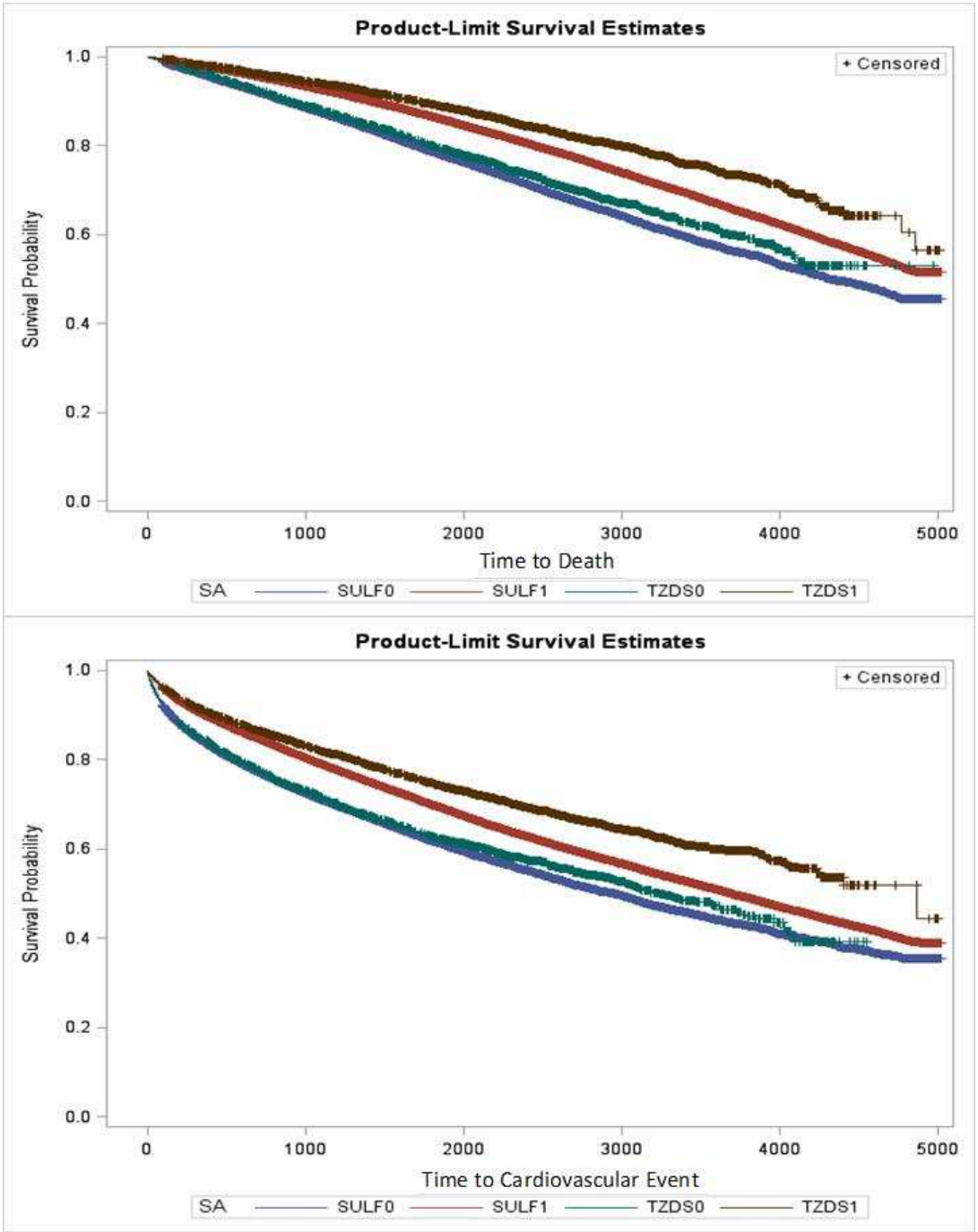


Figure 3.4. Unadjusted Kaplan-Meier curves for death (top) and cardiovascular event (bottom) for sulfonylurea (SULF) and TZD monotherapy and combination second-line treatments 0=switch; 1=augment

Sensitivity As-Treated Analysis.

The primary results of the as-treated analysis suggest similar elevated effects in the METF+SU cohort for both outcomes. Sulfonylurea use as a second-line therapy was associated with an increased hazard compared to second-line TZD use of all-cause mortality (HR=1.39, 95% CI [1.14, 1.70]) and composite cardiovascular events (HR=1.19, 95% CI [1.09, 1.30]).

In the secondary analyses, changing the analytical approach from an intent-to-treat to as-treated analysis were again similar, but moved the effects away from the null. In the final model, combination METF+SU had a lower risk of death compared to sulfonylurea monotherapy (HR=0.61, 95% CI [0.66, 0.57]), a similar risk to TZD monotherapy (HR=0.92, 95% CI [0.66, 1.27]), and a higher risk than METF+TZD combination therapy (HR=1.49, 95% CI [1.14, 1.96]).

For weighted adjusted estimates comparing sulfonylurea monotherapy to TZD mono- and combination therapies, sulfonylurea monotherapy had an elevated hazard relative to TZD monotherapy (HR=1.50, 95% CI [1.07, 3.34]) and against METF+TZD combination therapy (HR=2.43, 95% CI [1.17, 5.07]) (compare to non-significant finding in ITT analysis; reference group=sulfonylurea monotherapy). There was no significant difference in risk between METF+TZD and TZD therapies (HR=0.61, 95% CI [0.26, 1.43]; reference=TZD monotherapy).

Table 3.4. Primary as-treated analysis: Crude, propensity adjusted and weighted, and weighted adjusted hazard ratios for all-cause mortality and cardiovascular events comparing sulfonylurea to TZD treatment

Treatment cohort	Patients	Events	Person-years	Incidence rate	Crude HR (95% CI)	PS adjusted (95% CI)	Weighted, unadjusted (95% CI)*	Weighted, adjusted (95% CI)*†
2nd line treatment								
All-cause mortality								
Sulfonylurea	138,097	4,830	245,607	1.97	1.26 (1.07-1.48)	1.33 (1.13-1.57)	1.37 (1.12-1.68)	1.39 (1.14-1.70)
TZD	7,153	151	9,883	1.53	1.00 (reference)	1.00 (reference)	1.00 (reference)	1.00 (reference)
Cardiovascular composite event								
Sulfonylurea	138,097	19,113	220,508	8.67	1.11 (1.03-1.19)	1.14 (1.06-1.23)	1.19 (1.09-1.30)	1.19 (1.09-1.30)
TZD	7,153	761	9,065	8.39	1.00 (reference)	1.00 (reference)	1.00 (reference)	1.00 (reference)

CI, confidence interval; HR, hazard ratio; *PS*, propensity score

Note: Incidence rate per 10,000 person-years

*Weighted by the inverse probability of treatment

†Adjusted for demographic, utilization, laboratory values, medications, and comorbidities at baseline

Table 3.5. Secondary as-treated analysis: Crude, propensity adjusted and weighted, and weighted adjusted hazard ratios for all-cause mortality and cardiovascular events comparing second-line treatments as metformin combinations or monotherapy

Treatment cohort	Patients	Events	Person-years	Incidence rate	Crude HR (95% CI)	PS adjusted (95% CI)	Weighted, unadjusted (95% CI)*	Weighted, adjusted (95% CI)*†
2nd line treatment								
All-cause mortality								
Metformin+Sulfonylurea	103,181	1,689	141,946	1.19	1.00 (reference)	1.00 (reference)	1.00 (reference)	1.00 (reference)
Sulfonylurea only	34,916	1,798	61,766	2.91	2.43 (2.28-2.60)	1.60 (1.47-1.74)	1.85 (1.72-1.99)	1.63 (1.52-1.76)
TZD only	2,256	75	3,156	2.38	2.01 (1.59-2.53)	1.26 (0.99-1.59)	1.29 (0.93-1.77)	1.09 (0.79-1.51)
Metformin+TZD	4,897	46	5,303	0.87	0.74 (0.55-0.99)	0.65 (0.49-0.88)	0.59 (0.45-0.78)	0.67 (0.51-0.88)
Cardiovascular composite event								
Metformin+Sulfonylurea	103,181	9,549	131,760	7.25	1.00 (reference)	1.00 (reference)	1.00 (reference)	1.00 (reference)
Sulfonylurea only	34,916	6,320	53,273	11.86	1.70 (1.65-1.76)	1.17 (1.13-1.22)	1.29 (1.25-1.34)	1.24 (1.19-1.28)
TZD only	2,256	336	2,796	12.02	1.61 (1.44-1.79)	1.06 (0.95-1.19)	1.14 (0.98-1.31)	1.01 (0.87-1.16)
Metformin+TZD	4,897	328	4,984	6.58	0.84 (0.75-0.94)	0.76 (0.68-0.85)	0.67 (0.61-0.74)	0.77 (0.69-0.85)

CI, confidence interval; HR, hazard ratio; *PS*, propensity score

Note: Incidence rate per 10,000 person-years

*Weighted by the inverse probability of treatment

†Adjusted for demographic, utilization, laboratory values, medications, and comorbidities at baseline

In the final as-treated model for composite cardiovascular events, combination METF+SU had a lower risk of cardiovascular composite events compared to sulfonylurea monotherapy (HR=0.81, 95% CI [0.78, 0.84]), a similar risk to TZD monotherapy (HR=0.99, 95% CI [0.86, 1.15]), and a 30% higher risk than METF+TZD combination therapy (HR=1.30, 95% CI [1.18, 1.45]).

For weighted adjusted estimates comparing sulfonylurea monotherapy to TZD mono and combination therapies, sulfonylurea monotherapy had a higher risk against TZD monotherapy (HR=1.23, 95% CI [1.07, 2.07) and METF+TZD combination therapy (HR=1.61, 95% CI [1.04, 2.50]; compare to non-significant finding in ITT analysis) (reference group=sulfonylurea monotherapy). There was no significant difference in risk between TZD therapies (HR=0.76, 95% CI [0.45, 1.30]; reference=TZD monotherapy).

Discussion

With limited evidence-based information on long-term outcomes to guide appropriate therapy, this paper addresses critical gaps in existing knowledge surrounding the safety risks of sulfonylureas. New information regarding the long-term safety of sulfonylurea against TZD as second-line therapies is reported. This study also adds knowledge regarding the relationship that continuing or discontinuing existing metformin to a second-line medication has on long-term outcomes.

Findings from the primary analysis suggest that sulfonylurea second-line treatment is related to an increased risk of mortality and composite cardiovascular events compared to TZD second-line users. This study also helps to better understand the risks of adding medications to metformin and discontinuing metformin in favor of other oral

antidiabetics. Results suggest there is effect modification (see Table 3.6 for a summary). When second-line cohorts are divided into patients initially treated with metformin and subsequently add or switch to a new second-line therapy, patients who add sulfonylurea compared to those who switch to sulfonylurea monotherapy have a lower risk of mortality and composite cardiovascular events. This suggests it may be favorable to add instead of switch to a sulfonylurea. When compared to those who augmented TZD to existing metformin, METF+SU combination therapy had an elevated risk on both long-term outcomes in all models. While an elevated risk was found against METF+TZD, the

Table 3.6. Summary of risks comparing second-line cohorts that augment to metformin (combo therapy) or switch off metformin (monotherapy)

Cohort	Effect	Comparator (reference)
METF+SU	NEGATIVE	SU only
	POSITIVE	METF+TZD
	NULL	TZD only
SU only	POSITIVE	METF+SU
	INCONCLUSIVE1	METF+TZD
	POSITIVE	TZD only
METF+TZD	NEGATIVE	METF+SU
	INCONCLUSIVE2	SU only
	NULL	TZD only
TZD only	NULL	METF+SU
	NEGATIVE	SU only
	NULL	METF+TZD

+ elevated risk for cohort relative to comparator
- decreased risk for cohort relative to comparator
INCONCLUSIVE1 = elevated in as-treated analysis; null in intention-to-treat analysis
INCONCLUSIVE2 = lowered risk in as-treated analysis; null in intention-to-treat analysis
NULL=no difference in risk

results failed to find a difference in risk when METF+SU is compared to those who switched to TZD monotherapy. Specifically, while METF+SU combination therapy had a lower risk in both outcomes when compared to those who switched off metformin to TZD monotherapy in the crude models, it was not confirmed in the weighted adjusted models.

Amongst patients who switched off of metformin, the sulfonylurea cohort had higher risks compared to those who switched to TZD for both outcomes. Compared to TZD augmented to metformin, the as-treated analysis found second-line sulfonylurea monotherapy was worse (higher risk) than TZD augmented to metformin for both outcomes, but was not confirmed in the weighted adjusted model intention-to-treat analysis. With the exception noted, as-treated design models had similar conclusions to those found using the intention-to-treat design, with the magnitude of the effect being larger in the as-treated design models.

To date, the long-term safety of sulfonylurea compared to other agents has been unclear and insufficient. There have only been a few studies examining sulfonylurea against TZD as second-line agents after metformin. Two RCTs examined second-line use but were relatively short term with few events. Hamann et al. (2008) (only 52 weeks, two deaths in each group) and Matthews (2005) (two deaths in METF+SU group, zero deaths in METF+TZD group). In an observational cohort study Morgan et al. (2012) studied patients in the General Practice Research Database (data of patients from approximately 700 primary care practices in the United Kingdom) and found similar findings to this current study. Specifically, they found elevated effects on all-cause

mortality and cardiovascular composite events for patients who switched to sulfonylurea compared to METF+TZD and METF+SU second-line therapies. Also, they report an elevated effect for METF+SU compared to METF+Pioglizatone (a TZD).

Other studies have compared sulfonylureas to TZD therapies but often ignore the sequencing of therapies, combine augments and switches into the same cohort, ignore earlier therapies prior to study exposure (not incident users), or only assess medications as first-line therapies. Earlier studies examining sulfonylurea against TZD with at least one year of follow-up are described in Appendix Table B-6 and B-7.

Three randomized control trials examined all-cause mortality and cardiovascular events, but none specifically at sulfonylureas and TZDs as second-line therapies. The largest investigation, A Diabetes Outcome Progression Trial (ADOPT), failed to find an elevated risk for first-line sulfonylurea monotherapy compared to first-line TZD for all-cause mortality (crude relative risk=0.92, 95% CI [0.81, 1.04]) and cardiovascular events (crude relative risk=0.67, 95% CI [0.41, 1.08]). Yet there were issues with the study worth noting.

Most troublesome were that more patients withdrew in the sulfonylurea group (44% vs. 37%) and the sulfonylurea group had shorter follow-up (3.3 compared to 4 years) (Kahn et al., 2006). Two smaller first-line monotherapy trials reported on all-cause mortality but were fairly short-term, only 56 weeks (Jain et al., 2006) and 52 weeks (Hanefeld, Patwardhan, & Jones, 2007) with limited information on mortality. Jain et al. (2006) reported one death in the sulfonylurea group and no deaths in the TZD group. Hanefeld et al. (2007) reported zero deaths in both groups. Hanefeld and colleagues

(2007) also reported cardiovascular events finding twice as many events in the sulfonylurea group (22 from 251 sulfonylurea patients vs. 11 from 251 TZD patients; crude relative risk=2.00, 95% CI 1.01–3.99). One additional RCT examined cardiovascular morbidity. St. John Sutton (2002) conducted a small, 52-week RCT that included non-incident users and found similar incidence rates for both groups: 12 events from 99 patients in the sulfonylurea treatment group and 16 of 104 in the TZD group.

Evidence suggesting a larger sulfonylurea risk compared to TZD is primarily derived from observational cohort studies. Two observational cohort studies examined sulfonylurea vs. TZD monotherapy provide some evidence for an elevated risk for sulfonylurea therapy. Pantalone et al. (2009) conducted an observational study using patients from the Cleveland Clinic and found increased risk of death for sulfonylurea patients against pioglitazone (adjusted HR=1.69, 95% CI [1.23, 2.33]) but not for rosiglitazone (adjusted HR=1.37, 95% CI [0.98, 1.96]). They also found no increased risk for cardiovascular events. Wheeler (2013) studied mortality amongst new users on sulfonylurea and TZD monotherapy in the VA from 2005–2009 and found an elevated risk for two sulfonylurea agents relative to TZDs: glyburide HR=1.12, 95% CI [1.02, 1.23] and glipizide (HR=1.27, 95% CI [1.01, 1.59].

Fewer studies examined combination therapy. One RCT compared the effect of sulfonylurea and TZD metformin combinations therapies to each other, but was short-term and underpowered: Bakris et al. (2006) 32 weeks (no deaths in sulfonylurea group, one death in TZD group).

Our primary finding that second-line sulfonylurea treatment may be inferior in

terms of long-term safety risks vs. TZDs is consistent with an earlier observational study of veterans by Prentice et al. (2014) who examined sulfonylurea and TZD use after a metformin prescription was dispensed using VA data from 2000–2009. They found an elevated risk for mortality (HR=1.50, CI [1.31, 2.15]) and stroke/acute myocardial infarction (HR=1.15, 95% CI [0.80, 1.17]). However, this study combined augmenting and switching groups into the same cohort and did not examine the influence that augmenting and switching may have on long-term risks.

The present study has both limitations and strengths worth addressing. This is one of largest studies of its kind with the several years of follow-up. Also, confounding was addressed through several strategies that lead to a rigorous observational study design. In the VA, guidelines state that METF+SU is the preferred combination therapy in patients who no longer have adequate glycemic control on metformin, and suggest TZD combination therapy to be considered for patients unable to use a sulfonylurea "due to contraindications, adverse events, or risk for adverse events" (Management of Diabetes Mellitus Update Working Group, 2010).

While, confounding by indication can never entirely be ruled out, this study addressed observed confounding in several ways. First, strict inclusion criteria produced patients at similar stages of diabetes, which help to differentiate drug effects from disease effects better. Second, this study also used a variant of new user design to identify patients at similar progressions of a more advanced stage of diabetes. This variant of the new user design allowed for a better group of comparators whose underlying cardiovascular risks were similar. Third, propensity scores were used in an effort to

obtain less biased estimates of the average treatment effect. Specifically, propensity score weighting resulted in relatively few imbalances, which suggest that the study design and inclusion criteria worked well to create similar cohorts even before balancing. Performance suggest that the various ways the model was adjusted did not change the effect size estimates much provides some evidence that decreased model dependence was a result of using an IPTW approach. Fourth we were able to control for a rich set of factors that included disease severity, comorbidities, utilization, and other factors that may help explain the relationship between medication exposure and the outcomes. Though there still may be unmeasured confounding that explains why patients augment or switch off of metformin, effect sizes were similar to those found by Prentice (2014) who used an instrumental variables approach (using variation in provider-prescribing as the instrument) in a study of dual Medicare/VHA veterans.

Strict inclusion/exclusion criteria and the new user design likely generated a suitable comparison group at similar stages of diabetes but there may be a lack of generalization. Patients who did not tolerate metformin in short term are excluded, so this study cannot comment on the safety profile of patients who did not tolerate metformin in the short term. The same is true for those who did not tolerate their second-line medication in the short term. Also, this study consists of mostly a male population.

Both intention-to-treat and as-treated analyses were used to understand the bias that may occur when exposure is not properly captured. Both analyses yielded similar conclusions in most comparisons. These designs have different strengths and weaknesses. Intention-to-treat is susceptible to exposure misclassification as the follow-

up becomes longer, but it is simpler and does not make any assumptions about adherence to therapy. Intention-to-treat tends to reduce effects thus it is biased toward the null hypothesis (Hernán & Hernández-Díaz, 2012; Patorno et al., 2014). This is what was found in the current study. Conversely, as-treated analyses use informative censoring, which terminates exposure when therapy is augmented or discontinued, so it may bias estimates if censoring predicts future events. As-treated analyses result in more censored patients thus reducing the number of events. Both of these models do not account for time-varying exposure and covariates, however. Future work may explore the use of marginal structural models as an additional approach, though there is a time-varying assumption that treatment changes are independent of outcomes.

There were several other limitations. We only had access to several years of Medicare utilization data. However, cohorts were stratified by Medicare beneficiary status in the entire study period and results were similar to those who were not dually enrolled in Medicare.

Also, this study cannot comment on newer medications because of the lack of prescribing in data. Not many VA users were prescribed newer medications (e.g., DPP-4 inhibitors, GLP-1 agonists, SGLT-2 inhibitors) to conduct meaningful analyses up through 2012. This may be a byproduct of good adherence to the clinical guideline for use, since the formulary requires failure or contraindication of sulfonylurea or TZD, and these patients might be third-line users and were eliminated through the inclusion criteria of the study.

Caution should be made to encourage physicians to rush to newer medications.

These results cannot provide evidence on how sulfonylurea is related to newer oral antidiabetic medications. The long-term safety profile of newer drugs as second-line agents is not yet known.

Finally, this study does not consider other safety and effectiveness considerations that would be involved in making diabetes medication treatments decisions. They include therapy effectiveness (glycemic control, weight, lipid levels), safety/adverse events (e.g., hypoglycemia, pancreatitis, fractures), long-term effects (e.g., microvascular complications), and costs.

Conclusion

This study attempted to understand the complexities involved in prescribing antidiabetic therapies related to adding on and switching classes of medications, specifically sulfonylureas relative to other second-line therapy. Results have implications for EBM treatment decisions. While the primary analysis found that second-line sulfonylurea treatment might be inferior in terms of long-term safety risks vs. TZD, a closer look into whether first-line metformin is discontinued appeared to modify the results. This is one of the few studies to examine the risks associated with augmenting or switching from metformin to second-line sulfonylurea or TZD therapies. Research suggests switching off metformin onto sulfonylurea monotherapy may lead to elevated long-term risks compared to adding sulfonylurea to metformin. It also suggests that sulfonylurea used as an add-on therapy to metformin is related to an increased risk compared to when TZD is used as an add-on therapy. Future research should continue to disentangle the influence of augmenting and substituting second-line therapies, and

examine whether augmenting metformin leads to lower long-term risks for both TZD and sulfonylurea when compared to switching off metformin to other monotherapies.

CHAPTER 4
AN EVIDENCE BASED-MEDICINE FRIENDLY APPROACH TO EVALUATING THE COMPARATIVE SAFETY OF SULFONYLUREA RISK COMPARED TO TZD SECOND-LINE THERAPY FOR ALL-CAUSE MORTALITY

Abstract

Evidence based medicine (EBM)-friendly research methods that allow for more thoughtful and comprehensive assessments are needed. This paper reintroduces a method adapted from Brophy and Joseph (1995) that can help healthcare practitioners evaluate clinically relevant research in a way that supports evidence-based decision-making: leading to more comprehensive evaluations of evidence. Within this approach, the degree of uncertainty around pre-existing evidence, for example a meta-analysis derived estimate of prior studies, is varied to simulate differences in clinical evaluations of pre-existing evidence by modifying the influence of earlier study results (e.g., 100% use of the prior information, 50%, 10%, or 0%). Then using a Bayesian statistical approach, the learning process of combining existing evidence and updating it with new study data can be represented. The weight (i.e., use of prior information) that most closely resembles a clinician's own evaluation of the strength of the earlier evidence will determine how much it will influence the new conclusion. If prior evidence is down-weighted, it reflects a lack of confidence in the early studies and will contribute less information when forming the new conclusion.

This approach allows clinicians to draw their own conclusions by interpreting the results of a new study while also taking into account their evaluation of the pre-existing evidence. This approach requires evaluations of prior evidence to be stated explicitly and

allows clinicians to draw their own conclusions as well as observe the potential conclusions other clinicians may reach.

In this illustrative example, the evidence surrounding the question of the long-term safety risks of sulfonylureas and whether they should continue to be prescribed as an appropriate second-line therapy to treat type 2 diabetes given is explored. Results suggest that the extent to which pre-study results influenced the newly formed conclusions were minimal relative to the contribution that the new VA study data made in terms of information. Each post-study estimate reached a similar conclusion across each scenario, even when the extent and type of prior evidence used in the analyses was varied. Depending on which scenario was chosen, there was a 15%–19% increased risk of all-cause mortality associated with second-line sulfonylurea therapy relative to TZD second-line therapy following metformin. Implications of this method as well as future directions are discussed.

Introduction

The ability to evaluate existing research evidence on the therapeutic benefits and harms of therapy is a key element in evidence-based medicine (EBM) approaches of clinical decision-making (Sackett et al., 1996, 1996). According to the EBM framework, clinically relevant research is determined through evaluations of the strength and weaknesses of studies related to the treatment population. However, evaluations of scientific rigor (e.g., methodological quality of a study) and relevance of study results to particular patient populations can vary by clinician. As a result, clinical opinions regarding the same study results or body of evidence may be fragmented, and may not

lead to convergence of opinion, especially when the evidence appears inconsistent or inconclusive. Even amongst researchers disagreement can exist on the quality of a given study: the inter-rater agreement can often be poor to fair when applying valid survey instruments to assess the quality of a study. This is true in assessing both observational cohort and RCT studies (Armijo-Olivo et al., 2014; Armijo-Olivo, Stiles, Hagen, Biondo, & Cummings, 2012; Clark et al., 1999; Lo, Mertz, & Loeb, 2014).

Therefore, when conducting new studies, it may be helpful to utilize analytic methodologies that are complementary to the EBM framework, in particular, where differing evaluations of the pre-existing evidence are allowed to be made explicit and then taken into account when evaluating new study evidence. The purpose of this chapter is to demonstrate the utility of an underutilized approach by Brophy and Joseph (1995) that incorporates results from earlier studies and then combines it with the results of a new study to provide a more comprehensive evaluation of the evidence supporting a given hypothesis. To apply this method, the long-term risks of prescribing sulfonylureas as a second-line glucose lowering therapy are examined by comparing it to TZD therapy. This is a particularly appropriate research question to apply this method since clinicians have different opinions regarding the relative benefits and safety of sulfonylureas, as evidenced in the support of different guidelines recommending/cautioning against the use of sulfonylureas as second-line agents.

