MINI-SENTINEL MEDICAL PRODUCT ASSESSMENT

A PROTOCOL FOR ACTIVE SURVEILLANCE OF ACUTE MYOCARDIAL INFARCTION IN ASSOCIATION WITH USE OF ANTI-DIABETIC AGENTS

Version 6

January 27, 2016

Prior versions:
Version 5: January 7, 2015
Version 4: January 8, 2014
Version 3: October 21, 2012
Version 2: December 28, 2011
Version 1: October 26, 2010

Prepared by: Joe Selby, MD, MPH[1*], Marsha E. Reichman, PhD[2], David Graham, MD, MPH[2], Melissa Butler, PharmD, PhD[3], Christian Hampp, PhD[2], Mark Levenson, PhD[4], Mary Ross Southworth, PharmD[5], Darren Toh, ScD[6], Bruce Fireman, MA[1]

Author Affiliations: 1. Kaiser Permanente Northern California, Oakland, CA. 2. Office of Surveillance and Epidemiology, Center for Drug Evaluation and Research, Food and Drug Administration, Silver Spring, MD. 3. Kaiser Permanente Southeast, Atlanta, GA. 4. Office of Biostatistics, Office of Translation Sciences, Center for Drug Evaluation and Research, Food and Drug Administration, Silver Spring, MD. 5. Office of New Drugs, Center for Drug Evaluation and Research, Food and Drug Administration, Silver Spring, MD. 6. Department of Population Medicine, Harvard Medical School and Harvard Pilgrim Health Care Institute, Boston, MA. * Dr. Selby led the development of version 1 of the protocol, but did not participate in the implementation or subsequent revision of the protocol. His current affiliation is The Patient-Centered Outcomes Research Institute, Washington, DC.

The authors acknowledge the rich input and contributions of members of the working groups on Methods, Endpoints, Diabetes and Data as well as constructive comments and suggestions from numerous FDA staff, Data Partners and Mini-Sentinel Operations Center staff during the deliberations that helped develop this protocol.

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 numbers HHSF223200910006I, HHSF22301004T, and HHSF22301007T.

History of Modifications

Version	Date	Modification	By
V2	12/28/2011	• Modified the title of the protocol • Included an analysis plan for a simulated surveillance of sitagliptin • Updated Tables 2, 3, and 4 • Provided more detail about the analysis plan • Minor text modifications	Mini-Sentinel AMI Surveillance Workgroup
V3	10/21/2012	• Updated Tables 3 and 4 • Restored missing citations • Minor text modifications	Mini-Sentinel AMI Surveillance Workgroup
V4	1/8/2014	• Included minor edits and clarifications in response to comments from independent audit of analytic code • Removed Poisson analysis, a proposed secondary analysis • Revised the number of sequential looks • Updated the timeline	Mini-Sentinel AMI Surveillance Workgroup
V5	1/7/2015	• Included minor edits and clarifications in response to comments from independent audit of analytic code	Mini-Sentinel AMI Surveillance Workgroup
V6	1/27/2016	• Included an addendum that described a one-time analysis of hospitalized heart failure risk in association with DPP-4 inhibitors	Mini-Sentinel AMI Surveillance Workgroup

This protocol is modified periodically to document major changes made during protocol implementation.

MINI-SENTINEL MEDICAL PRODUCT ASSESSMENT

A Protocol for Active Surveillance of Acute Myocardial Infarction in Association with Use of Anti-Diabetic Agents

Table of Contents

I. BACKGROUND

Saxagliptin is an oral dipeptidyl peptidase-4 (DPP-4) inhibitor used in the treatment of type 2 diabetes mellitus. It was approved by the FDA in July 2009 as an adjunct to diet and exercise to improve glycemic control in adults with type 2 diabetes mellitus, and can be used as either monotherapy or combination therapy. As of October 2010, saxagliptin is one of two DPP-4 inhibitors marketed in the U.S.. DPP-4 inhibitors work by slowing the inactivation of the incretin hormones by the DPP-4 enzyme. Incretin hormones, including glucagon-like peptide-1 (GLP-1) and glucose-dependent insulinotropic polypeptide, are released into the bloodstream from the intestine and cause the release of insulin from the pancreatic beta-cells. GLP-1 also lowers the secretion of glucagons from pancreatic alpha cells, leading to reduced hepatic glucose production. The resulting increase and prolongation of incretin levels by DPP-4 inhibitors reduces both fasting and postprandial glucose concentrations in a glucose-dependent manner.

