# MINI-SENTINEL METHODS

# ANALYTICAL METHODS TO ASSESS ROBUSTNESS OF DRUG SAFETY MONITORING RESULTS

**Prepared by:** Joshua J. Gagne, PharmD, ScD,[1] Shirley V. Wang, PhD,[1] Matthew Fox, PhD,[2] Timothy Lash, PhD,[3] Wesley Eddings, PhD,[1] Sascha Dublin, MD, PhD,[4] Tobias Gerhard, PhD,[5] Bruce Psaty, MD, PhD,[6] Judy Maro, PhD,[7] Catherine Rogers, MPH,[7] Rita Ouellet-Hellstrom, PhD, MPH,[8] Simone Pinheiro, ScD, MSc,[8] Mark Levenson, PhD,[9] Richard Forshee,[10] Azadeh Shoaibi, PhD, MS, MHS,[11] Sebastian Schneeweiss, MD, ScD[1]

**Author Affiliations:** 1. Division of Pharmacoepidemiology and Pharmacoeconomics, Department of Medicine, Brigham and Women's Hospital, Boston, MA 2. Department of Epidemiology, Boston University School of Public Health, Boston University, Boston, MA 3. Department of Epidemiology, Rollins School of Public Health, Emory University, Atlanta, GA 4. Group Health Research Institute, Seattle, WA 5. Center for Health Services Research on Pharmacotherapy, Chronic Disease Management, and Outcomes, Institute for Health, Health Care Policy and Aging Research, Rutgers, The State University of New Jersey, New Brunswick, NJ, and Department of Pharmacy Practice and Administration, Ernest Mario School of Pharmacy, Rutgers University, Piscataway, NJ 6. Cardiovascular Health Research Unit, Departments of Medicine, Epidemiology, and Health Services, University of Washington, Seattle, WA, and Group Health Research Institute, Group Health Cooperative, Seattle, WA 7. Department of Population Medicine, Harvard Medical School and Harvard Pilgrim Healthcare Institute, Boston, MA 8. Office of Surveillance and Epidemiology, Center for Drug Evaluation and Research, US Food and Drug Administration, Silver Spring, MD 9. Office of Biostatistics, Center for Drug Evaluation and Research, US Food and Drug Administration, Silver Spring, MD 10. Office of Biostatistics and Epidemiology, Center for Biologics Evaluation and Research, US Food and Drug Administration, Silver Spring, MD 11. Office of Medical Policy, Center for Drug Evaluation and Research, US Food and Drug Administration, Silver Spring, MD

## September 4, 2015

Mini-Sentinel is a pilot project sponsored by the U.S. Food and Drug Administration (FDA) to inform and facilitate development of a fully operational active surveillance system, the Sentinel System, for monitoring the safety of FDA-regulated medical products. Mini-Sentinel is one piece of the Sentinel Initiative, a multi-faceted effort by the FDA to develop a national electronic system that will complement existing methods of safety surveillance. Mini-Sentinel Collaborators include Data and Academic Partners that provide access to health care data and ongoing scientific, technical, methodological, and organizational expertise. The Mini-Sentinel Coordinating Center is funded by the FDA through the Department of Health and Human Services (HHS) Contract number HHSF223200910006I.

# Analytical Methods to Assess Robustness of Drug Safety Monitoring Results

# I. INTRODUCTION

Analyses of secondary electronic healthcare data, such as those that comprise the Sentinel Distributed Database (SDD), can provide important and timely information about the safety of marketed medical products. The Sentinel program's routine analytic framework now includes Level 2 and 3 routine query tools that can perform self-control risk interval analyses, a propensity score-matched cohort analyses, multivariable risk adjustment, and inverse probability of treatment weighting. These programs enable the Food and Drug Administration (FDA) to implement a range of methods that are often used in medical product safety assessments and provide the capability to perform rigorous sequential analyses to investigate associations between medical products and outcomes as data describing the experience with these products accrue prospectively in the SDD. These tools are currently being used in several assessments of the safety of newly marketed medications.[1,2]

As with all observational analyses of administrative data, the results of Sentinel medical product safety assessments are vulnerable to certain potential biases, including bias due to confounding, patient selection, selective prescribing, and misclassification of variables, including exposures, outcomes, and covariates. While these biases can often be mitigated using appropriate design and analytic strategies,[3] it is not possible to guarantee control of all of these biases. Many quantitative bias analysis methods have been developed to assess the impact of potential residual biases on the results of observational analyses. For example, quantitative bias analysis methods can be used to correct effect estimates for potential outcome misclassification if the sensitivity and specificity of the outcome definition are known or can be reasonably estimated.[4] In addition, the potential impact of unmeasured confounding can be quantified or bounded if associations between unmeasured confounders and the exposure and outcome can be estimated or are assumed.[5]

As FDA increasingly uses Sentinel assessments to gather information on the safety of marketed medical products, it is critical to have the ability to quickly understand the robustness of the results of these assessments with respect to biases that can commonly occur in observational analyses. However, methods to evaluate the sensitivity of results to underlying design and analytic assumptions and to quantify the sensitivity of results to unguarded threats to validity have not yet been integrated into the Sentinel routine analytic framework.

# II. OBJECTIVES

The purpose of this Workgroup was to identify suitable methods for evaluating the robustness of findings from observational analyses of the safety of drugs using longitudinal healthcare databases. The intent is that these methods for assessing the robustness of findings will be integrated into the Sentinel routine analytic framework and made available for use in safety monitoring activities to support decision makers. In particular, the scope of work focused on methods that either make structural assumptions about confounding and misclassification mechanisms that can cause bias (e.g., correcting for imperfect measurement of outcomes) or on methods that enable empirical evaluation of the effect of varying

study parameters, such as the covariate assessment period and the exposure effect window. In particular, the Workgroup sought to identify and determine how to integrate methods that are applicable to the distributed data environment in which individual-level data reside with the Data Partners and only aggregate-level, de-identified data are transmitted beyond the Data Partners' firewalls.

The specific objectives of this Workgroup were to:

- Identify and select suitable sensitivity analyses and quantitative bias analysis methods for residual confounding and other biases that can occur in observational analyses of healthcare databases and that allow assessment of the robustness of such findings from these analyses;

- Evaluate, select, and determine how to integrate different quantitative bias analysis methods, such as simple bias analysis (i.e., quantification of bias from a single source), probabilistic bias analysis (which extends simple bias analysis by assigning probability distributions to the bias parameters, rather than using a few fixed values), and multiple bias modeling (i.e., simultaneous quantification of different biases), into the Sentinel routine analytic framework;
- Identify and make available the capability of performing selected sensitivity analyses alongside existing routine query tools;
- Evaluate the implementation of the sensitivity analysis and quantitative bias analysis approaches in a test case using data in the Sentinel Common Data Model (MSCDM) format;
- Develop documentation for the sensitivity analysis program module(s) to support routine use.