The basic steps of the Brophy and Joseph evidence synthesis approach are outlined in Table 4.1 and discussed in the following sections.

Table 4.1. List of steps in the Brophy and Joseph (1995) approach

Step	Description
1)	Review the existing evidence. Pool estimates, if appropriate, using evidence synthesis methods (e.g., meta-analysis) to summarize pre-existing information.
2)	For existing evidence, generate a range of plausible clinical evaluations by adjusting the degree of uncertainty around the pre-existing results to represent various scenarios of the strength of the evidence.
3)	Conduct a new study. Use Bayesian statistics to combine the new study estimate with each of the different clinical evaluations of the pre-existing evidence.
4)	Reevaluate evidence. Compare the post-study conclusion derived from each of the different pre-study clinical evaluations. Then, find the post-study conclusion that was derived from the pre-study clinical evaluation that most closely resembles one's own and base the updated conclusion on this estimate.

First an evaluation of all relevant prior evidence is needed. Oftentimes, if there are enough studies on the topic and pooling estimates across studies is appropriate, then prior results can be synthesized into statistical summaries, using analytic methods like meta-analysis.

Second, to reflect the confidence in precision of the meta-analysis estimate obtained in the first step, the uncertainty around the pooled estimate is varied. As a result, a range of probability distributions summarizing the prior evidence is generated. When applied to EBM, this would simulate the range of different clinical opinions that likely exists regarding the scientific rigor and relevancy of previous study results. This range should reflect the uncertainty and relevance of the evidence that exists amongst clinicians. The greater the uncertainty in the prior evidence, the less influence it will ultimately have when forming a new conclusion. In this paper, approximations of potential clinical opinions are created by weighting the prior evidence, using 100%, 50%,

or 10% of the prior information, where the amount of information is represented by the width of the confidence interval.

The decision for a clinician to weight the earlier evidence may stem from several types of concerns about the study's internal and external validity, including its ability to examine long-term risks or the ability to generalize to their specific patient panel. Other concerns often are aspects of the study methodology, such as not adequately minimizing the risk of bias more specifically controlling for confounding by indication. For example, if a clinician believes that previous studies were poorly designed to evaluate long-term risks with serious methodological flaws and should be completely ignored, then the model that uses only new study data and gives no influence to earlier evidence (0% of prior information) would be used to form their new conclusion.

The third step is to conduct a new study and use that information along with the earlier evidence to update existing knowledge. Combining pre-study evidence with new data is done using a Bayesian model. Since several clinical evaluations exist before the study (see step 2 in Table 4.1) there is a range of post-study conclusions reported – one derived from each of pre-study clinical evaluations.

The fourth step is then to reevaluate the evidence. First, a consideration of all of the different conclusions that have been derived from each pre-study clinical evaluation should be made. This allows for a broad understanding across all of the likely conclusions reached by clinicians who had judged the existing evidence differently coming into the study. Next, a clinician should form a new conclusion by identifying the pre-study evaluation that most closely resembles their own and then finding the

corresponding post-study evaluation derived from it.

Evidence Based Medicine Model of Clinical Practice

Since this approach is being applied to EBM, this section describes EBM and how the Brophy and Joseph method can be applied. EBM is commonly defined as "the integration of best research evidence with clinical expertise and patient values" (Rosenberg & Haynes with Sackett, Straus Sharon E. and RichardsonW. Scott, 2000) and involves "the conscientious, explicit, and judicious use of clinically relevant research in making decisions about the care of individual patients" (Sackett et al., 1996). This is the basis for the conceptual model developed by Sacket et al. (2000) and is depicted in the diagram presented in the Chapter 1 (Figure 1.1).

The integration of three main components are involved in most EBM conceptual models of decision-making: Best Research Evidence; Individual Clinical Expertise; and Patient Values and Preferences. First, the model acknowledges the importance of previous research in informing clinical practice. Best External Evidence involves evaluating evidence from systematic investigations that are clinically relevant. For diabetes research, this includes existing comparative effectiveness and safety evaluations. The gold standard in evidence is appropriately designed RCTs given their strong internal validity to infer causality. In EBM approaches, physicians are "urged to assume that the clinical effectiveness results from RCTs could be applied to their own patients, unless there was a good reason not to make this default assumption" (Charles et al., 2011; Straus et al., 2005). However, RCT designs have their limitations. Evidence solely from RCTs is inadequate for the questions in this dissertation for several reasons. First, the relatively

short follow-up and small sample sizes of RCTs make long-term evaluations difficult and underpowered (susceptible in making type II errors). Second, the generalizability is of concern. This is particularly true in populations that have multiple comorbidities — a common criterion for study exclusion. That may not reflect common characteristics in clinical practice.

Evidence from other study designs should be included to a certain degree. A key challenge for evaluators is the extent in which to incorporate evidence from other study designs that are quasi-experimental. These observational studies may introduce bias given that patients are not randomized into particular treatment groups, but are typically much larger in size and have longer follow-up duration than RCTS. Also, they may provide useful information about real clinical populations typically excluded from RCTs and how they may benefit/be harmed by treatments. A Bayesian framework allows previous results to be included when new evidence is obtained and allows decision-makers and clinicians to assess the pertinence of the findings to their own patient population. This is a useful, yet underutilized framework in comparative safety research and may help stakeholders evaluate the strength of evidence and inform clinical decisions.

A clinician's experience, education, knowledge, and skills are also needed in the decision making process. The Individual Clinical Expertise component refers to "the proficiency and judgment that individual clinicians acquire through clinical experience and clinical practice" (Sackett et al., 1996). Also, the extent to which external evidence is judged to apply to an individual patient may rely on clinical expertise.

EBM models also acknowledge that the patient plays an important role in the decision-making process. The Patient Values and Preferences component includes the patient's "own personal and unique concerns, expectations, and values" (Sackett et al., 1996). This component also includes the types of patient information that should be considered in treatment decisions (i.e., either treatment decisions made on behalf of the patient or with active participation of the patient in the process).

The three components do not have to be equally relevant in a given decision and determining the appropriate balance may not be an easy task. That is, any one component or components may be stronger in any particular decision.

Background

To demonstrate how the Brophy and Joseph approach can be applied in health services research, we explore whether sulfonylureas should be used as second-line diabetic therapy relative to TZD based on all-cause mortality risk. This is an appropriate application of the Brophy and Joseph approach since clinicians and their patients face increasingly complex decisions about which medications regimens to use to manage diabetes with inconclusive and inconsistent evidence-based information to guide therapy. Evidence-based medicine relies, in part, on relevant empirical evidence of the benefits and harms of medications relative to other treatment. A wide array of medications exists to manage glucose levels, resulting in an even greater number of sequencing possibilities to treat type 2 diabetes patients as the illness progresses.

Sulfonylurea medications have been used to treat patients with type 2 diabetes since the 1950s, but they are increasingly controversial and have attracted considerable

debate in recent years (Abrahamson, 2015; Genuth, 2015). While they are an inexpensive, widely used, and very effective therapy for glycemic control (Bennett, Wilson, et al., 2011; UKPDS-33, 1998), they are also linked to adverse outcomes, including higher rates of cardiovascular events and death.

Clinicians have different opinions regarding the relative benefits and safety of sulfonylureas, leading to the support of different guidelines recommendations. Thus, this topic appears to be an appropriate candidate to apply the Brophy and Joseph approach since there is documented differences in clinical opinion. While the current consensus is that metformin should be the initial first-line agent when diet and exercise are not sufficient, the debate is whether sulfonylureas should be used as a) a second-line agent after metformin, b) a third- or fourth-line treatments, or c) not at all (Abrahamson, 2015; Genuth, 2015).

Given the disparate clinical opinions and lack of direct head-to-head long-term comparisons evaluating sulfonylurea treatment relative to other oral antidiabetic medications, further research on sulfonylureas safety is needed. Since this is a clinical topic where prior clinical opinions regarding the evidence differ, it makes sense to use several different evaluations of existing evidence to simulate these varying clinical opinions.

Methods

Bayesian Framework Overview

This study combines evidence from relevant prior studies with results from a new study of veterans to evaluate mortality risks associated with sulfonylureas relative to

TZDs using a Bayesian framework. In general, Bayesian reasoning is a mathematical way to modify existing knowledge when new information is presented. Applying a Bayesian approach to research questions relies on Bayes' Theorem and its three elements: the prior probability distribution, the likelihood function, and the posterior probability distribution. Goldstein (2006) summarizes the Bayesian approach in the following way:

> The subjective Bayesian approach is based on a very simple collection of ideas. You are uncertain about many things in the world. You can quantify your uncertainties as probabilities [i.e., the prior distribution], for the quantities you are interested in... When data arrives [i.e., the likelihood ratio], Bayes theorem tells you how to move from your prior probabilities to new conditional probabilities [i.e., the posterior distribution] for the quantities of interest.

When applied to research, a prior probability distribution can be derived empirically by using results estimated in an earlier study or collection of similar studies. The Brophy and Joseph approach applies Bayesian reasoning to present conclusions of new results taking into account different prior probability distributions. This approach is different from a typical Bayesian analyses conducted in health services research and other disciplines. Bayesian analyses are conducted in most scientific investigations using non-informative prior probability distributions. That means they ignore the results of earlier studies by specifying that all values are equally possible. In other words, it gives no weight to earlier results. Even if prior evidence is incorporated, it is typically specified in only one way. Yet this is clearly problematic in areas where there is a lack of a clinical consensus regarding the strength of existing evidence — in particular, where clinical opinion is fragmented like in the long-term risks of sulfonylurea use. When clinical

opinions differ, a single analysis will only include one clinical opinion and cannot reflect the reality that clinicians do not agree on the quality and strength of all previous studies. Therefore, certain groups might disregard or highly question any conclusions reached because the analysis had used a prior probability distribution that fails to reflect their appraisal of earlier evidence.

As a result, it is more difficult for researchers to converge clinical opinions when the prior reflects only one of the existing clinical opinions. This is especially true in health services research where observational studies are commonly used to derive empirical evidence (e.g., comparative effectiveness and safety studies), but other disciplines/groups (e.g., practicing clinicians) may only value results from RCTs and disregard or strongly discount any evidence that does not implement the RCTs gold standard design, perhaps partly because of a lack of familiarity with the methods employed in these studies (e.g., inverse probability weighting, propensity scoring, instrumental variables). Yet observational data are able to include patients with more complex conditions with profiles of patients usually seen in clinical practice and can extend the generalizability but can also potentially introduce bias and confounding.

The major objective of this study is to apply the Brophy and Joseph method to demonstrate how to allow clinicians to determine what 'Best Research Evidence' means to their own practice by allowing judgments of rigor and relevancy of prior research to be included in the study and combined with new study data. Whether this allows for a more comprehensive investigation concerning the evidence supporting a differential risk in all-cause mortality for sulfonylurea use relative to other antidiabetic therapies is explored.

Meta-Analysis

The first step of the Brophy and Joseph approach is to summarize the existing research evidence. In this paper, we rely upon three meta-analyses that pool earlier study findings to summarize the relative hazard between sulfonylurea and TZD therapy. The three separate meta-analyses are stratified by type of study design: within RCTs only, within observational cohort studies only, and combined across both of these study designs.

The details of the meta-analyses are described below. The search strategy and study selection criteria follow the same methodology used in Study One. Briefly, all observational cohort and RCT studies that explicitly evaluated all-cause mortality with at least one year of follow-up were culled. Studies were then excluded if they met any of the following criteria: only included patients with serious conditions at baseline such as a history of major cardiovascular events or renal failure, consisted of only children (younger than 18 years of age), type I diabetes patients only, did not include an active comparator (e.g., diet/exercise, placebo), case-control design, research only on animals, and written in a language other than English. For observational studies, an attempt to address confounding in some way must have been implemented (e.g., matched in the design or model adjustment) and had to include basic demographic information (i.e., age, sex, and race) and relevant comorbidities at baseline (those adjusting for CVD risk at least). For the study in this chapter, only earlier studies specifically evaluating sulfonylurea or TZD therapy after prior metformin use are included. Any studies that used VA data and overlapped with the current study's cohorts were also excluded.

To obtain the overall pooled estimate of risk that combines across RCT and observational cohort studies, the 'bayesmh' command with random effect of study design in Stata 14.1 (StataCorp, 2015) was used. In this model, heterogeneity between observational cohort and RCT study designs is accounted for in the estimate. It is important to note that no other assumptions were made regarding the relative strength RCT has relative to observational cohort studies when pooling the overall estimate of risk. This is similar to the approach by Peters et al. (2005) and involves Markov chain Monte Carlo (MCMC) estimation using a Metropolis-Hastings algorithm and Gibbs sampling with vague prior distributions specified on unknown parameters. Convergence diagnostics suggest fairly rapid convergence with no trend in trace plots, low autocorrelation, and acceptance rates for the Metropolis-Hastings algorithm around 75% (well above the 10% rule of thumb) and efficiencies above one percent for all analyses.

For pooled estimates by study design, the inverse variance and DerSimonian-Laird methods are used to estimate and report fixed and random effects estimates of risk respectively using the METAN command in Stata 14.1 with a value of 1.0 representing the null hypothesis (StataCorp, 2015). Heterogeneity across the studies was assessed via the I^2 statistic, with values greater than 50% benchmarked as indicating substantial heterogeneity (Higgins et al., 2003). This statistic represents the percent of variance in the effect size attributable to heterogeneity with larger values indicating less overlap in confidence intervals across studies. A benefit of the statistic is that number of studies involved in the meta-analysis has little influence on the I^2 statistic unlike other estimates. Unless noted, the inverse variance fixed effect estimates are used as the basis to generate

the prior probability distributions for within RCTS only and within observational cohort studies only stratified analyses.

Developing a Range of Prior Probability Distributions

Three series of stratified analyses were conducted using the three pooled estimates of prior evidence and served as the basis to approximate the pre-existing evidence represented in the models: 1) the overall combined (prior evidence from both study designs), 2) Prior RCT studies only, and 3) Prior observational studies only. For each, prior evidence was weighted by using 10%, 50%, and 100% of the prior information. These different weights simulate a potential range of clinical opinions regarding the rigor and relevancy of the prior results. This is done by changing the uncertainty around the estimate. Using 100% of the information means that the variance estimated from the meta-analysis is used to define the prior probability distribution.

Weighting prior evidence involves modifying the width of the confidence interval around the prior estimate. Therefore, the point estimate will remains the same, but the variance of the distributions will become flatter (less certain) as the weight of the prior data is decreased. Lower weights will cause the prior evidence to have less of an impact on the new conclusion formed.

Conducting a New Study and Combining it With Pre-Existing Information

Each evaluation of pre-existing evidence (i.e., prior probability distribution) is combined with the new study results (i.e., the likelihood ratio) to evaluate the hypothesis that sulfonylurea has a greater risk of all-cause mortality compared to TZD as a second-

line therapy. This revised evaluation of the evidence is the new conclusion and is called the posterior probability distribution. In general, if the evidence supporting a given hypothesis is strong and the prior evidence is weak, then the posterior will more closely resemble the results from the new study (i.e., the likelihood ratio). Whereas, if the prior evidence is strong and the likelihood ratio is weak then the posterior will more closely resemble the prior distribution (see Appendix C-2 and C-3 for more information on Bayes estimation).

A description of the basic methodology and data sources of this new study is presented in Chapter 3. In brief, a comparative safety analysis using Veterans Health Administration (VA) medical and administrative observational data was conducted of diabetic veterans who were prescribed metformin as a first-line therapy, and subsequently prescribed sulfonylurea or TZD for long-term use. To reduce confounding several approaches were taken. First, a variation of a new user cohort design was used in which incident use of a second-line medication was implemented after prior metformin prescribing to ensure that patients were at similar stages in diabetes progression when the second-line therapy was prescribed.

To minimize against bias, statistical models were weighted by the inverse probability of treatment using the propensity score and adjusted for a rich set of potential confounders. An intent-to-treat analytic approach was used, which ignores future changes to exposure and only uses initial exposure. Patients were only censored when an event did not occur by the end of the study period. Cox proportional hazards models were estimated using the PHREG procedure in SAS 9.2 (SAS Institute Inc., 2010) with

the BAYES statement added to invoke a Bayesian analysis of the model.

To obtain summary statistics for each model parameter, posterior probability distributions were obtained using Markov chain Monte Carlo via Gibbs sampling (MCMC) with an informed normal prior distribution specified for the drug exposure regression coefficient and non-informative prior distributions specified on all other variable regression coefficients (normal distributions, each with mean=0 and variance=1E6). The analyses resulted in a posterior MCMC sample size of 5,000 iterations after 2,000 iterations of burn-in for each parameter estimate.

Since the number of iterations determines the precision of posterior summaries, MCMC standard errors were examined to ensure that posterior samples are accurate to three decimal places. To assess Markov chain convergence, several diagnostic tests were evaluated: the Geweke test, autocorrelations, and effective sample size all suggest the Markov chain rapidly converged with good mixing, allowing the distribution to be estimated relatively easily. All together the diagnostics indicates the model converged on all parameters, and therefore are able to use the summaries of the posterior distributions that resulted from each simulation.

Results

Meta-Analysis

The meta-analytic results summarizing the prior evidence are presented in Figure 4.1. There were two RCTs and five estimates from observational cohort studies included in the meta-analyses. Using only prior evidence derived from RCT designs, the results

fail to suggest a difference in hazards between sulfonylurea second-line therapy and TZD (effect size[2]= 1.58, 95% CI [0.31, 8.2]). Conversely, when using only observational cohort studies, the results suggest a significant 28% increase in the hazard for sulfonylureas relative to TZDs (effect size=1.28, 95% CI [1.15, 1.41]). Finally, when all studies are combined taking into account the variation coming from between study designs type (i.e., using a hierarchical Bayesian model), there is a 95% chance that the hazard ratio comparing sulfonylurea vs. TZD second-line therapies lies between 0.33 and 6.95 (Effect size=1.31, 95% Credible Interval [0.33, 6.95]).

Meta-Analysis Summary

A meta-analysis of pre-existing RCTs yielded imprecise estimates of risk categorized by wide confidence intervals that extend above and below the value of one (equal hazards). This translates into 95% confidence intervals that ranged from up to a 8.18 times greater risk for sulfonylurea vs. TZD to a 3.27 times lower risk for sulfonylurea vs. TZD. Since, frequentist analysis assumes all model parameters are fixed, all values in this interval are equally likely. Whereas, evidence using only observational cohort studies is relatively more precise with 95% confidence intervals above one, suggesting higher sulfonylurea risk compared to TZD therapy.

[2] Relative risk estimate for invariance variance fixed effect estimate

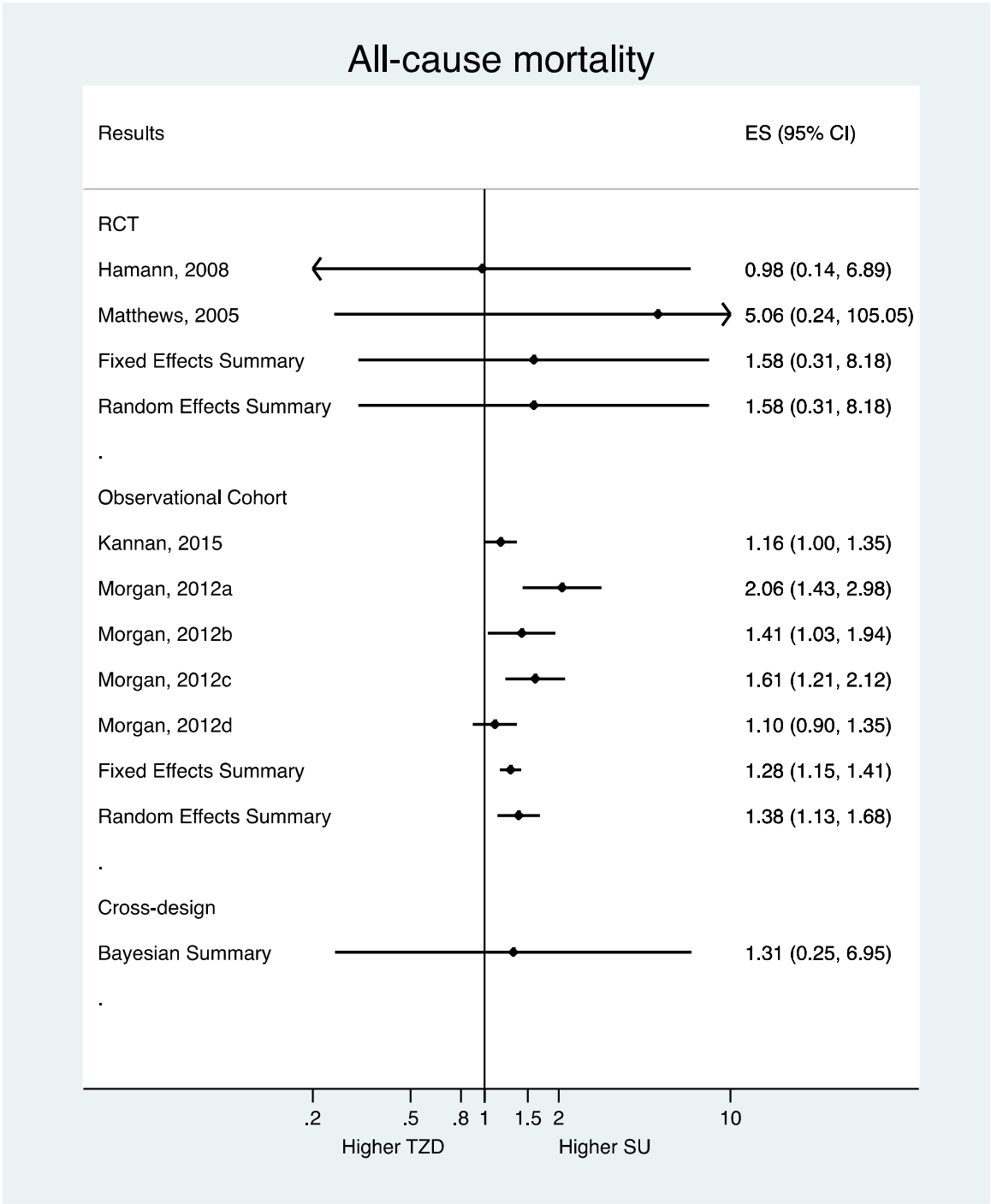


Figure 4.1. Meta-analysis of pre-existing evidence comparing second-line TZD use to sulfonylurea after first-line metformin on all-cause mortality by study design. Values greater than one suggest higher risk for sulfonylurea. Fixed and Random effects models are reported. For overall (cross-design) the results from a two-level Bayesian hierarchical model result are reported (bottom).

Weighting Prior Research Evidence

Figure 4.2 graphs the overall (combined across observational and RCT designs) meta-analysis estimates previously reported at the bottom of Figure 4.1 when the degree of uncertainty is varied around the point estimate using 10%, 50%, and 100% of the prior information. See Appendix C-1 for similar figures for RCT studies only and observational cohorts only. These different weights simulate a potential range of clinical opinions regarding the rigor and relevancy of the prior results, and are a function of varying the standard deviation. For example, if a clinician believes that the prior research was poorly designed with many limitations and is not relevant to their treatment population, then they would be less certain in the previous evidence and more likely to choose a weight of 10% or 0%, such that prior evidence holds little or no influence relative to the new study data. This was done by changing the uncertainty around the estimate. Using 100% of the information means that the variance estimated from the meta-analysis is not changed and is used to define the prior probability distribution (solid blue line). Notice here that the point estimate remains the same, but the variance of the distributions become more diffuse (less certain) as the weight of the prior data is decreased.

The concept of weighting the prior information is perhaps more tractable when presented in the form of quasi a-priori sample sizes. The quasi pre-study sample size reflects the relative contribution that the prior information contributes to the formation of the post-study conclusion relative to the new study. Suppose the quasi a-priori sample size is r, i.e., the information in the current evidence comes from a study of r individuals.

Weighting lets us reflect variation in opinions about the quality and relevance of the current evidence of risk by hypothetically using all *r* individuals or some percent of them (see Appendix C-2 and C-3 for the calculation and further discussion).