Saxagliptin provides modest reductions in hemoglobin A1c (0.5% to 0.8%) relative to what has been observed with other anti-diabetic agents (1% to 2% reduction for metformin and sulfonylureas, and 1.5% to 3.5% for insulin).[1, 2] As a whole, DPP-4 inhibitors appear to be well tolerated. They have neutral effects on weight while other anti-diabetic agents (e.g., sulfonylureas, insulin, thiazolidinediones) are associated with weight gain. Due to their glucose-dependent mechanism of action, DPP-4 inhibitors have a low risk of hypoglycemia which is similar to that of placebo in clinical trials. Saxagliptin is dosed once-daily and does not require dose titration. Saxagliptin can be used in patients with severe renal impairment (creatinine clearance <30 mL/min).

The cardiovascular disease (CVD) risk of long-term treatment with saxagliptin is unknown. Based on pre-approval clinical trials, there is no signal of increased cardiovascular risk with saxagliptin; in fact, the pooled data suggest that it could be modestly protective against CVD.[3] However, patients at risk for cardiovascular events may have been excluded from pre-market trials; therefore there is uncertainty of the level of risk in a more general population.

The FDA has recommended that newly approved medications for treatment of diabetes be thoroughly and systematically evaluated for risk of CVD.[4] In the case of saxagliptin, a large double-blind post-market randomized trial comparing saxagliptin use with placebo in patients with type 2 diabetes mellitus is being conducted.[5] The primary objective of this trial is to establish that the upper bound of the two-sided 95% confidence interval for the estimated risk ratio comparing CVD incidence with saxagliptin to that observed in the control group is less than 1.3. The trial will include a portion of the population at greater risk for cardiovascular adverse events. However, its results will not be available for more than five years.

In the meantime, the FDA intends to conduct post-market active surveillance using the exposure and outcome data from large, population-based clinical and claims databases. This protocol is proposed as an additional method to monitor and detect any potential increase in CVD associated with saxagliptin within Mini-Sentinel. More generally, the FDA seeks to learn from the surveillance activities described in this protocol about efficient approaches to active surveillance for other newly approved pharmaceuticals and for other endpoints.

II. OBJECTIVES

1. To develop a protocol for the active surveillance of acute myocardial infarction (AMI) in users of saxagliptin compared to users of comparator agents, based on prospective data obtained from large, population-based clinical and claims databases. The protocol should allow for repeated assessment of accumulating experience at a frequency commensurate with the rate of acquisition of new data.

2. To document the deliberations behind all decisions that support the final protocol (**Appendix A**).

3. To recommend specific data validation efforts that should be applied as the protocol is applied for surveillance.

4. To evaluate the validity and efficiency of several statistical approaches for the active surveillance analyses to inform future surveillance activities.

III. PROTOCOL DEVELOPMENT PROCEDURE

During the three months allocated for this protocol development, the Kaiser Permanente-led team convened four working groups: Methods, Endpoints, Diabetes and Data. Formal membership on the four working groups is shown in **Table 1**. Numerous additional individuals from the FDA and Mini-Sentinel also participated in working group deliberations. The deliberations of these working groups are documented in **Appendix A**. The Mini-Sentinel Data Core, joined by Drs. Selby and Butler and FDA, served as the Data working group for this project. Each working group met periodically as needed in open teleconference meetings. Meetings were joined by additional representatives of the FDA and the Mini-Sentinel Operations Center. In addition, the Methods and Endpoints working groups each met with Dr. Jerry Gurwitz who was heading a parallel Mini-Sentinel Workgroup on AMI validation. This workgroup was charged with conducting AMI validation work using data from the Mini-Sentinel Distributed Database (MSDD). Continuity across the four working groups was provided through attendance on all calls by Drs. Selby and Butler and project manager, Ms. Cathy Chou. Additional communication within groups was accomplished by email exchanges between calls. The work of the four working groups was synthesized and presented bi-weekly in teleconferences open to all Mini-Sentinel participants.

Table 1. Protocol Development Team

Working Group	Investigator	Institution
Overall Leads	Joe Selby MD MPH	Kaiser Permanente
	Marsha Reichman PhD	FDA
	Bruce Fireman MA	Kaiser Permanente
	Melissa Butler PharmD PhD	Kaiser Permanente
Methods Working Group	Melissa Butler PharmD PhD	Kaiser Permanente
	Daniel Canos PhD	FDA
	Andrea Cook PhD	HMORN
	Bruce Fireman MA	Kaiser Permanente
	David Graham MD MPH	FDA
	Marie Griffin MD MPH	Vanderbilt
	Lanh Green PharmD MPH	FDA
	Jennifer Nelson PhD	HMORN

