Published in final edited form as:

Curr Diab Rep. 2014 April ; 14(4): 478. doi:10.1007/s11892-014-0478-0.

Systematic Reviews to Ascertain the Safety of Diabetes Medications

Michael R. Gionfriddo, Pharm.D.^{1,2}, Oscar L. Morey-Vargas, M.D.^{1,3}, Juan Pablo Brito, MBBS^{1,3}, Aaron L. Leppin, M.D.¹, M. Hassan Murad, M.D.^{1,4,5}, and Victor M. Montori, M.D.^{1,3,*}

¹Knowledge and Evaluation Research Unit, Mayo Clinic, 200 First Street SW, Rochester, MN, USA

²Mayo Graduate School, Mayo Clinic, 200 First Street SW, Rochester, MN, USA

³Division of Endocrinology, Diabetes, Metabolism, and Nutrition, Mayo Clinic, 200 First Street SW, Rochester, MN, USA

⁴Division of Preventive, Occupational and Aerospace Medicine, Mayo Clinic, 200 First Street SW, Rochester, MN, USA

⁵Division of Health Care Policy and Research, Mayo Clinic, 200 First Street SW, Rochester, MN, USA

Abstract

Clinicians and patients with type 2 diabetes enjoy an expanding list of medications to improve glycemic control. With this expansion has come a flurry of concerns about the safety of these anti-hyperglycemic agents, concerns that affect judgments about the risk/benefit balance of therapy. Some of these safety signals have been identified through the synthesis of existing research evidence. Thus, it has become important for clinicians and clinical policymakers to understand the strengths and limitations of systematic reviews and meta-analyses in determining the safety of diabetes medications. In this paper, we highlight key safety concerns with diabetes medications and discuss the role evidence synthesis plays in each, with special attention to its strengths and limitations.

*Corresponding Author: Victor M. Montori M.D. Division of Endocrinology, Diabetes, Metabolism, and Nutrition, Mayo Clinic, 200 First Street SW, Rochester, MN 55905, USA. Tel.: 1-507-293-0175; fax: 1-507-538-0850. montori.victor@mayo.edu (V.M. Montori). Michael R. Gionfriddo Pharm.D. Knowledge and Evaluation Research Unit, Mayo Clinic, 200 First Street SW, Rochester, MN 55905, USA. Tel.: 1-507-266-1752; fax: 1-507-538-0850. Gionfriddo.michael@mayo.edu (M.R. Gionfriddo). Oscar L. Morey-Vargas M.D. Knowledge and Evaluation Research Unit, Mayo Clinic, 200 First Street SW, Rochester, MN 55905, USA. Tel.: 1-507-266-1746; fax: 1-507-538-0850. moreyvargas.oscar@mayo.edu (O.L. Morey-Vargas). Juan Pablo Brito MBBS. Knowledge and Evaluation Research Unit, Mayo Clinic, 200 First Street SW, Rochester, MN 55905, USA. Tel.: 1-507-266-1746; fax: 1-507-538-0850. brito.juan@mayo.edu (J.P. Brito). Aaron L. Leppin M.D. Knowledge and Evaluation Research Unit, Mayo Clinic, 200 First Street SW, Rochester, MN 55905, USA. Tel.: 1-507-266-1754; fax: 1-507-538-0850. leppin.aaron@mayo.edu (A.L. Leppin). M. Hassan Murad M.D. Knowledge and Evaluation Research Unit, Mayo Clinic, 200 First Street SW, Rochester, MN 55905, USA. Tel.: 1-507-266-1829; fax: 1-507-538-0850. murad.mohammad@mayo.edu (M.H. Murad).

Conflict of Interest

Michael R. Gionfriddo, Oscar L. Morey-Vargas, Juan Pablo Brito, Aaron L. Leppin, M. Hassan Murad, and Victor M. Montori declare that they have no conflict of interest.

Human and Animal Rights and Informed Consent

This article does not contain any studies with human or animal subjects performed by any of the authors.

Keywords

Systematic Review; Meta-Analysis; Diabetes; Safety; Diabetes Medications

Introduction

The past 20 years have seen an explosion in the number of available anti-hyperglycemic options, over a dozen, for patients with type 2 diabetes. This windfall of treatments for diabetes is timely as the incidence of diabetes has also greatly increased over the past two decades [1]. With the benefits in terms of choice and flexibility that arise from the development and approval of more options, have come revelations of unforeseen adverse effects. These revelations have negatively affected both the perceived balance of risk and benefit of these agents and trust in the industrial and regulatory handling of new diabetes medicines. Although hints of harm have often arisen from case reports and observational studies of use in clinical practice, in many notable cases the key signal was identified in evidence syntheses.

Systematic reviews (SR) are studies that systematically search, select, appraise, and summarize the available evidence concerning a focused question. Meta-analysis (MA) is the statistical procedure to pool results across studies. When used together these studies avoid selective review of the evidence, support conclusions based on an assessment of the confidence in estimates of effect, pool results statistically, and offer explanations for inconsistency of results across studies. This method of summarizing evidence has become exponentially common. In August 2013, we found 2050 evidence syntheses published in English and related to type 2 diabetes in the PubMed database; 717 were published from 2011 to August 2013. By comparison, 1224 were published in 2000-2010, and 110 in the 1990-2000 period.