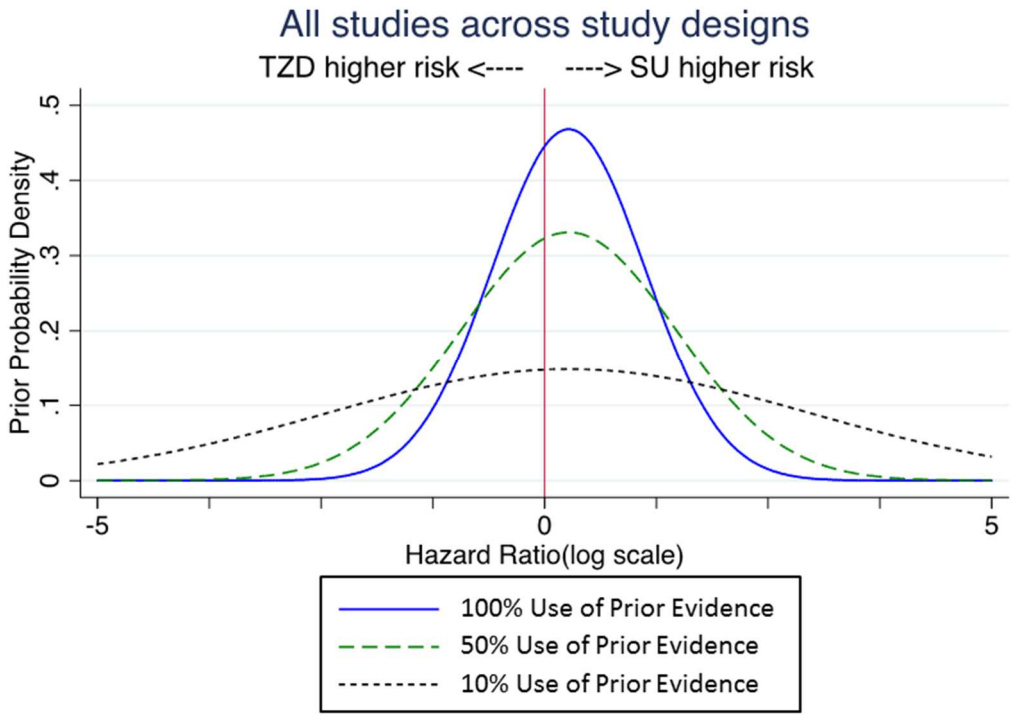


Figure 4.2. Prior probability density distributions plotting the hazard ratio (log scale) comparing sulfonylurea relative to TZD using weights of 100%, 50% and 10% of the earlier existing evidence across all study designs. Earlier data is synthesized and captured by the meta-analysis results.

Forming a New Conclusion After a New Study

Table 4.2 presents the estimated hazard ratio before and after the new study is conducted. These results are stratified by the study design of prior evidence (i.e., evidence across all study design types, RCTs only, observational cohorts only) and the degree of uncertainty in these prior estimates (0%, 10%, 50%, 100%). The quasi pre-study sample size is presented in the second column to illustrate the relative contribution

pre-existing evidence had on the post study conclusion relative to the new VA study (n=145,250). The third column presents the pre-existing estimate (i.e., prior probability distribution). Each value was obtained from the meta-analysis of earlier studies and weighted accordingly. Notice that the rows corresponding to 100% use of the prior evidence (i.e., unweighted) match those estimates presented in Figure 4.1. For the third column, the uncertainty surrounding the estimate becomes wider as the percent of prior evidence used in the analysis is decreased. The last column reports the post-study conclusion (i.e., the posterior probability distributions) after the new study findings are used to update the pre-existing estimate.

Table 4.2. Scenarios evaluating evidence before and after a new study. Note: Results are stratified by study design and amount of prior evidence used; Column 2 is the quasi sample size for the prior evidence, in parenthesis is the relative contribution prior evidence contributes to the post study conclusion; Column 3 is the prior probability distribution; Column 4 is the posterior probability distribution; Values above one suggest higher sulfonylurea risk relative to TZD; New VA study sample size *n*=145,250

Use of prior evidence	**Quasi a-priori sample size (relative contribution)**	**Before study:** HR (95% Credible Intervals)		**After study:** HR (95% Credible Intervals)
No Prior (0% of prior evidence)	0 (0.00%)	Flat prior	→	1.15 (1.09-1.23)
Overall (across study designs)				
100% of prior evidence	193 (0.13%)	1.31 (0.25-6.95)	→	1.15 (1.09-1.23)
50% of prior evidence	96 (0.07%)	1.31 (0.12-13.87)	→	1.15 (1.09-1.23)
10% of prior evidence	19 (0.01%)	1.31 (0.01-256.90)	→	1.15 (1.09-1.23)
By RCTs only				
100% of prior evidence	198 (0.14%)	1.58 (0.31-8.18)	→	1.15 (1.09-1.23)
50% of prior evidence	99 (0.07%)	1.58 (0.16-16.14)	→	1.15 (1.09-1.23)
10% of prior evidence	20 (0.01%)	1.58 (0.01-284.72)	→	1.15 (1.09-1.23)
By Observational Cohort only				
100% of prior evidence	57,295 (28.29%)	1.28 (1.15-1.41)	→	1.19 (1.13-1.25)
50% of prior evidence	27,987 (16.16%)	1.28 (1.10-1.47)	→	1.17 (1.11-1.24)
10% of prior evidence	5,481 (3.64%)	1.28 (0.93-1.75)	→	1.16 (1.09-1.23)

Taken altogether, the after study conclusion reached a similar conclusion across each scenario, even when stratified by prior study design and the extent to which prior evidence was used. This is reflected in the relatively wider uncertainty around the prior point estimates – effectively limiting its overall influence. When translated into sample sizes it is easy to see that the prior information contributed much less information than the new study did in the analysis, even when 100% of the prior data was used.

Without incorporating prior information, the conclusion made through a standard Frequentist analysis is that there was a significant elevated effect between the hazard rate for sulfonylurea compared to TZD (HR=1.15 (95% CI [1.08, 1.23]; $p<0.0001$; not shown in table). Using no external prior information in a Bayesian analysis (i.e., 0% use of the prior data) leads to a similar interpretation, with a posterior probability that sulfonylurea is inferior to TZD on mortality risk and is greater than *probability*>0.99 or that greater than 99% of the posterior distribution for the hazard ratio is above 1.0 (HR=1.15, 95% Credible Interval [1.09, 1.23]).

In this example, the conclusions reached when prior information is used are very similar. When weights of 10%, 50%, and 100% on the prior evidence across all designs (i.e., overall) is used, there remains a greater probability of sulfonylurea having higher mortality risk than TZD with each estimate being the same when rounded to two decimals places, HR=1.15, 95% Credible Interval [1.09, 1.23]. Here, the quasi pre-study sample size ranges from 193 when 100% of the prior data is used to 19 when only 10% is used compared to 145,250 in the new study. Thus, at most, the prior evidence is contributing only 0.13% of the information used to form the post-study conclusion, while

the remaining 99.87% is from the new VA study. Conclusions are again similar when only prior evidence from RCT is taken into account when rounded to two decimal places, HR=1.15, 95% Credible Interval [1.09, 1.23].

Only when prior observational cohort study evidence is used exclusively is there some movement in the effect and the certainty around it. That is because the relative contribution of the prior data (100%) is much larger when using only observational cohort prior evidence. Specifically, 28.29% of the post-study conclusion comes from the prior evidence (compared to 0.14% when only prior RCT data is used only).

As more data from the prior observational evidence is incorporated into the analysis, the hazard ratio becomes larger and the credible intervals smaller. In particular, the lower credible limit increases while the upper limit stays mostly the same as more of the prior evidence is incorporated. When 10% of the prior data is used, there's an estimated 16% increased risk HR=1.16, 95% Credible Interval [1.09, 1.23]. When the prior data is weighted 50%, the hazard ratio is 1.17, 95% Credible Interval [1.11, 1.24]. The effect becomes higher and most precise when using all of the information (100%) from the prior observational cohort evidence — at any point in time, the risk of dying while in the sulfonylurea therapy group is 19% higher than the hazard of TZD treatment group (HR=1.19, 95% Credible Interval [1.13, 1.25)]).

Sensitivity Analysis: Reducing the Precision of the Study Results

Since the new study results were relatively strong in comparison to the weaker prior evidence overall and within RCTs, the new conclusion closely resembles the results of the new study. But what would happen if the new results were not as strong?

A sensitivity analysis was run that increases the uncertainty in the new study estimates, by weighting the new study as if only 10% of the total sample (n=14,525) was collected. Now each person has 1/10 the influence it did in the original VA study analysis, which produces more uncertainty in the VA study estimates through wider credible intervals around the new study hazard ratio. This was implemented by adding the WEIGHT statement to the SAS PHREG procedure (SAS Institute Inc., 2010).

Table 4.3. Sensitivity analysis weighting the new study by 10 percent: Scenarios evaluating evidence before and after a new study. Note: Results are stratified by study design and amount of prior evidence used; Column 2 is the quasi sample size for the prior evidence, in parenthesis is the relative contribution prior evidence contributes to the post study conclusion; Column 3 is the prior probability distribution; Column 4 is the posterior probability distribution; Values above one suggest higher sulfonylurea risk relative to TZD; Note: New VA study sample size *n*=14,525

Use of prior evidence	Quasi a-priori sample size (relative contribution)	Before study: HR (95% Credible Intervals)		After study: HR (95% Credible Intervals)
No Prior (0% of prior evidence)	0 (0.00%)	Flat prior	→	1.17(0.96-1.42)
Overall (across study designs)				
100% of prior evidence	193 (1.31%)	1.31 (0.25-6.95)	→	1.16(0.96-1.41)
50% of prior evidence	96 (0.66%)	1.31 (0.12-13.87)	→	1.17(0.96-1.42)
10% of prior evidence	19 (0.13%)	1.31 (0.01-256.90)	→	1.16(0.96-1.42)
By RCTs only				
100% of prior evidence	198 (1.35%)	1.58 (0.31-8.18)	→	1.17(0.96-1.44)
50% of prior evidence	99 (0.68%)	1.58 (0.16-16.14)	→	1.16(0.96-1.42)
10% of prior evidence	20 (0.14%)	1.58 (0.01-284.72)	→	1.16(0.96-1.42)
By Observational Cohort only				
100% of prior evidence	57,295 (79.78%)	1.28 (1.15-1.41)	→	1.25(1.14-1.37)
50% of prior evidence	27,987 (65.83%)	1.28 (1.10-1.47)	→	1.23(1.10-1.39)
10% of prior evidence	5,481 (27.40%)	1.28 (0.93-1.75)	→	1.19(1.01-1.41)

Before and after study results are presented in Table 4.3. When ignoring prior study results (no prior data incorporated), there was a 17% increased risk of mortality with confidence intervals that now include values less than or equal to one (values suggesting equal or lower risk), HR=1.17, 95% Credible Interval [0.96, 1.42]. In the

Frequentist analysis, the conclusion is that we failed to find a significantly different hazard between sulfonylurea and TZD therapies. From a Bayesian perspective, 93.5% of the after study HR distribution is above the value of 1, suggesting a higher risk for sulfonylurea compared to TZD.

Even though the influence of the new VA study is attenuated, the prior evidence still does not influence the post study estimate much, except when using only the prior observational cohort evidence. When the prior observational evidence is weighted 10%, the after study probability of sulfonylurea mortality is higher than TZD mortality. Here, the prior evidence contributed 27.40% of the total information used to estimate the post-study conclusion. This shorter 95% credible interval no longer includes the value of 1 (as it did when no prior or RCT evidence was used), with 98.1% of the curve being above the HR of 1.0, (HR=1.19, 95% Credible Interval [1.01, 1.41]). When using 50% of the prior observational results (equivalent to a quasi pre-study sample size of 27,987 and a 65.83% relative contribution to the post-study estimate), there is a larger estimated effect (a 23% higher risk) with smaller credible intervals, in particular the lower credible interval moves further above the value of 1 while the upper limit stays fairly similar (probability of sulfonylurea >1=99.9%; HR=1.23, 95% Credible Intervals [1.10, 1.39]). The post-study conclusion is most similar when 100% of the prior data is used with higher risk and slightly smaller credible intervals, (probability of sulfonylurea >1=99.9%; HR= 1.25, 95% Credible Interval [1.14, 1.37], which is closest to the before study estimate since its relative contribution to the post-study conclusion is 79.78%.

Discussion

The ability to assess scientific rigor and determine what represents the best research evidence to clinicians is a key component in evidence-based practice approaches. Yet research methods in alignment with EBM practice that allow for more thoughtful and comprehensive assessments of evidence are lacking in health services research. This chapter reintroduced a research methodological approach that can facilitate EBM models of decision-making by allowing clinicians to more formally express their uncertainty in prior study evidence and account for that when evaluating new results.

In this example, the extent to which prior results influenced the new conclusion was relatively minimal. This should not be seen as a weakness to this approach, rather it serves to highlight the relatively small contribution prior evidence made when forming a new conclusion in this particular example, especially when all prior evidence and within RCT evidence was integrated. At its most influential, the relative information that was contributed by pre-existing evidence was only 0.13% when using all prior data, 0.14% when using prior RCT data only, and 28.29% when using prior observational data. In models using prior observational data only, going from zero percent use to 100% use of the prior evidence translated into a 4% change in hazard ratio (from 1.15 to 1.19) with increasingly smaller credible intervals.

When the new study was weakened in the sensitivity analysis, the prior evidence had a greater influence — prior evidence contributed 1.31%, 1.35%, 79.78% of the total evidence when all, RCT only, and observational cohort evidence only was incorporated,

respectively.

Conversely, the research method helped to underscore the strength of the new study results. This new VA study is one of the largest samples of older adults with diabetes and had several years of follow-up making it one of the biggest of its kind. Consequently, this led each new conclusion to closely resemble the results of the new study. If all clinicians agree the quality of the new study is strong, then clinical opinions should likely be similar post-study even though clinical evaluations of the pre-existing evidence were different. Thus this method has the potential to reduce the variation in treatment and enhance EBM decision-making. Therefore within an EBM framework, clinicians should have a similar base of evidence when individualizing treatment decisions. Since this method allows for a more comprehensive assessment of risk, clinicians can also have better conversations with their patients regarding the uncertainty around the long-term risk of taking sulfonylurea medication relative to other antidiabetic agents.

Study Strengths and Weaknesses.

There are several advantages to the approach taken in this paper. First, this approach provides a framework to help clinicians evaluate evidence and is potentially most useful where clinical opinions differ prior to running a new study. By presenting a range of conclusions that reflect the uncertainty clinicians have, it acknowledges openly that evaluations on the strength and generalizability of prior evidence might not be all the same.

Second, this approach requires clinicians to place a value on their uncertainty of

prior results. By being explicit, it makes assumptions transparent since it assigns a specific numeric value to the uncertainty and allows for a deeper discussion related to why the value was chosen. The explicit nature also offers the opportunity for clinicians to understand and examine other opinions. Then, when evaluating the evidence of a study, conclusions derived from the range of different priors evaluations can be compared, argued for and against, and discussed. By discussing the range of different evaluations, the underlying causes for these differences can be explored: a necessary step to resolving any differences in the future. Thus, this approach can provide a framework for clinicians to debate different conclusions and determine where future research should be focused in order to move the research closer to informing key decisions.

This approach also has the advantages of Bayesian analyses in general, like asking direct questions such as "Which treatment is superior?" or "What is the probability of a clinically meaningful treatment difference?" as well as overcomes methodological limitations of using p-values in hypothesis testing since credible intervals reflects the variability in the unknown parameter (Brophy & Joseph, 1995; Cohen, 1994; Freiman, Chalmers, Smith, & Kuebler, 1978; Kruschke, 2010; Trafimow, 2003; Woodworth, 2004).

This methodological approach also helps to understand the weaknesses in the entire body of evidence related to this hypothesis. In this example, information from RCTs contributes very little to the existing knowledge to this research question. Existing RCTs of oral antidiabetics have been, for the most part, poorly applied to evaluate long-term outcomes (designed instead to evaluate efficacy), often resulting in few events over

the longer term and also providing little evidence on long term outcomes (Bennett, Wilson, et al., 2011; Bolen et al., 2007). There is even fewer studies where RCTs have examined second-line therapies. This lead to imprecise estimates of pooled risk and with relatively little impact on the conclusions reached in this study. Instead, most of the prior evidence has been derived from observational studies, which have larger cohorts and longer follow-up than the RCTs. Indeed, when only observational cohort prior evidence was incorporated it did have an influence on the post study conclusions.

The example also highlights the relative strength of new study results. Here, the new study used rigorous methods for large cohort sizes with long-term follow-up. Confounding was addressed through several strategies that lead to a rigorous observational study design. It involved weighted propensity score models with rich confounder adjustment, strict inclusion criteria coupled with an incident user design, which likely generated a suitable comparison group at similar stages of diabetes (see Chapter 3 for more detail).

Future directions

There are several avenues of future research that can strengthen and refine this methodological approach. In this paper, there are no judgments made on the relative strength that one study design has over another when combining prior study results. In the overall pooled estimate, the only assumption was that the study designs (RCT vs. cohort) are inherently different and therefore the variance existing between study designs should be included. Since well-run RCTs are the gold standard, it's likely that those results should hold more weight than well-run observational designs. The question of

how much is still uncertain.

Future work should explore ways in which to synthesize evidence across different study designs. In particular, Bayesian multilevel models that use informed prior probability distributions that are formally specified to reflect the relative confidence in the accuracy of RCT result compared to observational results may be most beneficial. This would, in effect, assign less weight to designs that are more susceptible to bias (e.g., observational designs) relative to RCT designs. One potential research direction would be to survey research experts to obtain their informed judgments using a Delphi or group consensus approach. Empirically, these weights might also be informed with a meta-regression that examine how the effect varies by study design as has been suggested previously (Goodman, 2013).

Another related assumption made in this paper is that the evidence of the current study is strong and should not be weighted. Therefore, if all clinicians agree the quality of the new study is strong, then it suggests that opinions should likely be similar even though opinions of the earlier evidence are different. Yet this is unlikely, as clinicians might be less confident in the accuracy of results coming from a cohort design since it is not randomized and more susceptible to bias. Conclusions reached can be different if more uncertainty is added to the new VA study estimate.

The sensitivity analysis illustrates that to a certain extent and provided one way to weight new study data and add uncertainty around the study estimate. In that analysis the new study strength was weakened by making each patient worth 10% of what they did in the original sample. This led to less precise estimates of risk derived from the current

study, leading to post study credible intervals that included the value of one when prior evidence derived by RCT and overall was used. Despite weakening the new study, it was still more precise than pooled estimates derived from RCT and overall combined prior evidence.

Conclusion

An underutilized method to the EBM framework that allows for a more thoughtful consideration of the evidence was applied. This framework allows clinicians to question the scientific rigor and relevancy of previous study results when incorporating new study data to form new conclusions.

Results suggest an elevated effect in all models for second-line sulfonylurea compared to TZD, but highlight the need for better-designed studies to evaluate long-term safety outcomes. This paper does not make any assumptions as to the relative worth of one design over another. A critical question still to be explored is how much more is evidence derived from a well-designed RCT worth relative to a well-designed observational study? This will have several implications to this methodology, specifically whether the new study is weighted as well as the extent to which cohort studies contribute to the overall estimate of prior evidence.

CHAPTER 5 : CONCLUDING COMMENTS

A key element in evidence-based medicine (EBM) approaches is the ability for clinicians to evaluate the scientific rigor and relevance of research evidence. Clinicians must make increasingly difficult decisions about which antidiabetic drug regimens to manage patients with limited evidence-based information to guide appropriate therapy. Given the varying clinical opinions coupled with the lack of comparisons evaluating treatment relative to other antidiabetic medications, further research investigating the comparative safety research of therapeutic options and the development of better research methods was needed to support evidence-based decision-making.

Among the most pressing clinical decisions in diabetes treatment are which drugs should be used after metformin is no longer sufficient, and in particular, whether sulfonylureas should remain as a suitable second-line treatment. Sulfonylureas are widely prescribed but some studies suggest an increased long-term safety risks relative to other drug classes. Newer classes of drugs are available but with limited long-term evidence regarding their long-term safety.

In this dissertation, the extent to which sulfonylureas should be included in the clinical pathway of treatment decisions was evaluated in an effort to improve the evidence-base for clinical based decision-making around medication safety. Within each step of treatment, clinicians monitor short and long-term outcomes beyond how well current treatment manages a patient's glucose levels. However evidence to inform decisions on long-term adverse outcomes is limited. This dissertation provided a more comprehensive evaluation into the long-term risks of sulfonylureas relative to other

antidiabetic medications on two key long-term outcomes compared to other diabetes medications through three studies.

As evidence accumulates, it is helpful to summarize the benefits and harms associated with each antidiabetic medication. However, it is often unreasonable to expect clinicians to be able to systematically synthesize and assess the quality of all existing study results on their own. Therefore, meta-analysis is viewed as a critical component of evidence-based medicine (Herman, 2002; Jadad, Haynes, Hunt, & Browman, 2000; Sauerland & Seiler, 2005).

The aim of Study 1 was to summarize pre-existing evidence on the long-term safety risks associated with sulfonylurea use relative to other drug classes in an effort to offer clinicians a more comprehensive picture of sulfonylurea as a therapeutic option to manage diabetes. Using observational and experimental studies with at least one year of follow-up in patients without serious conditions at baseline, a series of meta-analyses were conducted to provide pooled estimates of the risks of cardiovascular events and death.

The results provide some evidence that sulfonylurea treatment is associated with elevated risk relative to other drug classes either when compared alone (as a monotherapy) or when used in combination with metformin. However, these findings are almost entirely derived from observational data, and are not confirmed by smaller, efficacy designed RCTs (with the exception of one analysis). Therefore when evidence is pooled using both types of studies design, there is high variability around effect estimates (wide credible intervals) as a result of the imprecise estimates reported from prior RCT

studies.

Findings from Study 1 also highlight the lack of evidence specifically examining sulfonylurea as a second-line therapy after metformin. It also reveals that even less is known about the influence that augmenting and switching from metformin may have on the risk of mortality and cardiovascular events. Study 2 attempted to address this gap in the existing literature by analyzing the comparative safety of second-line treatment in diabetic patients in the Veterans Health Administration. This study consisted of a large sample with long follow-up and used a rigorous design to address confounding.

Results suggested that second-line use of sulfonylureas was related to higher mortality and cardiovascular risks than TZD therapy, which is consistent with the few existing studies. It also suggests that whether or not metformin is discontinued may lead to differential hazards. In particular, it suggests switching to sulfonylurea instead of adding sulfonylurea to existing metformin may increase the risk of all-cause mortality and cardiovascular events, and that adding sulfonylurea is related to an elevated risk when compared to adding TZD to existing metformin.

While Study 1 intended to summarize prior evidence of sulfonylurea risks and Study 2 attempted to address weaknesses in the current comparative safety literature, clinicians may disagree over the quality of the evidence as well as its relevance to their own treatment population. Improvements in methods for evidence-based medicine that take this into account are needed.

Study 3 addressed this need by focusing on a research method that allows clinicians to vary the uncertainty around pre-existing evidence in a way that reflects a

clinician's appraisal of the quality of the evidence and relevancy to their own treatment population. This facilitates EBM by allowing clinicians to specify the extent to which prior results should be added to new study results when forming new evidence-based conclusions.

In our example, the extent to which prior results influenced the new conclusion was minimal relative to the new VA study data. Each post-study estimate reached a similar conclusion across each scenario, even when stratified by prior study design and prior weight. Here the research method helps to highlight that the prior evidence, especially when combined across designs or within RCTs only, had wide variability and the new study data was relatively stronger given its large sample size and precise study estimates.

Taken altogether, this dissertation provides a more comprehensive evaluation and deeper understanding of the comparative safety of sulfonylurea long-term risks and suggests a research method to aid EBM approaches. It also highlights the need for more evidence, and the need for more refined research methods to help clinician's evaluate evidence, which can serve as the foundation to make complex treatment decisions. Large RCTs with years of follow-up are needed but unlikely to be funded, yet observational studies using secondary databases provide another source of evidence despite methodological flaws and potential for bias. Like Study 2, comparative safety studies should continue to address particular gaps in the literature, specifically the use of medications as second-line agents following metformin and the influence that augmenting and switching off medications has on adverse outcomes.

This dissertation also explored the use of a method intended to support EBM. One of the benefits of this approach appears to be its ability to aid in interpreting study results in a way that yield a richer understanding and a deeper insight into evidence supporting a particular hypothesis. While promising as an EBM research method, there are several aspects its method that need be refined. Future work should focus on the relative strength that observational evidence contributes when compared to RCT evidence, as well as the ways in which evidence can be synthesized across observational cohort and RCT study designs.

Within medication safety, this methodological approach may also be applied to improve evidence synthesis for policy makers in setting health policy, managers making operational decision, and health services researchers in designing focused research to identify the safest treatments to deliver diabetes care.

In addition, this research method offers a potentially useful framework in other health services research areas, yet existing approaches are underutilized and innovative new applications are often left unexplored. For example, patient-centered outcomes research does not have adequate frameworks nor specific measures to assess how new information is incorporated into patient health beliefs. Assessments of how this methodology serves as a basis to understand how and to what extent patients modify their health beliefs when faced with new evidence-based information may prove valuable to patient centered outcomes research initiatives. Greater insight into the patient's belief and how they combine new information into existing belief may lead to enhanced information exchange between clinicians and patients, improved targeting of where

patient-provider communication can be focused, increased patient health self-improvement efforts, and better facilitated shared decision making processes when planning and delivering individualized care.

Thus, this method has the potential to be used in different, yet complementary ways, however, key initial research steps and priorities for developing and evaluating their utility in other HSR domains is still in its infancy. Future work should explore how this method may have broader applications beyond evidence-based decision making in an effort to advance methods in others health services research areas.