## III. IDENTIFICATION OF METHODS TO ASSESS THE ROBUSTNESS OF SURVEILLANCE RESULTS

The Workgroup identified potential sensitivity and quantitative bias analysis approaches by reviewing the literature and textbooks and by soliciting input from Workgroup members regarding existing and ready-to-implement methods that can be used to assess the robustness of medical product safety monitoring results produced by the Level 2 and 3 routine query tools within the Mini-Sentinel data environment. As the purpose of the sensitivity analyses is to assess the sensitivity of the routine query tool results to underlying design and analytic assumptions using existing data in the MSCDM, this project did not include approaches that require external data collection (e.g., medical record validation), which are being addressed in other Mini-Sentinel activities.

The Workgroup sought to identify approaches to address the following issues that are common to observational analysis of electronic healthcare data, including (Table 1):

**Table 1. Common biases and issues that threaten the validity of findings from observational medical product safety assessments**

| | |
|---|---|
| Confounding (including confounder misclassification, unmeasured confounding, unknown confounding, and other sources of unaddressed confounding) | Outcome misclassification |
| Exposure misclassification | Selection bias (including informative censoring) |
| Misspecification of the exposure risk window | Misspecification of the covariate assessment window |
| Misspecification of the length of the "wash out period" in incident user designs | Misspecification of time-varying hazards |

The Workgroup identified candidate approaches and categorized them based on a set of five criteria:

1. Routine query tool to which the approach applies (i.e., self-controlled risk interval, propensity score matching, cohort regression)
2. Source of bias/difference in results addressed (e.g., exposure misclassification, confounding, assumptions about risk window, etc.)
3. Whether the approach potentially alters the population evaluated, as differences in results between the sensitivity analysis and the primary analysis may be due to differences in characteristics of patients in the population if these differences modify the treatment effect
4. Practicability:
    a. Defined as how feasible it would be to accommodate the implementation of the approach within the Sentinel Operations Center (SOC) production schedule
    b. This was determined jointly by the Workgroup and SOC
5. Utility:
    a. Defined as the extent to which the approach would increase confidence in the robustness of results
    b. Utility was ranked on a score of 1 to 5 where 1 indicated not useful and 5 indicated useful
    c. All Workgroup members were asked to score each approach
    d. Final utility scores were determined by consensus on a Workgroup call

All of the identified approaches, including brief descriptions and information for each criterion are presented in APPENDIX B (Table B1). The approaches were grouped into two main categories: (1) sensitivity analyses and (2) quantitative bias analysis approaches. In the setting of the SDD, the sensitivity analyses have "front end" solutions and the quantitative bias analysis methods have "back end" solutions, based on whether the analyses are performed using distributed code that is run within each Data Partner or whether the analyses are conducted on aggregate data returned by the Data Partners, respectively.

## A. SENSITIVITY ANALYSES

Solutions for sensitivity analyses are "front end" approaches (see Table B1 for list) in that they require additional data from the Data Partner than what may be available from the primary analysis based on a single set of parameters. The main objective of these approaches is to assess the sensitivity of the primary findings to underlying assumptions made by the user. Ideally, these assumptions are based on the biology and pharmacology of the specific assessment, but are often imperfectly informed by this background information, and therefore not definitive to balance multiple considerations with potentially unknown impact on the results. For example, it is common to consider patients to be at risk for an outcome over the duration of their treatment episode. However, some drug-induced effects (e.g., anaphylaxis) manifest most commonly shortly following initial exposure whereas other effects manifest only after some induction period. Misspecifying the etiologically relevant period could mean that an important drug-outcome association is missed or substantially underestimated. Sensitivity analyses vary the length of the exposure risk window to elucidate whether the results of an assessment are sensitive to the user's underlying assumption about the specified risk window.

Often, these analyses also involve creation of cohorts that differ slightly from the primary analysis the exclusion of more patients than in the primary cohort because fewer patients will be identified as new users.

## B. QUANTITATIVE BIAS ANALYSIS APPROACHES

Quantitative bias analysis methods quantify residual bias in effect estimation by modeling structural assumptions about confounding and misclassification mechanisms that can cause bias. They also assess the impact of these potential mechanisms on changes in study results (Table B1). Such quantitative bias modeling approaches have been well described elsewhere.[4,6] Commonly used approaches involve correcting effect estimates for outcome misclassification and examining the impact of potential unmeasured confounding. For example, given the sensitivity and specificity of a claims-based outcome definition, one can estimate what an observed effect estimate would have been under perfect sensitivity and specificity. Similarly, assuming the prevalence of an unmeasured confounder and its associations with the exposure and outcome, and assuming independence from controlled covariates, one can use algebraic formulas to calculate what an effect estimate would have been if the unmeasured confounder were adequately controlled. Other approaches address issues such as selection bias and exposure misclassification. Methods have also been developed to address multiple sources of bias simultaneously.[4]

Several resources exist for implementing quantitative bias modeling methods, including those available in the textbook, *Applying Quantitative Bias Analysis to Epidemiologic Data*, by Lash, Fox, and Fink[4] and other resources made available or curated by the authors: https://sites.google.com/site/biasanalysis/ and variants of approaches to address unmeasured confounding: http://www.drugepi.org/dope-downloads/#Sensitivity Analysis. In addition, under contract with FDA, SciMetrika has developed a user-friendly software program for implementing quantitative bias analysis. The program implements a number of approaches, including simple bias analysis for outcome misclassification, exposure misclassification, and confounding for both cohort and self-controlled designs, as well as multiple and multi-dimensional bias approaches.

## IV. IMPLEMENTATION

The Workgroup worked closely with the SOC to implement several "front end" sensitivity analysis approaches in the Cohort Identification and Descriptive Analysis (CIDA) tool. Implementation focused on those approaches that apply to the PSM tool since this module had been integrated into the CIDA tool at the time that this work was conducted. Approaches selected for subsequent implementation in the routine analytic framework were those that were both practicable, based on MSOC's assessment, and had high utility scores (i.e., ≥ 4). The following approaches were selected:

- #1 Vary length of washout period in incident user designs
- #4 Vary start of exposure risk window
- #5 Very length of exposure risk window
- #6 Vary baseline covariate assessment periods
- #7 Vary the sets of confounders to adjust, such as using the high-dimensional propensity score (hdPS)

Appendix A describes where in the query request form each "front end" sensitivity analysis can be specified. As each sensitivity analysis varies existing parameters of the CIDA and PSM tool, the query request form was modified to enable the user to specify multiple values for these parameters in the primary analyses.

