



Published in final edited form as:

Epidemiology. 2017 November ; 28(6): 838–846. doi:10.1097/EDE.0000000000000709.

Sentinel Modular Program for Propensity-Score Matched Cohort Analyses: Application to Glyburide, Glipizide, and Serious Hypoglycemia

Meijia Zhou¹, Shirley V. Wang², Charles E. Leonard¹, Joshua J. Gagne², Candace Fuller³, Christian Hampf⁴, Patrick Archdeacon⁴, Sengwee Toh³, Aarthi Iyer³, Tiffany Siu Woodworth³, Elizabeth Cavagnaro³, Catherine A. Panozzo³, Sophia Axtman³, Ryan M. Carnahan⁵, Elizabeth A. Chrischilles⁵, and Sean Hennessy¹

¹Center for Pharmacoepidemiology Research and Training, Center for Clinical Epidemiology and Biostatistics, and Department of Biostatistics and Epidemiology, University of Pennsylvania Perelman School of Medicine, Philadelphia, Pennsylvania, USA

²Division of Pharmacoepidemiology and Pharmacoeconomics, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts, USA

³Department of Population Medicine, Harvard Medical School and Harvard Pilgrim Health Care Institute, Boston, Massachusetts, USA

⁴Center for Drug Evaluation and Research, US Food and Drug Administration, Silver Spring, Maryland, USA

⁵g

Abstract

Sentinel is a program sponsored by the US Food and Drug Administration to monitor the safety of medical products. We conducted a cohort assessment to evaluate the ability of the Sentinel Propensity Score Matching Tool to reproduce in an expedited fashion the known association between glyburide (versus glipizide) and serious hypoglycemia. Thirteen data partners that contribute to the Sentinel Distributed Database participated in this analysis. A pre-tested and customizable analytic program was run at each individual site. De-identified summary results from each Data Partner were returned and aggregated at the Sentinel Operations Center. We identified a total of 198,550 and 379,507 new users of glyburide and glipizide, respectively. The incidence of emergency department visits and hospital admissions for serious hypoglycemia was 19 per 1,000 person-years (95% confidence interval, 17.9, 19.7) for glyburide users and 22 (21.6, 22.7) for glipizide users. In cohorts matched by propensity score based on predefined variables, the hazard ratio (HR) for glyburide was 1.36 (1.24, 1.49) vs. glipizide. In cohorts matched on a high-dimensional propensity score based on empirically selected variables, for which the program ran to completion in five data partners, the HR was 1.49 (1.31, 1.70). In cohorts matched on propensity scores based on both pre-defined and empirically selected variables via the high-dimensional propensity score algorithm (the same five data partners), the HR was 1.51 (1.32,

1.71). These findings are consistent with the literature, and demonstrate the ability of the Sentinel Propensity Score Matching Tool to reproduce this known association in an expedited fashion. 3

Keywords

Sentinel modular program; propensity score matching; pharmacoepidemiology; hypoglycemia; glyburide; glipizide

INTRODUCTION

The Food and Drug Administration (FDA) Amendments Act of 2007 required the FDA to create the capability to perform active surveillance of the safety of approved medical products using routinely collected health information from at least 100 million people.¹ In response to this mandate, the FDA created the Mini-Sentinel pilot program. The pilot program has since evolved into a medical product safety surveillance system envisioned in the FDA Amendments Act. The system functions as a collaboration between the FDA and a consortium that includes an operations center, data partners, and academic institutions.² Sentinel utilizes a distributed data system in which 18 Data Partner sites maintain and regularly update patients' administrative claims and clinical information formatted in a common data model.³ A customizable set of pre-tested modular programs compatible with the common data model, known as the Sentinel Active Risk Identification and Analysis (ARIA) system, enables the FDA to perform analyses evaluating associations between medical products and pre-specified health outcomes of interest. These programs are run on the distributed database, and de-identified results are returned to the Sentinel Operations Center for aggregation, thus preserving the privacy of individual health plan members.⁴

The Sentinel Propensity Score Matching Tool is one of these customizable Sentinel modular programs, and enables the conduct of cohort analyses that are matched on propensity scores generated using either user-defined variables and/or variables identified by the automated, high-dimensional propensity score algorithm.^{5,6} While propensity score methods have been used in prior safety assessments conducted within Sentinel, past evaluations have required a detailed protocol, *de novo* analytic programs, and investigator-specified variables for the propensity score rather than an automated high-dimensional propensity score algorithm for identifying variables.

In contrast, the Sentinel Propensity Score Matching Tool allows investigators to: use a pre-existing template to define standard design options rather than writing a *de novo* protocol; utilize customizable, pre-existing analytic programs; enable investigator-specified covariates and/or a high-dimensional propensity score approach. Thus, the Sentinel Propensity Score Matching Tool has the potential to accelerate the process required to conduct comparative drug safety assessments within Sentinel, while also reducing the required resources. We sought to pilot test the ability of the Sentinel Propensity Score matching Tool to reproduce the well-documented association between glyburide and serious hypoglycemia, using glipizide as the comparator agent. Prior epidemiologic studies have found a 1.67 to 1.90-fold increased risk of serious hypoglycemia associated with glyburide versus glipizide,^{7,8} and a systematic review of randomized trials found a higher risk for glyburide than for other

insulin secretagogues, including glipizide.⁹ We therefore pilot tested the Sentinel Propensity Score Matching Tool by assessing its ability to reproduce this known association and to assess and identify issues related to the implementation of the tool.