APPENDIX A

A- 1: MEDLINE search terms

(diabetes mellitus, type 2[mh] OR (diabet*[tiab] AND (“non-insulin dependent”[tiab] OR type-2[tiab] OR "type II"[tiab] OR "type 2"[tiab])))

AND (sulfonylurea compounds [mh] OR sulfonylurea*[tiab] OR sulphonylurea*[tiab] OR glipizide[tiab] OR glyburide [tiab] OR glimepiride[tiab] OR glibenclamide[tiab])

AND English[lang]

NOT (animal[mh] NOT human[mh])

NOT (letter[pt] OR comment[pt] OR editorial[pt])

AND Humans[mh]

AND ("controlled clinical trial"[pt] OR "randomized controlled trial"[pt] OR "comparative study"[pt] OR "case control studies"[mh] OR "cohort studies"[mh])

A- 2. Meta-analytic results for outcomes. Note: D&L = DerSimonian and Laird random effects model; NA=Not applicable

METF+SU vs. METF

Outcome	Study	Study Effect Size (95% CI)	Bayesian Pooled Estimate (95% CI)	D&L Method (95% CI)	Inverse-Variance Method (95% CI)	I-squared	Egger's Test P-value (95% CI)
All-cause mortality	*Single RCT*						
	Ahrén, 2014	0.99 (0.10–9.38)					
	Obs Cohort						
	Johnson, 2002	0.81 (0.66–0.99)					
	Gulliford, 2004	0.95 (0.64–1.40)					
	Evans, 2006	0.60 (0.35–1.03)					
	Evans, 2006	2.47 (1.88–3.25)					
	Evans, 2006	2.16 (1.68–2.78)					
	Sillars, 2010	1.18 (0.81–1.72)					
	Currie, 2013	1.10 (1.02–1.18)					
	Pooled Estimate for Obs Cohort			1.21 (0.89–1.65)	1.15 (1.08–1.22)	91.81%	NA
	Overall Pooled Estimate		1.12 (0.16–7.59)	1.21 (0.89–1.64)	1.15 (1.08–1.22)	90.44%	NA
Cardio-vascular compo-site	*Obs Cohort*						
	Johnson, 2005	1.17 (0.92–1.48)					
	Sillars, 2010	0.83 (0.60–1.15)					
	Currie, 2013	1.10 (0.98–1.22)					

Outcome	Study	Study Effect Size (95% CI)	Bayesian Pooled Estimate (95% CI)	D&L Method (95% CI)	Inverse-Variance Method (95% CI)	I-squared	Egger's Test P-value (95% CI)
	Li, 2014	1.99 (1.07–3.70)					
	Pooled Estimate for Obs Cohort			1.11 (0.91–1.34)	1.10 (1.00–1.20)	54.86%	NA
Cardio-vascular death	*Obs Cohort*						
	Johnson, 2005	0.78 (0.52–1.15)					
	Evans, 2006	0.62 (0.25–1.53)					
	Evans, 2006	2.29 (1.45–3.61)					
	Evans, 2006	2.43 (1.61–3.66)					
	Sillars, 2010	1.49 (0.85–2.63)					
	Pooled Estimate for Obs Cohort			1.38 (0.80–2.37)	1.47 (1.18–1.82)	82.73%	NA

METF+SU vs. METF+DPP-4

Outcome	Study	Study Effect Size (95% CI)	Bayesian Pooled Estimate (95% CI)	D&L Method (95% CI)	Inverse-Variance Method (95% CI)	I-squared	Egger's Test P-value (95% CI)
All-cause mortality	*RCT*						
	Ferrannini, 2009	1.50 (0.25–8.98)					
	Filozof, 2010	1.03 (0.06–16.49)					
	Matthews, 2010	0.86 (0.29–2.55)					
	Seck, 2010	8.05 (1.01–64.20)					
	Gallwitz, 2012	1.00 (0.25–3.99)					
	Göke, 2013	0.50 (0.09–2.70)					
	Ahrén, 2014	2.95 (0.31–28.22)					
	DelPrato, 2014	1.68 (0.51–5.49)					
	Obs Cohort						
	Mogensen, 2014	1.54 (1.25–1.85)					
	Morgan, 2014a	1.50 (1.09–2.05)					
	Kannan, 2015	0.97 (0.76–1.23)					
	Ou, 2015	1.59 (1.39–1.82)					
	Pooled Estimate for RCT			1.29 (0.74–2.23)	1.29 (0.74–2.23)	0.00%	NA
	Pooled Estimate for Obs Cohort			1.39 (1.12–1.72)	1.45 (1.32–1.59)	76.70%	NA
	Overall Pooled Estimate		1.42 (0.67–2.81)	1.38 (1.15–1.65)	1.44(1.32–1.59)	41.12%	0.697608

Outcome	Study	Study Effect Size (95% CI)	Bayesian Pooled Estimate (95% CI)	D&L Method (95% CI)	Inverse-Variance Method (95% CI)	I-squared	Egger's Test P-value (95% CI)
Cardio-vascular compo-site	*RCT*						
	Filozof, 2010	1.77 (0.70–4.47)					
	Gallwitz, 2012	2.17 (1.10–4.35)					
	DelPrato, 2014	1.58 (0.72–3.47)					
	Obs Cohort						
	Mogensen, 2014	1.43 (1.18–1.75)					
	Morgan, 2014a	1.55 (1.08–2.23)					
	Ou, 2015	1.47 (1.20–1.82)					
	Pooled Estimate for RCT			1.86 (1.18–2.93)	1.86 (1.18–2.93)	0.00%	NA
	Pooled Estimate for Obs Cohort			1.46 (1.28–1.68)	1.46 (1.28–1.68)	0.00%	NA
	Overall Pooled Estimate		1.56 (0.77–3.52)	1.49 (1.31–1.70)	1.49 (1.31–1.70)	0.00%	NA
Cardio-vascular death	*RCT*						
	Ferrannini, 2009	0.50 (0.05–5.52)					
	Göke, 2010	1.99 (0.18–21.87)					
	Seck, 2010	5.03 (0.24–104.63)					
	Gallwitz, 2012	1.00 (0.14–7.14)					

Outcome	Study	Study Effect Size (95% CI)	Bayesian Pooled Estimate (95% CI)	D&L Method (95% CI)	Inverse-Variance Method (95% CI)	I-squared	Egger's Test P-value (95% CI)
	DelPrato, 2014	2.02 (0.51–8.04)					
	Single Obs Cohort						
	Mogensen, 2014	1.75 (1.25–2.50)					
	Pooled Estimate for RCT			1.55 (0.63–3.81)	1.55 (0.63–3.81)	0.00%	NA
	Overall Pooled Estimate		1.70 (0.56–4.81)	1.73 (1.24–2.40)	1.73 (1.24–2.40)	0.00%	NA

METF+SU vs. METF+GLP-1

Outcome	Study	Study Effect Size (95% CI)	Bayesian Pooled Estimate (95% CI)	D&L Method (95% CI)	Inverse-Variance Method (95% CI)	I-squared	Egger's Test P-value (95% CI)
All-cause mortality	*RCT*						
	Gallwitz, 2012b	1.01 (0.29–3.45)					
	Ahrén, 2014	0.98 (0.20–4.84)					
	Obs Cohort						
	Mogensen, 2014	1.30 (0.85–1.96)					
	Kannan, 2015	1.76 (0.93–3.33)					
	Pooled Estimate for RCT			1.00 (0.38–2.65)	1.00 (0.38–2.65)	0.00%	NA
	Pooled Estimate for Obs Cohort			1.42 (1.00–2.01)	1.42 (1.00–2.01)	0.00%	NA
	Overall Pooled Estimate		1.30 (0.37–3.80)	1.36 (0.98–1.89)	1.36 (0.98–1.89)	0.00%	NA

METF+SU vs. METF+MEGL

Outcome	Study	Study Effect Size (95% CI)	Bayesian Pooled Estimate (95% CI)	D&L Method (95% CI)	Inverse-Variance Method (95% CI)	I-squared	Egger's Test P-value (95% CI)
All-cause mortality	*RCT*						
	Gerich, 2005	1.05 (0.07–16.64)					
	Schwarz, 2008	2.63 (0.11–62.66)					
	Pooled Estimate for RCT			1.56 (0.19–12.54)	1.56 (0.19–12.54)	0.00%	NA

METF+SU vs. METF+SGLT-2

Outcome	Study	Study Effect Size (95% CI)	Bayes-ian Pooled Esti-mate (95% CI)	D&L Method (95% CI)	Inverse-Variance Method (95% CI)	I-squared	Egger's Test P-value (95% CI)
All-cause mortality	*RCT*						
	DelPrato, 2015	2.49 (0.49–12.75)					
	Leiter, 2015	0.67 (0.14–3.30)					
	Leiter, 2015	6.02 (0.25–147.47)					
	NCT01167881, 2015	0.99 (0.25–3.93)					
	Pooled Estimate for RCT			1.29 (0.55–3.01)	1.29 (0.55–3.01)	0.00%	NA

METF+SU vs. METF+TZD

Outcome	Study	Study Effect Size (95% CI)	Bayesian Pooled Estimate (95% CI)	D&L Method (95% CI)	Inverse-Variance Method (95% CI)	I-squared	Egger's Test P-value (95% CI)
All-cause mortality	*RCT*						
	Matthews, 2005	5.06 (0.24–105.05)					
	Hamann, 2008	0.98 (0.14–6.89)					
	Obs Cohort						
	Morgan, 2012	1.41 (1.03–1.94)					
	Morgan, 2012	1.10 (0.90–1.35)					
	Prentice, 2014	1.50 (1.09–2.09)					
	Kannan, 2015	1.16 (1.00–1.35)					
	Pooled Estimate for RCT			1.58 (0.31–8.18)	1.58 (0.31–8.18)	0.00%	NA
	Pooled Estimate for Obs Cohort			1.21 (1.07–1.37)	1.20 (1.08–1.34)	17.24%	NA
	Overall Pooled Estimate		1.22 (0.30–6.50)	1.20 (1.08–1.34)	1.20 (1.08–1.34)	0.00%	NA
Cardio-vascular compo-site	*Obs Cohort*						
	McAfee, 2007	1.27 (0.75–2.13)					
	Walker, 2008	1.12 (0.96–1.32)					
	Walker, 2008	1.08 (0.95–1.22)					

Outcome	Study	Study Effect Size (95% CI)	Bayesian Pooled Estimate (95% CI)	D&L Method (95% CI)	Inverse-Variance Method (95% CI)	I-squared	Egger's Test P-value (95% CI)
	Morgan, 2012	1.38 (0.93–2.06)					
	Morgan, 2012	1.19 (0.91–1.55)					
	Pooled Estimate for Obs Cohort			1.12 (1.03–1.23)	1.12 (1.03–1.23)	0.00%	NA

METF+SU vs. SU

Outcome	Study	Study Effect Size (95% CI)	Bayesian Pooled Estimate (95% CI)	D&L Method (95% CI)	Inverse-Variance Method (95% CI)	I-squared	Egger's Test P-value (95% CI)
All-cause mortality	*Single RCT*						
	UKPDS, 1998	1.60 (1.02–2.52)					
	Obs Cohort						
	Olsson, 2000	1.63 (1.27–2.09)					
	Johnson, 2002	0.63 (0.57–0.71)					
	Gulliford, 2004	1.06 (0.85–1.31)					
	Evans, 2006	0.42 (0.23–0.75)					
	Evans, 2006	1.73 (1.22–2.45)					
	Evans, 2006	1.51 (1.09–2.10)					
	Sillars, 2010	1.02 (0.76–1.37)					
	Morgan, 2012	0.69 (0.57–0.83)					
	Currie, 2013	0.63 (0.57–0.69)					
	Pooled Estimate for Obs Cohort			0.94 (0.73–1.22)	0.76 (0.71–0.80)	93.15%	NA
	Overall Pooled Estimate		1.06 (0.02–50.38)	0.98 (0.76–1.27)	0.77 (0.72–0.81)	92.92%	2.412320
Cardio-vascular compo-site	*Obs Cohort*						
	Johnson, 2005	0.96 (0.82–1.12)					

Outcome	Study	Study Effect Size (95% CI)	Bayesian Pooled Estimate (95% CI)	D&L Method (95% CI)	Inverse-Variance Method (95% CI)	I-squared	Egger's Test P-value (95% CI)
	Sillars, 2010	0.87 (0.66–1.15)					
	Morgan, 2012	0.63 (0.48–0.84)					
	Currie, 2013	0.79 (0.68–0.92)					
	Li, 2014	1.79 (0.68–4.74)					
	Pooled Estimate for Obs Cohort			0.84 (0.71–0.99)	0.84 (0.77–0.93)	58.91%	NA
Cardio-vascular death	*Obs Cohort*						
	Johnson, 2005	0.59 (0.45–0.78)					
	Evans, 2006	1.43 (0.83–2.47)					
	Evans, 2006	0.36 (0.14–0.97)					
	Evans, 2006	1.35 (0.75–2.41)					
	Sillars, 2010	0.98 (0.66–1.45)					
	Pooled Estimate for Obs Cohort			0.88 (0.57–1.35)	0.80 (0.66–0.97)	74.13%	NA

METF+SU vs. SU+TZD

Outcome	Study	Study Effect Size (95% CI)	Bayesian Pooled Estimate (95% CI)	D&L Method (95% CI)	Inverse-Variance Method (95% CI)	I-squared	Egger's Test P-value (95% CI)
Cardio-vascular composite	*Obs Cohort*						
	Walker, 2008	0.85 (0.72–1.01)					
	Walker, 2008	0.99 (0.83–1.18)					
	Pooled Estimate for Obs Cohort			0.92 (0.79–1.07)	0.92 (0.81–1.04)	30.98%	NA

SU vs. DDP-4

Outcome	Study	Study Effect Size (95% CI)	Bayesian Pooled Estimate (95% CI)	D&L Method (95% CI)	Inverse-Variance Method (95% CI)	I-squared	Egger's Test P-value (95% CI)
Cardio-vascular compo-site	Rosenstock, 2013	2.03 (0.19–22.20)					
	Single Obs Cohort						
	Berkowitz, 2014	1.17 (0.97–1.42)					
	Overall Pooled Estimate		1.25 (0.16–10.53)	1.18 (0.97–1.42)	1.18 (0.97–1.42)	0.00%	NA

SU vs. METF

Outcome	Study	Study Effect Size (95% CI)	Bayesian Pooled Estimate (95% CI)	D&L Method (95% CI)	Inverse-Variance Method (95% CI)	I-squared	Egger's Test P-value (95% CI)
All-cause mortality	*Single RCT*						
	Kahn, 2006	1.01 (0.62–1.65)					
	Obs Cohort						
	Johnson, 2002	1.28 (1.09–1.54)					
	Evans, 2006	1.43 (1.15–1.77)					
	Pantalone, 2009	1.85 (1.56–2.17)					
	Tzoulaki, 2009	1.24 (1.14–1.35)					
	Corrao, 2011	1.37 (1.26–1.49)					
	Schramm, 2011	1.05 (0.94–1.16)					
	Schramm, 2011	1.32 (1.24–1.40)					
	Schramm, 2011	1.27 (1.17–1.38)					
	Schramm, 2011	1.19 (1.11–1.28)					
	Sullivan, 2011	1.09 (0.83–1.42)					
	Currie, 2013	1.75 (1.64–1.86)					
	Wheeler, 2013	1.38 (1.27–1.50)					
	Wheeler, 2013	1.55 (1.43–1.67)					
	Morgan, 2014b	1.50 (1.37–1.65)					
	Pooled Estimate for Obs Cohort			1.37 (1.26–1.48)	1.38 (1.35–1.41)	90.73%	0.671953

Outcome	Study	Study Effect Size (95% CI)	Bayesian Pooled Estimate (95% CI)	D&L Method (95% CI)	Inverse-Variance Method (95% CI)	I-squared	Egger's Test P-value (95% CI)
	Overall Pooled Estimate		1.30 (0.46–3.05)	1.36 (1.25–1.47)	1.38 (1.35–1.41)	90.12%	0.567854
Cardio-vascular compo-site	*Single RCT*						
	Kahn, 2006	0.71 (0.44–1.16)					
	Obs Cohort						
	Johnson, 2005	1.22 (1.02–1.47)					
	McAfee, 2007	1.30 (1.04–1.61)					
	Pantalone, 2009	1.06 (0.95–1.18)					
	Schramm, 2011	1.06 (0.95–1.18)					
	Schramm, 2011	1.21 (1.14–1.29)					
	Schramm, 2011	1.17 (1.07–1.28)					
	Schramm, 2011	1.12 (1.04–1.21)					
	Sullivan, 2011	0.97 (0.79–1.18)					
	Currie, 2013	1.39 (1.25–1.55)					
	Hung, 2013	1.68 (1.15–2.45)					
	Hung, 2013	3.23 (2.50–4.17)					
	Berkowitz, 2014	1.16 (1.04–1.29)					
	Li, 2014	1.11 (0.52–2.35)					

Outcome	Study	Study Effect Size (95% CI)	Bayesian Pooled Estimate (95% CI)	D&L Method (95% CI)	Inverse-Variance Method (95% CI)	I-squared	Egger's Test P-value (95% CI)
	Morgan, 2014b	1.07 (0.88–1.30)					
	Pooled Estimate for Obs Cohort			1.24 (1.13–1.36)	1.18 (1.15–1.22)	85.18%	0.294266
	Overall Pooled Estimate		1.02 (0.13–6.12)	1.22 (1.11–1.34)	1.18 (1.15–1.22)	84.76%	0.517152
Cardio-vascular death	*Single RCT*						
	Kahn, 2006	1.52 (0.25–9.06)					
	Obs Cohort						
	Johnson, 2005	1.32 (1.00–1.72)					
	Evans, 2006	1.70 (1.18–2.45)					
	Schramm, 2011	1.05 (0.91–1.20)					
	Schramm, 2011	1.28 (1.18–1.38)					
	Schramm, 2011	1.25 (1.12–1.40)					
	Schramm, 2011	1.14 (1.03–1.25)					
	Sullivan, 2011	1.17 (0.72–1.91)					
	Pooled Estimate for Obs Cohort			1.21 (1.12–1.31)	1.21 (1.16–1.27)	50.63%	NA
	Overall Pooled Estimate		1.25 (0.29–5.67)	1.21 (1.12–1.31)	1.21 (1.16–1.27)	42.68%	NA

SU vs. METF+TZD

Outcome	Study	Study Effect Size (95% CI)	Bayesian Pooled Estimate (95% CI)	D&L Method (95% CI)	Inverse-Variance Method (95% CI)	I-squared	Egger's Test P-value (95% CI)
All-cause mortality	*Obs Cohort*						
	Morgan, 2012	2.06 (1.43–2.98)					
	Morgan, 2012	1.61 (1.21–2.12)					
	Pooled Estimate for Obs Cohort			1.77 (1.39–2.24)	1.76 (1.41–2.20)	11.90%	NA
Cardio-vascular compo-site	*Obs Cohort*						
	Morgan, 2012	1.87 (1.27–2.76)					
	Morgan, 2012	2.18 (1.34–3.56)					
	Pooled Estimate for Obs Cohort			1.99 (1.47–2.70)	1.99 (1.47–2.70)	0.00%	NA

SU vs. TZD

Outcome	Study	Study Effect Size (95% CI)	Bayesian Pooled Estimate (95% CI)	D&L Method (95% CI)	Inverse-Variance Method (95% CI)	I-squared	Egger's Test P-value (95% CI)
All-cause mortality	*RCT*						
	Tan, 2004	3.59 (0.15–87.00)					
	Jain, 2006	5.00 (0.24–103.62)					
	Kahn, 2006	0.92 (0.57–1.49)					
	Obs Cohort						
	Pantalone, 2009	1.69 (1.23–2.33)					
	Pantalone, 2009	1.37 (0.98–1.96)					
	Tzoulaki, 2009	1.16 (0.82–1.63)					
	Wheeler, 2013	1.13 (0.90–1.42)					
	Wheeler, 2013	1.27 (1.01–1.59)					
	Pooled Estimate for RCT			0.99 (0.62–1.58)	0.99 (0.62–1.58)	0.00%	NA
	Pooled Estimate for Obs Cohort			1.28 (1.12–1.47)	1.28 (1.13–1.45)	13.73%	NA
	Overall Pooled Estimate		1.20 (0.46–2.60)	1.26 (1.11–1.43)	1.26 (1.12–1.42)	6.97%	NA
Cardio-vascular compo-site	*RCT*						
	St. John Sutton, 2002	0.58 (0.20–1.69)					

Outcome	Study	Study Effect Size (95% CI)	Bayesian Pooled Estimate (95% CI)	D&L Method (95% CI)	Inverse-Variance Method (95% CI)	I-squared	Egger's Test P-value (95% CI)
	Jain, 2006	2.00 (1.01–3.99)					
	Kahn, 2006	0.67 (0.41–1.08)					
	Perriello, 2006	0.53 (0.14–2.09)					
	Obs Cohort						
	McAfee, 2007	1.22 (0.98–1.61)					
	Walker, 2008	1.10 (0.94–1.29)					
	Walker, 2008	1.12 (0.94–1.34)					
	Hsiao, 2009	0.97 (0.61–1.54)					
	Hsiao, 2009	0.65 (0.54–0.78)					
	Pantalone, 2009	0.96 (0.79–1.16)					
	Pantalone, 2009	1.11 (0.88–1.41)					
	Berkowitz, 2014	1.10 (0.89–1.37)					
	Pooled Estimate for RCT			0.87 (0.45–1.67)	0.87 (0.61–1.24)	61.39%	NA
	Pooled Estimate for Obs Cohort			1.01 (0.87–1.18)	1.00 (0.93–1.07)	75.11%	NA
	Overall Pooled Estimate		0.98 (0.53–1.70)	0.99 (0.85–1.16)	0.99 (0.92–1.07)	69.83%	0.871221
Cardio-vascular death	*RCT*						

Outcome	Study	Study Effect Size (95% CI)	Bayesian Pooled Estimate (95% CI)	D&L Method (95% CI)	Inverse-Variance Method (95% CI)	I-squared	Egger's Test P-value (95% CI)
	Tan, 2004	3.59 (0.15–87.00)					
	Jain, 2006	3.00 (0.12–73.29)					
	Kahn, 2006	1.52 (0.25–9.06)					
	Pooled Estimate for RCT			2.04 (0.50–8.29)	2.04 (0.50–8.29)	0.00%	NA

A- 3. Newcastle-Ottawa Scale for assessing the quality of Observational cohort studies and an additional item on industry funding

Study	Representativeness of the Exposed Cohort	Selection of the Non-Exposed Cohort	Ascertainment of Exposure	Outcome Was Not Present at Study Start	Comparability of cohorts on the basis of the design or analysis	Assessment of Outcome	Was Follow-Up Long Enough for Outcomes to Occur?	Adequacy of Follow -up of Cohorts	Total	Industry Funded
Berkowitz, 2014	**b**	**a**	**a**	**a**	**a,b**	**b**	**a**	d	7	Yes
Corrao, 2011	**b**	**a**	**a**	**a**	**a**	**b**	**a**	**a**	8	No
Currie, 2013	**a**	**a**	**a**	**a**	**a,b**	**b**	**a**	**b**	9	No
Evans, 2006	**a**	**a**	**a**	**a**	**a,b**	**b**	**a**	**b**	9	No
Gulliford, 2004	**a**	**a**	**a**	**a**	**a,b**	**b**	**a**	**b**	9	No
Hsiao, 2009	**a**	**a**	**a**	**a**	**a,b**	**b**	**a**	**a**	9	No
Hung, 2013	**b**	**a**	**a**	**a**	**a,b**	**b**	**a**	**a**	9	No
Johannes, 2007	**b**	**a**	**a**	**a**	**a,b**	**b**	**a**	**b**	9	Yes
Johnson, 2002	**a**	**a**	**a**	**a**	**a,b**	**b**	**a**	**b**	9	No
Johnson, 2005	**a**	**a**	**a**	**a**	**a,b**	**b**	**a**	**a**	9	No
Kannan, 2015	**b**	**a**	**a**	**a**	**a,b**	**b**	**a**	**a**	9	Unclear
Li, 2014	**b**	**a**	**b**	**a**	**a,b**	**b**	**a**	**b**	9	Yes
McAfee, 2007	**b**	**a**	**a**	**a**	**a,b**	**b**	**a**	**b**	9	Yes
Mogensen, 2014	**b**	**a**	**a**	**a**	**a,b**	**b**	**a**	**a**	9	Unclear
Morgan, 2012	**a**	**a**	**a**	**a**	**a,b**	**b**	**a**	**b**	9	Yes
Morgan, 2014a	**a**	**a**	**a**	**a**	**a,b**	**b**	**a**	**b**	9	Yes
Morgan, 2014b	**a**	**a**	**a**	**a**	**a,b**	**b**	**a**	**b**	9	Yes
Olsson, 2000	**b**	**a**	**a**	**a**	**a,b**	**b**	**a**	**a**	9	No
Ou, 2015	**a**	**a**	**a**	**a**	**a,b**	**b**	**a**	**a**	9	No
Pantalone, 2009	**b**	**a**	**a**	**a**	**a,b**	**b**	**a**	**a**	9	Yes
Prentice, 2014	**b**	**a**	**a**	**a**	**a,b**	**b**	**a**	**a**	9	No
Schramm, 2011	**a**	**a**	**a**	**a**	**a,b**	**b**	**a**	**a**	9	No
Sillars, 2010	**b**	**a**	**a**	**a**	**a,b**	**b**	**a**	**a**	9	No
Sullivan, 2011	**b**	**a**	**b**	**a**	**a,b**	**a**	**a**	**b**	9	Yes
Tzoulaki, 2009	**a**	**a**	**a**	**a**	**a,b**	**b**	**a**	**a**	9	No
Walker, 2008	**b**	**a**	**a**	**a**	**a,b**	**b**	**a**	**a**	9	Yes

A- 4. Items from Jadad Scale for randomized trials and an additional item on industry funding

