Supplemental Material

Table S1. Definitions of comorbid conditions and medications, on the basis of codes and prescriptions in 730 days before treatment intensification

ACEI = angiotensin-converting enzyme inhibitor; ARB = angiotensin-receptor blocker; COPD = chronic obstructive pulmonary disease; CPT = Current Procedural Terminology; ICD-9- CM = International Classification of Diseases, Ninth Revision; MI = myocardial infarction; TIA = transient ischemic attack If medications are combinations of 2 drug classes then a patient is recorded as using both medications. * Each co-morbid condition was defined as present if there was 1 specified inpatient or 2 specified outpatient codes separated by 30 days, or 1 specified procedure code or prescription for a medication defining that comorbid condition in the 730 days before treatment intensification.

Table S2. Details for the construction of the propensity score model

The pre-matching cohort was composed of all eligible persons who initiated metformin or sulfonylurea for diabetes and met the study's inclusion criteria. The matched cohort was formed by matching metformin users to sulfonylurea users with similar propensity scores. The propensity score (PS) is defined as the probability of sulfonylurea use, given a particular pattern of baseline covariates (Table 2). We estimated the PS using a logistic regression model in which the dependent variable was 1 for patients who used sulfonylurea and 0 for metformin users and used restricted cubic splines (3 knots) for continuous covariates in the model. The PS model is designed to be non-parsimonious and highly flexible to capture all observable confounding by indication. Indicator variables denoting missingness were included in the PS model, allowing the PS to balance missingness patterns between the exposures and control for potentially informative missingness. Multiply imputed PS model coefficients were aggregated using Rubin's rules and the aggregated model used to generate PS values. The PS model is displayed in Appendix Table 2. The PS model yielded a C statistic of 0.71. When used to facilitate matching, the success of the PS model is determined by the covariate balance achieved in the matched cohort. Table S1 and Figures S2 and S3 demonstrate the mean standardized differences before and after propensity score matching. Indicating good balance after matching, all standardized differences have an absolute value < 0.1. An important condition for propensity score methods is that every cohort member have a nontrivial probability of having received either of the study therapies (positivity). Our matching procedure excluded sulfonylurea patients for whom very few similar metformin users existed. Unmatched sulfonylurea patients primarily were older, had higher number of co-morbidities and had a higher serum creatinine at the time of drug initiation (See Table S3 for characteristics of unmatched patients). The matching was performed on the log odds of the propensity scores using an 8:1 digit greedy match algorithm.

Logistic regression model for the probability of initiating Sulfonylurea (N=65,986 matches)

Table S4. Description and Characteristics of the weighted analysis cohort

The weighted analyses were performed using inverse probability of treatment weights (IPTW). As opposed to a matched analysis which balances the baseline covariate distributions by selecting a subset of patients from each exposure, a weighted analysis balances the covariate distributions by assigning various weights to the patients in one exposure such that the weighted group now resembles the other group. When comparing metformin and sulfonylurea users, the sulfonylurea users were weighted so that their distribution of covariates resembled that of the metformin users. This was achieved by using stabilized IPTW such that metformin users receive a weight of 1 and sulfonylurea users a weight of $e/(1-e_i)$, where e_i is the probability of patient i receiving metformin given their covariates. This creates a pseudo-cohort that uses all of the eligible patients. In simple terms, the older, less healthy sulfonylurea users (who are over-abundant relative to metformin) are down-weighted to match the metformin distribution and the younger, healthier sulfonyurea users are up-weighted to match the metformin population. The sum of the metformin users' weights will equal the number of metformin users because they each received a weight of 1. The sum of the sulfonylurea users' weights will approximate the number of metformin users because the sulfonylurea users are being weighted to approximate that group. The sum will not equal the number of metformin users exactly because the IPTW rely on modeling the exposure and thus provide an approximate solution. Like with matching, the success of the weighting in achieving a well-balanced pseudo-cohort can be seen in the table of patient characteristics and plot of standardized differences. Also like matching, the weighted analysis may be used with or without additional direct covariate adjustment. The analysis that does not use additional covariate adjustment estimates the average sulfonylurea versus metformin effect in a population of metformin users - our control group. This is referred to as the average treatment effect among controls (ATC). In the metformin versus thiazolidinedione comparison, the smaller thiazolidinedione group could not be easily up-weighted to approximate the much larger metformin group; however, the metformin group could be easily down-weighted to approximate the thiazolidinedione users. Hence, we used the thiazolidinedione users as the stabilizing population and estimated the average treatment effect among the treated population (ATT).