In this paper we will review recent safety concerns with diabetes medications in which evidence syntheses played a key role and use these examples to highlight the strengths and limitations of SRs and MAs as well as the special challenges that clinicians face when trying to determine the safety of diabetes medications.

Metformin and Lactic Acidosis

Metformin, a biguanide, is widely used in the treatment of diabetes and is considered a first line treatment for most patients. Despite a relatively high incidence of bothersome gastrointestinal side effects it is generally considered safe. However, for many years there were lingering concerns about the potential for metformin to cause lactic acidosis. This concern was due to a relatively high incidence of lactic acidosis with phenformin, another biguanide [2]. There were case reports of metformin causing lactic acidosis, but the causality of the association was confounded by the fact that most cases occurred in patients with other causes of lactic acidosis. Therefore, investigators from the Cochrane Collaboration set out to determine the true incidence of lactic acidosis with metformin. What they found was surprising; among prospective trials and observational cohorts, they did not find a single case of lactic acidosis. This SR provided a comprehensive summary of the body of evidence

and provided clinicians with information they can use to inform their practice. In this case, this SR reassured metformin users that this medicine does not appreciably increase the risk of lactic acidosis.

This review and its results can be trusted for a number of reasons. The Cochrane Collaboration always pre-specifies a protocol that is subject to peer review before initiating their reviews. This is an important step as it increases the transparency of the process, the quality and pertinence of the review to its end users, and allows for reproducibility. It also increases trust in the review as any change in the data collection or analysis plans would be noted. Such changes may not be uncommon and could be problematic when they are triggered by inconvenient results.[3] In this way, SRs differ from the more common narrative reviews (i.e., commentaries, editorials, expert topic reviews) in that the former are conducted following an explicit protocol that includes a clearly stated set of objectives, eligibility criteria for studies, a systematic search that attempts to identify all studies that would meet the eligibility criteria, and an explicit and reproducible methodology to apply these criteria and select the studies.[4] Of note, this SR summarized the results of 148 studies containing nearly 63,000 patients. An individual clinician simply does not have time to review and summarize this amount of information. Thus, SRs are particularly useful because they allow access to a concise summary of the body of evidence. These advantages make SRs the preferred tools for guideline panels, policy makers, and clinicians [5].

Agents targeting peroxisome proliferation-activating receptors and cardiovascular risk

Although anti-hyperglycemic agents that act on the nuclear peroxisome proliferation-activating receptor proteins are very effective, they have been plagued by safety signals which have led to restrictions in their use (rosiglitazone and cardiovascular (CV) events), withdrawal from the market (troglitazone and liver failure) or failure to reach the market (muraglitazar and CV events).

Muraglitazar—A Food and Drug Administration (FDA) panel had recommended muraglitazar for approval. As the approval process continued, Nissen et al., using data from phase II and III trials, conducted a MA to determine the CV safety of muraglitazar [6]. They found that in the muraglitazar-treated patients, death, myocardial infarction (MI), or stroke occurred in 35 of 2374 (1.47%) patients and in 9 of 1351 (0.67%) patients in the control group [relative risk (RR) 2.23; 95% confidence interval (CI), 1.07-4.66]. After the publication of these findings, the FDA recommended against approval and the manufacturer ceased further development of muraglitazar. It is interesting to note that Nissen et al. did not perform a SR, but instead based his analysis on briefing documents supplied to the FDA. A MA can be done on any dataset; confidence in its results comes from the way the dataset was assembled. A SR process seeks to produce an unbiased selection of pertinent studies and make these data available to statistical pooling. A MA that results from a SR will be more credible than one that results from an ad-hoc collection of trials. In this case, the assumption is that the trial reports made available to the FDA and the basis for the MA was a complete and unbiased set.

Rosiglitazone—Rosiglitazone was approved for the treatment of diabetes in 1999. Following its approval, its uptake into practice was rapid [7]. However, its approval was based upon the surrogate endpoints of glucose and glycated hemoglobin reduction, the only benefit outcomes the FDA has ever used to approve anti-hyperglycemic agents [8]. As a result, trials designed to seek drug approval did not have the precision required to confidently assess other outcomes of importance to patients with diabetes, such as cardiovascular events. Based on this rationale, Steven Nissen and Kathy Wolski (2 of the 3 authors of the rosiglitazone analysis) conducted a MA to examine the CV safety of rosiglitazone [9]. In their analysis, they included 42 trials obtained from the FDA and industry documents as well as two randomized trials that had been recently published. They found that rosiglitazone, when compared against either placebo or active comparator, was associated with an increased risk of myocardial infarction [odds ratio (OR) 1.43, 95% CI 1.03-1.98] and perhaps death from cardiovascular causes (OR 1.64, 95% CI 0.98-2.74). A few months later, Singh et al. published, not just a MA, but a SR of the same question, i.e. was rosiglitazone associated with an increased risk of CV events.[10] Unlike Nissen and Wolski's MA, Singh et al. chose to restrict their analysis to trials lasting one year or more (Nissen and Wolski restricted their analysis to trials lasting 24 weeks or more) and as a result of this restriction included only 4 trials and found that while rosiglitazone was associated with a 42% increase in the risk myocardial infarction (RR 1.42, 95% CI 1.06-1.91), it was not associated with CV mortality (RR 0.9 95% CI 0.63-1.26). In 2007, the FDA reviewed these data and decided that the evidence was sufficient to warrant a black boxed warning about myocardial ischemia for rosiglitazone.[11] Later, in 2010 they further restricted the use of rosiglitazone, while in Europe, the regulatory body banned the use of rosiglitazone altogether. Also in 2010, Nissen published an update of his earlier MA.[12] This update included 56 trials with a total of over 35,000 randomized patients, nearly 20,000 of which were randomized to rosiglitazone. The analysis (with and without the Rosiglitazone Evaluated for Cardiovascular Outcomes and Regulation of Glycemia in Diabetes (RECORD) trial results) showed the same relationship that the earlier MAs had found, namely an association of rosiglitazone with myocardial infarction (with RECORD OR 1.28, 95% CI 1.02-1.63, without RECORD OR 1.39 95% CI 1.02-1.89) and no significant association between rosiglitazone and CV mortality (with RECORD OR 1.03, 95% CI 0.78-1.36, without RECORD OR 1.46 95% CI 0.92-2.33).