The Workgroup also worked closely with SciMetrika and FDA members of FDA's Center for Biologics Evaluation and Research, who have been collaborating on a project to implement a number of quantitative bias analysis methods described above into a user-friendly software tool. As part of this collaboration, some Workgroup members received and reviewed the SAS code that conducts the analyses within the tool. Particular focus was given to ensuring that the outputs of the Level 2 and 3 tools that are available for use in Sentinel were compatible as inputs into the SciMetrika tool. The SciMetrika tool enables analyses based on both aggregate data from 2x2 tables as well as those based on individual-level data. As Mini-Sentinel seeks to minimize sharing of individual-level data, the Workgroup focused on approaches that use aggregate-level information. The SCRI and PSM tools are particularly well suited to the aggregate-level data analyses since both yield forms of 2x2 tables that are adjusted for confounding. As part of the collaboration and through this Workgroup, SciMetrika has provided the SAS code to SOC as a prototype for conducting quantitative bias analysis within Sentinel.

## V. TEST CASE

The Workgroup used the example of lisinopril versus beta-blockers and angioedema as a known positive test case. The primary analysis used specifications that were created during the development and beta-testing of the PSM module. These specifications were based on the analysis by Toh and colleagues conducted in the SDD.[7]

## A. EXPOSURE AND OUTCOME PAIR

The exposure of interest was defined as new use of lisinopril with no prior use of lisinopril, any other angiotensin converting enzyme inhibitors, or beta-blockers in the prior 183 days. We identified initiators of beta-blockers, with no prior use of beta-blockers or any angiotensin converting enzyme inhibitor, as the comparator group. In the primary analysis, patients were followed from the date of initiation (i.e., the index date) until outcome occurrence, discontinuation of the index medication, death, or disenrollment from the health care plan. The angioedema outcome was defined as an International Classification of Disease code 995.1 (angioedema) recorded in any position during an outpatient, inpatient, or emergency department encounter.

## B. DATA

The WG used data from a large commercial insurer covering the period 2008 to 2013. These data were converted into the MSCDM.

## C. CONFOUNDING ADJUSTMENT

Six covariates were included in the propensity score model in the primary analysis: age at index date, prior allergic reactions, diabetes, heart failure, ischemic heart disease, and use of prescription non-steroidal anti-inflammatory drugs. Except for age, these covariates were assessed in the 183 days preceding each patient's index date. We matched lisinopril initiators to beta-blocker initiators in a 1:1 ratio using the propensity score in the primary analysis.

## D. ROBUSTNESS METHODS APPLIED

### 1. Sensitivity analyses with "front end" solutions

We conducted five empirical sensitivity analyses using "front end" approaches. We first conducted a sensitivity analysis in which we increased the number of confounders included in the propensity score from 6 to 200 by using the high-dimensional propensity score (hdPS) algorithm.[8] The hdPS algorithm identifies potential confounders, or proxies thereof, by assessing their empirical associations between the exposure and outcome and selecting those with the highest potential for causing bias. We used the "exposure-based" hdPS implementation in which only the prevalence of the potential confounders and their associations with the exposure were considered.

In the second sensitivity analysis, we extended the new user washout and covariate assessment windows from 183 days to 365 days as a strategy to reduce exposure and confounder misclassification. Using a longer washout period ensures that patients identified as new users had only more distant exposure if they had any at all. Extending the covariate assessment window enables more complete ascertainment of chronic conditions that may have been recorded previously but not in more recent medical encounters. The tradeoff with these modifications is that requiring patients be continuously enrolled in the data for 365 days rather than 183 days usually reduces the size of the eligible analysis population.

In the third sensitivity analysis scenario, we capped the exposure risk window at the first 30 days of follow-up. Such a sensitivity analysis can be useful in different ways depending on the specific exposure-outcome relation of interest. If a drug is not believed to have a short-term effect on a given outcome, an analysis restricted to the period shortly after the index date can elucidate whether there are unaccounted for differences between the treatment groups that give rise to immediate differences in outcomes.[9] If the risk of the outcome is believed to vary over time following treatment initiation (as with lisinopril and angioedema), varying the exposure risk window can also shed light on the highest risk period. Furthermore, analyses of short-term outcomes can be useful to limit bias due to time-varying confounding and informative censoring. Ideally, clinical knowledge would be used to guide the selection and interpretation of the most relevant sensitivity analyses for a given safety assessment.

We also conducted two additional sensitivity analyses in which we simultaneously implemented two of the approaches described above. First we extended the washout and covariate assessment windows to 365 days and implemented hdPS. Second, we implemented hdPS with a 183-day washout and covariate assessment windows, but restricted follow-up to a maximum of 30 days.

## 2. Quantitative bias analyses with "back end" approaches

We applied two single bias analysis strategies, addressing bias due to outcome misclassification and to potential unmeasured confounding. For the outcome misclassification analysis, we used published validation studies to obtain classification parameters for the angioedema outcome definition. No published validation studies examined sensitivity and specificity of the algorithm, but several studies reported positive predictive values (PPVs) ranging from 90% to 95% for records that could be used for validation.[10-12] We used the minimum and maximum of this range to calculate an approximate specificity of the definition in test case population. Assuming a cumulative incidence of 0.00203, which corresponds to the cumulative incidence of angioedema in the primary matched cohort described below, and assuming a sensitivity of 1.0, a PPV of 90% corresponds to a specificity of 0.999777494 and a PPV of 95% corresponds to a specificity of 0.999891798. Note that assuming a sensitivity of 1.0 is conservative since, for a given PPV and a fixed cumulative incidence, higher sensitivity implies lower specificity. Assuming non-differential sensitivity and specificity of outcome classification between exposure groups, we examined five different scenarios for each value of specificity in which we varied the sensitivity across the values 0.2, 0.4, 0.6, 0.8, and 1.0. We also considered two additional scenarios – one in which the specificity was 1.0 and the sensitivity was 0.80 and the other in which the specificity was 0.99 and the sensitivity was 1.0.

To assess the extent to which unmeasured and independent confounding could have explained the observed result, we modeled an unknown confounder (which could represent a single confounder [e.g., race] or set of confounders) by assuming a prevalence of 0.2 in the lisinopril group and a prevalence of 0.1 in the beta-blocker group (i.e., a confounder-exposure association of 2.0). We also assumed that the confounder doubled the risk of the outcome (i.e., confounder-outcome association of 2.0). By assuming a positive association with both exposure and outcome, we ensured that the confounder caused upward bias, consistent with the direction of the observed effect.

## E. RESULTS

### 1. Primary results

We identified 385,649 lisinopril initiators and 274,977 beta-blocker initiators eligible for the primary analysis based on 183-day continuous enrollment and new user washout windows. Matching on the pre-defined propensity score with a maximum caliper of 0.025 units of the propensity score resulted in 231,520 matched pairs (84% of the beta-blocker initiators). APPENDIX C (Figure C1) and APPENDIX D (Figure D1) display the distribution of the pre-defined propensity scores before and after matching, respectively. APPENDIX E (Figure E1) provides a screenshot of the standard PSM output depicting the distributions of pre-defined variables between the matched cohorts.
Two separate primary results were obtained and used in subsequent analyses. The first result was derived from a Cox proportional hazards model that is built into the PSM tool where it is used to aggregate data from multiple Data Partners into a single stratified model. By conducting a time-to-event analysis, this model uses person-time data in the formation of risk sets. The hazard ratio (HR) from this model was 2.32 (95% confidence interval [CI], 1.94 to 2.79) and was used as the reference result for the "front end" sensitivity analysis approaches.