Table 1. Protocol Development Team		
Working Group	**Investigator**	**Institution**
	Bruce Psaty MD*	U Washington
	Judity Racoosin MD MPH	FDA
	Marsha Reichman PhD	FDA
	Melissa Robb RN	FDA
	Sebastian Schneeweiss MD ScD	HMORN
	Joe Selby MD MPH	Kaiser Permanente
	Mary Ross Southworth PharmD	FDA
	Judy Staffa PhD RPh	FDA
	Darren Toh ScD	HMORN
	Diane Wysowski PhD	FDA
	Hui Talia Zhang ScD	FDA
Endpoints Working Group	Allen Brinker MD	FDA
	Melissa Butler PharmD PhD	Kaiser Permanente
	Amy Egan MD MPH	FDA
	Alan Go MD	Kaiser Permanente
	Hylton Joffe MD	FDA
	Joe Selby MD MPH	Kaiser Permanente
Diabetes Mellitus Working Group	Amy Egan MD MPH	FDA
	William Herman MD	U Michigan
	Hylton Joffe MD	FDA
	Judity Racoosin MD MPH	FDA
	Marsha Reichman PhD	FDA
	Melissa Robb RN	FDA
	Joe Selby MD MPH	Kaiser Permanente
Data working group	Nicolas Beaulieu MA	HMORN
	Denise Boudreau PhD	HMORN
	Lesley Curtis PhD	Duke University
	Greg Daniel PhD	HealthCore
	Marie Griffin MD MPH	Vanderbilt
	Vinit Nair BPharm MS	Humana
	Marsha Raebel PharmD	Kaiser Permanente
	Judity Racoosin MD MPH	FDA
	Marsha Reichman PhD	FDA
	Melissa Robb RN	FDA
	Mark Weiner MD	U Pennsylvania
Operations Center	Darren Toh ScD	HMORN
* Dr. Psaty participated on an ad hoc basis without reimbursement.		

Early in the protocol development, a data request was submitted to the Data Core. This request was modified several times in iterative discussions between Methods, Endpoints, Diabetes and Data working groups. The final request is shown in **Appendix B**. The request was designed to be run against the MSDD. Requested data included estimates of the number of individuals who met criteria for type 2 diabetes mellitus by age group from each Data Partner based on the two years of data in the MSDD; the numbers of new users of saxagliptin and three of the four proposed comparators (sitagliptin, pioglitazone, long-acting or combination insulin) observed in the last 12 months of the 2008-2009 period; patterns of concurrent anti-diabetic medication use for each of these new-user groups; and the proportions of all diabetic patients in each age group and Data Partner who appeared to experience an AMI, or a hospitalization for acute coronary syndrome (ACS) without infarction, during 2009. Results of

the data request are also presented in **Appendix B** and were used to inform the work of the Methods and Endpoints working groups.

During protocol development, an opportunity was recognized to conduct a simulated surveillance exercise similar to that proposed here. Sitagliptin, the first DPP-4 inhibitor, was approved in October 2006. The Data Partners have agreed that, using the MSDD, it would be possible to conduct surveillance activities using already available data for the period from 2006 through the first quarter of 2010, a period almost identical in length to that proposed here. From this activity, which can be completed during the first nine months of the proposed saxagliptin surveillance, we can quickly evaluate some of the implicit assumptions and unknowns in this saxagliptin active surveillance protocol. The simulated surveillance protocol is included as **Appendix C** and is referenced several occasions in this protocol.

IV. PROTOCOL FOR ACTIVE SURVEILLANCE OF AMI ASSOCIATED WITH USE OF SAXAGLIPTIN

A. SURVEILLANCE DESIGN

This active surveillance will use "new user" parallel cohort design.[6, 7] New users of saxagliptin will be identified from the MSDD using as many of the Data Partners as can comply with protocol requirements, and beginning in August 2009. New users of four comparator anti-diabetic medications will also be identified. These cohorts of new users will be followed longitudinally in four separate analyses in which the occurrence of AMI is compared between new users of saxagliptin and new users of one comparator. In each comparison, previous users of saxagliptin or of the comparator being studied are excluded. This new-user approach is chosen over a strategy that would include all identified users (i.e., prevalent as well as incident users) because it more closely resembles a clinical trial and it ensures that all persons included in the analysis were considered eligible and appropriate for initiation of a new anti-diabetic therapy at the start of follow-up. It allows patient characteristics to be measured before the start of therapy so that they are not influenced by the therapy choice. Early, as well as delayed, effects of the drug of interest can be uniformly identified with no risk of a "survivor user" bias, wherein only the survivors of early therapy are studied. Duration of use at cohort entry does not differ between those exposed to saxagliptin and the comparators and thus does not require measurement or adjustment. The cohorts of new users will be separated immediately into subgroups with and without a prior history of CVD.