Before we continue with the rosiglitazone saga, it is worth mentioning that despite their different approaches with their inherent strengths and limitations, the Nissen [9, 12] and Singh [10] syntheses reached the same conclusion: rosiglitazone increases the risk for myocardial infarction by about 30-40%. Both studies highlight the ability of MAs to increase the precision of an estimate of effect through statistical pooling of results. In individual studies the confidence interval surrounding the estimate of risk of myocardial infarction was imprecise and did not reach statistical significance. For example, the hazard ratio (HR) for myocardial infarction in the RECORD trial was 1.14 (95% CI 0.80–1.63).[13] However, as shown above, when the results of similar trials were pooled, the number of events available for analysis increased as did our ability to detect signal from noise. As a result an important safety signal for rosiglitazone was detected and resulted in important changes in the risk/benefit balance for rosiglitazone and impacted its use in practice.

A limitation of the first of the two MAs by Nissen et al. [9] was that the quality of the studies included in the analysis was not assessed. In order to properly understand the results of an evidence synthesis, it is important for clinicians to appraise the quality of the evidence synthesized. If the methodological quality of the studies is inadequate then the findings of reviews of this material may also be compromised: garbage-in-garbage-out [14]. To that end, clinicians need to consider features associated with confidence in the estimates of effect: the quality of included studies, their consistency, magnitude and precision of the effect, and directness in relation to the question of interest to decision makers. Table 1 shows the approach used by the GRADE Working Group to ascertain the confidence in the estimates based on these factors [15].

With the evidence syntheses by Nissen [9, 12] and Singh [10] one would think the saga of rosiglitazone would come to an end. In June of 2013, the FDA held a meeting to re-visit the CV safety of rosiglitazone [11]. At the meeting, they evaluated a re-adjudication of the RECORD trial. RECORD compared patients treated with rosiglitazone and metformin or a sulfonylurea, versus metformin and a sulfonylurea [13]. After more than 5 years of follow-up the investigators found no significant association between rosiglitazone arm CV death (HR 0.84, 95% CI 0.59-1.18), myocardial infarction (HR 1.14, 95% CI 0.80-1.63), or stroke (HR 0.72, 95% CI 0.49-1.06). They did however find that rosiglitazone was associated with heart failure (HR 2.10, 95% CI 1.35–3.27), a finding that echoed an earlier finding by Singh, who in his review [10] found that rosiglitazone was also associated with an increased risk of heart failure (RR 2.09; 95% CI 1.52-2.88). Effective interventions to reduce cardiovascular risk were more commonly used in the rosiglitazone arm of the RECORD trial, a situation that often emerges in open label trials. These trials may also introduce bias prior to the ascertainment of outcomes: some RECORD case report forms were differentially submitted for blinded adjudication [16]. In November of 2013, the FDA removed most of the restrictions for rosiglitazone prescription following an independent assessment of the RECORD trial data the results of which contributed to reduce the FDA's concerns about this drug's cardiovascular safety [17].

This scenario highlights the need to examine the whole body of evidence as opposed to single studies, which may mislead the clinician [14]. An example where evaluation of single studies can lead to misleading conclusions is illustrated by the Proteus phenomenon, a situation where early studies of a drug or other intervention show extreme estimates of effect compared to studies published later [18, 19]. In this context, making a decision based on the first or only study conducted, without waiting for subsequent research to replicate or confirm those early findings might lead a clinician to make an invalid conclusion about an intervention's safety or efficacy. This phenomenon ensures that we will need to rely on ongoing or cumulative reviews of the outcomes of using anti-hyperglycemic drugs beyond approval to determine their true value. Evaluating the entire body of evidence often permits a cancellation of the "noise" and provides a firmer sense of what is true signal. It is important to again note that while SRs can improve on the primary studies by virtue of improving their precision, they can help detect other shortcomings such as bias or indirectness (a study that studies patients, interventions or outcomes that are not relevant to the decision maker). While SRs cannot overcome these shortcomings they offer the full

picture, the body of evidence, with its strengths and limitations, facilitating debate about the state of the science [14, 20].

Pioglitazone—We have now seen that rosiglitazone increases the risk of myocardial infarction and heart failure in patient with diabetes, but is this a class effect? Nissen and colleagues examined this question in a MA that used patient-level time-to-event data from randomized trials.[21] They found that pioglitazone compared with control was associated with a reduction in the composite outcome of death, MI, or stroke (HR 0.82; 95% CI 0.72-0.94). Therefore it appears that pioglitazone may actually reduce the risk of this composite outcome. Despite this favorable effect, they did find that the increased risk of heart failure is a class effect: compared to control, pioglitazone is associated with a 41% increase in the risk of serious heart failure (HR 1.41 95% CI 1.14-1.76).