The second approach used counts of the numbers of matched lisinopril and beta-blocker initiators and the number of individuals in each group who experienced an angioedema outcome during follow-up. The 2x2 table below displays the count data (Figure 1). While this test case used data from a single Data Partner, a similar table can be created by aggregating 2x2 table data from fixed ratio-matched cohort across multiple Data Partners.

**Figure 1. 2x2 table data of matched lisinopril and beta-blocker initiators and the number of individuals in each group who experienced an angioedema outcome during follow-up**

| | Lisinopril | Beta-blockers |
|---|---|---|
| Angioedema | 664 | 276 |
| No angioedema | 230856 | 231244 |
| Total | 231520 | 231520 |

The resulting risk ratio from these data was 2.41 (95% CI, 2.09 to 2.77) and was used as the reference result for the "back end" quantitative bias analysis solutions. This risk ratio is adjusted for the same variables that went into the propensity score model but differs slightly from the HR from the Cox model likely because it does not account for differences in follow-up time between the exposure groups. Notably, the mean person-time at risk in the lisinopril group was 229 days (standard deviation [sd], 264 days) and in the beta-blocker group was 212 (sd, 259) days. It is expected that not accounting for the longer average follow-up time in the lisinopril group would result in a higher estimate of association, as was observed.

## 2. Results of sensitivity analyses with "front end" solutions

The results of the five empirical sensitivity analysis scenarios are presented in Table **2**. Across all five scenarios, the fact that the hazard ratios were larger than that from the primary analysis (i.e., 2.32; 95% C, 1.94 to 2.79), indicates that potential biases due to unadjusted confounding and exposure misclassification may have slightly biased the primary estimate downward. Restricting follow-up to the first 30 days following treatment initiation resulted in the largest change in the hazard ratio, and suggests that the risk of angioedema may be highest shortly following treatment initiation, which has been observed in other studies.[7] Only the fifth scenario, combining hdPS and changing the risk window to the first 30 days following initial exposure, yielded an estimate of association (HR, 3.50; 95% CI 2.59 to 4.75) that was meaningfully different from the primary analysis result.

**Table 2. Results from three "front end" sensitivity analyses varying parameters from primary analysis**

| Scenario | Description | HR (95% CI) |
|---|---|---|
| Reference | Primary analysis | 2.32 (1.94 to 2.79) |
| 1 | Additional confounding adjustment with empirical variables identified hdPS | 2.49 (2.05 to 3.02) |
| 2 | Extend new users washout and covariate assessment windows to 365 days | 2.62 (2.13 to 3.21) |
| 3 | Change in risk window to first 30 days following initial exposure | 2.98 (2.25 to 3.95) |
| 4 | 1 and 2 above simultaneously | 2.52 (2.06 to 3.14) |
| 5 | 1 and 3 above simultaneously | 3.50 (2.59 to 4.75) |

hdPS, high-dimensional propensity score; HR, hazard ratio; CI, confidence interval

It is important to note that scenarios that extended the new user washout and covariate assessment windows reduced the size of the eligible and analysis populations. The number of eligible lisinopril initiators changed from 385,649 to 286,559 (74% of original) and the number of eligible beta-blocker initiators dropped from 274,977 to 203,185 (74% of original). The number of matched pairs dropped from 231,520 in the primary pre-defined propensity score matched analysis to 169,835 (73% of original) with the 365-day windows. To the extent that this results in an analysis population with distributions of effect modifiers (if any exist) that differ from the distributions in the analysis population, differences in hazard ratios may be attributable to effect measure modification rather than to differences in bias control.

## 3. Results of quantitative bias analysis with "back end" solutions

### a. Outcome misclassification

The results of the 12 scenarios with different combinations of assumed sensitivity and specificity are displayed in Table 3. All scenarios except for Scenario 12 indicate that, based on the assumed values, the observed risk ratio of 2.41 (95% CI, 2.09 to 2.77) was an underestimate of the true risk ratio. The

effect estimate was robust to assumptions of even very low sensitivity (e.g., 0.2). Specificity had a larger effect on the variability of estimates. At a specificity of 0.99977404 (corresponding to a PPV of 90%), the corrected risk ratio was 2.73, across the range of sensitivity values and, at a specificity of 0.999891798 (corresponding to a PPV of 95%), the risk ratio was 2.55.

Outcome definitions with specificity of 1.0 yield perfectly valid effect estimates on the ratio scale, regardless of the non-differential sensitivity. This was confirmed in Scenario 11, in which the specificity was set to 1.0 and the sensitivity was 0.8. Seemingly small decrements in specificity appeared to have a major impact on the robustness of the primary effect estimate. With perfect sensitivity and a non-differential specificity of 0.99, the effect estimate was not only lower, but it even crossed the null (i.e., 0.81) in Scenario 12. Although a specificity of 0.99 appears high, it corresponds to a PPV of only 17% given the low outcome incidence in the test case, and would mean that 1 in 100 persons truly without angioedema would be incorrectly classified as having angioedema.

It should be noted that the results of this analysis are applicable only to relative measures of effect. While perfect specificity yields valid relative effect estimates regardless of sensitivity, outcome measures with sensitivity of 1.0 yield perfectly valid effect estimations on the absolute scale (e.g., risk differences). For example, consider an exposed group with a true incidence rate of 10 per 1,000 person-years and an unexposed group with a true incidence rate of 5 outcomes per 1,000 person-years. The true incidence rate difference would be 5 per 1,000 person-years. An outcome definition with perfect specificity but a sensitivity of 0.80 would yield incidence rates of 8 per 1,000 and 4 per 1,000 person-years, respectively, and an incidence rate difference of 4 outcomes per 1,000 person-years. A definition with perfect sensitivity but imperfect specificity would still identify the 10 outcomes in the exposed group and the 5 outcomes in the unexposed group, but it might identify, say, two additional outcomes in each group. Nevertheless, the resulting difference in the rate ratios of 12 per 1,000 and 7 per 1,000 person-years would still be 5 per 1,000 person-years.