B. CHOICE OF COMPARATORS

New users of four comparator anti-diabetic agents will be identified, followed, and compared in separate analyses to new users of saxagliptin. The comparators are: sitagliptin, long-acting insulin, pioglitazone, and second-generation sulfonylureas (glimepiride, glipizide, and glyburide). These four are chosen because they represent common alternative agents to saxagliptin. Sitagliptin is the only other approved DPP-4 inhibitor at the time of protocol development and would be used in patients very similar to those starting saxagliptin. There is no evidence that risk for AMI is increased by use of sitagliptin.[3] Long-acting insulin is often initiated after failure of two-drug oral therapy to adequately control hemoglobin A1c, a clinical situation in which saxagliptin may also be considered. Insulin initiators are likely to have more severe and long-standing diabetes, probably with higher average hemoglobin A1c values at baseline. Pioglitazone is selected as the representative of the thiazolidinedione class of agents because it may have somewhat lower risk for CVD endpoints than rosiglitazone and because it is likely to be initiated much more frequently than rosiglitazone in coming months. Second-generation

sulfonylureas are also proposed as the fourth comparator because prescribing data provided by the Data Core suggest that saxagliptin and sitagliptin are sometimes being used as second-line therapy with metformin. Recent treatment guidelines also suggest a role as second-line therapy.[8]

No single comparator is ideal. Sitagliptin could share a "class effect" with saxagliptin that leads to increased risk for AMI. The relative effects of the other comparators on AMI risk are also not well known. Under these circumstances, it is preferable to examine the saxagliptin experience from several perspectives. Whether there are differences in risk for AMI between initiators of sitagliptin, pioglitazone, or a sulfonylurea is not known with certainty, although none are strongly suspected of increasing AMI risk. Thus, a signal suggesting an increased risk with saxagliptin in comparison to any of these would be of concern, precisely because saxagliptin is likely to be used in place of each of these agents. With respect to insulin, we anticipate that persons placed on long-acting insulin may well be at increased risk for AMI relative to initiators of these other agents, by virtue of having more severe or longer duration of diabetes, or a diabetes complication such as early renal disease, retinopathy, neuropathy or CVD, each of which is associated with increased risk for AMI. Therefore, any signal of a higher risk for AMI in saxagliptin users relative to insulin initiators would be of great concern.

C. COHORT IDENTIFICATION

To identify new users of saxagliptin and comparators, a prior observation period of at least 12 months is needed so that earlier use can be ruled out. This prior enrollment period must be accompanied by prescription drug coverage throughout so that previous use of the same agent can be detected with confidence. Although the MSDD will provide nearly two years of exposure history prior to August 1, 2009 for many patients, turnover within health plans and systems is sufficient so that requiring two full years of prior enrollment in each case would eliminate many potential new users. Therefore, a one-year period of continuous enrollment (i.e., a one-year period with no gap in enrollment longer than 31 days), with no evidence of previous use of saxagliptin and no evidence of previous use of the comparator in question, is required. Previous use of a comparator agent is of concern only in defining new users of the agent and in excluding those new saxagliptin users who have previously used it from comparisons with the agent. Since enrollment information is often available only at the monthly level in most Data Partner sites, "continuous" enrollment will be defined as having a gap no longer than 31 days.

Table 2 lists the search steps to be applied by each Data Partner for identifying patients eligible to be included in the analyses. The initial search using this strategy will identify new users beginning on August 1, 2009 and will likely cover at least a one-year period at the initial data collection. Thereafter, searches will be repeated quarterly, commensurate with the update frequency of the MSDD. This strategy is intended to identify new users by excluding anyone with prior use and to assure that all users truly have diabetes.

Table 2. Search Strategy to Identify New Users of Anti-Diabetic Medications in the Mini-Sentinel Distributed Database

a) Identify all currently enrolled individuals with pharmacy coverage and a dispensing of saxagliptin or any comparator on or after August 1, 2009.

b) Exclude any individual who does not have at least one year of continuous enrollment (i.e., no enrollment gap > 31 days) with prescription drug coverage immediately preceding the dispensing date.

c) Exclude any individual who had a prior dispensing of saxagliptin or the comparator of interest (for each pairwise comparison) during that 12-month period.

d) Exclude (and save records for) any individual with a history of hospital discharge for a principal diagnosis of acute myocardial infarction (ICD-9-CM codes 410.x0 or 410.x1) in the 60-day period prior to date of first dispensing of saxagliptin or comparator.