Thiazolidinediones and fractures

In 2006, evidence emerged that TZD's (specifically rosiglitazone) may be associated with an increased risk of fractures.[22] In 2008 a SR was conducted to determine the extent of this risk.[23] The authors of this study identified and analyzed 10 randomized trials involving nearly 14,000 patients. They found that compared with control, TZD's (no breakdown by agent was possible) increased the risk of fractures overall (OR 1.45, 95% CI 1.18–1.79), but there was inconsistency in this effect. The increased risk of fractures with TZDs was significantly greater in women (OR 2.23, 95% CI 1.65–3.01) than in men (OR 1.00, 95% CI 0.73–1.39). The practical implications of this appear to be that clinicians should be cautious when prescribing TZD's to women, especially older postmenopausal women, as they are already at an increased risk of fracture. We can trust these results because they emerge from a comprehensive search strategy which included several databases as well as websites from regulatory bodies, manufacturer's information sheets, and registers of clinical trials. This may help prevent certain forms of bias in the identification and selection of studies and gives the clinician confidence that the authors did not miss any studies as a result of an inadequate search strategy [24-26].

Thiazolidinediones and bladder cancer

In addition to being associated with heart failure and fragility fractures, pioglitazone has been implicated in increasing the risk for bladder cancer. This concern emerged after an increase in the number of cases of bladder cancer was reported among participants exposed to pioglitazone in the PROactive trial [27]. Some regulatory agencies responded by removing it from the market while others, like the FDA, chose to issue warnings and recommend close monitoring [28-30]. Following on this initial evidence, SRs were conducted to clarify this relationship [30, 31]. Colmers et al. in a SR that included both observational and randomized trials found that pioglitazone was associated with an increased risk of bladder cancer [30]. The strength of this association was dependent on study design with the strongest association appearing in case control studies (OR 4.2, 95% CI 2.82-6.52), moderate (yet nonsignificant) in randomized trials (RR 2.36, 95% CI 0.91-6.13), and weakest in cohort studies (RR 1.22, 95% CI 1.07-1.39). Of note, no association between rosiglitazone and bladder cancer was found. Ferwana et al found similar results in a SR of observational and randomized trials, estimating a pooled HR of developing

bladder cancer of 1.23, 95% CI 1.09–1.39 [31]. Figure 1 offers a forest plot – a visual representation of the point estimates and confidence intervals for the included studies in a MA – derived from Ferwana et al.[31] and a guide on how to interpret these plots. Observational studies are subject to a variety of biases, including confounding and selection bias, which limits our confidence in their findings. Thus, their inclusion in the SR would reduce our confidence in their findings, if their results are different from those in the randomized trial. Ferwana et al. also note that one of the included studies may suffer from outcome ascertainment bias as bladder cancer may have been more actively sought out in the observational studies. These examples once again show how evidence syntheses can improve the precision of estimates of effect and increase our confidence in the effect, but cannot overcome challenges in the primary studies.

Glucagon-like Peptide 1-Based Therapies, Pancreatitis, And Pancreatic Cancer

Recent evidence suggests that diabetes medications that target glucagon-like peptide 1 (GLP-1) (dipeptidyl peptidase-4 inhibitors (DPP-4i), and GLP-1 mimetics) may be associated with pancreatitis and pancreatic cancer.[32, 33] Singh et al. in a recent observational study[32] found that these therapies increased the risk of hospitalization due to pancreatitis by 107% (OR 2.07 95% CI 1.36-3.13), a finding that was replicated by Elashoff et al. using the FDA's database of adverse events.[33] They found that, compared to control, sitagliptin and exenatide were associated with a large increase in the risk of pancreatitis (OR 6.7, 95% CI 4.6-10 and 10.7, 95% CI 7.75-15.1, respectively). In addition, Elashoff found a significant association between these medications and pancreatic cancer (OR for sitagliptin 2.7, OR for exenatide 2.9).

In 2013, Monami et al. conducted a SR that only included randomized trials with the goal to clarify the effect of DPP-4i on the risk of pancreatitis and pancreatic cancer [34]. They found that the overall risk of pancreatitis or pancreatic cancer was not different between DPP 4is and control (OR 0.93, 95% CI 0.51–1.69, and OR 0.72, 95% CI 0.32-1.61, respectively). These results are different from those emerging from observational studies. This difference could result from the selection bias and confounding that may affect the observational studies (e.g. obese patients are at higher risk for pancreatitis are also preferentially prescribed these drugs because they are less likely to promote weight gain than alternatives). It could also reflect careful selection of participants for participation into the trials or biased reporting of outcomes [35]. At the time of this writing, the evidence cannot support with confidence a claim of safety of GLP-1 based therapies with regard to pancreatic complications.