Study	Randomization	Double-blinded	Dropouts described	Industry funded
Ahrén, 2014	Yes	Yes	Yes	Yes
Del Prato, 2014	Yes	Yes	Yes	Yes
Del Prato, 2015	Yes	Yes	No	Yes
Ferrannini, 2009	Yes	Yes	Yes	Yes
Filozof, 2010	Yes	Yes	Yes	Yes
Gallwitz, 2012	Yes	Yes	Yes	Yes
Gallwitz, 2012b	Yes	No	Yes	Yes
Gerich, 2005	Yes	Yes	Yes	Yes
Göke, 2010	Yes	Yes	Yes	Yes
Göke, 2013	Yes	Yes	Yes	Yes
Hamann, 2008	Yes	Yes	Yes	Yes
Jain, 2006	Yes	Yes	Yes	Unclear
Kahn, 2006	Yes	Yes	Yes	Yes
Leiter, 2015	Yes	Yes	Yes	Yes
Matthews, 2005	Yes	Yes	Yes	Yes
Matthews, 2010	Yes	Yes	Yes	Yes
NCT01167881, 2015	Yes	Yes	Yes	Yes
Perriello, 2006	Yes	Yes	Yes	Yes
Rosenstock, 2006	Yes	Yes	Yes	Yes
Rosenstock, 2013	Yes	Yes	Yes	Yes
Schwarz, 2008	Yes	Yes	Yes	Yes
Seck, 2010	Yes	Yes	Yes	Yes
St. John Sutton, 2002	Yes	No	Yes	Yes
Tan, 2004	Yes	No	Yes	Yes
UKPDS, 1998	Yes	No	Yes	Yes

APPENDIX B

B- 1. Selected covariate means and frequency at baseline by drug therapy

	Sulfonylurea after Metformin (n=138,097)	TZD after Metformin (n=7,153)
Sampling Factors		
Index date	10/21/06	10/7/05
Augmented to first-line metformin	74.72%	68.46%
1st year of diabetes	2002.97 (4.29)	2002.01 (2.91)
Time (days) on metformin before augment/switch to 2nd line therapy	759.02 (766.19)	798.73 (666.05)
Diabetes duration	3.13 (2.55)	3.04 (2.28)
Demographic and Social Factors (before index date)		
Age	62.44 (10.67)	64.18 (10.81)
Sex	96.37%	95.67%
Race		
White, non-hispanic	75.79	77.87
African American, non-Hispanic	13.51	8.74
Hispanic, white	4.46	6.79
Other	6.24	6.6
Marital status		
Widow	5.8	5.2
Single or divorced	32.96	23.82
Married	61.24	70.98
Income	$33,271 (58,215)	$42,863 (74,440)
income, unknown	5.96%	9.44%
Utilization and access to care (before index date)		
Medicare beneficiary at any time prior to index date	13.78%	27.15%
In Medicare denominator in year of index date or in prior year	20.92%	26.72%
Medicare part C beneficiary any time in 2 years before index date	3.01%	5.10%
# of outpatient visit days in VA in 6 months before index date	8.24 (9.06)	6.83 (7.34)
# of inpatient days in VA in 6 months before index date	0.09 (0.38)	0.05 (0.28)
# of urgent or emergency room visit days in VA in 6 months before index date	0.32 (0.92)	0.19 (0.62)
# of outpatient visit days in Medicare 6 months before index date	1.1 (4.19)	2.72 (6.58)
# of inpatient days in Medicare in 6 months before index date	0.04 (0.34)	0.08 (0.49)
Laboratory and Physical Exam Measures (before index date)		
HbA1c, most recent in 12 months before index date	7.95 (1.72)	7.36 (1.43)
HbA1c, unknown - none in 12 months before index date	12.13%	18.80%
Glucose, most recent in 12 months before index date	177.15 (78.37)	155.15 (58.1)
Glucose, unknown - none in 12 months before index date	19.44%	14.79%
LDL, most recent in 12 months before index date	95 (35.55)	94.25 (34.49)
LDL, unknown - none in 12 months before index date	22.45%	26.94%
Total serum cholesterol , most recent in 12 months before index date	175.59 (45.78)	175.35 (45.03)
Total serum cholesterol , unknown - none in 12 months before index date	14.35%	18.37%
eGFR, using most recent creatinine in 12 months before index date	80.97 (23.7)	77.32 (22.63)
eGFR unknown - none in 12 months before index date	13.15%	14.41%
Microalbumin/creatinine ratio, most recent in 12 months before index date	135.56 (1892.74)	72.89 (1075.49)
Microalbumin/creatinine ratio, unknown - none in 12 months before index date	74.67%	76.09%
Alanine amino transferase, most recent in 12 months before index date	36.13 (28.96)	29.3 (17.25)
Alanine amino transferase, unknown - none in 12 months before index date	25.46%	22.59%

	Sulfonylurea after Metformin (n=138,097)	TZD after Metformin (n=7,153)
Medications (before index date)		
Antihypertensives	82.96%	83.95%
Cardiovascular medications	89.41%	91.88%
Statins	60.93%	64.46%
Other lipid lowering medications	66.68%	70.96%
NSAIDs	20.63%	20.73%
NSAIDs cox-2 inhibitors	0.39%	0.96%
Platelet aggregation inhibitor	2.51%	0.70%
Anticoagulant prescriptions	2.39%	0.31%
Antidepressants	26.20%	26.07%
Comorbidities (before index date)		
Retinopathy	8.79%	7.93%
Nephropathy	4.60%	6.12%
Neuropathy	16.02%	17.96%
Cerebrovascular disease	6.32%	7.84%
Cardiovascular disease	33.16%	37.29%
Peripheral vascular disease	0.24%	0.35%
Metabolic syndrome	0.47%	0.36%
Macrovascular disease	91.27%	91.64%
Other heart disease	12.58%	14.71%
Microvascular disease	28.66%	30.00%
Mental health	29.05%	24.88%
Substance abuse	23.40%	21.05%
Neurological disorders	2.70%	2.82%

	SU monotherapy after 1st line Metformin (n=34916)	METF+SU after 1st line Metformin (n=103,181)	TZD monotherapy after 1st line Metformin (n=34916)	METF+TZD after 1st line Metformin (n=4,897)
Sampling Factors				
Index date	6/16/07	8/1/06	10/18/05	10/2/05
1st year of diabetes	2003.50 (3.90)	2002.79 (4.41)	2002.06 (3.06)	2001.98 (2.84)
Time (days) on metformin before augment/switch to 2nd line therapy	808.90 (806.53)	742.14 (751.31)	767.97 (694.71)	812.90 (652.00)
Diabetes duration	3.21 (2.77)	3.10 (2.47)	3.02 (2.37)	3.04 (2.24)
Demographic and Social Factors (before index date)				
Age	65.97 (11.25)	61.25 (10.19)	67.32 (11.05)	62.73 (10.38)
Sex	96.06%	96.48%	94.68%	96.12%
Race				
White, non-hispanic	76.99	75.38	78.28	77.68
African American, non-Hispanic	14.05	13.33	9.53	8.37
Hispanic, white	3.91	4.65	6.69	6.84
Other	5.05	6.64	5.50	7.11
Marital status				
Widow	7.86	5.10	6.78	4.47
Single or divorced	30.36	33.83	23.54	23.95
Married	61.78	61.06	69.68	71.57
Income	$36,641 ($64,183)	$32,131 ($56,007)	$43,046 ($75,433)	$42,779 ($73,986)
income, unknown	6.26%	5.85%	9.62%	9.35%
Utilization and Access to Care (before index date)				
Medicare beneficiary at any time prior to index date	16.20%	12.97%	31.60%	25.10%
In Medicare denominator in year of index date or in prior year	29.38%	18.06%	35.28%	22.77%
Medicare part C beneficiary any time in 2 years before index date	4.23%	2.60%	7.05%	4.21%
# of outpatient visit days in VA in 6 months before index date	8.63 (9.17)	8.10 (9.02)	7.20 (6.82)	6.66 (7.56)
# of inpatient days in VA in 6 months before index date	0.12 (0.48)	0.08 (0.34)	0.06 (0.31)	0.05 (0.26)
# of urgent or emergency room visit days in VA in 6 months before index	0.35 (0.93)	0.31 (0.91)	0.20 (0.62)	0.19 (0.62)
# of outpatient visit days in Medicare 6 months before index date	1.55 (5.31)	0.95 (3.72)	3.43 (7.30)	2.39 (6.20)
# of inpatient days in Medicare in 6 months before index date	0.08 (0.49)	0.03 (0.27)	0.11 (0.54)	0.07 (0.46)
Laboratory and Physical Exam Measures (before index date)				
HbA1c, most recent in 12 months before index date	7.20 (1.40)	8.20 (1.75)	6.89 (1.13)	7.57 (1.51)
HbA1c, unknown - none in 12 months before index date	11.18%	12.45%	19.06%	18.68%
glucose, most recent in 12 months before index date	152.05 (61.39)	186.72 (81.94)	141.45 (47.11)	161.61 (61.57)
glucose, unknown - none in 12 months before index date	12.10%	21.92%	13.48%	15.40%
LDL, most recent in 12 months before index date	94.35 (35.04)	95.23 (35.73)	96.52 (35.43)	93.19 (33.99)
LDL, unknown - none in 12 months before index date	19.66%	23.40%	26.60%	27.10%
Total serum cholesterol , most recent in 12 months before index date	172.07 (43.60)	176.81 (46.45)	176.00 (43.26)	175.05 (45.84)
Total serum cholesterol , unknown - none in 12 months before index	13.04%	14.79%	17.69%	18.68%
eGFR, using most recent creatinine in 12 months before index date	68.66 (24.67)	85.24 (21.79)	66.10 (23.20)	82.59 (20.32)
eGFR unknown - none in 12 months before index date	11.59%	13.67%	13.30%	14.93%
microalbumin/creatinine ratio, most recent in 12 months before index	151.50 (1741.78)	130.21 (1940.71)	110.32 (1382.90)	56.71 (911.35)
microalbumin/creatinine ratio, unknown - none in 12 months before	74.85%	74.60%	77.13%	75.62%
alanine amino transferase, most recent in 12 months before index date	33.69 (35.66)	37.06 (25.89)	27.70 (18.40)	30.04 (16.64)
alanine amino transferase, unknown - none in 12 months before index	18.54%	27.80%	22.65%	22.56%

	SU monotherapy after 1st line Metformin (n=34916)	METF+SU after 1st line Metformin (n=103,181)	TZD monotherapy after 1st line Metformin (n=34916)	METF+TZD after 1st line Metformin (n=4,897)
Medications (before index date)				
Antihypertensives	87.20%	81.52%	85.28%	83.34%
Cardiovascular medications	92.03%	88.53%	92.07%	91.79%
Statins	63.10%	60.19%	62.37%	65.43%
Other lipid lowering medications	68.28%	66.14%	69.02%	71.86%
NSAIDs	20.79%	20.58%	20.61%	20.79%
NSAIDs cox-2 inhibitors	0.55%	0.34%	1.11%	0.90%
Platelet aggregation inhibitor	3.58%	2.15%	0.71%	0.69%
Anticoagulant prescriptions	3.67%	1.96%	0.35%	0.29%
Antidepressants	27.31%	25.82%	25.35%	26.40%
Comorbidities (before index date)				
Retinopathy	7.52%	9.22%	8.60%	7.62%
Nephropathy	9.13%	3.06%	11.79%	3.51%
Neuropathy	16.13%	15.98%	20.21%	16.93%
Cerebrovascular disease	8.32%	5.64%	10.33%	6.70%
Cardiovascular disease	39.79%	30.92%	42.55%	34.86%
Peripheral vascular disease	0.24%	0.24%	0.40%	0.33%
Metabolic syndrome	0.32%	0.52%	0.09%	0.49%
Macrovascular disease	93.56%	90.49%	93.97%	90.57%
Other heart disease	17.22%	11.00%	19.99%	12.27%
Microvascular disease	29.99%	28.21%	35.06%	27.67%
Mental health	28.03%	29.39%	24.96%	24.85%
Substance abuse	22.89%	23.57%	21.50%	20.85%
Neurological disorders	3.00%	2.60%	3.86%	2.35%

B- 2 Standardized mean differences from 176 baseline characteristics between patients on sulfonylurea vs. TZD second-line therapy, before and after inverse probability of treatment weighting (sorted by weighted standardized mean difference)

Figure **B- 2** describes the similarities in baseline characteristics between cohorts before and after weighting. Applying propensity score via IPTW improved balance in the observed covariates. Before weighting, 4 of 176 covariates had a SMD above 25% (29 covariates were >10%) suggesting some covariates are not well balanced between sulfonylurea and TZDS cohort. The largest imbalances were for the proportion of Medicare beneficiaries (27% TZD vs. 14% sulfonylurea), the number of Medicare outpatient visits in the last 6 months (2.7 TZDs vs. 1.1 sulfonylurea), and A1C level (7.39 TZD vs. 7.95 sulfonylurea). After weighting, zero covariates were above the recommended guideline of 25% (7 were above 10%) suggesting cohorts were fairly balanced. However, since there may be some imbalance (i.e., SMDs above 10% but below 25%) in some of the covariates after weighting, final models were also adjusted to estimate the effect of sulfonylurea relative to TZDs.

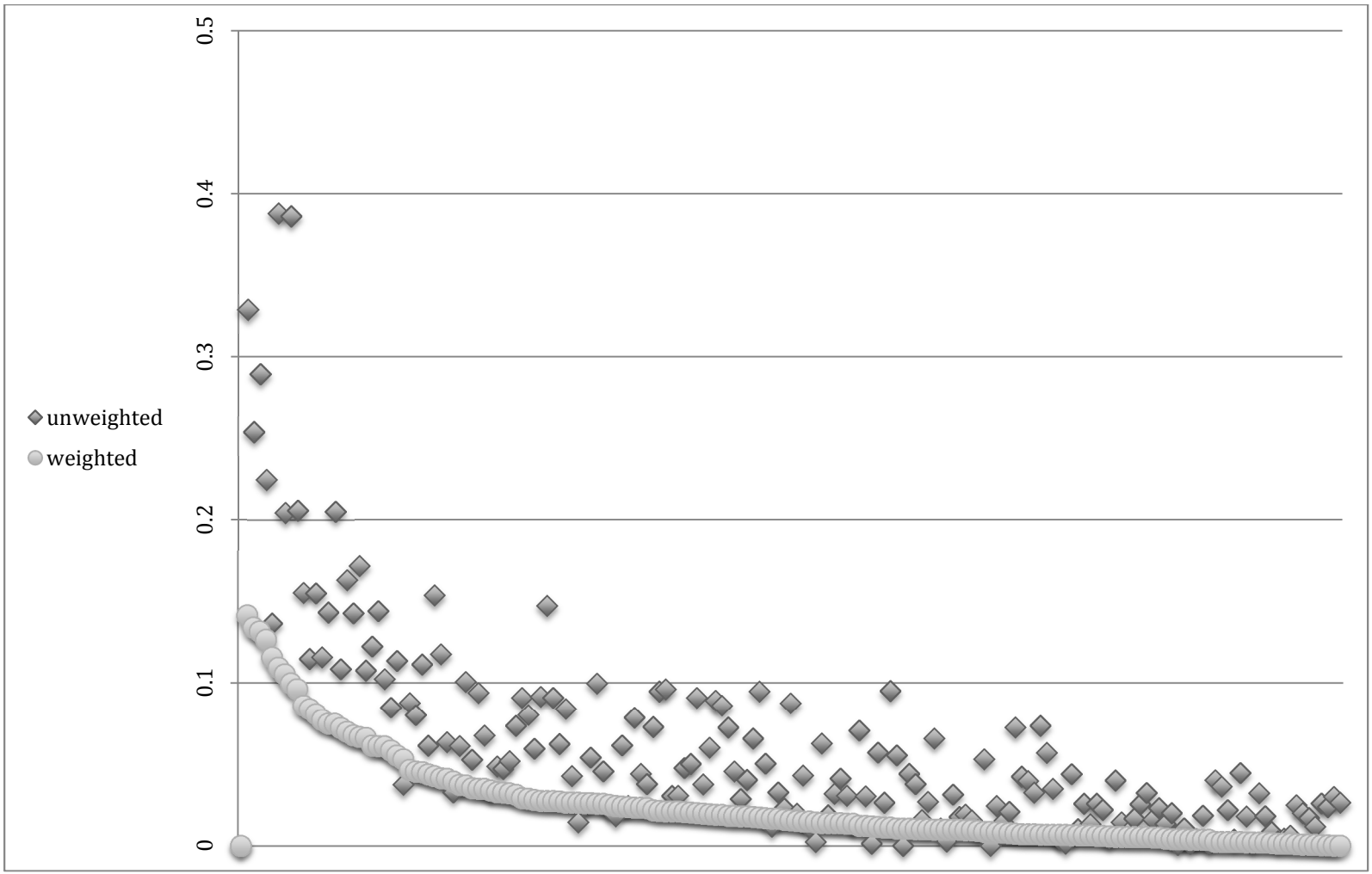


Figure B- 2. Standardized mean differences from 176 baseline characteristics between patients on sulfonylurea vs. TZD second-line therapy, before and after inverse probability of treatment weighting (sorted by weighted standardized mean difference)

Figure.B- 3. Intention-to-treat design stratified analysis: Hazard ratios for all-cause mortality and cardiovascular events comparing sulfonylurea to TZD second-line treatment

Treatment cohort	Patients	Events	Person-years	Incidence rate	Crude HR (95% CI)	PS adjusted (95% CI)	Weighted, unadjusted (95% CI)*	Weighted, adjusted (95% CI)*†
2nd line treatment								
Not Medicare beneficiaries:								
All-cause mortality								
Sulfonylurea	119,063	18,868	586,215	3.22	1.30 (1.21-1.39)	1.19 (1.11-1.28)	1.25 (1.15-1.35)	1.16 (1.08-1.26)
TZD	5,211	798	32,584	2.45	1.00 (reference)	1.00 (reference)	1.00 (reference)	1.00 (reference)
Cardiovascular composite event								
Sulfonylurea	119,063	33,537	494,057	6.79	1.27 (1.2-1.34)	1.17 (1.11-1.23)	1.21 (1.15-1.29)	1.15 (1.09-1.22)
TZD	5,211	1,482	28,622	5.18	1.00 (reference)	1.00 (reference)	1.00 (reference)	1.00 (reference)
Medicare beneficiaries:								
All-cause mortality								
Sulfonylurea	19,034	9,518	133,635	7.12	1.27 (1.18-1.38)	1.22 (1.13-1.32)	1.22 (1.11-1.35)	1.13 (1.02-1.25)
TZD	1,942	706	13,064	5.40	1.00 (reference)	1.00 (reference)	1.00 (reference)	1.00 (reference)
Cardiovascular composite event								
Sulfonylurea	19,034	12,771	86,803	14.71	1.35 (1.27-1.44)	1.25 (1.17-1.33)	1.27 (1.17-1.38)	1.02 (0.94-1.11)
TZD	1,942	1,032	9,612	10.74	1.00 (reference)	1.00 (reference)	1.00 (reference)	1.00 (reference)

CI , confidence interval; HR, hazard ratio; *PS* , propensity score

Note: Incidence rate per 10,000 person-years

*Weighted by the inverse probability of treatment

†Adjusted for demographic, utilization, laboratory values, medications, and comorbidities at baseline

Treatment cohort	Patients	Events	Person-years	Incidence rate	Crude HR (95% CI)	PS adjusted (95% CI)	Weighted, unadjusted (95% CI)*	Weighted, adjusted (95% CI)*†
2nd line treatment								
No prior cardiovascular disease:								
All-cause mortality								
Sulfonylurea	92,300	12,900	476,743	2.71	1.19 (1.10-1.28)	1.18 (1.09-1.27)	1.20 (1.09-1.31)	1.13 (1.03-1.24)
TZD	4,486	655	29,121	2.25	1.00 (reference)	1.00 (reference)	1.00 (reference)	1.00 (reference)
Cardiovascular composite event								
Sulfonylurea	92,300	20,999	425,293	4.94	1.18 (1.11-1.26)	1.16 (1.09-1.23)	1.16 (1.08-1.24)	1.08 (1.01-1.16)
TZD	4,486	1,103	26,497	4.16	1.00 (reference)	1.00 (reference)	1.00 (reference)	1.00 (reference)
Prior cardiovascular disease:								
All-cause mortality								
Sulfonylurea	45,797	15,486	243,108	6.37	1.23 (1.14-1.31)	1.20 (1.12-1.29)	1.22 (1.13-1.33)	1.17 (1.07-1.27)
TZD	2,667	849	16,526	5.14	1.00 (reference)	1.00 (reference)	1.00 (reference)	1.00 (reference)
Cardiovascular composite event								
Sulfonylurea	45,797	25,309	155,567	16.27	1.27 (1.20-1.34)	1.23 (1.17-1.30)	1.27 (1.19-1.35)	1.18 (1.11-1.26)
TZD	2,667	1,411	11,737	12.02	1.00 (reference)	1.00 (reference)	1.00 (reference)	1.00 (reference)

CI , confidence interval; HR, hazard ratio; *PS* , propensity score

Note: Incidence rate per 10,000 person-years

*Weighted by the inverse probability of treatment

†Adjusted for demographic, utilization, laboratory values, medications, and comorbidities at baseline

Treatment cohort	Patients	Events	Person-years	Incidence rate	Crude HR (95% CI)	PS adjusted (95% CI)	Weighted, unadjusted (95% CI)*	Weighted, adjusted (95% CI)*†
2nd line treatment								
Younger than 65:								
All-cause mortality								
Sulfonylurea	82,664	9,453	442,771	2.13	1.31 (1.18-1.44)	1.17 (1.06-1.30)	1.24 (1.11-1.38)	1.16 (1.04-1.29)
TZD	3,743	406	25,409	1.60	1.00 (reference)	1.00 (reference)	1.00 (reference)	1.00 (reference)
Cardiovascular composite event								
Sulfonylurea	82,664	20,371	374,815	5.43	1.22 (1.15-1.31)	1.14 (1.07-1.22)	1.18 (1.10-1.27)	1.12 (1.04-1.20)
TZD	3,743	953	22,185	4.30	1.00 (reference)	1.00 (reference)	1.00 (reference)	1.00 (reference)
65 and older:								
All-cause mortality								
Sulfonylurea	55,433	18,933	277,079	6.83	1.25 (1.17-1.32)	1.17 (1.10-1.25)	1.20 (1.11-1.29)	1.15 (1.07-1.24)
TZD	3,410	1,098	20,238	5.43	1.00 (reference)	1.00 (reference)	1.00 (reference)	1.00 (reference)
Cardiovascular composite event								
Sulfonylurea	55,433	25,937	206,045	12.59	1.24 (1.18-1.31)	1.18 (1.12-1.24)	1.21 (1.14-1.29)	1.12 (1.05-1.19)
TZD	3,410	1,561	16,049	9.73	1.00 (reference)	1.00 (reference)	1.00 (reference)	1.00 (reference)

CI , confidence interval; HR, hazard ratio; *PS* , propensity score

Note: Incidence rate per 10,000 person-years

*Weighted by the inverse probability of treatment

†Adjusted for demographic, utilization, laboratory values, medications, and comorbidities at baseline

Figure B- 4. As-treated design stratified analyses: Hazard ratios for all-cause mortality and cardiovascular events comparing sulfonylurea to TZD second-line treatment

Treatment cohort	Patients	Events	Person-years	Incidence rate	Crude HR (95% CI)	PS adjusted (95% CI)	Weighted, unadjusted (95% CI)*	Weighted, adjusted (95% CI)*†
2nd line treatment								
Not Medicare beneficiaries:								
All-cause mortality								
Sulfonylurea	119,063	3,199	202,095	1.58	1.34 (1.07-1.69)	1.25 (0.99-1.57)	1.40 (1.08-1.82)	1.34 (1.03-1.74)
TZD	5,211	77	6,648	1.16	1.00 (reference)	1.00 (reference)	1.00 (reference)	1.00 (reference)
Cardiovascular composite event								
Sulfonylurea	119,063	12,995	186,591	6.96	1.34 (1.20-1.49)	1.18 (1.06-1.32)	1.35 (1.20-1.52)	1.30 (1.15-1.46)
TZD	5,211	351	6,333	5.54	1.00 (reference)	1.00 (reference)	1.00 (reference)	1.00 (reference)
Medicare beneficiaries:								
All-cause mortality								
Sulfonylurea	19,034	1,631	43,512	3.75	1.62 (1.28-2.04)	1.59 (1.25-2.01)	1.63 (1.19-2.24)	1.53 (1.11-2.10)
TZD	1,942	74	3,235	2.29	1.00 (reference)	1.00 (reference)	1.00 (reference)	1.00 (reference)
Cardiovascular composite event								
Sulfonylurea	19,034	6,118	33,917	18.04	1.37 (1.24-1.51)	1.25 (1.13-1.39)	1.26 (1.11-1.43)	1.01 (0.88-1.15)
TZD	1,942	410	2,732	15.01	1.00 (reference)	1.00 (reference)	1.00 (reference)	1.00 (reference)

CI, confidence interval; HR, hazard ratio; *PS*, propensity score
Note: Incidence rate per 10,000 person-years
*Weighted by the inverse probability of treatment
†Adjusted for demographic, utilization, laboratory values, medications, and comorbidities at baseline