Table S5. Sensitivity Analyses evaluating the hazard of heart failure in first 180 days of use of Sulfonylurea versus Metformin and Thiazolidinedione versus Metformin, using new-user design and inverse probability treatment weighted analysis

*****For the weighted analysis comparing sulfonylurea to metformin, the sulfonylurea population is weighted by their characteristics to more closely resemble the younger and healthier metformin population. For the weighted analysis comparing thiazolidinedione to metformin, the metformin population is weighted by their characteristics to more closely resemble the older population. Refer to Table S4 for details.

† The N at risk for the analysis of the first 180 days is larger then for the primary analysis cohort because it includes all patients from the primary unmatched cohort and also includes people who were excluded during the 180 day lag period for being non persistent; not having a full 180 days of follow-up; those who died; or were censored for reaching the threshold creatinine.

Table S6. Analysis of sensitivity to unmeasured confounding¹

We evaluated the risk of heart failure in the presence of an unobserved confounder with a relative hazard of 2.3 for heart failure risk, and various prevalence levels of the confounder by exposure group. The primary analysis yielded a greater risk of heart failure with sulfonylurea use over metformin use; HR (95% CI): 1.32 (1.21, 1.43). The bolded numbers correspond to the necessary differential prevalence of such a confounder between exposure groups that could account for study results being the result of such confounding.

The observed risk of prior heart failure history with the primary outcome was an HR of 2.3. For an unmeasured confounder of this strength to tip the primary finding of this paper into statistical non-significance, it would need to be independent of the observed covariates and 17% more prevalent among sulfonylurea users if the prevalence among metformin users was 0%. If the prevalence in metformin users was between 20-50%, it would need to be 21-27% **more** prevalent. If the prevalence in metformin users was 70%, an unmeasured confounder of this strength could not tip the analysis into statistical non-significance. Due to the heterogeneous prescribing practices in the VHA during the study period, selection bias of this degree was not observed. There were no differences in prevalence of this magnitude among the observed covariates in the full (pre-matching) cohort (Table 1).

Figure S1. Study Design Schematic

Below is an example patient who initiated Metformin after having 180 days free of any antidiabetic drug. Sulfonylurea and Thiazolidinedione person time are tracked in the same manner. **Main analysis:**

Persistent exposure required: Gaps (red bars) of up to 90 days are allowed in order to refill the regimen. Patients are censored at addition of another drug or no medication refills within 90 days. **Sensitivity analyses:**

Persistent exposure not required: In this approach patients are analyzed as users of their incident prescribed regimen regardless of switching, stopping or additions (akin to intent to treat analysis) **First 180 days:** This person time has a high likelihood of exposure misclassification and was excluded in above two analyses. This analysis of the first 180 days includes both those that do and don't refill their prescriptions as well as those who make other regimen changes, such as switching regimens and stopping medications. The resultant exposure misclassification, if non-differential would make it harder to show differences between treatment regimens.

Propensity Scores

Figure S3. Mean Standardized difference plot comparing metformin versus sulfonylurea

variable

Supplemental References:

1. Schneeweiss S, Glynn RJ, Tsai EH, Avorn J, Solomon DH. Adjusting for unmeasured confounders in pharmacoepidemiologic claims data using external information: the example of COX2 inhibitors and myocardial infarction. *Epidemiology*. 2005;16:17-24.

2. Austin PC. An Introduction to Propensity Score Methods for Reducing the Effects of Confounding in Observational Studies. *Multivariate Behav Res*. 2011;46:399-424.

3. Austin PC, Stuart EA. Moving towards best practice when using inverse probability of treatment weighting (IPTW) using the propensity score to estimate causal treatment effects in observational studies. *Stat Med*. 2015;34:3661-79.