**Table 3. Results of quantitative bias analysis correcting for outcome misclassification (Reference Risk Ratio, 2.41)**

| Scenario | Assumed specificity | Assumed sensitivity | Corrected risk ratio |
|---|---|---|---|
| 1 | 0.999777494 | 0.2 | 2.73 |
| 2 | 0.999777494 | 0.4 | 2.73 |
| 3 | 0.999777494 | 0.6 | 2.73 |
| 4 | 0.999777494 | 0.8 | 2.73 |
| 5 | 0.999777494 | 1.0 | 2.73 |
| 6 | 0.999891798 | 0.2 | 2.55 |
| 7 | 0.999891798 | 0.4 | 2.55 |
| 8 | 0.999891798 | 0.6 | 2.55 |
| 9 | 0.999891798 | 0.8 | 2.55 |
| 10 | 0.999891798 | 1.0 | 2.55 |
| 11 | 1.0 | 0.8 | 2.41 |
| 12 | 0.99 | 1.0 | 0.81 |

### b. Unmeasured confounding

Applying quantitative bias analysis assuming an unknown confounder that doubles the risk of angioedema and has a prevalence of 0.2 in the lisinopril group and 0.1 in the beta-blocker group reduced the risk ratio from the observed value of 2.41 to 2.21. It is important to note that few risk factors for angioedema are known except for drugs that target the renin-angiotensin-aldosterone system and race. In this test case, race is a potentially important unmeasured confounder since it is not well captured in the database that was used.

## 1. Summary of test case findings

Leaving aside the outcome misclassification analysis with a specificity of 0.99, since the PPV was only 17%, the robustness assessment using both empirical sensitivity analysis and quantitative bias analysis methods yielded corrected effect estimates that ranged from 2.21 to 3.50. Given the high PPV of the outcome definition, the low incidence of the outcome in the test case population, and the focus on relative measures of effect, the primary results were robust to varying degrees of sensitivity of the outcome definition, assuming a valid bias model. While the results were robust to the assumptions about the prevalence and magnitude of upward unknown potential confounding, empirical sensitivity analyses that increased the level of confounding adjustment moved the point estimate upward, suggesting that the additional confounding that they adjusted for was in a downward direction.

# VI. RECOMMENDATIONS AND FUTURE WORK

Automated sensitivity analyses and application of quantitative bias analysis modeling methods should become a routine part of the Sentinel routine analytic framework. Results of such analyses can be used to quickly assess the robustness of results arising from Sentinel assessments, including ruling out safety alerts that may be likely attributable to biases that can commonly occur in medical product safety surveillance activities that use electronic healthcare data. Future enhancements to existing tools for conducting routine analyses should focus on increasing the computational efficiency of conducting empirical sensitivity analyses alongside primary analyses.

Additional areas for methodological work include understanding the complementary roles of medical record validation and quantitative bias analysis for outcome misclassification in Sentinel databases; in particular, quantitative bias analysis modeling may be useful for understanding when sufficient medical records have been reviewed to enable stable correction of outcome misclassification. Quantitative bias analysis for outcome misclassification might also consider whether the classification parameters are the same for records that are present and those that are missing. In addition, future work might consider extending methods to address exposure misclassification issues that may be specific to active comparator studies in which four (or more) exposure levels exist (e.g., exposed to drug A, exposed to drug B, neither drug A nor drug B, exposed to both drug A and drug B), but only two are considered in the analysis (e.g., exposed to drug A and exposed to drug B). Existing methods focus on two-level exposure situations (e.g., exposed versus unexposed).

Finally, careful clinical and methodological consideration may be needed to reconcile differences between results from the primary analysis and those from sensitivity and quantitative bias analyses. In the case of the quantitative bias analysis for outcome misclassification, assuming a specificity of 0.99 resulted in a point estimate that was on the opposite side of the null from the results of the primary analysis and all other sensitivity and quantitative bias analyses. In the context of these other results and the fact that the PPV corresponding to this specificity is only 17% for such a rare outcome, it is likely that this result is not plausible.

## VII. ACKNOWLEDGEMENTS

We thank Sophia Axtman and Susan Forrow for effective project management and support, Darryl Cooney, MStat, for multiple demonstrations of and discussions about the SciMetrika tool, and April Duddy, SM, for assessments of the practicability of implementation of different approaches and for leading the implementation of the sensitivity analysis methods in the Query Request Form.

## VIII. REFERENCES


1. Carnahan R, Gagne JJ, Nelson J, et al. Mini-Sentinel Prospective Routine Observational Monitoring Program Tools (PROMPT): rivaroxaban surveillance. 2014.

2. Prospective routine observational monitoring of mirabegron. 2014. Accessed September 4, 2015. http://www.mini-sentinel.org/work_products/Assessments/Mini-Sentinel_PROMPT_Mirabegron-Surveillance-Plan.pdf

3. Gagne JJ, Fireman B, Ryan PB, et al. Design considerations in an active medical product safety monitoring system. Pharmacoepidemiology and drug safety 2012;21 Suppl 1:32-40.

4. Lash TL, Fox MP, Aliza K. Applying quantitative bias analysis to epidemiologic data. 1 ed. New York: Springer-Verlag; 2009.

5. Schneeweiss S. Sensitivity analysis and external adjustment for unmeasured confounders in epidemiologic database studies of therapeutics. Pharmacoepidemiology and drug safety 2006;15:291-303.

6. Lash TL, Fox MP, MacLehose RF, Maldonado G, McCandless LC, Greenland S. Good practices for quantitative bias analysis. International journal of epidemiology 2014;43:1969-85.

7. Toh S, Reichman ME, Houstoun M, et al. Comparative risk for angioedema associated with the use of drugs that target the renin-angiotensin-aldosterone system. Archives of internal medicine 2012;172:1582-9.

8. 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;20:512-22.

9. Rothman KJ. Drug-eluting versus bare-metal stents in acute myocardial infarction. The New England journal of medicine 2009;360:301; author reply -2.

10. Brown NJ, Ray WA, Snowden M, Griffin MR. Black Americans have an increased rate of angiotensin converting enzyme inhibitor-associated angioedema. Clinical pharmacology and therapeutics 1996;60:8-13.

11. Brown NJ, Snowden M, Griffin MR. Recurrent angiotensin-converting enzyme inhibitor--associated angioedema. Jama 1997;278:232-3.

12. Miller DR, Oliveria SA, Berlowitz DR, Fincke BG, Stang P, Lillienfeld DE. Angioedema incidence in US veterans initiating angiotensin-converting enzyme inhibitors. Hypertension 2008;51:1624-30.