Treatment cohort	Patients	Events	Person-years	Incidence rate	Crude HR (95% CI)	PS adjusted (95% CI)	Weighted, unadjusted (95% CI)*	Weighted, adjusted (95% CI)*†
2nd line treatment								
No prior cardiovascular disease:								
All-cause mortality								
Sulfonylurea	92,300	2,022	161,514	1.25	1.20 (0.93-1.54)	1.22 (0.95-1.58)	1.29 (0.95-1.74)	1.33 (0.98-1.80)
TZD	4,486	63	6,183	1.02	1.00 (reference)	1.00 (reference)	1.00 (reference)	1.00 (reference)
Cardiovascular composite event								
Sulfonylurea	92,300	6,124	154,939	3.95	1.02 (0.89-1.16)	1.05 (0.92-1.20)	1.10 (0.94-1.28)	1.12 (0.96-1.31)
TZD	4,486	235	5,957	3.94	1.00 (reference)	1.00 (reference)	1.00 (reference)	1.00 (reference)
Prior cardiovascular disease:								
All-cause mortality								
Sulfonylurea	45,797	2,808	84,093	3.34	1.38 (1.11-1.70)	1.43 (1.15-1.77)	1.45 (1.11-1.91)	1.47 (1.12-1.94)
TZD	2,667	88	3,700	2.38	1.00 (reference)	1.00 (reference)	1.00 (reference)	1.00 (reference)
Cardiovascular composite event								
Sulfonylurea	45,797	12,989	65,569	19.81	1.28 (1.17-1.39)	1.28 (1.17-1.40)	1.32 (1.19-1.47)	1.28 (1.15-1.43)
TZD	2,667	526	3,108	16.93	1.00 (reference)	1.00 (reference)	1.00 (reference)	1.00 (reference)

CI, confidence interval; HR, hazard ratio; *PS*, propensity score
Note: Incidence rate per 10,000 person-years
*Weighted by the inverse probability of treatment
†Adjusted for demographic, utilization, laboratory values, medications, and comorbidities at baseline

Treatment cohort	Patients	Events	Person-years	Incidence rate	Crude HR (95% CI)	PS adjusted (95% CI)	Weighted, unadjusted (95% CI)*	Weighted, adjusted (95% CI)*†
2nd line treatment								
Younger than 65:								
All-cause mortality								
Sulfonylurea	82,664	1,307	137,285	0.95	1.90 (1.27-2.85)	1.77 (1.18-2.65)	1.75 (1.13-2.71)	1.74 (1.13-2.70)
TZD	3,743	24	4,847	0.50	1.00 (reference)	1.00 (reference)	1.00 (reference)	1.00 (reference)
Cardiovascular composite event								
Sulfonylurea	82,664	7,254	127,874	5.67	1.25 (1.10-1.43)	1.18 (1.03-1.35)	1.21 (1.04-1.39)	1.23 (1.06-1.42)
TZD	3,743	222	4,632	4.79	1.00 (reference)	1.00 (reference)	1.00 (reference)	1.00 (reference)
65 and older:								
All-cause mortality								
Sulfonylurea	55,433	3,523	108,322	3.25	1.26 (1.06-1.51)	1.24 (1.03-1.48)	1.30 (1.03-1.63)	1.30 (1.03-1.63)
TZD	3,410	127	5,036	2.52	1.00 (reference)	1.00 (reference)	1.00 (reference)	1.00 (reference)
Cardiovascular composite event								
Sulfonylurea	55,433	11,859	92,634	12.80	1.15 (1.05-1.25)	1.11 (1.02-1.21)	1.20 (1.08-1.34)	1.16 (1.04-1.29)
TZD	3,410	539	4,433	12.16	1.00 (reference)	1.00 (reference)	1.00 (reference)	1.00 (reference)

CI, confidence interval; HR, hazard ratio; *PS*, propensity score
Note: Incidence rate per 10,000 person-years
*Weighted by the inverse probability of treatment
†Adjusted for demographic, utilization, laboratory values, medications, and comorbidities at baseline

Figure B- 5. As-treated design stratified analyses: Hazard ratios for all-cause mortality and cardiovascular events comparing sulfonylurea to TZD second-line treatment

Treatment cohort	Patients	Events	Person-years	Incidence rate	Crude HR (95% CI)	PS adjusted (95% CI)	Weighted, unadjusted (95% CI)*	Weighted, adjusted (95% CI)*†
2nd line treatment								
Not Medicare beneficiaries:								
All-cause mortality								
Metformin+Sulfonylurea	89,802	1,203	119,912	1.00	1.00 (reference)	1.00 (reference)	1.00 (reference)	1.00 (reference)
Sulfonylurea only	29,261	1,166	48,581	2.40	2.39 (2.20-2.59)	1.66 (1.51-1.84)	1.86 (1.71-2.03)	1.63 (1.49-1.78)
TZD only	1,543	37	2,040	1.81	1.81 (1.31-2.52)	1.25 (0.90-1.75)	1.07 (0.68-1.70)	1.06 (0.67-1.68)
Metformin+TZD	3,668	24	3,665	0.65	0.66 (0.44-0.98)	0.61 (0.41-0.91)	0.60 (0.43-0.85)	0.74 (0.52-1.04)
Cardiovascular composite event								
Metformin+Sulfonylurea	89,802	6,525	113,391	5.75	1.00 (reference)	1.00 (reference)	1.00 (reference)	1.00 (reference)
Sulfonylurea only	29,261	4,273	43,496	9.82	1.75 (1.69-1.82)	1.27 (1.21-1.33)	1.38 (1.32-1.43)	1.28 (1.23-1.34)
TZD only	1,543	148	1,911	7.75	1.31 (1.12-1.55)	0.94 (0.80-1.11)	0.95 (0.77-1.17)	0.95 (0.77-1.16)
Metformin+TZD	3,668	154	3,545	4.34	0.70 (0.60-0.83)	0.66 (0.56-0.77)	0.61 (0.53-0.70)	0.71 (0.62-0.82)
Medicare beneficiaries:								
All-cause mortality								
Metformin+Sulfonylurea	13,379	486	22,034	2.21	1.00 (reference)	1.00 (reference)	1.00 (reference)	1.00 (reference)
Sulfonylurea only	5,655	632	13,185	4.79	2.19 (1.94-2.47)	1.62 (1.41-1.88)	1.87 (1.64-2.13)	1.68 (1.47-1.92)
TZD only	713	38	1,115	3.41	1.54 (1.11-2.15)	1.11 (0.79-1.56)	1.27 (0.81-1.99)	1.18 (0.75-1.86)
Metformin+TZD	1,229	22	1,638	1.34	0.61 (0.39-0.93)	0.55 (0.36-0.84)	0.47 (0.29-0.77)	0.54 (0.33-0.89)
Cardiovascular composite event								
Metformin+Sulfonylurea	13,379	3,024	18,370	16.46	1.00 (reference)	1.00 (reference)	1.00 (reference)	1.00 (reference)
Sulfonylurea only	5,655	2,047	9,777	20.94	1.41 (1.33-1.49)	1.16 (1.09-1.24)	1.20 (1.13-1.28)	1.21 (1.13-1.29)
TZD only	713	188	885	21.24	1.19 (1.03-1.38)	0.95 (0.82-1.11)	1.01 (0.83-1.23)	1.15 (0.94-1.40)
Metformin+TZD	1,229	174	1,439	12.09	0.66 (0.57-0.77)	0.62 (0.53-0.72)	0.62 (0.53-0.72)	0.92 (0.78-1.08)

CI, confidence interval; HR, hazard ratio; *PS*, propensity score

Note: Incidence rate per 10,000 person-years

*Weighted by the inverse probability of treatment

†Adjusted for demographic, utilization, laboratory values, medications, and comorbidities at baseline

Treatment cohort	Patients	Events	Person-years	Incidence rate	Crude HR (95% CI)	PS adjusted (95% CI)	Weighted, unadjusted (95% CI)*	Weighted, adjusted (95% CI)*†
2nd line treatment								
No prior cardiovascular disease:								
All-cause mortality								
Metformin+Sulfonylurea	71,278	786	98,094	0.80	1.00 (reference)	1.00 (reference)	1.00 (reference)	1.00 (reference)
Sulfonylurea only	21,022	671	36,584	1.83	2.27 (2.05-2.52)	1.66 (1.46-1.88)	1.83 (1.64-2.05)	1.70 (1.52-1.90)
TZD only	1,296	26	1,864	1.39	1.75 (1.18-2.59)	1.24 (0.83-1.84)	1.14 (0.68-1.92)	1.03 (0.61-1.73)
Metformin+TZD	3,190	23	3,431	0.67	0.85 (0.56-1.28)	0.78 (0.51-1.18)	0.63 (0.42-0.96)	0.70 (0.47-1.06)
Cardiovascular composite event								
Metformin+Sulfonylurea	71,278	3,205	95,031	3.37	1.00 (reference)	1.00 (reference)	1.00 (reference)	1.00 (reference)
Sulfonylurea only	21,022	1,693	34,921	4.85	1.46 (1.38-1.55)	1.22 (1.13-1.30)	1.23 (1.15-1.31)	1.21 (1.14-1.29)
TZD only	1,296	95	1,771	5.36	1.58 (1.29-1.94)	1.29 (1.05-1.59)	1.31 (1.03-1.66)	1.17 (0.92-1.49)
Metformin+TZD	3,190	104	3,348	3.11	0.90 (0.74-1.09)	0.86 (0.71-1.04)	0.75 (0.63-0.89)	0.80 (0.67-0.96)
Prior cardiovascular disease:								
All-cause mortality								
Metformin+Sulfonylurea	31,903	903	43,852	2.06	1.00 (reference)	1.00 (reference)	1.00 (reference)	1.00 (reference)
Sulfonylurea only	13,894	1,127	25,182	4.48	2.17 (1.98-2.36)	1.56 (1.40-1.74)	1.78 (1.62-1.96)	1.58 (1.43-1.75)
TZD only	960	49	1,291	3.79	1.85 (1.39-2.46)	1.28 (0.95-1.72)	1.35 (0.90-2.03)	1.13 (0.75-1.69)
Metformin+TZD	1,707	23	1,872	1.23	0.60 (0.40-0.91)	0.54 (0.36-0.82)	0.59 (0.40-0.86)	0.66 (0.45-0.96)
Cardiovascular composite event								
Metformin+Sulfonylurea	31,903	6,344	36,730	17.27	1.00 (reference)	1.00 (reference)	1.00 (reference)	1.00 (reference)
Sulfonylurea only	13,894	4,627	18,351	25.21	1.52 (1.47-1.58)	1.23 (1.18-1.29)	1.34 (1.28-1.39)	1.29 (1.24-1.35)
TZD only	960	241	1,025	23.51	1.29 (1.13-1.47)	1.02 (0.89-1.16)	1.01 (0.84-1.20)	0.96 (0.80-1.14)
Metformin+TZD	1,707	224	1,636	13.69	0.73 (0.64-0.84)	0.68 (0.60-0.78)	0.69 (0.60-0.78)	0.77 (0.68-0.87)

CI, confidence interval; HR, hazard ratio; *PS*, propensity score

Note: Incidence rate per 10,000 person-years

*Weighted by the inverse probability of treatment

†Adjusted for demographic, utilization, laboratory values, medications, and comorbidities at baseline

Treatment cohort	Patients	Events	Person-years	Incidence rate	Crude HR (95% CI)	PS adjusted (95% CI)	Weighted, unadjusted (95% CI)*	Weighted, adjusted (95% CI)*†
2nd line treatment								
Younger than 65:								
All-cause mortality								
Metformin+Sulfonylurea	66,548	607	88,170	0.69	1.00 (reference)	1.00 (reference)	1.00 (reference)	1.00 (reference)
Sulfonylurea only	16,116	346	26,094	1.33	1.92 (1.69-2.20)	1.72 (1.48-2.01)	1.73 (1.51-1.97)	1.64 (1.43-1.88)
TZD only	895	5	1,208	0.41	0.60 (0.25-1.45)	0.54 (0.22-1.30)	0.41 (0.15-1.16)	0.44 (0.16-1.24)
Metformin+TZD	2,848	14	2,915	0.48	0.70 (0.41-1.19)	0.68 (0.40-1.16)	0.85 (0.58-1.26)	0.85 (0.58-1.26)
Cardiovascular composite event								
Metformin+Sulfonylurea	66,548	4,167	83,563	4.99	1.00 (reference)	1.00 (reference)	1.00 (reference)	1.00 (reference)
Sulfonylurea only	16,116	1,850	23,589	7.84	1.62 (1.54-1.71)	1.28 (1.21-1.37)	1.31 (1.24-1.39)	1.30 (1.23-1.37)
TZD only	895	70	1,140	6.14	1.22 (0.96-1.54)	0.95 (0.75-1.21)	0.91 (0.70-1.18)	0.94 (0.72-1.22)
Metformin+TZD	2,848	118	2,820	4.18	0.78 (0.65-0.94)	0.74 (0.62-0.89)	0.75 (0.64-0.87)	0.77 (0.66-0.90)
65 and older:								
All-cause mortality								
Metformin+Sulfonylurea	36,633	1,082	53,776	2.01	1.00 (reference)	1.00 (reference)	1.00 (reference)	1.00 (reference)
Sulfonylurea only	18,800	1,452	35,672	4.07	2.02 (1.87-2.18)	1.58 (1.44-1.75)	1.81 (1.66-1.97)	1.70 (1.55-1.86)
TZD only	1,361	70	1,948	3.59	1.80 (1.41-2.29)	1.37 (1.06-1.75)	1.48 (1.05-2.07)	1.38 (0.98-1.94)
Metformin+TZD	2,049	32	2,389	1.34	0.67 (0.47-0.96)	0.62 (0.44-0.88)	0.51 (0.34-0.76)	0.54 (0.36-0.81)
Cardiovascular composite event								
Metformin+Sulfonylurea	36,633	5,382	48,198	11.17	1.00 (reference)	1.00 (reference)	1.00 (reference)	1.00 (reference)
Sulfonylurea only	18,800	4,470	29,684	15.06	1.41 (1.35-1.46)	1.12 (1.07-1.18)	1.23 (1.17-1.28)	1.24 (1.19-1.30)
TZD only	1,361	266	1,656	16.06	1.37 (1.21-1.54)	1.06 (0.93-1.20)	1.14 (0.96-1.35)	1.10 (0.93-1.30)
Metformin+TZD	2,049	210	2,165	9.70	0.80 (0.70-0.92)	0.74 (0.65-0.85)	0.68 (0.59-0.78)	0.77 (0.67-0.89)

CI, confidence interval; HR, hazard ratio; *PS*, propensity score

Note: Incidence rate per 10,000 person-years

*Weighted by the inverse probability of treatment

†Adjusted for demographic, utilization, laboratory values, medications, and comorbidities at baseline

Table B- 6. Summary of previous studies examining all-cause mortality for sulfonylurea vs. TZD with and without metformin (reference group=comparator; NR=not reported)

All-cause mortality

Study (year)	Effect size (95% CI)	SU events	SU n	Comparator events	Comparator n
METF+SU vs. METF+TZD					
RCT (n=2)					
Hamann, 2008	0.98 (0.14-6.89)	2	301	2	294
Matthews, 2005	5.06 (0.24-105.05)	2	313	0	317
Obs Cohort (n=4)					
Kannan, 2015	1.16 (1.00-1.35)	NR	9,419	NR	1846
Morgan, 2012	1.10 (0.90-1.35)	624	15,377	141	4677
Morgan, 2012	1.41 (1.03-1.94)	624	15,377	45	2525
Prentice, 2014	1.50 (1.09-2.09)	NR	73,726	NR	7210
METF+SU vs. SU					
RCT (n=1)					
UKPDS,1998	1.60 (1.02-2.52)	47	268	31	269
Obs Cohort (n=5)					
Currie,2013	0.63 (0.57-0.69)	1,152	23,049	2,269	16218
Evans,2006	0.42 (0.23-0.75)	NR	113	567	3331
Evans,2006	1.73 (1.22-2.45)	NR	985	567	3331
Evans,2006	1.51 (1.09-2.10)	NR	1,252	567	3331
Gulliford,2004	1.06 (0.85-1.31)	127	1,868	1,030	8488
Johnson,2002	0.63 (0.57-0.71)	635	4,683	750	3033
Morgan,2012	0.69 (0.57-0.83)	624	15,377	239	2244
Olsson,2000	1.63 (1.27-2.09)	NR	169	NR	741
Sillars,2010	1.02 (0.76-1.37)	94	216	150	317
METF+SU vs. TZD					
Obs Cohort (n=4)					
Evans, 2006	0.42 (0.23-0.75)	NR	113	567	3331
Johnson, 2002	0.63 (0.57-0.71)	635	4,683	750	3033
Olsson, 2000	1.63 (1.27-2.09)	NR	169	NR	741
Sillars, 2010	1.02 (0.76-1.37)	94	216	150	317
SU vs. METF+TZD					
Obs Cohort (n=2)					
Morgan, 2012	1.61 (1.21-2.12)	239	2,244	141	4677
Morgan, 2012	2.06 (1.43-2.98)	239	2,244	45	2525
SU vs. TZD					
RCT (n=3)					
Jain, 2006	5.00 (0.24-103.62)	2	251	0	251
Kahn, 2006	0.92 (0.57-1.49)	31	1,441	34	1456
Tan, 2004	3.59 (0.15-87.00)	0	109	1	91
Obs Cohort (n=5)					
Pantalone, 2009	1.37 (0.98-1.96)	NR	7,427	NR	1079
Pantalone, 2009	1.69 (1.23-2.33)	NR	7,427	NR	1508
Tzoulaki, 2009	1.16 (0.82-1.63)	1,379	58,095	34	8442
Wheeler, 2013	1.13 (0.90-1.42)	912	28,156	88	3753
Wheeler, 2013	1.27 (1.01-1.59)	1,121	28,957	88	3753

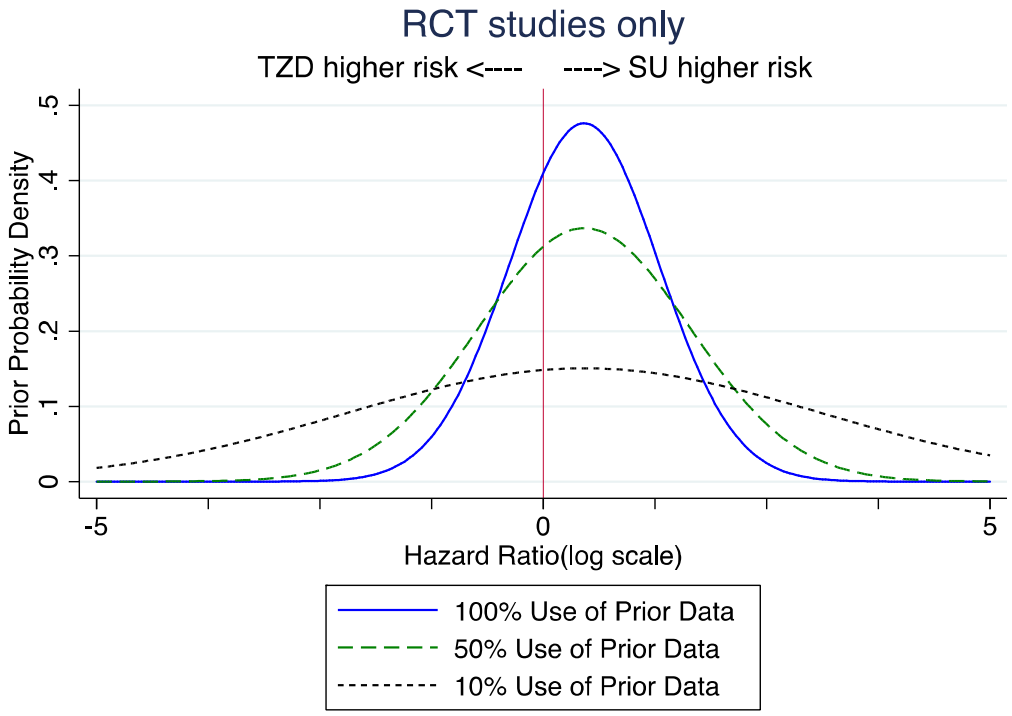
Table B- 7. Summary of previous studies examining cardiovascular disease composite events for sulfonylurea vs. TZD with and without metformin (reference group=comparator; NR=not reported)

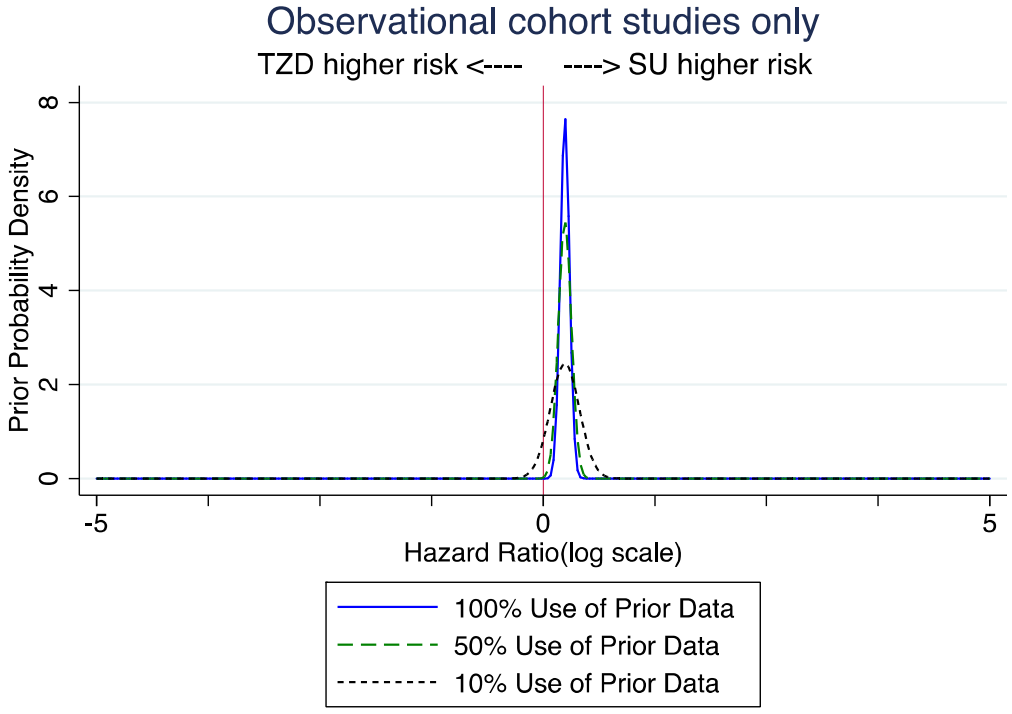
Cardiovascular composite events

Study (year)	Effect size (95% CI)	SU events	SU n	Comparator events	Comparator n
METF+SU vs. METF+TZD					
Obs Cohort (n=5)					
McAfee, 2007	1.27 (0.75-2.13)	36	1,362	24	1362
Morgan, 2012	1.19 (0.91-1.55)	323	15,377	79	4677
Morgan, 2012	1.38 (0.93-2.06)	323	15,377	27	2525
Walker, 2008	1.08 (0.95-1.22)	854	79,004	222	26885
Walker, 2008	1.12 (0.96-1.32)	854	79,004	135	17282
METF+SU vs. SU					
RCT (n=1)					
Rosenstock, 2006	5.22 (0.25-107.60)	0	116	2	111
Obs Cohort (n=3)					
Currie, 2013	0.79 (0.68-0.92)	538	23,049	624	16218
Li, 2014	1.79 (0.68-4.74)	33	NR	13	
Morgan, 2012	0.63 (0.48-0.84)	323	15,377	77	2244
METF+SU vs. TZD					
Obs Cohort (n=3)					
Johannes, 2007	0.98 (0.83-1.15)	NR	12,570	NR	12570
Johnson, 2005	0.96 (0.82-1.12)	264	1,081	541	2138
Sillars, 2010	0.87 (0.66-1.15)	104	216	159	317
SU vs. METF+TZD					
Obs Cohort (n=2)					
Morgan, 2012	1.87 (1.27-2.76)	77	2,244	79	4677
Morgan, 2012	2.18 (1.34-3.56)	77	2,244	27	2525
SU vs. TZD					
RCT (n=4)					
Jain, 2006	2.00 (1.01-3.99)	22	251	11	251
Kahn, 2006	0.67 (0.41-1.08)	41	1,441	62	1456
Perriello, 2006	0.53 (0.14-2.09)	3	137	6	146
St. John Sutton, 2002	0.58 (0.20-1.69)	5	99	9	104
Obs Cohort (n=8)					
Berkowitz, 2014	1.10 (0.89-1.37)	624	3,570	150	948
Hsiao, 2009	0.65 (0.54-0.78)	7,491	97,651	266	2093
Hsiao, 2009	0.97 (0.61-1.54)	7,491	97,651	44	495
McAfee, 2007	1.22 (0.98-1.61)	191	8,977	152	8977
Pantalone, 2009	0.96 (0.79-1.16)	NR	7,427	NR	1508
Pantalone, 2009	1.11 (0.88-1.41)	NR	7,427	NR	1079
Walker, 2008	1.10 (0.94-1.29)	551	48,376	160	16302
Walker, 2008	1.12 (0.94-1.34)	551	48,376	122	12440

APPENDIX C

Table C- 1. Summary of prior probability distributions (log scale)





C- 2. Understanding the influence of prior results: Approximate Bayes estimation

In the simplest Normal-Normal Bayesian model, shrinkage is defined as:

$$B = r/(r+n)$$

where…

r = quasi prior sample size

n = new study sample size

Although we do not have a linear relationship between the posterior and the new study data for the hazard ratio, we can say approximately that in model form (with square brackets defining [mean, variance]), this is…

Y |mu, sigma ~ $[h_{va}, \hat{\sigma}_{VA}/n]$ (for likelihood)

Mu|mu_{prior}, sigma ~$[h_{prior}, \hat{\sigma}_{VA}/r = \hat{\sigma}_{VA}/n \{1\text{-}B/B\}]$ (for prior)

Mu|mu_{post}, sigma, data ~ $[(1\text{-}B)\, h_{va} + B(h_{prior}), \hat{\sigma}_{VA}/n \{1\text{-}B\}]$ (for posterior)

To estimate mean and variance, both quasi pre-study sample size and then shrinkage needs to be calculated. To calculate quasi pre-study sample size (r), follow two steps…

Step 1: first find the variance from the VA study data $\hat{\sigma}_{VA}^2$, using known information (i.e., the study hazard ratio, its upper 95% confidence interval, and sample size):

$$\hat{\sigma}_{VA}^2 = \ln\left(\frac{h_{VA,upper}}{h_{VA}}\right)\left\{\frac{\sqrt{n}}{1.96}\right\}$$

where…

$h_{VA,upper}$ = upper 95% confidence interval for the study hazard ratio

h_{VA} = hazard ratio related to the new study

1.96= z-score approximating the 95% confidence interval

Step 2: From here, use $\hat{\sigma}_{VA}$ and the confidence interval for the hazard ratio of the prior evidence to get a quasi sample size equivalent for the prior evidence (*r*) (this entails the reasonable assumption that the standard deviation of the ln(h) is the same for the prior and the likelihood):

$$r = \frac{\hat{\sigma}_{VA}^2}{\left(\ln\left(\frac{h_{prior,upper}}{h_{prior}}\right)\left\{\frac{1}{1.96}\right\}\right)^2}$$

where…

$h_{prior,upper}$ = upper 95% confidence interval for the prior evidence hazard ratio

h_{prior} = hazard ratio of the prior evidence

1.96= z-score approximating the 95% confidence interval

$\hat{\sigma}_{VA}^2$ = variance from the study data

Since quasi sample size is now known, shrinkage can be calculated: $B = r/(r+n)$ and, therefore the posterior hazard ratio and variance can be calculated.