e) Exclude any individual who does not have at least one dispensing of an anti-diabetic medication* or at least one diagnosis of diabetes (regardless of type of encounter) during the 12-month period.

f) Exclude women with possible gestational diabetes on the date of their first dispensing of saxagliptin or comparator. Women are presumed to have gestational diabetes on this date if i) they have any type of encounter with a pregnancy diagnosis during the prior year and ii) they do not have an encounter with a diagnosis or procedure indicating that the last pregnancy in the prior year ended (due to childbirth, spontaneous abortion, etc, identified by ICD-9-CM codes 630-679, V22, V23, and V28) more than three months prior to the index dispensing.

g) Include only the first eligible treatment episode if the individual has more than one eligible treatment episode for saxagliptin or a specific comparator drug.

h) Stratify all remaining eligible individuals into two strata: those with a diagnosis or procedure consistent with CVD (see **Table 3**) and those without.

* The list of anti-diabetic medications does not include short-acting insulin to reduce chances of including type 1 diabetic patients.

Patients with a recent history (in the 60 days prior to the first dispensing) of hospitalization for AMI are excluded from the analysis because of their extreme risk for recurrence, likelihood that confounding factors may vary significantly from those for other patients, the high potential for residual confounding, and because they are excluded from the post-market randomized trial required by the FDA.[9] However, these patients will be retained in a separate dataset by each Data Partner for possible future analysis. All remaining eligible patients are divided immediately into those with and those without a history of a diagnosis or procedure consistent with CVD during the 12 months prior to the first dispensing, using the identification criteria in **Table 3**.

Table 3. Diagnoses and Procedures Indicative of a History of Cardiovascular Disease	
Diagnosis or procedure	**Codes***
Prior AMI[†]	ICD9D: 410
Other ischemic heart disease	ICD9D: 411–414
Other heart disease	ICD9D: 402.01, 402.11, 402.91, 420-429, 440
Stroke	ICD9D: 430–434, 436
Peripheral arterial disease	ICD9D: 443.9
Coronary revascularization procedures	
Coronary artery bypass graft	ICD9D: 996.03, V45.81 ICD9P: 36.1X, 36.2 CPT4: 33510-33514, 33516-33523, 33525, 33528, 33530, 33533-33536, 33560, 33570, 33572, 33575, 35600 HCPCS: S2205-S2209
Percutaneous coronary intervention	ICD9D: V45.82 ICD9P: 0.66, 17.55, 36.01-36.09, 37.22, 37.23, 88.5x CPT4: 92973, 92974, 92977, 92980, 92981, 92982, 92984, 92987, 92995, 92996 HCPCS: G0290, G0291
Carotid revascularization procedures	
Carotid endarterectomy, stenting, angioplasty, or atherectomy	ICD9P: 00.61, 00.63, 38.11, 38.12 CPT4: 35301, 35390, 35501, 35601, 35901, 0075T, 0076T, 37215, 37216 HCPCS: S2211
Carotid bypass	ICD9P: 39.28
Lower extremity revascularization	
Lower extremity endarterectomy, stenting, angioplasty, or atherectomy	ICD9P: 38.18, 38.19 CPT4: 35454, 35456, 35459, 35470, 35473, 35474, 35482, 35483, 35492, 35493, 35495, 37207, 37208, 37220-37235
Lower extremity bypass	ICD9P: 39.25, 39.29 CPT4: 35351, 35355, 35361, 35363, 35371, 35372, 35521, 35533, 35541, 35546, 35548, 35549, 35551, 35556, 35558, 35563, 35565, 35566, 35570, 35571, 35582, 35583, 35585, 35587, 35621, 35623, 35637, 35638, 35641, 35646, 35647, 35651, 35654, 35656, 35661, 35663, 35665, 35666, 35671, 35681-35683, 35879
Lower extremity amputation	ICD9P: 84.10-84.17 CPT4: 27295, 27590-27592, 27598, 27880-27882, 27888, 27889, 28800, 28805, 28810, 28820, 28825

* Use only codes associated with visits (inpatient or outpatient). All diagnoses and procedures are sought for the 12-month period prior to first dispensing. ICD9D: ICD-9-CM diagnosis codes; ICD9P: ICD-9 procedure codes; CPT4: Current Procedural Terminology codes; HCPCS: Healthcare Common Procedure Coding System codes.