# IX. APPENDIX A

## A. SPECIFICATION OF SENSITIVITY ANALYSES IN QUERY REQUEST FORM

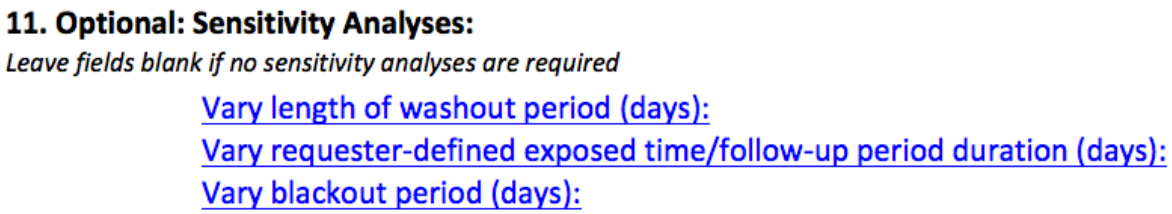
Items below refer to the most recent version of Mini-Sentinel Query Request Form available at the time of writing, located at: http://www.mini-sentinel.org/work_products/Data_Activities/Mini-Sentinel-Routine_Query_Request_Form.xlsm

Upon selecting a Level 2 or Level 3 analysis in the "1. Study Design" tab in the Query Request Form, the user is asked to specify a number of parameters related to exposures and follow-up, including the exposure(s) of interest, how exposed time will be determined, and a blackout period, which is an induction period during which follow-up is ignored.

In the "2c. L2 L3 Exposures & Follow-up" tab, the third (i.e., "Specify how exposed time will be determined") and fifth (i.e., "For assessments with require-defined exposed time") parameters enable users to specify analyses that vary the length of the exposure risk window. By selecting "Define number of days" in #3 and specifying the number of follow-up days in #5, the user can indicate and vary the length of the risk window.

At the bottom of the "2c. L2 L3 Exposures & Follow-up" tab, the user has the option of specifying one or more sensitivity analyses to vary the length of the washout period, vary the requestor-defined exposed time/follow-up period duration, and vary the blackout period (see Figure A1).

**Figure A1. Options for Sensitivity Analyses in the Query Request Form**




On the Outcomes tab, the user is also given the option to vary the length of the outcome washout period prior to the start of follow-up. On the Analysis PS Match tab, the user can opt to activate hdPS, enabling adjustment for any number of empirically-identified variables. One can also specify sensitivity analyses that vary the matching ratio and the matching caliper.

# X. APPENDIX B

**Table B1. Potential sensitivity analyses for Mini-Sentinel routine surveillance activities**

| | Relevant Level 2/3 module | Suggested sensitivity analysis | Comments/suggestions | Source of #bias/difference in results addressed | Potentially alters analysis population? | Implementation/Practicability | Utility (i.e., to what extent will it increase confidence in the robustness of results?) [1-5; 1 = not useful, 5 = highly useful] |
|---|---|---|---|---|---|---|---|
| **"Front end" sensitivity analyses – i.e., analyses performed within each Data Partner** | | | | | | | |
| 1 | PSM, GEE, IPTW | Vary length of washout period in incident user designs | Create macro parameter for washout period and ability to loop through multiple values | Exposure misclassification and confounding | Yes | Implemented for PSM | 4/5 |
| 2 | PSM, GEE, IPTW | Use incident user definitions based on both dispensations and days supply in washout period | Create parameter to define no prior use of index medication based on prescription dispensations in the baseline period and based on days supply (potentially with stockpiling) and enable program to perform both options | Exposure misclassification and confounding | Yes | Possible at a later date | 1 |
| 3 | PSM, GEE, IPTW | Vary exposure definition from single to multiple prescriptions | Incorporate option for requiring patients have at least two prescriptions to define exposure (with start of exposure risk window tied to the second prescription) | Exposure misclassification | Yes | Possible at a later date | 1 |
| 4 | PSM, GEE, IPTW | Vary start of exposure risk window | Incorporate parameter to enable induction period between start of exposure and start of exposure risk window (with ability to either include or exclude index date in exposure risk window) and ability to loop through multiple values | Assumptions about risk window | No | Implemented for PSM | 4/5 |
| 5 | PSM, GEE, IPTW | Vary length of exposure risk window | Incorporate parameter to extend exposure risk window from either the start of exposure (i.e., "ITT") or end of continuous exposure (i.e., "as treated") and ability to loop through multiple values | Assumptions about risk window | No | Implemented for PSM | 5 |
| 6 | PSM, GEE, IPTW | Vary baseline covariate assessment period | Create macro parameter for length of baseline covariate assessment period (with ability to either include or exclude index date in baseline period) and ability to loop through multiple values | Confounding (confounder misclassification) | Yes | Implemented for PSM | 4 |