The posterior estimated hazard ratio is:

$$h_{post} = (1-B)h_{VA} + B\left(h_{prior}\right)$$

where …

h_{VA}= hazard ratio from the new data

h_{prior} = hazard ratio of the prior probability distribution.

Variance of the prior and posterior distributions are calculated as:

Prior: $\hat{\sigma}^2_{prior} = \hat{\sigma}^2_{VA} * ((1 - B)/B)$

Posterior: $\hat{\sigma}^2_{post} = \hat{\sigma}^2_{VA} * (1 - B)$

Here the multiplier for the prior variance, (1-*B*)/*B*, is large if *B* is very small. *B* is very small when the information in the current study (*n*) is large compared to the information (*r*) in the prior. The multiplier for the posterior variance, (1-*B*), reflects the combined information from the data and the prior. If *B* is tiny, the only information mostly coming from the new data, and the posterior variance will be approximately that of the data. However, if the prior is providing substantial information (through the value of *r*), the final estimate of interest will have an interval much narrower than that from the data alone because it combines the *n* and *r*.

Table C- 3: Calculating quasi pre-study sample size, posterior hazard ratio, and variance

Using the formulas described in Appendix C-2 with the data from the observational (100%) prior estimate and corresponding confidence intervals as an example, quasi pre-study sample size, shrinkage, posterior hazard ratio and variance, and prior variance can be calculated as follows:

To calculate quasi pre-study sample size (two steps):

Step 1:

$$\hat{\sigma}_{VA} = \ln\left(\frac{h_{VA,upper}}{h_{VA}}\right)\left\{\frac{\sqrt{n}}{1.96}\right\} = \ln\left(\frac{1.2266}{1.1543}\right)\left\{\frac{\sqrt{145{,}250}}{1.96}\right\} = 11.813$$

Step 2:

$$r = \frac{\hat{\sigma}_{VA}^2}{\left(\ln\left(\frac{h_{prior,upper}}{h_{prior}}\right)\left\{\frac{1}{1.96}\right\}\right)^2} = \frac{11.8130^2}{\left(\ln\left(\frac{1.410}{1.280}\right)\left\{\frac{1}{1.96}\right\}\right)^2} = 57295$$

To calculate shrinkage:

$$B = \frac{r}{r+n} = \frac{57295}{57295 + 145250} = 0.2829$$

To approximate the post-study (posterior) hazard ratio:

$$h_{post} = (1-B)h_{VA} + B\left(h_{prior}\right)$$
$$= (1 - 0.2829) * 1.1543 + (00.2829 * 1.2800) \sim 1.1898$$

To approximate the prior and posterior variance:

Prior variance: $\hat{\sigma}_{prior}^2 = \hat{\sigma}_{VA}^2$ *(1-*B*)/*B*, so $11.8130^2 * \left(\frac{1-0.2829}{(0.2829)}\right) \sim 353.7731$

Posterior variance: $\hat{\sigma}_{post}^2 = \hat{\sigma}_{VA}^2$*(1-*B*), so $11.8130^2 * (1 - 0.2829) \sim 100.0737$

REFERENCES

Abrahamson, M. J. (2015). Should Sulfonylureas Remain an Acceptable First-Line Add-on to Metformin Therapy in Patients With Type 2 Diabetes? Yes, They Continue to Serve Us Well! *Diabetes Care*, *38*(1), 166–169. https://doi.org/10.2337/dc14-1945

Ahrén, B., Johnson, S. L., Stewart, M., Cirkel, D. T., Yang, F., Perry, C., … HARMONY 3 Study Group. (2014). HARMONY 3: 104-week randomized, double-blind, placebo- and active-controlled trial assessing the efficacy and safety of albiglutide compared with placebo, sitagliptin, and glimepiride in patients with type 2 diabetes taking metformin. *Diabetes Care*, *37*(8), 2141–2148. https://doi.org/10.2337/dc14-0024

Altman, D. G., & Bland, J. M. (2003). Interaction revisited: the difference between two estimates. *BMJ: British Medical Journal*, *326*(7382), 219. https://doi.org/10.1136/bmj.326.7382.219

Armijo-Olivo, S., Ospina, M., Costa, B. R. da, Egger, M., Saltaji, H., Fuentes, J., … Cummings, G. G. (2014). Poor Reliability between Cochrane Reviewers and Blinded External Reviewers When Applying the Cochrane Risk of Bias Tool in Physical Therapy Trials. *PLoS One*, *9*(5), e96920. https://doi.org/10.1371/journal.pone.0096920

Armijo-Olivo, S., Stiles, C. R., Hagen, N. A., Biondo, P. D., & Cummings, G. G. (2012). Assessment of study quality for systematic reviews: a comparison of the Cochrane Collaboration Risk of Bias Tool and the Effective Public Health Practice Project Quality Assessment Tool: methodological research. *Journal of Evaluation in Clinical Practice*, *18*(1), 12–18. https://doi.org/10.1111/j.1365-2753.2010.01516.x

Austin, P. C. (2011). An Introduction to Propensity Score Methods for Reducing the Effects of Confounding in Observational Studies. *Multivariate Behavioral Research*, *46*(3), 399–424. https://doi.org/10.1080/00273171.2011.568786

Austin, P. C. (2013). The performance of different propensity score methods for estimating marginal hazard ratios. *Statistics in Medicine*, *32*(16), 2837–2849. https://doi.org/10.1002/sim.5705

Austin, P. C. (2014). The use of propensity score methods with survival or time-to-event outcomes: reporting measures of effect similar to those used in randomized experiments. *Statistics in Medicine*, *33*(7), 1242–1258. https://doi.org/10.1002/sim.5984

Austin, P. C., & Schuster, T. (2014). The performance of different propensity score methods for estimating absolute effects of treatments on survival outcomes: A

simulation study. *Statistical Methods in Medical Research*, 0962280213519716. https://doi.org/10.1177/0962280213519716

Austin, P. C., & Stuart, E. A. (2015). Moving towards best practice when using inverse probability of treatment weighting (IPTW) using the propensity score to estimate causal treatment effects in observational studies. *Statistics in Medicine*, *34*(28), 3661–3679. https://doi.org/10.1002/sim.6607

Azoulay, L., Schneider-Lindner, V., Dell'Aniello, S., Schiffrin, A., & Suissa, S. (2010). Combination therapy with sulfonylureas and metformin and the prevention of death in type 2 diabetes: a nested case-control study. *Pharmacoepidemiology and Drug Safety*, *19*(4), 335–342. https://doi.org/10.1002/pds.1834

Bakris, G. L., Ruilope, L. M., McMorn, S. O., Weston, W. M., Heise, M. A., Freed, M. I., & Porter, L. E. (2006). Rosiglitazone reduces microalbuminuria and blood pressure independently of glycemia in type 2 diabetes patients with microalbuminuria. *Journal of Hypertension*, *24*(10), 2047–2055. https://doi.org/10.1097/01.hjh.0000244955.39491.88

Bennett, W. L., Maruthur, N. M., Singh, S., Segal, J. B., Wilson, L. M., Chatterjee, R., … Bolen, S. (2011). Comparative Effectiveness and Safety of Medications for Type 2 Diabetes: An Update Including New Drugs and 2-Drug Combinations. *Annals of Internal Medicine*, *154*(9), 602. https://doi.org/10.7326/0003-4819-154-9-201105030-00336

Bennett, W. L., Wilson, L. M., Bolen, S., Maruthur, N., Singh, S., Chatterjee, R., … Segal, J. B. (2011). *Oral Diabetes Medications for Adults With Type 2 Diabetes: An Update*. Rockville (MD): Agency for Healthcare Research and Quality (US). Retrieved from http://www.ncbi.nlm.nih.gov/books/NBK55754/

Berkowitz, S. A., Krumme, A. A., Avorn, J., Brennan, T., Matlin, O. S., Spettell, C. M., … Choudhry, N. K. (2014). Initial choice of oral glucose-lowering medication for diabetes mellitus: a patient-centered comparative effectiveness study. *JAMA Internal Medicine*, *174*(12), 1955–1962. https://doi.org/10.1001/jamainternmed.2014.5294

Bolen, S., Feldman, L., Vassy, J., Wilson, L., Hsin-Chieh Yeh, Marinopoulos, S., … Brancati, F. L. (2007). Systematic Review: Comparative Effectiveness and Safety of Oral Medications for Type 2 Diabetes Mellitus. *Annals of Internal Medicine*, *147*(6), 386–W77.

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to Meta-Analysis* (1 edition). Chichester, U.K.: Wiley.

Bo, S., Ciccone, G., Rosato, R., Villois, P., Appendino, G., Ghigo, E., & Grassi, G. (2012). Cancer mortality reduction and metformin: A retrospective cohort study

in type 2 diabetic patients. *Diabetes, Obesity and Metabolism*, *14*(1), 23–29. https://doi.org/10.1111/j.1463-1326.2011.01480.x

Brookhart, M. A., Wyss, R., Layton, J. B., & Stürmer, T. (2013). Propensity Score Methods for Confounding Control in Nonexperimental Research. *Circulation: Cardiovascular Quality and Outcomes*, *6*(5), 604–611. https://doi.org/10.1161/CIRCOUTCOMES.113.000359

Brophy, J. M., & Joseph, L. (1995). Placing trials in context using Bayesian analysis. GUSTO revisited by Reverend Bayes. *JAMA: The Journal of the American Medical Association*, *273*(11), 871–875.

Brownstein, J. S., Murphy, S. N., Goldfine, A. B., Grant, R. W., Sordo, M., Gainer, V., … Kohane, I. S. (2010). Rapid identification of myocardial infarction risk associated with diabetes medications using electronic medical records. *Diabetes Care*, *33*(3), 526–531. https://doi.org/10.2337/dc09-1506

Centers for Disease Control and Prevention. (2014). *National diabetes statistics report: estimates of diabetes and its burden in the United States, 2014*. Atlanta, GA: US Department of Health and Human Services. Retrieved from https://www.cdc.gov/diabetes/data/statistics/2014statisticsreport.html

Centers for Disease Control and Prevention. (2017). Data Access – National Death Index. Retrieved February 24, 2017, from https://www.cdc.gov/nchs/ndi/

Charles, C., Gafni, A., & Freeman, E. (2011). The evidence-based medicine model of clinical practice: scientific teaching or belief-based preaching? *Journal of Evaluation in Clinical Practice*, *17*(4), 597–605. https://doi.org/10.1111/j.1365-2753.2010.01562.x

Chien, H.-H., Chang, C.-T., Chu, N.-F., Hsieh, S.-H., Huang, Y.-Y., Lee, I.-T., … Sheu, W. H. H. (2007). Effect of glyburide-metformin combination tablet in patients with type 2 diabetes. *Journal of the Chinese Medical Association*, *70*(11), 473–480. https://doi.org/10.1016/S1726-4901(08)70044-3

Clark, H. D., Wells, G. A., Huët, C., McAlister, F. A., Salmi, L. R., Fergusson, D., & Laupacis, A. (1999). Assessing the quality of randomized trials: reliability of the Jadad scale. *Controlled Clinical Trials*, *20*(5), 448–452.

Cohen, J. (1994). The earth is round ($p < .05$). *American Psychologist*, *49*(12), 997–1003. https://doi.org/10.1037/0003-066X.49.12.997

Cole, S. R., & Hernán, M. A. (2008). Constructing Inverse Probability Weights for Marginal Structural Models. *American Journal of Epidemiology*, *168*(6), 656–664. https://doi.org/10.1093/aje/kwn164

Corrao, G., Romio, S. A., Zambon, A., Merlino, L., Bosi, E., & Scavini, M. (2011). Multiple outcomes associated with the use of metformin and sulphonylureas in type 2 diabetes: a population-based cohort study in Italy. *European Journal of Clinical Pharmacology*, *67*(3), 289–299. https://doi.org/10.1007/s00228-010-0939-6

Currie, C. J., Poole, C. D., Evans, M., Peters, J. R., & Morgan, C. L. (2013). Mortality and other important diabetes-related outcomes with insulin vs. other antihyperglycemic therapies in type 2 diabetes. *The Journal of Clinical Endocrinology and Metabolism*, *98*(2), 668–677. https://doi.org/10.1210/jc.2012-3042

Curtis, L. H., Hammill, B. G., Eisenstein, E. L., Kramer, J. M., & Anstrom, K. J. (2007). Using Inverse Probability-Weighted Estimators in Comparative Effectiveness Analyses With Observational Databases: *Medical Care*, *45*(Suppl. 2), S103–S107. https://doi.org/10.1097/MLR.0b013e31806518ac

DeFronzo, R. A., & Goodman, A. M. (1995). Efficacy of Metformin in Patients with Non-Insulin-Dependent Diabetes Mellitus. *New England Journal of Medicine*, *333*(9), 541–549. https://doi.org/10.1056/NEJM199508313330902

Del Prato, S., Camisasca, R., Wilson, C., & Fleck, P. (2014). Durability of the efficacy and safety of alogliptin compared with glipizide in type 2 diabetes mellitus: a 2-year study. *Diabetes, Obesity & Metabolism*, *16*(12), 1239–1246. https://doi.org/10.1111/dom.12377

Del Prato, S., Nauck, M., Durán-Garcia, S., Maffei, L., Rohwedder, K., Theuerkauf, A., & Parikh, S. (2015). Long-term glycaemic response and tolerability of dapagliflozin versus a sulphonylurea as add-on therapy to metformin in patients with type 2 diabetes: 4-year data. *Diabetes, Obesity & Metabolism*, *17*(6), 581–590. https://doi.org/10.1111/dom.12459

Egger, M., Davey Smith, G., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *BMJ: British Medical Journal*, *315*(7109), 629–634.

Eurich, D. T., Majumdar, S. R., McAlister, F. A., Tsuyuki, R. T., & Johnson, J. A. (2005). Improved clinical outcomes associated with metformin in patients with diabetes and heart failure. *Diabetes Care*, *28*(10), 2345–2351.

Evans, J. M. M., Ogston, S. A., Emslie-Smith, A., & Morris, A. D. (2006). Risk of mortality and adverse cardiovascular outcomes in type 2 diabetes: a comparison of patients treated with sulfonylureas and metformin. *Diabetologia*, *49*(5), 930–936. https://doi.org/10.1007/s00125-006-0176-9

Ferrannini, E., Fonseca, V., Zinman, B., Matthews, D., Ahrén, B., Byiers, S., … Dejager, S. (2009). Fifty-two-week efficacy and safety of vildagliptin vs. glimepiride in patients with type 2 diabetes mellitus inadequately controlled on metformin monotherapy. *Diabetes, Obesity & Metabolism*, *11*(2), 157–166. https://doi.org/10.1111/j.1463-1326.2008.00994.x

Filozof, C., & Gautier, J.-F. (2010). A comparison of efficacy and safety of vildagliptin and gliclazide in combination with metformin in patients with Type 2 diabetes inadequately controlled with metformin alone: a 52-week, randomized study. *Diabetic Medicine*, *27*(3), 318–326. https://doi.org/10.1111/j.1464-5491.2010.02938.x

Freiman, J. A., Chalmers, T. C., Smith, H., & Kuebler, R. R. (1978). The Importance of Beta, the Type II Error and Sample Size in the Design and Interpretation of the Randomized Control Trial. *New England Journal of Medicine*, *299*(13), 690–694. https://doi.org/10.1056/NEJM197809282991304

Gallwitz, B., Rosenstock, J., Rauch, T., Bhattacharya, S., Patel, S., von Eynatten, M., … Woerle, H.-J. (2012). 2-year efficacy and safety of linagliptin compared with glimepiride in patients with type 2 diabetes inadequately controlled on metformin: a randomised, double-blind, non-inferiority trial. *Lancet*, *380*(9840), 475–483. https://doi.org/10.1016/S0140-6736(12)60691-6

Garber, Abrahamson, M. J., Barzilay, J. I., Blonde, L., Bloomgarden, Z. T., Bush, M. A., … Umpierrez, G. E. (2017). Consensus Statement by The American Association of Clinical Endocrinologists and American College of Endocrinology on the Comprehensive Type 2 Diabetes Management Algorithm – 2017 Executive Summary. *Endocrine Practice*, *23*(2), 207–238. https://doi.org/10.4158/EP161682.CS

Garber, A., Henry, R., Ratner, R., Garcia-Hernandez, P. A., Rodriguez-Pattzi, H., Olvera-Alvarez, I., … Bode, B. (2009). Liraglutide versus glimepiride monotherapy for type 2 diabetes (LEAD-3 Mono): a randomised, 52-week, phase III, double-blind, parallel-treatment trial. *Lancet*, *373*(9662), 473–481. https://doi.org/10.1016/S0140-6736(08)61246-5

Garber, A. J., Donovan, D. S., Dandona, P., Bruce, S., & Park, J.-S. (2003). Efficacy of Glyburide/Metformin Tablets Compared with Initial Monotherapy in Type 2 Diabetes. *The Journal of Clinical Endocrinology & Metabolism*, *88*(8), 3598–3604. https://doi.org/10.1210/jc.2002-021225

Garrido, M. M., Kelley, A. S., Paris, J., Roza, K., Meier, D. E., Morrison, R. S., & Aldridge, M. D. (2014). Methods for Constructing and Assessing Propensity Scores. *Health Services Research*, *49*(5), 1701–1720. https://doi.org/10.1111/1475-6773.12182

Geiss LS, Wang J, Cheng YJ, & et al. (2014). Prevalence and incidence trends for diagnosed diabetes among adults aged 20 to 79 years, United States, 1980–2012. *JAMA: The Journal of the American Medical Association*, *312*(12), 1218–1226. https://doi.org/10.1001/jama.2014.11494

Genuth, S. (2015). Should Sulfonylureas Remain an Acceptable First-Line Add-on to Metformin Therapy in Patients With Type 2 Diabetes? No, It's Time to Move On! *Diabetes Care*, *38*(1), 170–175. https://doi.org/10.2337/dc14-0565

Gerich, J., Raskin, P., Jean-Louis, L., Purkayastha, D., & Baron, M. A. (2005). PRESERVE-beta: two-year efficacy and safety of initial combination therapy with nateglinide or glyburide plus metformin. *Diabetes Care*, *28*(9), 2093–2099.

Go, A. S., Mozaffarian, D., Roger, V. L., Benjamin, E. J., Berry, J. D., Blaha, M. J., … Turner, M. B. (2014). Heart Disease and Stroke Statistics—2014 Update A Report From the American Heart Association. *Circulation*, *129*(3), e28–e292. https://doi.org/10.1161/01.cir.0000441139.02102.80

Göke, B., Gallwitz, B., Eriksson, J. G., Hellqvist, Å., & Gause-Nilsson, I. (2013). Saxagliptin vs. glipizide as add-on therapy in patients with type 2 diabetes mellitus inadequately controlled on metformin alone: long-term (52-week) extension of a 52-week randomised controlled trial. *International Journal of Clinical Practice*, *67*(4), 307–316.

Göke, B., Gallwitz, B., Eriksson, J., Hellqvist, A., Gause-Nilsson, I., & D1680C00001 Investigators. (2010). Saxagliptin is non-inferior to glipizide in patients with type 2 diabetes mellitus inadequately controlled on metformin alone: a 52-week randomised controlled trial. *International Journal of Clinical Practice*, *64*(12), 1619–1631. https://doi.org/10.1111/j.1742-1241.2010.02510.x

Goldstein, M. (2006). Subjective Bayesian Analysis: Principles and Practice. *Bayesian Analysis*, *1*(3), 403–420. https://doi.org/10.1214/06-BA116

Goodman, S. N. (2013). Bayesian Methods for Evidence Evaluation. *Circulation*, *127*(24), 2367–2369. https://doi.org/10.1161/CIRCULATIONAHA.113.003193

Greevy, R. A., Huizinga, M. M., Roumie, C. L., Grijalva, C. G., Murff, H., Liu, X., & Griffin, M. R. (2011). Comparisons of persistence and durability among three oral antidiabetic therapies using electronic prescription-fill data: the impact of adherence requirements and stockpiling. *Clinical Pharmacology and Therapeutics*, *90*(6), 813–819. https://doi.org/10.1038/clpt.2011.228

Gulliford, M., & Latinovic, R. (2004). Mortality in type 2 diabetic subjects prescribed metformin and sulphonylurea drugs in combination: cohort study. *Diabetes/Metabolism Research and Reviews*, *20*(3), 239–245. https://doi.org/10.1002/dmrr.457

Hamann, A., Garcia-Puig, J., Paul, G., Donaldson, J., & Stewart, M. (2008). Comparison of fixed-dose rosiglitazone/metformin combination therapy with sulphonylurea plus metformin in overweight individuals with Type 2 diabetes inadequately controlled on metformin alone. *Experimental and Clinical Endocrinology & Diabetes*, *116*(1), 6–13. https://doi.org/10.1055/s-2007-984441

Hanefeld, M., Patwardhan, R., & Jones, N. P. (2007). A one-year study comparing the efficacy and safety of rosiglitazone and glibenclamide in the treatment of type 2 diabetes. *Nutrition, Metabolism, and Cardiovascular Diseases*, *17*(1), 13–23. https://doi.org/10.1016/j.numecd.2005.12.003

Harbord, R. M., Harris, R. J., & Sterne, J. A. C. (2009). Updated tests for small-study effects in meta-analyses. *Stata Journal*, *9*(2), 197–210.

Harder, V. S., Stuart, E. A., & Anthony, J. C. (2010). Propensity score techniques and the assessment of measured covariate balance to test causal associations in psychological research. *Psychological Methods*, *15*(3), 234–249. https://doi.org/10.1037/a0019623

Hemmingsen, B., Schroll, J. B., Lund, S. S., Wetterslev, J., Gluud, C., Vaag, A., … Almdal, T. (2013). Sulphonylurea monotherapy for patients with type 2 diabetes mellitus. *The Cochrane Database of Systematic Reviews*, *4*, CD009008. https://doi.org/10.1002/14651858.CD009008.pub2

Hemmingsen, B., Schroll, J. B., Wetterslev, J., Gluud, C., Vaag, A., Sonne, D. P., … Almdal, T. (2014). Sulfonylurea versus metformin monotherapy in patients with type 2 diabetes: a Cochrane systematic review and meta-analysis of randomized clinical trials and trial sequential analysis. *CMAJ Open*, *2*(3), E162–E175. https://doi.org/10.9778/cmajo.20130073

Hermann, L. S., Scherstén, B., Bitzén, P. O., Kjellström, T., Lindgärde, F., & Melander, A. (1994). Therapeutic comparison of metformin and sulfonylurea, alone and in various combinations. A double-blind controlled study. *Diabetes Care*, *17*(10), 1100–1109.