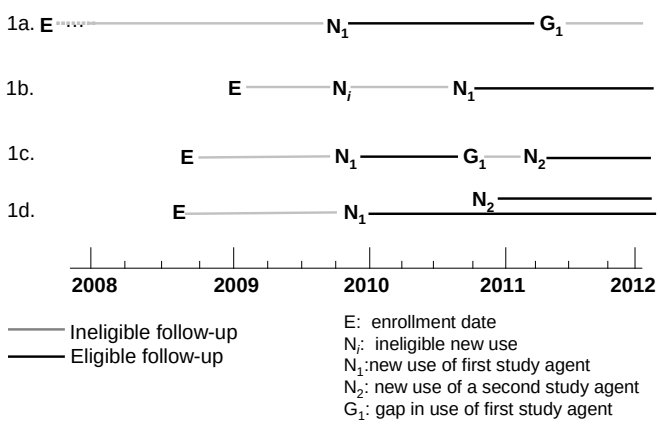
† We exclude patients with a principal hospital discharge diagnosis indicating AMI within 60 days of new use.

Note that patients could be selected as new users of more than one medication of interest during the course of surveillance. A patient could be followed in two comparator cohorts at the same time since these are essentially separate analyses. Two comparators could begin on the same day (e.g., the combination medications including pioglitazone and a sulfonylurea), but it is more likely that a second comparator would be added at some point during follow-up. Obviously, a patient cannot be in the saxagliptin cohort and a comparator cohort at the same time. However, patients could switch during follow-up from one comparator to another, or from a comparator to saxagliptin.

Figure 1 illustrates the appropriate identification of new users. Patient 1a was found to be a new user of either saxagliptin or a comparator in late 2009, had more than a year of enrollment at that point and continued using (and followed-up until a gap in use in mid-2011). Patient 1b appeared to be a new user of a drug of interest (a comparator) in late 2009 but had less than 12 months follow-back. Later, in 2010, new use of another comparator agent was observed. More than 12 months follow-back were now available and the patient continued under observation into 2012. Patient 1c was an eligible new user of a comparator in late 2009, stopped using that agent in late 2010, but initiated a different agent (either saxagliptin or another comparator) in 2011 and then contributed follow-up for that agent. Patient 1d initiated a comparator in late 2009. One year later, this patient began use of another comparator without stopping use of the first agent. That person contributed follow-up into 2012 in two comparator cohorts.

Figure 1 shows the selection and follow-up for new users.

Figure 1. Selection and Follow-up of New Users



Only one comparison-specific exclusion will be implemented. Patients with a history of congestive heart failure (ICD-9-CM: 404.x1, 402.x1, 404.x3, and 428) noted during the baseline period will be excluded from the saxagliptin versus pioglitazone comparison because congestive heart failure is a contraindication to use of pioglitazone.

D. CALCULATION OF PROPENSITY SCORES AND MATCHING

Confounding is a principal threat to the validity of observational drug safety assessments. New users of saxagliptin are likely to differ from those who initiate other anti-diabetic agents on numerous demographic and clinical characteristics that may also be predictive of AMI. Many such covariates, including age, sex, comorbidities, and use of other medications can be found in the MSDD. We propose three complementary approaches to adjust for potential confounders, two design-based and one analysis-based. Each approach will be implemented separately in subgroups defined by Data Partner and by presence versus absence of CVD.

a) Matching 1:1 by an exposure propensity score (PS)

b) Stratification by a disease risk score (DRS)

c) "Conventional" multivariable regression modeling at end of surveillance

Propensity scores (PS)[10-12] and disease risk scores (DRS)[13, 14] are two approaches for reducing large numbers of variables (covariates) into single summary scores variables that can be used to match saxagliptin users with comparators (as proposed for the PS) or to stratify users and comparators into groupings with comparable risk for developing AMI (as proposed for the DRS). Summary scores can make adjustment and estimation feasible compared with standard multivariable regression modeling when the number of events is low in relation to the number of covariates, as it will be, especially early on in surveillance.[15]

We identified a list of covariates for inclusion in each PS and DRS (**Table 4**). These covariates are to be sought in the MSDD for the 12 months immediately preceding the date of the first eligible dispensing of saxagliptin or a comparator drug. Each diagnosis, procedure or drug use is assumed to be absent for a person if not found in the MSDD. That is, there are no "missing" variables. Comorbidities are identified either from hospital discharges or from data related to outpatient visits or prescription fills. Diagnoses linked to laboratory tests or other non-visit-related utilization will not be included because the validity of these diagnoses is less well established and because they are not available in all Data Partners. Similarly, data on patient race/ethnicity is not uniformly available and is therefore not included in the primary PS and DRS models.