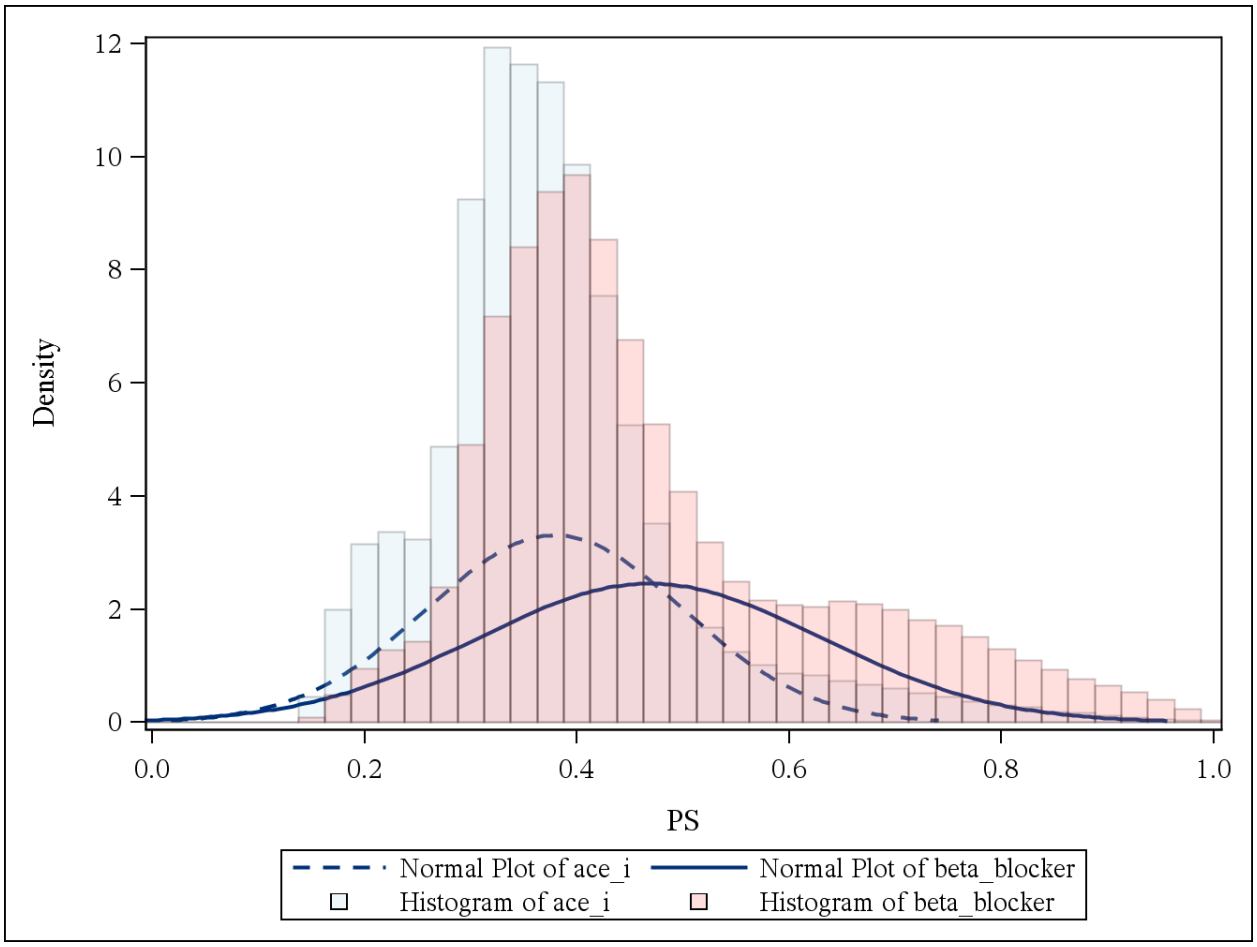
| | Relevant Level 2/3 module | Suggested sensitivity analysis | Comments/suggestions | Source of #bias/difference in results addressed | Potentially alters analysis population? | Implementation/Practicability | Utility (i.e., to what extent will it increase confidence in the robustness of results?) [1-5; 1 = not useful, 5 = highly useful] |
|---|---|---|---|---|---|---|---|
| **"Front end" sensitivity analyses – i.e., analyses performed within each Data Partner** | | | | | | | |
| 7 | SCRI, PSM, GEE, IPTW | Vary the number of confounders adjusted for | Compare "fully" adjusted results to age- and sex-adjusted results and to results adjusted for empirically-identified variables | Confounding | Yes (could result in differences in PS overlap and/or different matches) | Implemented for PSM | 4 |
| 8 | PSM | Vary matching strategy (where applicable) | Implement parameters to vary: (a) fixed vs. variable ratio; (b) maximum ratio (e.g., 1:1 vs. 2:1; n:1 variable ratio matching with no cap vs. n:1 with a maximum of 5:1); (c) matching caliper (e.g., different size calipers [e.g., 0.025 vs. 0.05 on the PS scale] and different scales (e.g., PS vs. s.d.'s of the log PS); (d) parallel versus sequential matching strategies; and ability to loop through options | Confounding | Yes | Partially implemented for PSM; others possible at a later date | 2/3 |
| 9 | PSM, IPTW | Trimming on the propensity score | Implement option for trimming on the propensity score before matching, stratifying, or weighting | Confounding (lack of exchangeability in different areas of PS distribution) | Yes | Possible at a later date | Better suited to be a feature rather than a sensitivity analysis |
| 10 | IPTW | Vary weighting strategy (where applicable) | Incorporate options to weight to the marginal population or the treated population | Effect measure modification/extreme weights | Yes (source population might be same, but weighting would be to a different population) | Possible at a later date | 2 |
| 11 | SCRI | Vary start of exposure time (or "hazard") window | Create macro parameter for start of exposure time window and ability to loop through multiple values | Assumptions about risk window | No | Possible at a later date | 4/5 |
| 12 | SCRI | Vary length of exposure time ("or hazard") window | Create macro parameter for end of exposure time window and ability to loop through multiple values | Assumptions about risk window | No | Possible at a later date | 4/5 |
| 13 | SCRI | Vary start of control time (or "referent") window | Create macro parameter for start of control time window and ability to loop through multiple values | Confounding | No | Possible at a later date | 4/5 |
| 14 | SCRI | Vary length of control time ("or | Create macro parameter for end of control time | Confounding | No | | 4/5 |

| | Relevant Level 2/3 module | Suggested sensitivity analysis | Comments/suggestions | Source of #bias/difference in results addressed | Potentially alters analysis population? | Implementation/Practicability | Utility (i.e., to what extent will it increase confidence in the robustness of results?) [1-5; 1 = not useful, 5 = highly useful] |
|---|---|---|---|---|---|---|---|
| **"Front end" sensitivity analyses – i.e., analyses performed within each Data Partner** | | | | | | | |
| | | referent") window | window and ability to loop through multiple values | | | Possible at a later date | |
| **"Back end" quantitative bias analyses – i.e., analyses performed on aggregate data** | | | | | | | |
| 16 | SCRI, PSM, GEE, IPTW | Unmeasured confounding | Assess how strong an unobserved and uncorrelated confounder would need to be to explain the findings | Confounding | No | Available with SciMetrika tool | 4/5 |
| 17 | SCRI, PSM, GEE, IPTW | Outcome misclassification | Make structural assumptions about the specificity and sensitivity of the outcome definition and examine how the results would change with a perfect definition | Outcome misclassification | No | Available with SciMetrika tool | 4/5 |
| 18 | SCRI, PSM, GEE, IPTW | Exposure misclassification | Make structural assumptions about exposure misclassification (due, for example, to free samples, assumptions about splitting tables/ skipping days, etc) | Exposure misclassification | No | Available with SciMetrika tool | 4/5 |
| 19 | PSM, GEE, IPTW | Time-varying hazards | Display effect estimates as function of duration of use | Assumptions about risk window | No | Possible at a later date | 4/5 |

GEE, generalized estimating equation tool; IPTW, inverse probability of treatment weighting tool; PSM, propensity score-matching tool; SCRI, self-control risk interval tool