Herman, W. H. (2002). Evidence-Based Diabetes Care. *Clinical Diabetes*, *20*(1), 22–23. https://doi.org/10.2337/diaclin.20.1.22

Hernán, M. A., & Hernández-Díaz, S. (2012). Beyond the intention-to-treat in comparative effectiveness research. *Clinical Trials*, *9*(1), 48–55. https://doi.org/10.1177/1740774511420743

Higgins, J. P. T., & Green, S. (2011). *Cochrane Handbook for Systematic Reviews of Interventions* (5.1.0 [updated March 2011]). John Wiley & Sons. Retrieved from www.handbook.cochrane.org

Higgins, J. P. T., Thompson, S. G., Deeks, J. J., & Altman, D. G. (2003). Measuring inconsistency in meta-analyses. *BMJ: British Medical Journal*, *327*(7414), 557–560. https://doi.org/10.1136/bmj.327.7414.557

Holman, R. R., Paul, S. K., Bethel, M. A., Matthews, D. R., & Neil, H. A. W. (2008). 10-Year Follow-up of Intensive Glucose Control in Type 2 Diabetes. *New England Journal of Medicine*, *359*(15), 1577–1589. https://doi.org/10.1056/NEJMoa0806470

Hsiao, F.-Y., Huang, W.-F., Wen, Y.-W., Chen, P.-F., Kuo, K. N., & Tsai, Y.-W. (2009). Thiazolidinediones and cardiovascular events in patients with type 2 diabetes mellitus: a retrospective cohort study of over 473,000 patients using the National Health Insurance database in Taiwan. *Drug Safety*, *32*(8), 675–690. https://doi.org/10.2165/00002018-200932080-00006

Hung, Y.-C., Lin, C.-C., Wang, T.-Y., Chang, M.-P., Sung, F.-C., & Chen, C.-C. (2013). Oral hypoglycaemic agents and the development of non-fatal cardiovascular events in patients with type 2 diabetes mellitus. *Diabetes/Metabolism Research and Reviews*, *29*(8), 673–679. https://doi.org/10.1002/dmrr.2444

International Diabetes Federation Guideline Development Group. (2014). Global Guideline for Type 2 Diabetes. *Diabetes Research and Clinical Practice*, *104*(1), 1–52. https://doi.org/10.1016/j.diabres.2012.10.001

Inzucchi, S. E., Bergenstal, R. M., Buse, J. B., Diamant, M., Ferrannini, E., Nauck, M., … Matthews, D. R. (2012). Management of Hyperglycemia in Type 2 Diabetes: A Patient-Centered Approach Position Statement of the American Diabetes Association (ADA) and the European Association for the Study of Diabetes (EASD). *Diabetes Care*, *35*(6), 1364–1379. https://doi.org/10.2337/dc12-0413

Inzucchi, S. E., Bergenstal, R. M., Buse, J. B., Diamant, M., Ferrannini, E., Nauck, M., … Matthews, D. R. (2015). Management of Hyperglycemia in Type 2 Diabetes, 2015: A Patient-Centered Approach: Update to a Position Statement of the American Diabetes Association and the European Association for the Study of Diabetes. *Diabetes Care*, *38*(1), 140–149. https://doi.org/10.2337/dc14-2441

Jadad, A. R., Haynes, R. B., Hunt, D., & Browman, G. P. (2000). The Internet and evidence-based decision-making: a needed synergy for efficient knowledge management in health care. *CMAJ: Canadian Medical Association Journal*, *162*(3), 362–365.

Jadad, A. R., Moore, R. A., Carroll, D., Jenkinson, C., Reynolds, D. J., Gavaghan, D. J., & McQuay, H. J. (1996). Assessing the quality of reports of randomized clinical trials: is blinding necessary? *Controlled Clinical Trials*, *17*(1), 1–12.

Jain, R., Osei, K., Kupfer, S., Perez, A. T., & Zhang, J. (2006). Long-term safety of pioglitazone versus glyburide in patients with recently diagnosed type 2 diabetes mellitus. *Pharmacotherapy*, *26*(10), 1388–1395. https://doi.org/10.1592/phco.26.10.1388

Johnson, J. A., Majumdar, S. R., Simpson, S. H., & Toth, E. L. (2002). Decreased mortality associated with the use of metformin compared with sulfonylurea monotherapy in type 2 diabetes. *Diabetes Care*, *25*(12), 2244–2248.

Johnson, J. A., Simpson, S. H., Toth, E. L., & Majumdar, S. R. (2005). Reduced cardiovascular morbidity and mortality associated with metformin use in subjects with Type 2 diabetes. *Diabetic Medicine*, *22*(4), 497–502. https://doi.org/10.1111/j.1464-5491.2005.01448.x

Kahler, K. H., Rajan, M., Rhoads, G. G., Safford, M. M., Demissie, K., Lu, S.-E., & Pogach, L. M. (2007). Impact of oral antihyperglycemic therapy on all-cause mortality among patients with diabetes in the Veterans Health Administration. *Diabetes Care*, *30*(7), 1689–1693. https://doi.org/10.2337/dc06-2272

Kahn, S. E., Haffner, S. M., Heise, M. A., Herman, W. H., Holman, R. R., Jones, N. P., … ADOPT Study Group. (2006). Glycemic durability of rosiglitazone, metformin, or glyburide monotherapy. *New England Journal of Medicine*, *355*(23), 2427–2443. https://doi.org/10.1056/NEJMoa066224

Kannan, S., Pantalone, K. M., Matsuda, S., Wells, B. J., Karafa, M., & Zimmerman, R. S. (2015). Risk of overall mortality and cardiovascular events in patients with type 2 diabetes on dual drug therapy including metformin: A large database study from the Cleveland Clinic. *Journal of Diabetes*. https://doi.org/10.1111/1753-0407.12301

King, G., & Nielsen, R. (2016). *Why Propensity Scores Should Not Be Used for Matching*. Retrieved from http://scholar.harvard.edu/files/gking/files/psnot.pdf

Kruschke, J. K. (2010). What to believe: Bayesian methods for data analysis. *Trends in Cognitive Sciences*, *14*(7), 293–300. https://doi.org/10.1016/j.tics.2010.05.001

Leiter, L. A., Yoon, K.-H., Arias, P., Langslet, G., Xie, J., Balis, D. A., … Meininger, G. (2015). Canagliflozin provides durable glycemic improvements and body weight reduction over 104 weeks versus glimepiride in patients with type 2 diabetes on metformin: a randomized, double-blind, phase 3 study. *Diabetes Care*, *38*(3), 355–364. https://doi.org/10.2337/dc13-2762

Lipska KJ, & Krumholz HM. (2014). Comparing diabetes medications: Where do we set the bar? *JAMA Internal Medicine*, *174*(3), 317–318. https://doi.org/10.1001/jamainternmed.2013.13433

Li, Y., Hu, Y., Ley, S. H., Rajpathak, S., & Hu, F. B. (2014). Sulfonylurea use and incident cardiovascular disease among patients with type 2 diabetes: prospective cohort study among women. *Diabetes Care*, *37*(11), 3106–3113. https://doi.org/10.2337/dc14-1306

Lo, C. K.-L., Mertz, D., & Loeb, M. (2014). Newcastle-Ottawa Scale: comparing reviewers' to authors' assessments. *BMC Medical Research Methodology*, *14*, 45. https://doi.org/10.1186/1471-2288-14-45

Management of Diabetes Mellitus Update Working Group. (2010). *VA/DoD Clinical Practice Guidelines for the Management of diabetes Mellitus*. Washington, DC: Veterans Health Administration and Department of Defense. Retrieved from http://www.healthquality.va.gov/guidelines/CD/diabetes/DM2010_FUL-v4e.pdf

Marathe, P. H., Gao, H. X., & Close, K. L. (2017). American Diabetes Association Standards of Medical Care in Diabetes 2017. *Journal of Diabetes*. https://doi.org/10.1111/1753-0407.12524

Matthews, D. R., Charbonnel, B. H., Hanefeld, M., Brunetti, P., & Schernthaner, G. (2005). Long-term therapy with addition of pioglitazone to metformin compared with the addition of gliclazide to metformin in patients with type 2 diabetes: a randomized, comparative study. *Diabetes/Metabolism Research and Reviews*, *21*(2), 167–174. https://doi.org/10.1002/dmrr.478

Matthews, D. R., Dejager, S., Ahren, B., Fonseca, V., Ferrannini, E., Couturier, A., … Zinman, B. (2010). Vildagliptin add-on to metformin produces similar efficacy and reduced hypoglycaemic risk compared with glimepiride, with no weight gain: results from a 2-year study. *Diabetes, Obesity & Metabolism*, *12*(9), 780–789. https://doi.org/10.1111/j.1463-1326.2010.01233.x

McAfee, A. T., Koro, C., Landon, J., Ziyadeh, N., & Walker, A. M. (2007). Coronary heart disease outcomes in patients receiving antidiabetic agents. *Pharmacoepidemiology and Drug Safety*, *16*(7), 711–725. https://doi.org/10.1002/pds.1443

Meinert, C. L., Knatterud, G. L., Prout, T. E., & Klimt, C. R. (1970). A study of the effects of hypoglycemic agents on vascular complications in patients with adult-onset diabetes. II. Mortality results. *Diabetes*, *19*, Suppl:789–830.

Miller, D. R., & Pogach, L. (2008). Longitudinal Approaches to Evaluate Health Care Quality and Outcomes: The Veterans Health Administration Diabetes Epidemiology Cohorts. *Journal of Diabetes Science and Technology (Online)*, *2*(1), 24–32.

Mogensen, U. M., Andersson, C., Fosbøl, E. L., Schramm, T. K., Vaag, A., Scheller, N. M., … Køber, L. (2014). Cardiovascular safety of combination therapies with

incretin-based drugs and metformin compared with a combination of metformin and sulphonylurea in type 2 diabetes mellitus: A retrospective nationwide study. *Diabetes, Obesity & Metabolism*, *16*(10), 1001–1008. https://doi.org/10.1111/dom.12314

Morgan, C. L., Mukherjee, J., Jenkins-Jones, S., Holden, S. E., & Currie, C. J. (2014a). Association between first-line monotherapy with sulphonylurea versus metformin and risk of all-cause mortality and cardiovascular events: a retrospective, observational study. *Diabetes, Obesity and Metabolism*, *16*(10), 957–962. https://doi.org/10.1111/dom.12302

Morgan, C. L., Mukherjee, J., Jenkins-Jones, S., Holden, S. E., & Currie, C. J. (2014b). Combination therapy with metformin plus sulphonylureas versus metformin plus DPP-4 inhibitors: association with major adverse cardiovascular events and all-cause mortality. *Diabetes, Obesity and Metabolism*, *16*(10), 977–983. https://doi.org/10.1111/dom.12306

Morgan, C. L., Poole, C. D., Evans, M., Barnett, A. H., Jenkins-Jones, S., & Currie, C. J. (2012). What next after metformin? A retrospective evaluation of the outcome of second-line, glucose-lowering therapies in people with type 2 diabetes. *The Journal of Clinical Endocrinology and Metabolism*, *97*(12), 4605–4612. https://doi.org/10.1210/jc.2012-3034

Nathan, D. M., Buse, J. B., Kahn, S. E., Krause-Steinrauf, H., Larkin, M. E., Staten, M., … GRADE Study Research Group. (2013). Rationale and design of the glycemia reduction approaches in diabetes: a comparative effectiveness study (GRADE). *Diabetes Care*, *36*(8), 2254–2261. https://doi.org/10.2337/dc13-0356

National Institute for Health and Clinical Excellence. (2014). *Type 2 Diabetes: The management of type 2 diabetes - NICE Clinical Guideline 87*. London: Royal College of Physicians (UK). Retrieved from http://www.ncbi.nlm.nih.gov/books/NBK53885/

Olsson, J., Lindberg, G., Gottsäter, M., Lindwall, K., Sjöstrand, A., Tisell, A., & Melander, A. (2000). Increased mortality in Type II diabetic patients using sulphonylurea and metformin in combination: a population-based observational study. *Diabetologia*, *43*(5), 558–560. https://doi.org/10.1007/s001250051343

Ou, S.-M., Shih, C.-J., Chao, P.-W., Chu, H., Kuo, S.-C., Lee, Y.-J., … Chen, Y.-T. (2015). Effects on Clinical Outcomes of Adding Dipeptidyl Peptidase-4 Inhibitors Versus Sulfonylureas to Metformin Therapy in Patients With Type 2 Diabetes Mellitus. *Annals of Internal Medicine*, *163*(9), 663–672. https://doi.org/10.7326/M15-0308

Pantalone, K. M., Kattan, M. W., Yu, C., Wells, B. J., Arrigain, S., Jain, A., … Zimmerman, R. S. (2009). The risk of developing coronary artery disease or

congestive heart failure, and overall mortality, in type 2 diabetic patients receiving rosiglitazone, pioglitazone, metformin, or sulfonylureas: a retrospective analysis. *Acta Diabetologica*, *46*(2), 145–154. https://doi.org/10.1007/s00592-008-0090-3

Patorno, E., Patrick, A. R., Garry, E. M., Schneeweiss, S., Gillet, V. G., Bartels, D. B., … Seeger, J. D. (2014). Observational studies of the association between glucose-lowering medications and cardiovascular outcomes: addressing methodological limitations. *Diabetologia*, *57*(11), 2237–2250. https://doi.org/10.1007/s00125-014-3364-z

Perriello, G., Pampanelli, S., Di Pietro, C., Brunetti, P., & Italian Pioglitazone Study Group. (2006). Comparison of glycaemic control over 1 year with pioglitazone or gliclazide in patients with Type 2 diabetes. *Diabetic Medicine*, *23*(3), 246–252. https://doi.org/10.1111/j.1464-5491.2006.01801.x

Peters, J. L., Rushton, L., Sutton, A. J., Jones, D. R., Abrams, K. R., & Mugglestone, M. A. (2005). Bayesian methods for the cross-design synthesis of epidemiological and toxicological evidence. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, *54*(1), 159–172. https://doi.org/10.1111/j.1467-9876.2005.00476.x

Prentice, J. C., Conlin, P. R., Gellad, W. F., Edelman, D., Lee, T. A., & Pizer, S. D. (2014). Capitalizing on prescribing pattern variation to compare medications for type 2 diabetes. *Value in Health*, *17*(8), 854–862. https://doi.org/10.1016/j.jval.2014.08.2674

Qaseem A, Barry MJ, Humphrey LL, Forciea M, & for the Clinical Guidelines Committee of the American College of Physicians. (2017). Oral pharmacologic treatment of type 2 diabetes mellitus: A clinical practice guideline update from the american college of physicians. *Annals of Internal Medicine*, *166*(4), 279–290. https://doi.org/10.7326/M16-1860

Qaseem, A., Humphrey, L. L., Sweet, D. E., Starkey, M., & Shekelle, P. (2012). Oral Pharmacologic Treatment of Type 2 Diabetes Mellitus: A Clinical Practice Guideline From the American College of Physicians. *Annals of Internal Medicine*, *156*(3), 218–231. https://doi.org/10.7326/0003-4819-156-3-201202070-00011

Reiber, G. E., Koepsell, T. D., Maynard, C., Haas, L. B., & Boyko, E. J. (2004). Diabetes in Nonveterans, Veterans, and Veterans Receiving Department of Veterans Affairs Health Care. *Diabetes Care*, *27*(Suppl. 2), b3–b9. https://doi.org/10.2337/diacare.27.suppl_2.B3

Ridderstråle, M., Andersen, K. R., Zeller, C., Kim, G., Woerle, H. J., Broedl, U. C., & EMPA-REG H2H-SU trial investigators. (2014). Comparison of empagliflozin and glimepiride as add-on to metformin in patients with type 2 diabetes: a 104-

week randomised, active-controlled, double-blind, phase 3 trial. *The Lancet. Diabetes & Endocrinology*, *2*(9), 691–700. https://doi.org/10.1016/S2213-8587(14)70120-2

Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, *70*(1), 41–55. https://doi.org/10.1093/biomet/70.1.41

Rosenbaum, P. R., & Rubin, D. B. (1984). Reducing Bias in Observational Studies Using Subclassification on the Propensity Score. *Journal of the American Statistical Association*, *79*(387), 516–524. https://doi.org/10.1080/01621459.1984.10478078

Rosenberg, W., & Haynes, R. B. with Sackett, D. L., Straus Sharon E. and Richardson W. Scott. (2000). *Evidence-Based Medicine: How to Practice and Teach EBM* (2 edition). Edinburgh; New York: Churchill Livingstone.

Rosenstock, J., Wilson, C., & Fleck, P. (2013). Alogliptin versus glipizide monotherapy in elderly type 2 diabetes mellitus patients with mild hyperglycaemia: a prospective, double-blind, randomized, 1-year study. *Diabetes, Obesity & Metabolism*, *15*(10), 906–914. https://doi.org/10.1111/dom.12102

Roumie, C. L., Hung, A. M., Greevy, R. A., Grijalva, C. G., Liu, X., Murff, H. J., … Griffin, M. R. (2012). Comparative effectiveness of sulfonylurea and metformin monotherapy on cardiovascular events in type 2 diabetes mellitus: a cohort study. *Annals of Internal Medicine*, *157*(9), 601–610. https://doi.org/10.7326/0003-4819-157-9-201211060-00003

Rubin, D. B. (2007). The design versus the analysis of observational studies for causal effects: parallels with the design of randomized trials. *Statistics in Medicine*, *26*(1), 20–36. https://doi.org/10.1002/sim.2739

Ryan, D. H., Espeland, M. A., Foster, G. D., Haffner, S. M., Hubbard, V. S., Johnson, K. C., … Look AHEAD Research Group. (2003). Look AHEAD (Action for Health in Diabetes): design and methods for a clinical trial of weight loss for the prevention of cardiovascular disease in type 2 diabetes. *Controlled Clinical Trials*, *24*(5), 610–628.

Sackett, D. L., Rosenberg, W. M. C., Gray, J. A. M., Haynes, R. B., & Richardson, W. S. (1996). Evidence based medicine: what it is and what it isn't. *BMJ: British Medical Journal*, *312*(7023), 71–72. https://doi.org/10.1136/bmj.312.7023.71

SAS Institute Inc. (2010). PROC PHREG: Bayesian Analysis :: SAS/STAT(R) 9.22 User's Guide. Retrieved December 8, 2016, from https://support.sas.com/documentation/cdl/en/statug/63347/HTML/default/viewer.htm#statug_phreg_sect004.htm

Sauerland, S., & Seiler, C. M. (2005). Role of systematic reviews and meta-analysis in evidence-based medicine. *World Journal of Surgery*, *29*(5), 582–587. https://doi.org/10.1007/s00268-005-7917-7

Schramm, T. K., Gislason, G. H., Vaag, A., Rasmussen, J. N., Folke, F., Hansen, M. L., … Torp-Pedersen, C. (2011). Mortality and cardiovascular risk associated with different insulin secretagogues compared with metformin in type 2 diabetes, with or without a previous myocardial infarction: a nationwide study. *European Heart Journal*, *32*(15), 1900–1908. https://doi.org/10.1093/eurheartj/ehr077

Schwarz, S. L., Gerich, J. E., Marcellari, A., Jean-Louis, L., Purkayastha, D., & Baron, M. A. (2008). Nateglinide, alone or in combination with metformin, is effective and well tolerated in treatment-naïve elderly patients with type 2 diabetes. *Diabetes, Obesity & Metabolism*, *10*(8), 652–660. https://doi.org/10.1111/j.1463-1326.2007.00792.x

Scott, R., Wu, M., Sanchez, M., & Stein, P. (2007). Efficacy and tolerability of the dipeptidyl peptidase-4 inhibitor sitagliptin as monotherapy over 12 weeks in patients with type 2 diabetes. *International Journal of Clinical Practice*, *61*(1), 171–180. https://doi.org/10.1111/j.1742-1241.2006.01246.x

Seaquist, E. R., Anderson, J., Childs, B., Cryer, P., Dagogo-Jack, S., Fish, L., … Vigersky, R. (2013). Hypoglycemia and Diabetes: A Report of a Workgroup of the American Diabetes Association and The Endocrine Society. *Diabetes Care*, *36*(5), 1384–1395. https://doi.org/10.2337/dc12-2480

Seck, T., Nauck, M., Sheng, D., Sunga, S., Davies, M. J., Stein, P. P., … for the Sitagliptin Study 024 Group. (2010). Safety and efficacy of treatment with sitagliptin or glipizide in patients with type 2 diabetes inadequately controlled on metformin: a 2-year study. *International Journal of Clinical Practice*, *64*(5), 562–576. https://doi.org/10.1111/j.1742-1241.2010.02353.x

Seino, Y., Min, K. W., Niemoeller, E., & Takami, A. (2012). Randomized, double-blind, placebo-controlled trial of the once-daily GLP-1 receptor agonist lixisenatide in Asian patients with type 2 diabetes insufficiently controlled on basal insulin with or without a sulfonylurea (GetGoal-L-Asia). *Diabetes, Obesity & Metabolism*, *14*(10), 910–917. https://doi.org/10.1111/j.1463-1326.2012.01618.x

Sillars, B., Davis, W. A., Hirsch, I. B., & Davis, T. M. E. (2010). Sulphonylurea-metformin combination therapy, cardiovascular disease and all-cause mortality: the Fremantle Diabetes Study. *Diabetes, Obesity & Metabolism*, *12*(9), 757–765. https://doi.org/10.1111/j.1463-1326.2010.01230.x

Simpson, S. H., Majumdar, S. R., Tsuyuki, R. T., Eurich, D. T., & Johnson, J. A. (2006). Dose-response relation between sulfonylurea drugs and mortality in type 2 diabetes mellitus: a population-based cohort study. *CMAJ: Canadian Medical*

Association Journal = Journal de l'Association Medicale Canadienne, *174*(2), 169–174. https://doi.org/10.1503/cmaj.050748

StataCorp. (2015). Stata Statistical Software (Version Release 14). College Station, TX: StataCorp LP.

St John Sutton, M., Rendell, M., Dandona, P., Dole, J. F., Murphy, K., Patwardhan, R., … Freed, M. (2002). A comparison of the effects of rosiglitazone and glyburide on cardiovascular function and glycemic control in patients with type 2 diabetes. *Diabetes Care*, *25*(11), 2058–2064.

Straus, S. E., Richardson, W. S., Glasziou, P., & Haynes, R. B. (2005). Evidence-based medicine: How to practice and teach EBM. Retrieved from http://works.bepress.com/paul_glasziou/25

Stuart, E. A., Lee, B. K., & Leacy, F. P. (2013). Prognostic score-based balance measures can be a useful diagnostic for propensity score methods in comparative effectiveness research. *Journal of Clinical Epidemiology*, *66*(8 Suppl), S84–S90.e1. https://doi.org/10.1016/j.jclinepi.2013.01.013

Sullivan, D., Forder, P., Simes, J., Whiting, M., Kritharides, L., Merrifield, A., … FIELD Study Investigators. (2011). Associations between the use of metformin, sulphonylureas, or diet alone and cardiovascular outcomes in 6005 people with type 2 diabetes in the FIELD study. *Diabetes Research and Clinical Practice*, *94*(2), 284–290. https://doi.org/10.1016/j.diabres.2011.07.028

Tan, M. H., Johns, D., Strand, J., Halse, J., Madsbad, S., Eriksson, J. W., … GLAC Study Group. (2004). Sustained effects of pioglitazone vs. glibenclamide on insulin sensitivity, glycaemic control, and lipid profiles in patients with Type 2 diabetes. *Diabetic Medicine*, *21*(8), 859–866. https://doi.org/10.1111/j.1464-5491.2004.01258.x

Trafimow, D. (2003). Hypothesis testing and theory evaluation at the boundaries: Surprising insights from Bayes's theorem. *Psychological Review*, *110*(3), 526–535. https://doi.org/10.1037/0033-295X.110.3.526

Tzoulaki, I., Molokhia, M., Curcin, V., Little, M. P., Millett, C. J., Ng, A., … Elliott, P. (2009). Risk of cardiovascular disease and all cause mortality among patients with type 2 diabetes prescribed oral antidiabetes drugs: retrospective cohort study using UK general practice research database. *BMJ: British Medical Journal* , *339*, b4731.

UKPDS-33. (1998). Intensive blood-glucose control with sulphonylureas or insulin compared with conventional treatment and risk of complications in patients with type 2 diabetes (UKPDS 33). UK Prospective Diabetes Study (UKPDS) Group. *Lancet*, *352*(9131), 837–853.

UKPDS-34. (1998). Effect of intensive blood-glucose control with metformin on complications in overweight patients with type 2 diabetes (UKPDS 34). UK Prospective Diabetes Study (UKPDS) Group. *Lancet*, *352*(9131), 854–865.

Valentine, J. C., Pigott, T. D., & Rothstein, H. R. (2009). How Many Studies Do You Need? A Primer on Statistical Power for Meta-Analysis. *Journal of Educational and Behavioral Statistics*. https://doi.org/10.3102/1076998609346961

Veterans Health Administration. (2011). Close to 25 percent of VA Patients Have Diabetes - Veterans Health Administration [General Information]. Retrieved October 1, 2015, from http://www.va.gov/health/NewsFeatures/20111115a.asp

Walker, A. M., Koro, C. E., & Landon, J. (2008). Coronary heart disease outcomes in patients receiving antidiabetic agents in the PharMetrics database 2000–2007. *Pharmacoepidemiology and Drug Safety*, *17*(8), 760–768. https://doi.org/10.1002/pds.1598

Wells, G., Shea, B., O'Connell, D., Peterson, J., Welch, V., Losos, M., & Tugwell, P. (2000). The Newcastle-Ottawa Scale (NOS) for assessing the quality of nonrandomised studies in metaanalyses. Retrieved from http://www.medicine.mcgill.ca/rtamblyn/Readings%5CThe%20Newcastle%20-%20Scale%20for%20assessing%20the%20quality%20of%20nonrandomised%20studies%20in%20meta-analyses.pdf

Wheeler, S., Moore, K., Forsberg, C. W., Riley, K., Floyd, J. S., Smith, N. L., & Boyko, E. J. (2013). Mortality among veterans with type 2 diabetes initiating metformin, sulfonylurea or rosiglitazone monotherapy. *Diabetologia*, *56*(9), 1934–1943. https://doi.org/10.1007/s00125-013-2958-1

Woodworth, G. G. (2004). *Biostatistics: A Bayesian Introduction* (1st ed.). Wiley-Interscience.

Xiao, Y., Moodie, E. E. M., & Abrahamowicz, M. (2013). Comparison of Approaches to Weight Truncation for Marginal Structural Cox Models. *Epidemiologic Methods*, *2*(1), 1–20. https://doi.org/10.1515/em-2012-0006

Zhang, Y., Hong, J., Chi, J., Gu, W., Ning, G., & Wang, W. (2014). Head-to-head comparison of dipeptidyl peptidase-IV inhibitors and sulfonylureas: A meta-analysis from randomized clinical trials. *Diabetes/Metabolism Research and Reviews*, *30*(3), 241–256. https://doi.org/10.1002/dmrr.2482

CURRICULUM VITAE