Emerging Challenges and Opportunities for Using Systematic Reviews and Meta-Analyses to Evaluate the Safety of Diabetes Medications

Challenges

Conducting SRs and MAs with the goal of examining safety present special challenges [36]. In clinical trials, adverse effects are often collected on standardized report forms. How these

report forms are constructed determines the nature and amount of adverse events recorded. For example, an adverse effect is more likely to be reported if participants in a trial are asked specifically to report about that event's occurrence [37]. When event data collection relies solely on self-reporting, the measured occurrence will be significantly less. Furthermore, adverse events are sometimes classified by organ system, resulting in an unclear picture of the true adverse effect profile for an intervention. For example, when side effects are reported as "gastrointestinal," they may actually include entities as mild as nausea and vomiting or as serious as pancreatitis. This lack of granularity can present a challenge in evidence synthesis because, unless the individual data are available, it can become difficult to accurately represent the true side effect profile of an intervention.

Determining the safety of new drugs is often difficult because of the rarity of certain serious adverse events (e.g. bladder cancer with pioglitazone). On average, before drugs are approved they are tested on approximately 2000 patients [38]. This presents a challenge when trying to determine whether or not a specific drug causes a certain adverse event. For example if a drug is tested in a sample of 2000 patients prior to approval, the odds of seeing an adverse event that occurs in 1%, 0.1%, 0.01%, and 0.001% of exposed patients are 100%, 86%, 18%, and 2% respectively.

To help increase our knowledge about the safety (and effectiveness) profile of new and existing drugs, Phase IV studies, commonly known as post-marketing studies can and should be conducted. Phase IV studies are long term studies where a clinical event (as opposed to a surrogate endpoint) is the primary endpoint.[39] In addition to Phase IV studies, observational, post-marketing surveillance is also carried out to gather data from a variety of sources (individual clinicians, patients, databases, etc.). Unfortunately, the current system for monitoring and reporting adverse events after marketing approval is not optimal. The FDA had taken steps to improve post-marketing surveillance with its Sentinel Initiative. [40] The Sentinel Initiative aims to develop an active surveillance system to identify safety signals for FDA approved products. However, this initiative is still in its early stages and it will be many years until it is fully functional.

Together, Phase IV studies and post-marketing surveillance, along with Phase III trials, can be subjected to synthesis to help paint a clearer picture of the risk/benefit balance of an intervention. These analyses have led to Black Box warnings and drug withdrawals (e.g. rosiglitazone's Black Box warning for heart failure). However, SRs are constrained by the data that is available and cannot reverse underlying biases. Specifically, post-marketing surveillance is fraught with challenges because of its observational nature. For example, an adverse event may be reported, but its causal link to the drug might be difficult to prove. Furthermore, serious adverse events are more likely to be reported in this scheme as compared to more common, but relatively benign, side effects.[41, 42] This is unfortunate as the presence of seemingly benign side effects can greatly affect the risk/benefit balance of administering a particular drug to a particular patient.

Individual Patient Data Meta-Analysis

"Traditional" MAs generally consist of the pooling of aggregate data from a number of unique studies. Individual patient data meta-analyses pool data from individual participants

of included trials. This has the potential advantage of improving the quality of the data and increasing the precision in the estimate of effect particularly in the assessment of subgroups defined by patient characteristics. This type of MA has been used for more than 20 years, but they still represent a minority of the meta-analyses completed (when we limited our earlier search to individual patient data meta-analysis we found only 5 studies, one of which is a protocol). To conduct these analyses, investigators require time, resources, expertise and, most importantly, access to the data through either data repositories mandated by funding agencies and through collaboration with trialists. MAs focused on safety concerns in diabetes should consider using individual patient data MAs to improve the quality of evidence syntheses.

Early Phase Trials and Pre-Clinical Data

Increasing the scope of the body of evidence may also improve one's evaluation of the safety of interventions for diabetes. This can be accomplished by systematically including early phase trials (Phase I and II) and data from animal studies in safety assessments. While the intervention may have different effects in animal models, they may provide early safety signals which should be monitored for in later stage trials. Results from early phase trials should be reported in the literature, or at the very least, stored in a publically accessible database. In this way, SRs can gain confidence in the pooled effects from clinical trials if there is a chain of evidence that provides support for the findings. Perhaps more important for the safety of our patients, complete access to preclinical and early clinical data may provide, when pooled, evidence of a safety signal that later stage trials can be designed to specifically seek. Transparency in the process of developing anti-hyperglycemic drugs from inception to the market is one of the ways in which drugs can be made safer and trust in the enterprise may be recovered.

Conclusion

With the continual emergence of new therapies to treat diabetes and the rush to design and conduct the smallest and briefest trials possible for approval, the use of SRs and MAs to evaluate the safety of these medications will continue to be important. We have seen that individual studies alone are not enough to determine the safety of diabetes medications and that relying on them may lead to an overestimation of safety. SRs and MAs represent a way to credibly and transparently assess the entire body of evidence and produce a more precise estimate of effect than what is possible with single studies. Despite these strengths SRs and MAs must be interpreted in light of their limitations, namely, that synthesizing evidence does not fix biases introduced in the conduct and analyses of the underlying studies and cannot overcome the decision of authors, funders, and publishers to not publish the results of trials that fail to advance their interests. Corrupted in these ways, the data going into the MA will be of poor quality, reducing our confidence in the data that comes out. By understanding the role of SRs and MAs in evaluating the safety of diabetes medications as well as being aware of the challenges and opportunities for SRs and MAs in this capacity, the astute clinician can better serve their patients.

Acknowledgments

This publication was supported by CCaTS Grant Number TL1 TR000137 from the National Center for Advancing Translational Science (NCATS). Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the NIH.