METHODS

Data

As of July 2014, the Sentinel Distributed Database comprised data from 18 Data Partner sites covering approximately 178 million individuals cumulatively from January 2000 through January 2014.¹⁰ The Sentinel Distributed Database includes demographics, enrollment, diagnosis, procedure, outpatient dispensing, and laboratory data. Thirteen data partners participated in this assessment: Aetna, Blue Bell, PA; HealthCore, Inc., Alexandria, VA; Group Health Research Institute, Seattle, WA; Harvard Pilgrim Health Care Institute, Boston, MA; HealthPartners Institute, Saint Paul, Minnesota; Meyers Primary Care Institute, Worcester, MA; Marshfield Clinic Research Foundation, Marshfield, WI; Humana, Inc., Miramar, FL; Kaiser Permanente Colorado, Denver, CO; Kaiser Permanente Hawaii, Honolulu, HI; Kaiser Permanente Northern California, Oakland, CA; Kaiser Permanente Northwest, Portland, OR; and Optum, Inc., Waltham, MA. Four of the 13 data partners are national insurers and the remaining nine are regional insurers.¹⁰ All of the data partners listed contribute claims data to the Sentinel Distributed Database and several also contribute information from electronic medical records. Sentinel has been deemed a public health activity under the auspices of the FDA and thus not under the purview of institutional review boards.^{11, 12}

Study population

We performed a cohort study of individuals aged 18 years or older who initiated glyburide or glipizide between January 1, 2008 and September 30, 2014. The index date was the date of the first dispensing of glipizide or glyburide during the study period. Individuals were excluded if there was evidence of a hypoglycemia event in the 30 days before cohort entry or if any of the following insulin secretagogues were dispensed in the 183 days before cohort entry (i.e., the baseline period): glyburide, glipizide, chlorpropamide, tolbutamide, tolazamide, glimepiride, nateglinide, repaglinide, or acetoexamide. To be eligible, individuals had to be continuously enrolled in a plan with both medical and drug coverage during the baseline period, during which gaps in enrollment up to 45 days were allowed. Exposure episodes were defined using outpatient pharmacy dispensing days supplied to create a sequence of continuous exposure. Exposure episodes were considered continuous if gaps in days supplied were 14 days or less. A stockpiling algorithm was used to account for dispensings for the same generic name with overlapping days of supply.¹³ Any overlap of supply between dispensings was corrected by pushing the start date of the second dispensing to occur following the end of the days supplied for the first dispensing.⁴ Only the first episode for each person was included in the analysis.

Follow-up began on the day the first dispensing of interest and continued until the first occurrence of any the following: 1) serious hypoglycemia, as defined below; 2) death (Inpatient discharge disposition, including expired at discharge, is captured at all data

partners. Data partners have a variety of other methods for capturing death data, such as Social Security Administration data, state death records, or internal data sources. Sentinel uses all death information available and provided by data partners in Sentinel activities.); 3) 14 days after the end of exposure episode; 4) filling a prescription for a secretagogue other than that identified upon cohort entry; 5) disenrollment from the health plan or; 6) reaching the end of available data for that health plan.

Covariates

We assessed the following pre-defined covariates during the baseline period. The 12 basic covariates for the program that were automatically included were as follows: age, sex, time period, year of exposure, combined Charlson/Elixhauser comorbidity score,¹⁴ as well as seven measures of healthcare utilization intensity: number of unique generic drugs dispensed, dispensed prescriptions, inpatient hospital encounters, non-acute institutional encounters, emergency department encounters, ambulatory encounters and other ambulatory encounters such as telemedicine and email consults.¹⁵ In addition, based on prior studies, we specified five covariates: history of serious hypoglycemia,¹⁶ chronic kidney disease¹⁷ (see eTable 1; <http://links.lww.com/EDE/B229> for algorithms used to identify these covariates), use of insulin, metformin, or non-secretagogue antidiabetic drugs. These variables were included in the propensity score models described below.

Outcome

The primary outcome was serious hypoglycemia defined as an any-position emergency department or first-listed inpatient diagnosis for serious hypoglycemia as defined by *International Classification of Diseases, 9th revision, Clinical Modification* (ICD-9-CM) codes 251.0 hypoglycemic coma, 251.1 other specified hypoglycemia, 251.2 hypoglycemia unspecified, or 250.8× diabetes with other specified manifestations. Outcomes identified by 250.8× were not included if they occurred with one of the following diagnoses: 259.8 other specified endocrine disorders, 272.7 lipidoses, 681.xx cellulitis and abscess of finger and toe, 682.xx other cellulitis and abscess, 686.9 unspecified local infection of skin and subcutaneous tissue, 707.1× ulcer of lower limbs, except decubitus ulcer, 707.2× pressure ulcer stages, 707.8 chronic ulcer of other specified sites, 707.9 chronic ulcer of unspecified site, 709.3 degenerative skin disorders, 730.0× acute osteomyelitis, 730.1× chronic osteomyelitis, 730.2× unspecified osteomyelitis, 731.8 other bone involvement in diseases classified elsewhere. The emergency department and inpatient components of this algorithm have positive predictive values of 89%¹⁶ and 78%,¹⁸ respectively. The secondary outcome was defined by the emergency department component only. The discharge diagnoses were utilized in the definitions described above.