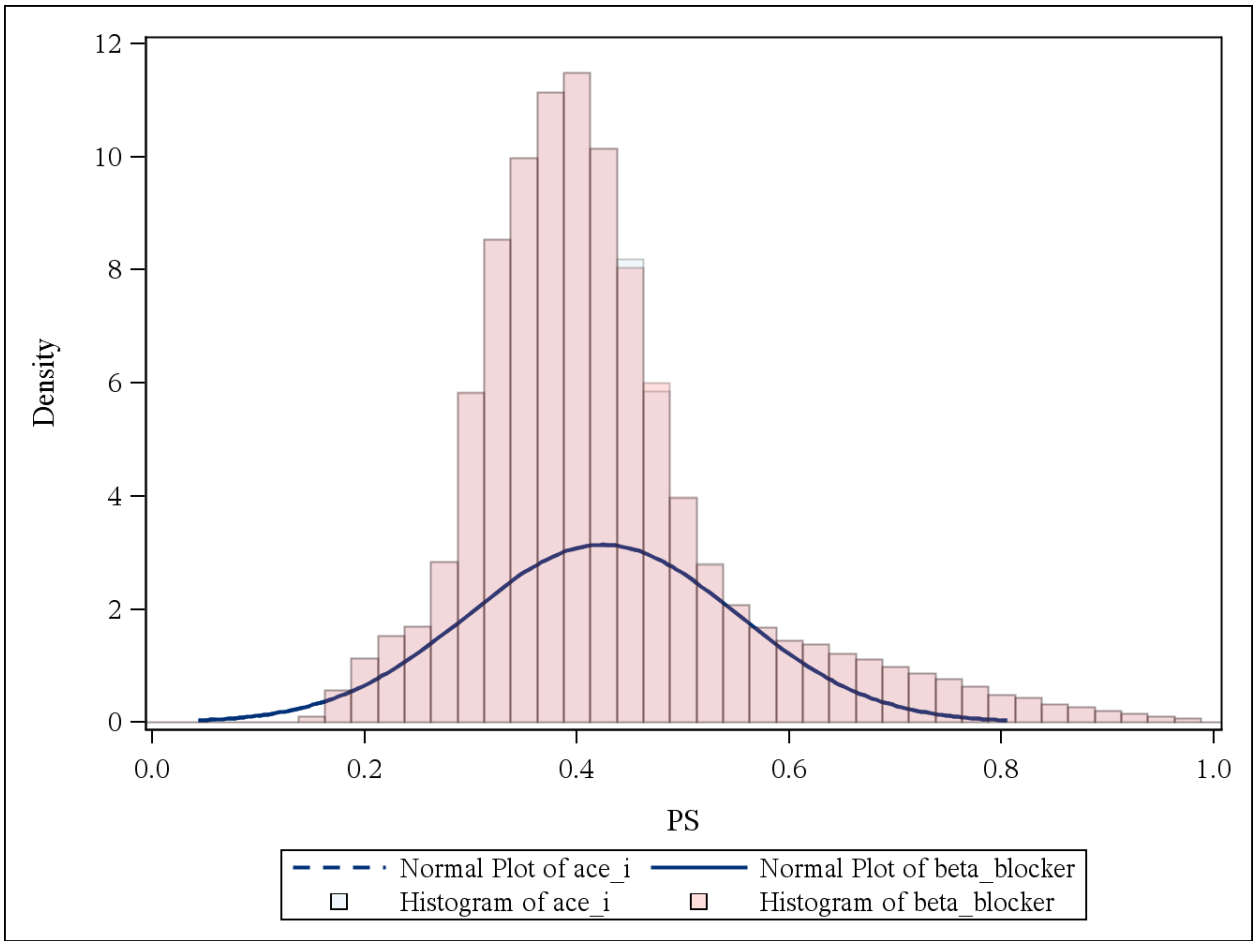
# XI. APPENDIX C

**Figure C1. Distribution of pre-defined propensity score among unmatched initiators of lisinopril and beta-blockers**



# XII. APPENDIX D

**Figure D1. Distribution of pre-defined propensity score among matched initiators of lisinopril and beta-blockers**



# XIII. APPENDIX E

**Figure E1. Baseline characteristics of patients in primary matched test case cohort**

| Cohort of New Initiators of lisinopril and beta-blockers (Matched Predefined, PS Caliper = .025) | | | | |
|---|---|---|---|---|
| | **Primary Analysis** | | **Covariate Balance** | |
| | N (%) | N (%) | Absolute Difference | Standardized Difference |
| **Characteristic** | lisinopril | beta_blocker | | |
| Number of patients (Percent of Cohort Matched) | 231,520 (60.0 %) | 231,520 (84.2 %) | | |
| Number of Events While on Therapy | 664 (0.3 %) | 276 (0.1 %) | 0.2 | 0.0 |
| Days at risk | 228.7 ( 262.4) | 213.6 ( 260.2) | 15.1 | 0.1 |
| | | | | |
| **Patient Characteristics** | | | | |
| Age, mean year (standard deviation) | 57.4 ( 8.1) | 57.3 ( 8.4) | 0.2 | 0.0 |
| 45-54 years | 98,590 (42.6 %) | 102,240 (44.2 %) | -1.6 | 0.0 |
| 55-64 years | 102,290 (44.2 %) | 97,480 (42.1 %) | 2.1 | 0.0 |
| 65-74 years | 21,524 (9.3 %) | 21,918 (9.5 %) | -0.2 | 0.0 |
| 75-84 years | 7,328 (3.2 %) | 7,883 (3.4 %) | -0.2 | 0.0 |
| 85-99 years | 1,788 (0.8 %) | 1,999 (0.9 %) | -0.1 | 0.0 |
| | | | | |
| Gender (F) | 124,845 (53.9 %) | 119,340 (51.5 %) | 2.4 | 0.0 |
| | | | | |
| **Recorded history of:** | | | | |
| Combined Comorbidity Score | 0.1 ( 1.3) | 0.1 ( 1.3) | 0.0 | 0.0 |
| Allergic Reactions | 25,675 (11.1 %) | 25,169 (10.9 %) | 0.2 | 0.0 |
| Diabetes | 33,874 (14.6 %) | 36,152 (15.6 %) | -1.0 | 0.0 |

| Cohort of New Initiators of lisinopril and beta-blockers (Matched Predefined, PS Caliper = .025) | | | | |
|---|---|---|---|---|
| | **Primary Analysis** | | **Covariate Balance** | |
| | N (%) | N (%) | Absolute Difference | Standardized Difference |
| Heart Failure | 4,633 (2.0 %) | 5,557 (2.4 %) | -0.4 | 0.0 |
| Ischemic Heart Disease | 18,891 (8.2 %) | 20,415 (8.8 %) | -0.6 | 0.0 |
| | | | | |
| **Recorded use of:** | | | | |
| NSAID | 32,293 (13.9 %) | 32,044 (13.8 %) | 0.1 | 0.0 |
| | | | | |
| **Health Service Utilization Intensity:** | | | | |
| Number of Unique Generics Dispensed | 4.1 ( 3.9) | 4.0 ( 3.8) | 0.0 | 0.0 |
| Number of Filled Rx | 9.5 ( 11.1) | 9.4 ( 10.8) | 0.1 | 0.0 |
| Number of inpatient hospital encounters (IP) | 0.1 ( 0.4) | 0.1 ( 0.4) | 0.0 | 0.0 |
| Number of non-acute institutional encounters (IS) | 0.1 ( 0.9) | 0.1 ( 1.1) | 0.0 | 0.0 |
| Number of emergency room encounters (ED) | 0.3 ( 1.6) | 0.3 ( 1.6) | 0.0 | 0.0 |
| Number of ambulatory encounters (AV) | 6.5 ( 8.4) | 6.6 ( 7.8) | -0.1 | 0.0 |
| Number of other ambulatory encounters (OA) | 0.3 ( 1.4) | 0.3 ( 1.6) | 0.0 | 0.0 |
| | | | | |
| Mahalanobis Distance | 0.006 | | | |