References

1. Increasing prevalence of diagnosed diabetes--United States and Puerto Rico, 1995-2010. *MMWR Morb Mortal Wkly Rep.* 2012; 61(45):918–21. [PubMed: 23151951]
2. Salpeter SR, Greyber E, Pasternak GA, Salpeter EE. Risk of fatal and nonfatal lactic acidosis with metformin use in type 2 diabetes mellitus. *Cochrane Database Syst Rev.* 2010; (4) CD002967. 10.1002/14651858.CD002967.pub4
3. Silagy CA, Middleton P, Hopewell S. Publishing protocols of systematic reviews: comparing what was done to what was planned. *JAMA.* 2002; 287(21):2831–4. [PubMed: 12038926]
4. Green, S.; H, J.; Alderson, P.; Clarke, M.; Mulrow, CD.; Oxman, AD. Introduction. In: Higgins, JPT.; G, S., editors. *Cochrane Handbook for Systematic Reviews of Interventions* Version 5.1.0. The Cochrane Collaboration; 2011.
5. Lavis JN. How can we support the use of systematic reviews in policymaking? *PLoS Med.* 2009; 6(11):e1000141. 10.1371/journal.pmed.1000141 [PubMed: 19936223]
6. Nissen SE, Wolski K, Topol EJ. Effect of rosiglitazone on death and major adverse cardiovascular events in patients with type 2 diabetes mellitus. *JAMA.* 2005; 294(20):2581–6. 10.1001/jama.294.20.joc50147 [PubMed: 16239637]
7. Mehta SN, Goldfine AB, Abrahamson MJ, DiVincenzo R, Laffel LM. Changing prescribing patterns of type 2 diabetes medications from 2002-2010: an electronic health record-based evaluation. *J Diabetes Sci Technol.* 2013; 7(1):119–22. [PubMed: 23439167]
8. Psaty BM, Furberg CD. Rosiglitazone and cardiovascular risk. *N Engl J Med.* 2007; 356(24):2522–4. 10.1056/NEJMe078099 [PubMed: 17517854]
9. Nissen SE, Wolski K. Effect of rosiglitazone on the risk of myocardial infarction and death from cardiovascular causes. *N Engl J Med.* 2007; 356(24):2457–71. 10.1056/NEJMoA072761 [PubMed: 17517853]
10. Singh S, Loke YK, Furberg CD. Long-term risk of cardiovascular events with rosiglitazone: a meta-analysis. *JAMA.* 2007; 298(10):1189–95. 10.1001/jama.298.10.1189 [PubMed: 17848653]
11. Hiatt WR, Kaul S, Smith RJ. The Cardiovascular Safety of Diabetes Drugs - Insights from the Rosiglitazone Experience. *N Engl J Med.* 2013. 10.1056/NEJMp1309610
12. Nissen SE, Wolski K. Rosiglitazone revisited: an updated meta-analysis of risk for myocardial infarction and cardiovascular mortality. *Arch Intern Med.* 2010; 170(14):1191–201. This meta-analysis included the largest trial of cardiovascular safety for rosiglitazone and confirmed an increased risk of myocardial infarction and heart failure associated with rosiglitazone use. 10.1001/archinternmed.2010.207 [PubMed: 20656674]
13. Home PD, Pocock SJ, Beck-Nielsen H, Curtis PS, Gomis R, Hanefeld M, et al. Rosiglitazone evaluated for cardiovascular outcomes in oral agent combination therapy for type 2 diabetes (RECORD): a multicentre, randomised, open-label trial. *Lancet.* 2009; 373(9681):2125–35. This is the largest randomized trial looking at the cardiovascular safety of rosiglitazone and is at the center of the controversy surrounding the cardiovascular safety of rosiglitazone. 10.1016/S0140-6736(09)60953-3 [PubMed: 19501900]
14. Egger M, Smith GD, Altman D. *Systematic reviews in health care: meta-analysis in context.* BMJ books. 2008
15. Balshem H, Helfand M, Schunemann HJ, Oxman AD, Kunz R, Brozek J, et al. GRADE guidelines: 3. Rating the quality of evidence. *J Clin Epidemiol.* 2011; 64(4):401–6. 10.1016/j.jclinepi.2010.07.015 [PubMed: 21208779]
16. Department of Health and Human Services. Advisory committee meeting for NDA 21071 Avandia (rosiglitazone maleate) tablet; July 13 and 14, 2010; www.fda.gov/downloads/AdvisoryCommittees/CommitteesMeetingMaterials/Drugs/

[EndocrinologicandMetabolicDrugsAdvisoryCommittee/UCM218493.pdf](#). This document details the 2010 FDA review of rosiglitazone, within which the limitations of the RECORD trial are highlighted