Statistical analysis

The Sentinel Cohort Identification and Descriptive Analysis Tool is the foundation of the Sentinel modular program routine querying system, and is integrated with other modular programs including the Sentinel Propensity Score Matching Tool. The Sentinel Cohort Identification and Descriptive Analysis Tool was used to identify and extract the cohort of new users of glyburide and glipizide from the Sentinel Distributed Database according to the inclusion and exclusion criteria described above.⁴ The tool utilized the exposure and follow-

up time cohort identification strategy to identify new users of glyburide and glipizide, determine exposed time using drug dispensing days of supply, and look for serious hypoglycemia events during the exposed time period. The Sentinel Cohort Identification and Descriptive Analysis Tool also extracted covariates of interest during the specified time window for the propensity score model and output analytic datasets for the Sentinel Propensity Score Matching Tool. We then used the Sentinel Propensity Score Matching Tool to conduct three analyses. The first analysis utilized a propensity score model that included only investigator-specified variables, the second utilized a propensity score model that included variables empirically identified by an automated high-dimensional propensity score algorithm^{19,20} and the third utilized a propensity score model that included both investigator-specified variables and those empirically identified by the high-dimensional propensity score algorithm.^{19,20} For the high-dimensional propensity score, up to 100 baseline covariates from each of five dimensions (drug claims, ICD-9-CM diagnoses, ICD-9 procedures, Healthcare Common Procedure Coding System [HCPCS] procedures, and Current Procedural Terminology [CPT] procedures) were initially evaluated. From this pool of candidate variables, the high-dimensional propensity score algorithm automatically selected up to 200 covariates or the count of new users in the smaller exposure group if the count was less than 200. Covariates were ranked based on covariate-exposure associations.²⁰ Zero-cell correction was added to each cell in the covariate outcome 2×2 table when there were cells with 0 patients, thus the covariate and outcome associations were consistently computable at sites with few events.

The Sentinel Propensity Score Matching Tool calculates the propensity scores, identifies matched cohorts based on propensity scores, and performs an analysis of the matched cohorts using proportional hazards regression²¹ yielding hazard ratios (HRs) and 95% confidence intervals (CIs) for the association between serious hypoglycemia and glyburide (vs. glipizide). Propensity score estimation and matching were performed separately within each Data Partner site. Sites returned de-identified data files containing propensity scores, treatment group (i.e., glyburide or glipizide), a binary outcome indicator (i.e., serious hypoglycemia), and the number of days of follow-up between index date and outcome or censoring date. The Sentinel Operations Center aggregated data and used a Cox proportional hazards model to estimate a site-adjusted HR and 95% CIs in the unmatched population and a separate model to estimate an adjusted HR and 95% CI in the 1:1 propensity score matched cohort. All Cox models were stratified by site. The conditional models fit to the propensity score matched cohorts were further stratified on matched pair in which follow-up time was truncated for patients in the matched pair when either person of the pair was censored or had an event, resulting in equal person time for the two groups.

Glyburide and glipizide users were matched in a 1:1 ratio using a nearest-neighbor matching algorithm with a maximum matching caliper of 0.025²² on the propensity score scale. We examined the distribution of propensity score values between glyburide and glipizide cohorts pooled across data partners, and compared baseline characteristics between pooled cohorts before and after propensity score matching using standardized mean differences. A standardized mean difference 0.10 or -0.10 was used to indicate potential imbalance.²³

RESULTS

All 13 data partners ran pre-specified propensity score and high-dimensional propensity score models to completion. Five data partners ran high-dimensional propensity score models without any convergence warnings or high-dimensional propensity score code issues, two had high-dimensional propensity score code issues that caused errors in selecting covariates into the high-dimensional propensity score model, and six had “questionable convergence” warnings. Unmatched table 1s and propensity score distribution figures were available from all data partners for both pre-specified and high-dimensional propensity score models (data not shown). The Sentinel Propensity Score Matching Tool prevented further execution of analyses based on the high-dimensional propensity scores at data partners with questionable convergence. All 13 data partners returned results for unmatched pre-specified propensity scores and high-dimensional propensity scores models as well as results for matched pre-specified propensity score models, and seven returned results for matched high-dimensional propensity score models. Two of the seven sites were affected by an issue in the high-dimensional propensity score related codes and thus were excluded from further high-dimensional propensity score analyses. The issue affected proper covariate selections by the high-dimensional propensity score algorithm such that the entire dimensions of clinical codes could be omitted from being considered by the algorithm for selection into the high-dimensional propensity score model.

Predefined propensity score analyses

Baseline characteristics are shown in Table 1 (all 13 data partners) and Table 2 (the five data partners that returned results for matched high-dimensional propensity score models and completed the high-dimensional propensity score models without errors). In the unmatched cohorts, a total of 198,550 glyburide and 379,507 glipizide new users contributed 89,719 and 244,094 person-years of observation (Table 3), respectively. The median length of follow-up was 79 days in glyburide users and 114 days in glipizide users. In the unmatched cohorts, the incidence rate for the primary definition of serious hypoglycemia was lower for glyburide (19 per 1,000 person-years; 95% CI: 17.9, 19.7) than for glipizide users (22 per 1,000 person-years; 95% CI: 21.6, 22.7). However, after stratification by site, the HR for glyburide vs. glipizide was 1.11 (1.05, 1.18). The conditional analysis using the propensity score based on investigator-defined variables ran successfully in all 13 data partners, yielding a HR of 1.36 (1.24, 1.49). The unadjusted and adjusted HRs at each site are presented in eTable 2; <http://links.lww.com/EDE/B229>. All sites had crude HRs greater than 1, except for data partners 9 and 11.