Table 4. Baseline Covariates to be Included in Propensity Score and Disease Risk Score Model

FOR BOTH STRATA (With, without prior CVD)

Demographics	**Codes***
Age at 1st cohort entry Sex Residence in nursing home (or stay in other non-hospital institution) during prior year	"Residence in nursing home" is determined from the Encounter Type, Discharge Status and Admitting_Source values in the MSDD Encounter table (including nursing home, skilled nursing facility, hospice, rehabilitation center, overnight non-hospital dialysis, and other non-hospital institutional stays).
Utilization Measures	
Any hospitalization within prior 30 days Any hospitalization 31-365 days Any ED visit within prior 30 days Any ED visit 31-365 days before Number of outpatient visits in prior year Number of unique medications dispensed in prior year	"Unique medications" are determined by counting unique 9-digit NDC codes in the MSDD Dispensing table. Combination drugs are not broken into their constituents. (The study drugs and covariate drugs are identified in more detail than the 11-digit NDC codes, as specified elsewhere)
Co-morbid conditions	
Asthma	493
Cancer (excluding non-melanoma skin cancer)	140-209 (excluding 173 and 209.4-209.6)
Chronic kidney disease (excluding ESRD)	585.1–585.4 HCPCS: G0420, G0421, G8487, G8771
Chronic obstructive pulmonary disease	491, 492, 496
Dementia	290.0–290.4, 291.2, 292.82, 294.0, 294.1, 294.8, 331.0–331.2, 331.7–331.9, 797
Depression	296.2, 296.3, 300.4, 311
End stage renal disease (ESRD)	458.21, 585.5, 585.6, 996.56, 996.68, 996.73, V42.0, V45.1, V56

Table 4. Baseline Covariates to be Included in Propensity Score and Disease Risk Score Model	
	ICD9P: 38.95, 39.27, 39.42, 39.43, 39.53, 39.93, 39.94, 39.95, 54.98, 55.6 CPT4: 36145, 36800, 36810, 36815, 36825, 36830-36833, 50323, 50325, 50327-50329, 50340, 50341, 50360, 50365, 50366, 90918-90925, 90935, 90937, 90939-90944, 90945, 90947, 90951-90969, 90970, 90976-90979, 90982-90985, 90989, 90993, 90995, 90996, 90997, 90998, 90999, 93990, 99512 HCPCS: A4653, A4656, A4657, A4670-A4674, A4680, A4706-A4709, A4712, A4714, A4719, A4720-A4726, A4728, A4730, A4736, A4737, A4740, A4750, A4755, A4760, A4765, A4766, A4770, A4771, A4773, A4774, A4802, A4860, A4870, A4890, A4911, A4913, A4918, A4928, A4929, C1881, E1500, E1520, E1530, E1540, E1550, E1560, E1570, E1575, E1580, E1600, E1610, E1615, E1620, E1625, E1634-E1639, E1699, G0257, G0308-G0327, G8727, G9013, G9014, J0635, J0636, S2065, S9335, S9339
Fracture	733.1, 733.93-733.98, 805–815 (excluding 807.5 and 807.6), 818-825, 827, 828, V54.13, V54.23 ICD9P: 79.01-79.03, 79.05-79.07, 79.11-79.13, 79.15-79.17, 79.21-79.23, 79.25-79.27, 79.31-79.33, 79.35-79.37, 79.61-79.63, 79.65-79.67, 81.65, 81.66
Heart failure	404.x1, 402.x1, 404.x3, and 428
HIV/AIDS	042, 043, 044, 795.71, V08
Hyperlipidemia or lipid disorder	272.0, 272.1, 272.2, 272.4
Hypertension	401–405 (excluding 402.01, 402.11, 402.91)
Hypoglycemia	250.8, 251.0–251.2
Obesity (or weight gain)	278.0, 793.91, V85.3, V85.4 (783.1)
Osteoporosis	733.0, V17.81, V82.81
Peripheral neuropathy	250.6, 337.1, 354, 355, 357.2
Tobacco use	305.1, V15.82
Concurrent Medication Use	
Other anti-diabetic agents (current at baseline – yes/no for each medication class)	Any medications other than the comparators; entered as class
Alpha-glucosidase inhibitors Biguanides Other DPP-4 inhibitors GLP-1 analogues Insulin Meglitinides Sulfonylureas Other Thiazolidinediones Any of the above	Compound drugs are split into their constituents for these calculations.
Anti-diabetic agents dispensed in prior year (yes/no for each possible medication class)	
Alpha-glucosidase inhibitors Biguanides Other DPP-4 inhibitors GLP-1 analogues Insulin Meglitinides	Compound drugs are split into their constituents for these calculations.