17. Food and Drug Administration. FDA requires removal of certain restrictions on the diabetes drug Avandia. Nov 25. 2013 <http://www.fda.gov/NewsEvents/Newsroom/PressAnnouncements/ucm376516.htm>
18. Ioannidis JP, Trikalinos TA. Early extreme contradictory estimates may appear in published research: the Proteus phenomenon in molecular genetics research and randomized trials. *J Clin Epidemiol*. 2005; 58(6):543–9.10.1016/j.jclinepi.2004.10.019 [PubMed: 15878467]
19. Ioannidis JP. Evolution and translation of research findings: from bench to where? *PLoS Clin Trials*. 2006; 1(7):e36.10.1371/journal.pctr.0010036 [PubMed: 17111044]
20. Murad M, Montori VM. Synthesizing evidence: Shifting the focus from individual studies to the body of evidence. *JAMA*. 2013; 309(21):2217–8. This paper expounds the reasons why clinicians should examine the body of evidence as rather than individuals studies. 10.1001/jama.2013.5616 [PubMed: 23736731]
21. Lincoff AM, Wolski K, Nicholls SJ, Nissen SE. Pioglitazone and risk of cardiovascular events in patients with type 2 diabetes mellitus: a meta-analysis of randomized trials. *JAMA*. 2007; 298(10):1180–8.10.1001/jama.298.10.1180 [PubMed: 17848652]
22. Kahn SE, Haffner SM, Heise MA, Herman WH, Holman RR, Jones NP, et al. Glycemic Durability of Rosiglitazone, Metformin, or Glyburide Monotherapy. *New England Journal of Medicine*. 2006; 355(23):2427–43.10.1056/NEJMoa066224 [PubMed: 17145742]
23. Loke YK, Singh S, Furberg CD. Long-term use of thiazolidinediones and fractures in type 2 diabetes: a meta-analysis. *CMAJ*. 2009; 180(1):32–9.10.1503/cmaj.080486 [PubMed: 19073651]
24. Felson DT. Bias in meta-analytic research. *J Clin Epidemiol*. 1992; 45(8):885–92. [PubMed: 1624971]
25. Savoie I, Helmer D, Green CJ, Kazanjian A. Beyond Medline: reducing bias through extended systematic review search. *Int J Technol Assess Health Care*. 2003; 19(1):168–78. [PubMed: 12701949]
26. Sampson M, McGowan J. Errors in search strategies were identified by type and frequency. *J Clin Epidemiol*. 2006; 59(10):1057–63.10.1016/j.jclinepi.2006.01.007 [PubMed: 16980145]
27. Dormandy JA, Charbonnel B, Eckland DJ, Erdmann E, Massi-Benedetti M, Moules IK, et al. Secondary prevention of macrovascular events in patients with type 2 diabetes in the PROactive Study (PROspective pioglitAzone Clinical Trial In macroVascular Events): a randomised controlled trial. *Lancet*. 2005; 366(9493):1279–89.10.1016/s0140-6736(05)67528-9 [PubMed: 16214598]
28. Lewis JD, Ferrara A, Peng T, Hedderson M, Bilker WB, Quesenberry CP Jr, et al. Risk of bladder cancer among diabetic patients treated with pioglitazone: interim report of a longitudinal cohort study. *Diabetes Care*. 2011; 34(4):916–22. doi:10.2337/dc.10-1068. [PubMed: 21447663]
29. Piccinni C, Motola D, Marchesini G, Poluzzi E. Assessing the association of pioglitazone use and bladder cancer through drug adverse event reporting. *Diabetes Care*. 2011; 34(6):1369–71. doi: 10.2337/dc.10-2412. [PubMed: 21515844]
30. Colmers IN, Bowker SL, Majumdar SR, Johnson JA. Use of thiazolidinediones and the risk of bladder cancer among people with type 2 diabetes: a meta-analysis. *CMAJ*. 2012; 184(12):E675–83.10.1503/cmaj.112102 [PubMed: 22761478]
31. Ferwana M, Firwana B, Hasan R, Al-Mallah MH, Kim S, Montori VM, et al. Pioglitazone and risk of bladder cancer: a meta-analysis of controlled studies. *Diabet Med*. 2013; 30(9):1026–32.10.1111/dme.12144 [PubMed: 23350856]
32. Singh S, Chang HY, Richards TM, Weiner JP, Clark JM, Segal JB. Glucagonlike peptide 1-based therapies and risk of hospitalization for acute pancreatitis in type 2 diabetes mellitus: a population-based matched case-control study. *JAMA Intern Med*. 2013; 173(7):534–9.10.1001/jamainternmed.2013.2720 [PubMed: 23440284]
33. Elashoff M, Matveyenko AV, Gier B, Elashoff R, Butler PC. Pancreatitis, pancreatic, and thyroid cancer with glucagon-like peptide-1-based therapies. *Gastroenterology*. 2011; 141(1):150–6.10.1053/j.gastro.2011.02.018 [PubMed: 21334333]

34. Monami M, Dicembrini I, Mannucci E. Dipeptidyl peptidase-4 inhibitors and pancreatitis risk: a meta-analysis of randomized clinical trials. *Diabetes Obes Metab*. 2013;10.1111/dom.12176
35. Cohen D. Has pancreatic damage from glucagon suppressing diabetes drugs been underplayed? *BMJ*. 2013; 346 This an investigative report written by an editor from the BMJ that highlights the concerns around the safety of glucagon suppressing drugs. 10.1136/bmj.f3680
36. Murad MH, Coto-Yglesias F, Wang AT, Sheidaee N, Mullan RJ, Elamin MB, et al. Clinical review: Drug-induced hypoglycemia: a systematic review. *J Clin Endocrinol Metab*. 2009; 94(3): 741–5. 10.1210/jc.2008-1416 [PubMed: 19088166]
37. Meinert, CL. *ClinicalTrials: Design, Conduct and Analysis*. Oxford University Press; USA: 2012.
38. Duijnhoven RG, Straus SM, Raine JM, de Boer A, Hoes AW, De Bruin ML. Number of patients studied prior to approval of new medicines: a database analysis. *PLoS Med*. 2013; 10(3):e1001407. 10.1371/journal.pmed.1001407 [PubMed: 23526887]
39. Meinert, CL. *Clinical Trials Dictionary: Terminology and usage recommendations*. John Wiley & Sons, Inc.; 2012.
40. U.S. Food and Drug Administration. [19 September 2013] FDA's Sentinel Initiative. 2013. <http://www.fda.gov/Safety/FDASentinelInitiative/default.htm>
41. Heeley E, Riley J, Layton D, Wilton LV, Shakir SAW. Prescription-event monitoring and reporting of adverse drug reactions. *The Lancet*. 2001; 358(9296):1872–3. [http://dx.doi.org/10.1016/S0140-6736\(01\)06898-2](http://dx.doi.org/10.1016/S0140-6736(01)06898-2).
42. Hazell L, Shakir SAW. Under-Reporting of Adverse Drug Reactions: A Systematic Review. *Drug Safety*. 2006; 29(5):385. [PubMed: 16689555]