High-dimensional propensity score analyses

In the unmatched high-dimensional propensity score analysis, ten out of 13 data partners (including data partners 9 and 11, which had unadjusted HRs < 1) produced a cluster of persons with high-dimensional propensity scores near 1.0 in users of glyburide but not glipizide (see eFigures 1 and 2; <http://links.lww.com/EDE/B229> for examples of the histograms, and eFigure 3 and 4; <http://links.lww.com/EDE/B229> for the histograms of data partner 9 and 11, respectively), indicating that the high-dimensional propensity score identified a group of glyburide users with a near certain predicted probability of receiving

glyburide vs. glipizide. The cluster of glyburide users with high-dimensional propensity score near 1.0 did not appear to be related to convergence warnings, as five of the seven data partners without convergence warnings and five of six data partners with convergence warnings had patients with predicted probability of glyburide exposure near 1.0. The majority of the top ten variables that most strongly predicted glyburide exposure within each Data Partner were related to pregnancy, such as ICD-9-CM diagnosis code 648.83 for abnormal glucose tolerance of mother, antepartum, and CPT code 76811 or 76805 for ultrasound for pregnancy (see eTable 3 to 6; <http://links.lww.com/EDE/B229> for lists of top ten covariates selected by the high-dimensional propensity score algorithm). No glipizide users were identified within maximum allowable caliper distance for glyburide-treated persons with high-dimensional propensity score values of near 1.0. For Data Partner 9 and 11, the distributions of propensity scores were skewed and we found a match for only 43% to 64% of glyburide users. In the matched cohorts, all predefined covariates were balanced except for the baseline metformin use in the model matched only on covariates selected by the high-dimensional propensity score algorithm. In the five data partners in which the program completed the high-dimensional propensity score analysis, the HR matched on high-dimensional propensity scores was 1.49 (1.31, 1.70) and the HR matched on a propensity score that included both pre-defined variables and those identified using the high-dimensional propensity score algorithm was 1.51 (1.32, 1.71).

The HRs based on the secondary definition of serious hypoglycemia that included only emergency department cases were slightly higher than those using the primary outcome definition (eTable 7; <http://links.lww.com/EDE/B229>).

DISCUSSION

The primary goal of this assessment was to pilot-test the ability of the Sentinel Propensity Score Matching Tool to reproduce the known association between glyburide (vs. glipizide) and serious hypoglycemia. In the analyses matched on three different types of propensity scores, the incidence rate of serious hypoglycemia was 1.36 to 1.51-fold higher in users of glyburide vs. glipizide. These findings are consistent with the results of prior studies.^{7,8,9,18} Thus, we were able to reproduce this well-known association using the Sentinel Propensity Score Matching Tool, which enables the conduct of two-group comparative cohort evaluations in a privacy-preserving distributed data environment using an input specification form rather than a protocol, and customizable modular programs rather than *de novo* statistical programs. We have previously used the same tool to replicate another known association between angiotensin-converting enzyme inhibitor use and angioedema.²⁴

We observed a somewhat smaller HR (1.36- to 1.51-fold increased risk) compared to previous epidemiological studies (1.67- to 1.90-fold increased risk).^{7,8} This difference may be due to differences in the outcome definition and a younger study population. van Staa and colleagues conducted a cohort study enrolling subjects at least 20 years of age and over 61% of the study population were older than 65 years.⁷ Moreover, the study was conducted in the United Kingdom and hypoglycemia was ascertained using the Oxford Medical Information Systems (OXMIS) code. Shorr et al performed a cohort study specifically in subjects aged 65 years and above using data from the Tennessee Medicaid Program.⁸

Although we observed a lower crude incidence rate among glyburide users vs. glipizide users in the unmatched cohorts using data from 13 sites, the site adjusted HR indicated elevated risk of serious hypoglycemia for glyburide vs. glipizide. The observed difference in the direction of association was due to stratification by data partners. The site-specific estimates suggested an elevated risk for glyburide vs. glipizide at each site except for data partners 9 and 11.

Examination of the site-specific results, including the distribution of propensity score and the top ten empirically selected covariates by the high-dimensional propensity score suggested that the two sites had a large group of glyburide users who were pregnant women. Given the relatively high percentage of pregnant glyburide users, there may be residual confounding by pregnancy at these sites in the analyses that matched on a propensity score based solely on predefined covariates. Since the Sentinel Distributed Database is updated regularly, analyses with pregnant women excluded could not be repeated in the identical study population.

The existence of a subgroup of patients with a high-dimensional propensity score of close to 1.0 led to the identification of pregnant women as a group who, at least in some data partners, were essentially always prescribed glyburide in preference to glipizide. This was supported by the fact that the top ten empirically selected covariates at these sites (see eTable 3 to 6; <http://links.lww.com/EDE/B229>) were all related to pregnancy, a variable that were not specified by the investigator a priori. This reflects the real world utilization of glyburide, which is commonly used for the treatment of gestational diabetes.²⁵ If essentially all women with gestational diabetes take glyburide rather than glipizide, it may be inadvisable to try to infer a causal effect of glyburide vs. glipizide in this subgroup. Nevertheless, in the analyses that used only investigator-specified variables in the propensity score (which did not include markers of pregnancy), it is likely that pregnant women using glyburide were matched to non-pregnant glipizide users. Since pregnancy was not included in the analysis that was matched on the propensity score that included only investigator-defined variables, the difference between the investigator-defined propensity score and high-dimensional propensity score results may be due to either residual confounding by pregnancy (if it is a confounder) or to non-collapsibility of the HR. This illustrates a potential advantage of the high-dimensional propensity score approach: the identification of potential covariates that were not pre-specified by the investigators, which may ultimately help to improve confounding adjustment. However, in this case the difference between the investigator-defined propensity score and high-dimensional propensity score results was small, at least in part because pregnancy was not common in the overall cohort. In instances when an evaluation is performed to inform causal inferences about the exposures rather than to assess the performance of an analytic tool (i.e., nearly all instances), identification of subgroups in which nearly all individuals receive one treatment can signal the need to exclude or stratify based on that variable.

Our study included data from 13 data partners. The observed differences in sample sizes, events rate and effect estimates across data partners indicate the potential of database heterogeneity. Despite the differences, using the pooled stratified analysis approach, the HRs in the pooled analysis were highly consistent across different propensity score models.

The Sentinel Propensity Score Matching Tool includes three options of estimating propensity scores. The predefined covariates allows investigators to specify covariates based on prior knowledge, while the high-dimensional propensity score option allows an automated algorithm to identify a list of empirically selected covariates based on the potential for confounding the exposure-outcome association. In this study, the HRs were broadly consistent across all three types of propensity score models, supporting the robustness of the results. In the cohorts matched on covariates selected by the high-dimensional propensity score algorithm, the algorithm failed to achieve balance on baseline metformin use, one of the predefined variables. In the model matched on both predefined variables and variables selected by the high-dimensional propensity score, all predefined covariates were balanced. This indicates high-dimensional propensity score alone may not be sufficient to achieve balance for all predefined covariates when these variables are not included in the propensity score model. However, it is likely that metformin did not have a large empirical association with the outcome; otherwise, it would have been identified and included in the high-dimensional propensity score model.

Another goal of this study was to assess and identify practical issues encountered during the implementation of these tools and to identify issues for future enhancements. At the time of this pilot test, the Sentinel Propensity Score Matching Tool included a defensive coding strategy that prevented matching on propensity scores and further analysis when there were warnings about model convergence. In this pilot study, high-dimensional propensity score models ran without convergence warnings in seven out of 13 data partners; the remaining six data partners had warnings regarding questionable convergence of high-dimensional propensity score models.

These six sites were of smaller size compared to others, so insufficient sample size is a likely explanation for this questionable convergence. No matched tables or propensity score distribution figures were created and returned to the Sentinel Operations Center from these six data partners. This experience helped the Sentinel Operations Center and development team recognize that changes were needed for the next update to the Sentinel Propensity Score Matching Tool. Without the matched tables and distribution figures for the propensity scores with potentially questionable convergence, investigators and Sentinel Operations Center staff were unable to assess balance on important confounders after matching. By modifying the defensive coding strategy in an updated version Sentinel Propensity Score Matching Tool, in the future, matched tables and figures using propensity scores from models with warnings about convergence will be returned for review and assessment of balance or other anomalies. In this assessment, we used the default value of 200 for the number of empirically identified variables to include in the high-dimensional propensity score models at each Data Partner. It is possible that using a smaller number would have avoided potential convergence issues in this assessment. Since the Sentinel Distributed Database is updated regularly, this assessment could not be repeated in the identical study population.

At the time of this analysis, the high-dimensional propensity score portion of the Sentinel Propensity Score Matching Tool was written in Java. In this evaluation, we identified an issue that when multiple high-dimensional propensity score packages are running

concurrently on the same machine, the code implemented in Java may have strained computing resources, thereby, improperly selected variables into the high dimensional propensity score models. We removed the two sites that were affected by this issue from the pooled results of the high-dimensional propensity score models. The code has been rewritten in SAS and a new, JAVA-free version of the tool has been released.

The Sentinel Propensity Score Matching Tool and the data to which it was applied have several limitations. First, it controls for baseline but not time-varying covariates. Second, as other studies using claim data, it uses ICD-9 codes to identify outcomes and covariates, which are subject to misclassification. This limitation is mitigated by our use of validated algorithms with high positive predictive values. A limitation of distributed data environments such as Sentinel, in which data are updated regularly and locked copies are not maintained, is that iterative analyses to elucidate unexpected findings may need to be performed in non-identical study populations, which could introduce an additional source of variability across analyses.

In conclusion, the results of our assessment broadly demonstrated the ability of the Sentinel Propensity Score Matching Tool to successfully reproduce the known association between glyburide versus glipizide and serious hypoglycemia in the Sentinel Distributed Database, while identifying characteristics of the tool that needed to be improved.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

The authors would like to gratefully acknowledge the contributions of the following organizations: Aetna, HealthCore, Inc., Group Health Research Institute, Harvard Pilgrim Health Care Institute, HealthPartners Institute, Meyers Primary Care Institute, Marshfield Clinic Research Foundation, Humana, Inc., Kaiser Permanente Colorado, Kaiser Permanente Hawaii, Kaiser Permanente Northern California, Kaiser Permanente Northwest, and Optum, Inc.

Sources of financial support

The Mini-Sentinel program is funded by the Food and Drug Administration through contract HHSF2232009100061 from the US Department of Health and Human Services

REFERENCES

1. Behrman RE, Benner JS, Brown JS, McClellan M, Woodcock J, Platt R. Developing the Sentinel System--a national resource for evidence development. *N Engl J Med*. 2011 2 10;364(6):498–9. doi: 10.1056/NEJMp1014427 Epub 2011 Jan 12. [PubMed: 21226658]
2. Platt R, Carnahan R. The U.S. Food and Drug Administration's Mini-Sentinel Program. *Pharmacoepidemiol Drug Saf*. 2012 1; 21 Suppl 1: 1–303. doi: 10.1002/pds.3230.
3. Curtis LH, Weiner MG, Boudreau DM, Cooper WO, Daniel GW, Nair VP, Raebel MA, Beaulieu NU, Rosofsky R, Woodworth TS, Brown JS. Design considerations, architecture, and use of the Mini-Sentinel distributed data system. *Pharmacoepidemiol Drug Saf*. 2012 1;21 Suppl 1:23–31. doi: 10.1002/pds.2336. [PubMed: 22262590]
4. Sentinel Operations Center. Sentinel Modular Programs. Querying Tools: Overview of Functionality and Technical Documentation. Version 2.1.1. 2 2016 http://mini-sentinel.org/work_products/Data_Activities/Sentinel-Routine_Querying_System-Documentation.pdf [Assessed July 27, 2016].