Figure 1.

Forest Plots: What they are and how to interpret them. Forest plots are the most common and efficient way to display meta-analytic data. Understanding how to interpret forest plots permits quick comprehension and appraisal of the source and nature of synthesized data. Here are the basics of interpretation: Each horizontal row of raw data corresponds to an individual study. The graphical contribution of each study's data is displayed within the forest plot to the right of the dataset, generally in the form of a box centered on a horizontal line. The location of the box horizontally indicates the point estimate of the study's measured effect. In many cases, the size of the box is representative of the size of the study, and thus the proportion of data contributed to the total. The horizontal line is the 95% confidence interval; its width reflects precision and its location (i.e. whether or not it crosses a relative risk of 1, etc.) reflects significance. The pooled data is represented on the forest plot by a diamond, which can be thought of as the mathematical weighted average of all the individual studies. The point on the diamond with greatest vertical dimension corresponds to the point estimate of pooled effect; the width of the diamond corresponds to the 95% confidence interval for the pooled data. Example: The forest plot pictured here is from a recent review assessing the impact of pioglitazone usage on bladder cancer risk [31]. Notice that outcomes are reported as a hazard ratio of exposed/unexposed diabetics for the development of bladder cancer. Also note that the authors conducted a subgroup analysis of the data to assess the effect of treatment duration on the outcome. In the subgroup with <12 months pioglitazone use, for example, no significant effect was seen (the diamond is practically centered on a hazard ratio of 1). Whereas in the subgroups with 12-23 and ≥ 24 months of pioglitazone exposure a significant cancer risk was seen. Finally, notice that the longer the duration of exposure, the more consistent and precise the effect estimate becomes (the confidence intervals narrow and the point estimates seem to align). Heterogeneity is an important concept to consider when attempting to pool data from multiple studies. A full explanation of

heterogeneity is beyond the scope of this paper, but it is useful to know how to interpret the I^2 statistic. The I^2 statistic is an estimate of the percent chance that the observed variability is due to a true difference in treatment effect. Note that in each subgroup here, the I^2 is 0%, indicating that chance alone provides a satisfactory explanation for the variability in the effect estimate. Generally, $I^2 < 25\%$ is considered low heterogeneity, $I^2 50\%-75\%$ is considered moderate heterogeneity, and $I^2 > 75\%$ is considered high heterogeneity. Note that low heterogeneity is desired for any pooled data (studies within individual subgroups), but in the case of between group comparisons a high I^2 statistic is often desirable (i.e. we want the % of variability beyond chance to be high). For example, in the test for subgroup differences shown, the I^2 is 68.5%, indicating that the observed difference in effect between groups is likely due to a true treatment effect; this supports the validity of the conclusion that the duration of pioglitazone therapy is directly related to the risk of bladder cancer. (With permission from: Ferwana M, Firwana B, Hasan R, Al-Mallah MH, Kim S, Montori VM et al. Pioglitazone and risk of bladder cancer: a meta-analysis of controlled studies. *Diabet Med.* 2013;30(9):1026-32. doi:10.1111/dme.12144) [31].

Table 1

Rating the quality of the body of evidence

<p>DECREASE MY CONFIDENCE</p> <p>STUDY LIMITATION</p> <p>The higher the risk of bias the lower the confidence in evidence is. Evidence generated by randomized controlled trials is usually considered to be of high quality, whereas evidence generated by non-randomized studies is considered to be of lower quality because there is no assurance of prognostic balance.</p> <p>INCONSISTENCY</p> <p>If different studies attempt to answer the same question but the magnitude of intervention effects differs across studies the confidence in the overall results is lower. High variability is evaluated using I^2 statistic, which represents the proportion of inconsistency in the results that is not attributable to chance or random error.</p> <p>IMPRECISION</p> <p>A wide confidence interval around the pooled estimate suggests an imprecise result, which can cause uncertainty about the result.</p> <p>INDIRECTNESS</p> <p>When significant differences exist between the population, the intervention, or the outcomes measured in included studies and those under consideration.</p> <p>INCREASE MY CONFIDENCE</p> <p>LARGE MAGNITUDE OF EFFECT</p> <p>Evidence from a study with large effect would increase our confidence in the evidence (i.e. $RR > 5$ or $RR < 0.2$)</p> <p>DOSE-RESPONSE GRADIENT</p> <p>Dose-response gradient increase confidence in the estimates of observational studies and thereby increase the assigned quality of evidence.</p>