5. Rassen JA, Schneeweiss S. Using high-dimensional propensity scores to automate confounding control in a distributed medical product safety surveillance system. *Pharmacoepidemiol Drug Saf.* 2012 1;21 Suppl 1:41–9. doi: 10.1002/pds.2328 [PubMed: 22262592]
6. Mini-Sentinel. Routine Querying Tools (Modular Programs). 12 2014 http://mini-sentinel.org/data/activities/modular_programs/details.aspx?ID=166 [Assessed January 12, 2016].
7. van Staa T, Abenhaim L, Monette J. Rates of hypoglycemia in users of sulfonylureas. *J Clin Epidemiol.* 1997 6;50(6):735–41. [PubMed: 9250272]
8. Shorr RI, Ray WA, Daugherty JR, Griffin MR. Individual sulfonylureas and serious hypoglycemia in older people. *J Am Geriatr Soc.* 1996 7;44(7):751–5. [PubMed: 8675920]
9. Gangji AS, Cukierman T, Gerstein HC, Goldsmith CH, Clase CM. A systematic review and meta-analysis of hypoglycemia and cardiovascular events: a comparison of glyburide with other secretagogues and with insulin. *Diabetes Care.* 2007 2;30(2):389–94. [PubMed: 17259518]
10. Mini-Sentinel Data Core. Mini-Sentinel Distributed Database Year Five Summary Report. 10 2015 http://www.mini-sentinel.org/work_products/Data_Activities/Mini-Sentinel_Year-5-Distributed-Database-Summary-Report.pdf [Assessed July 27, 2016].
11. Forrow S, Campion DM, Herrinton LJ, Nair VP, Robb MA, Wilson M, Platt R. The organizational structure and governing principles of the Food and Drug Administration's Mini-Sentinel pilot program. *Pharmacoepidemiol Drug Saf.* 2012 1;21 Suppl 1:12–7. doi: 10.1002/pds.2242. [PubMed: 22262588]
12. McGraw D, Rosati K, Evans B. A policy framework for public health uses of electronic health data. *Pharmacoepidemiol Drug Saf.* 2012 1;21 Suppl 1:18–22. doi: 10.1002/pds.2319. [PubMed: 22262589]
13. Mini-Sentinel Operations Center. Mini-Sentinel Modular Programs: Modular Programs 3: Frequency of select events during exposure to a drug/procedure group of interest. Version 7.5. 9 2014 http://www.mini-sentinel.org/work_products/Data_Activities/Mini-Sentinel-Modular-Program-3-Documentation.pdf [Assessed September 24, 2015].
14. Gagne JJ, Glynn RJ, Avorn J, Levin R, Schneeweiss S. A combined comorbidity score predicted mortality in elderly patients better than existing scores. *J Clin Epidemiol.* 2011 7;64(7):749–59. doi: 10.1016/j.jclinepi.2010.10.004 Epub 2011 Jan 5. [PubMed: 21208778]
15. Schneeweiss S, Seeger JD, Maclure M, Wang PS, Avorn J, Glynn RJ. Performance of comorbidity scores to control for confounding in epidemiologic studies using claims data. *Am J Epidemiol.* 2001 11 1;154(9):854–64. [PubMed: 11682368]
16. Ginde AA, Blanc PG, Lieberman RM, Camargo CA Jr. Validation of ICD-9-CM coding algorithm for improved identification of hypoglycemia visits. *BMC Endocr Disord.* 2008 4 1;8:4. doi: 10.1186/1472-6823-8-4.
17. Quan H, Li B, Saunders LD, Parsons GA, Nilsson CI, Alibhai A, Ghali WA;IMECCHI Investigators. Assessing validity of ICD-9-CM and ICD-10 administrative data in recording clinical conditions in a unique dually coded database. *Health Serv Res.* 2008 8;43(4):1424–41. [PubMed: 18756617]
18. Schelleman H, Bilker WB, Brensinger CM, Wan F, Hennessy S. Anti-infectives and the risk of severe hypoglycemia in users of glipizide or glyburide. *Clin Pharmacol Ther.* 2010 8;88(2):214–22. doi: 10.1038/clpt.2010.74 Epub 2010 Jun 30. [PubMed: 20592722]
19. Schneeweiss S, Rassen JA, Glynn RJ, Avorn J, Mogun H, Brookhart MA. High-dimensional propensity score adjustment in studies of treatment effects using health care claims data. *Epidemiology.* 2009 7;20(4):512–22. doi: 10.1097/EDE.0b013e3181_a663cc. [PubMed: 19487948]
20. Rassen JA, Glynn RJ, Brookhart MA, Schneeweiss S. Covariate selection in high-dimensional propensity score analyses of treatment effects in small samples. *Am J Epidemiol.* 2011 6 15;173(12):1404–13. doi: 10.1093/aje/kwr001 Epub 2011 May 20. [PubMed: 21602301]
21. Cox DR. Regression models and life-tables. *Journal of the Royal Statistical Society.* 1972; Series B (Methodological) 34: 187–220.
22. Austin PC. Optimal caliper widths for propensity-score matching when estimating differences in means and differences in proportions in observational studies. *Pharm Stat.* 2011 Mar-Apr;10(2): 150–61. doi: 10.1002/pst.433. [PubMed: 20925139]

23. Normand ST, Landrum MB, Guadagnoli E, Ayanian JZ, Ryan TJ, Cleary PD, McNeil BJ. Validating recommendations for coronary angiography following acute myocardial infarction in the elderly: a matched analysis using propensity scores. *J Clin Epidemiol.* 2001 4;54(4):387–98 [PubMed: 11297888]
24. Gagne JJ, Han X, Hennessy S, Leonard CE, Chrischilles EA, Carnahan RM, Wang SV, Fuller C, Iyer A, Katcoff H, Woodworth TS, Archdeacon P, Meyer TE, Schneeweiss S, Toh S. Successful Comparison of US Food and Drug Administration Sentinel Analysis Tools to Traditional Approaches in Quantifying a Known Drug-Adverse Event Association. *Clin Pharmacol Ther.* 2016 11;100(5):558–564. doi: 10.1002/cpt.429. [PubMed: 27416001]
25. Camelo Castillo W, Boggess K, Stürmer T, Brookhart MA, Benjamin DK Jr, Jonsson Funk M. Trends in glyburide compared with insulin use for gestational diabetes treatment in the United States, 2000–2011. *Obstet Gynecol.* 2014 6;123(6):1177–84. doi: 10.1097/AOG.000000000000285. [PubMed: 24807336]

Statement about availability of data and code for replication

Sentinel uses a distributed data approach in which Data Partners (DPs) maintain physical and operational control over electronic health data in their existing environments after transforming their data into a common data model. This analysis utilized the Sentinel distributed database and standardized data querying tools. Code for Sentinel standardized data querying tools, query specifications, and related documentation are shared via the Sentinel website, which allows for transparency and potential replicability of this study on other data sources. Due to its distributed nature, Sentinel generally does not save, maintain, or post individual level datasets. Sentinel DPs update data at varying intervals and retain a limited number of iterations of their historical data, which may affect replication of this assessment.

Table 1. Baseline characteristics of glyburide users and glipizide users from 13 Sentinel data partners.

Covariates	Unmatched			Matched on propensity score		
	Glyburide (N=198,553)	Glipizide (N=379,508)	Standardized difference	Glyburide (N=173,657)	Glipizide (N=173,657)	Standardized difference
New Users^a						
Sex, female	101,307 (51.0%)	166,997 (44.0%)	0.141	78,013 (44.9%)	80,048 (46.1%)	-0.024
Mean age (SD)	55.0 (13.9)	59.7 (12.7)	-0.353	57.9 (12.8)	57.7 (12.5)	0.016
Recorded history of						
Chronic kidney disease	11,640 (5.9%)	42,418 (11.2%)	-0.190	11,579 (6.7%)	12,794 (7.4%)	-0.027
Serious Hypoglycemia	4,532 (2.3%)	18,754 (4.9%)	-0.143	4,433 (2.6%)	4,657 (2.7%)	-0.008
Insulin	14,903 (7.1%)	34,011 (9.0%)	-0.069	13,692 (7.9%)	13,988 (8.1%)	-0.006
Metformin	66,450 (33.5%)	183,560 (48.4%)	-0.303	65,680 (37.8%)	67,446 (38.8%)	-0.021
Other antidiabetic drugs	28,462 (14.3%)	55,113 (14.5%)	0.005	28,248 (16.3%)	28,908 (16.6%)	-0.010
Combined Charlson/Elixhauser comorbidity score (SD)	0.4 (1.7)	0.7 (2.0)	-0.151	0.5 (1.7)	0.6 (1.8)	-0.040
Health Service Utilization Intensity						
Number of unique drugs dispensed (SD)	5.3 (4.3)	6.3 (4.6)	-0.218	5.6 (4.4)	5.7 (4.5)	-0.030
Number of dispensed prescriptions (SD)	12.6 (13.1)	14.2 (13.5)	-0.121	13.6 (13.7)	14.0 (13.9)	-0.026
Number of inpatient hospital encounters (SD)	0.1 (0.5)	0.2 (0.6)	-0.094	0.1 (0.5)	0.1 (0.5)	0.000
Number of non-acute institutional encounters (SD)	0.1 (0.9)	0.1 (1.0)	0.034	0.1 (1.0)	0.1 (1.0)	0.000
Number of emergency room encounters (SD)	0.3 (0.7)	0.3 (0.8)	0.029	0.3 (0.8)	0.3 (0.8)	0.000
Number of ambulatory encounters (SD)	6.9 (7.6)	6.3 (7.8)	0.082	6.5 (7.4)	6.8 (8.7)	-0.037
Number of other ambulatory encounters (SD) ^b	2.0 (3.5)	2.7 (4.4)	-0.167	1.7 (3.3)	1.9 (3.8)	-0.036

SD – Standard deviation.

^aPlease note, one Data Partner removed a small number of users that moved to administrative services-only plans. Information from these users is included in Tables 1 and 2, but removed from subsequent tables.

^bOther ambulatory encounters included telemedicine and email consults, etc.

The baseline characteristics of glyburide users and glipizide users from five Sentinel data partners with high dimensional propensity score models converged and completed without errors.

Table 2.

Covariates	Unmatched			Matched on predefined covariates			Matched on hGPS only			Matched on predefined covariates & hGPS		
	Glyburide (N=15,111) (n)	Glipizide (N=12,053) (n)	Standardized difference	Glyburide (N=12,053) (n)	Glipizide (N=12,053) (n)	Standardized difference	Glyburide (N=116,53) (n)	Glipizide (N=116,53) (n)	Standardized difference	Glyburide (N=116,64) (n)	Glipizide (N=116,64) (n)	Standardized difference
New Users^a												
Gender (Female)	69,491 (50.0%)	75,873 (41.7%)	0.165	52,058 (43.3%)	53,888 (44.8%)	-0.031	48,028 (41.1%)	47,968 (41.0%)	0.001	47,931 (41.1%)	47,733 (40.9%)	0.003
Mean age (SD)	52.8 (14.1)	57.0 (12.5)	-0.311	55.5 (12.9)	55.2 (12.4)	0.021	56.2 (12.4)	56.2 (12.4)	0.000	56.2 (12.4)	56.2 (12.4)	0.003
Recorded history of												
Chronic Kidney Disease	4,704 (3.4%)	11,470 (6.3%)	-0.136	4,648 (3.9%)	5,483 (4.6%)	-0.035	4,604 (3.9%)	5,402 (4.6%)	-0.034	4,605 (3.9%)	4,677 (4.0%)	-0.003
Serious Hypoglycemia	2,555 (1.8%)	4,367 (2.4%)	-0.039	2,465 (2.0%)	2,547 (2.1%)	-0.005	2,446 (2.1%)	2,470 (2.1%)	-0.001	2,438 (2.1%)	2,456 (2.1%)	-0.001
Insulin	8,552 (6.1%)	14,669 (8.1%)	-0.075	8,272 (6.9%)	8,427 (7.0%)	-0.005	8,095 (6.9%)	8,736 (7.5%)	-0.021	8,107 (7.0%)	7,964 (6.8%)	0.005
Metformin	44,603 (32.1%)	79,171 (45.5%)	-0.236	43,977 (36.5%)	45,16 (37.5%)	-0.020	42,705 (36.5%)	49,924 (42.7%)	-0.126	42,794 (36.7%)	42,656 (36.6%)	0.002
Other antidiabetic drugs	22,311 (16.0%)	39,383 (21.6%)	-0.144	22,137 (18.4%)	22,791 (18.9%)	-0.014	22,098 (18.9%)	24,727 (21.1%)	-0.056	22,121 (19.0%)	22,171 (19.0%)	-0.001
Combined Charlson/Elixhauser comorbidity score (SD)	0.3 (1.5)	0.4 (1.8)	-0.103	0.3 (1.5)	0.4 (1.6)	-0.040	0.3 (1.6)	0.3 (1.6)	0.002	0.3 (1.6)	0.3 (1.6)	0.000
Health Service Utilization Intensity												
Number of unique generic drugs dispensed (SD)	4.9 (4.1)	5.7 (4.5)	-0.194	5.1 (4.3)	5.3 (4.3)	-0.031	5.1 (4.3)	5.4 (4.2)	-0.060	5.1 (4.3)	5.1 (4.2)	0.000
Number of dispensed prescriptions (SD)	11.8 (12.6)	14.5 (14.1)	-0.201	12.6 (13.0)	13.0 (13.3)	-0.030	12.7 (13.1)	13.3 (13.0)	-0.045	12.7 (13.1)	12.7 (13.0)	0.001
Number of inpatient hospital encounters (SD)	0.1 (0.5)	0.2 (0.5)	-0.125	0.1 (0.5)	0.1 (0.5)	0.000	0.1 (0.5)	0.1 (0.5)	0.000	0.1 (0.5)	0.1 (0.5)	0.000
Number of non-acute institutional	0.1 (1.0)	0.1 (1.2)	-0.055	0.1 (1.0)	0.1 (1.0)	0.000	0.1 (1.1)	0.1 (1.1)	0.001	0.1 (1.1)	0.1 (1.1)	-0.001

Table 3. Incidence rates and hazard ratios of emergency department visits and hospital admissions for hypoglycemia.

Exposure	New users ^a	Person years at risk	Serious hypoglycemia events	Incidence rate per 1,000 person years	Hazard Ratio (95% CI)
Data from 13 data partners					
Unmatched ^b					
Glyburide	198,550	89,719	1,685	19	1.11 (1.05–1.18)
Glipizide	379,507	244,094	5,406	22	
Predefined covariates - Unconditional model ^c					
Glyburide	173,655	83,108	1,633	20	1.35 (1.26, 1.45)
Glipizide	173,656	99,834	1,393	14	
Predefined covariates - Conditional model ^d					
Glyburide	173,655	38,986	1,064	27	1.36 (1.24, 1.49)
Glipizide	173,656	38,986	784	20	
Data from 5 data partners in which the hdPS model converged and completed without errors					
Unmatched ^b					
Glyburide	139,113	58,075	905	16	1.26 (1.16, 1.38)
Glipizide	181,911	94,941	1,079	11	
Predefined covariates - Unconditional model ^c					
Glyburide	120,334	53,366	859	16	1.41 (1.27, 1.56)
Glipizide	120,335	61,552	666	11	
Predefined covariates - Conditional model ^d					
Glyburide	120,334	24,708	568	23	1.42 (1.25, 1.62)
Glipizide	120,335	24,708	399	16	
hdPS - Unconditional model ^c					
Glyburide	116,930	52,816	870	17	1.50 (1.36, 1.66)
Glipizide	116,931	62,526	644	10	
hdPS - Conditional model ^d					
Glyburide	116,930	24,494	581	24	1.49 (1.31, 1.70)

Exposure	New users ^a	Person years at risk	Serious hypoglycemia events	Incidence rate per 1,000 person years	Hazard Ratio (95% CI)
Glipizide	116,931	24,498	389	16	
Predefined covariates and hdPS - Unconditional model ^c					
Glyburide	116,639	52,713	868	17	1.49 (1.34, 1.65)
Glipizide	116,641	61,778	644	10	
Predefined covariates and hdPS - Conditional model ^d					
Glyburide	116,639	24,332	575	24	1.51 (1.32, 1.71)
Glipizide	116,641	24,333	382	16	

hdPS – High dimensional propensity score; SD – Standard deviation.

^aPlease note, one Data Partner removed a small number of users that moved to administrative services only plans. Information from these users is included in Tables 1 and 2, but removed from subsequent tables.

^bPlease note, all models were stratified by Data Partner.

^cThe conditional models were stratified by the matched pair.

^dThe unconditional models were not stratified by the matched pair.