Table 4. Baseline Covariates to be Included in Propensity Score and Disease Risk Score Model	
Sulfonylureas Other Thiazolidinediones Any of the above	
Anti-hypertensive agents use at baseline or in prior year (coded separately) – yes/no for each medication class	
Angiotensin-converting enzyme inhibitors Alpha blockers Angiotensin receptor blockers Beta blockers Calcium channel blockers Direct vasodilators Loop diuretics Potassium sparing diuretics Thiazide diuretics Any of the above	Compound drugs are split into their constituents for these calculations.
Lipid-lowering agents at baseline or in prior year (coded separately)	
ADDITIONAL VARIABLES FOR STRATUM WITH PRIOR CVD	
Prior AMI[†]	410
Other ischemic heart disease	411–414
Other heart disease	402.01, 402.11, 402.91, 420-429, 440
Stroke (narrow)[§]	430, 431, 433.x1, 434.x1, 436
Stroke (broad)[§]	430–434, 436
Peripheral arterial disease	443.9
Coronary revascularization procedures	
Coronary artery bypass graft	ICD9D: 996.03, V45.81 ICD9P: 36.1X, 36.2 CPT4: 33510-33514, 33516-33523, 33525, 33528, 33530, 33533-33536, 33560, 33570, 33572, 33575, 35600 HCPCS: S2205-S2209
Percutaneous coronary intervention	ICD9D: V45.82 ICD9P: 0.66, 17.55, 36.01-36.09, 37.22, 37.23, 88.5x CPT4: 92973, 92974, 92977, 92980, 92981, 92982, 92984, 92987, 92995, 92996 HCPCS: G0290, G0291
Carotid revascularization procedures	
Carotid endarterectomy, stenting, angioplasty, or atherectomy	ICD9P: 00.61, 00.63, 38.11, 38.12 CPT4: 35301, 35390, 35501, 35601, 35901, 0075T, 0076T, 37215, 37216 HCPCS: S2211
Carotid bypass	ICD9P: 39.28
Lower Extremity revascularization	
Lower extremity endarterectomy, stenting, angioplasty, or atherectomy	ICD9P: 38.18, 38.19 CPT4: 35454, 35456, 35459, 35470, 35473, 35474, 35482, 35483, 35492, 35493, 35495, 37207, 37208, 37220-37235
Lower extremity bypass	ICD9P: 39.25, 39.29 CPT4: 35351, 35355, 35361, 35363, 35371, 35372, 35521, 35533, 35541, 35546, 35548, 35549, 35551, 35556, 35558, 35563, 35565, 35566, 35570, 35571, 35582, 35583, 35585,

Table 4. Baseline Covariates to be Included in Propensity Score and Disease Risk Score Model	
	35587, 35621, 35623, 35637, 35638, 35641, 35646, 35647, 35651, 35654, 35656, 35661, 35663, 35665, 35666, 35671, 35681-35683, 35879
Lower extremity amputation	ICD9P: 84.10-84.17 CPT4: 27295, 27590-27592, 27598, 27880-27882, 27888, 27889, 28800, 28805, 28810, 28820, 28825

* Use only codes associated with visits (inpatient or outpatient). All diagnoses and procedures are sought for the 12-month period prior to first dispensing. ICD9D: ICD-9-CM diagnosis codes; ICD9P: ICD-9 procedure codes; CPT4: Current Procedural Terminology codes; HCPCS: Healthcare Common Procedure Coding System codes.
† We exclude patients with a principal hospital discharge diagnosis indicating AMI within 60 days of new use.
§ The narrowly defined stroke includes the principal discharge diagnoses. The more broadly defined stroke includes diagnoses associated with an inpatient or outpatient encounter.

Covariates were selected *a priori* on the basis of their availability in the MSDD, a probability of having an association with risk for AMI, and a likely prevalence of at least 1% in the population. Covariates include any anti-diabetic medications the patient may be taking except for the new use of saxagliptin or the specific comparator of interest. We do not propose a search for covariates that predict choice of anti-diabetic agent but not AMI risk because these are likely to vary substantially across Data Partners, and because adjustment for variables that predict exposure (treatment choice) but not outcome can increase the variance and introduce bias.[16, 17]

PS will be developed locally by each Data Partner using a centrally written program. Four separate PSs will be developed, to estimate the probability of saxagliptin use versus use of each comparator. Estimation of each PS will be performed separately within subgroups defined by presence or absence of prior CVD. Only new users of saxagliptin and a specific comparator are included in each analysis and any prior users of either saxagliptin (among comparators) or of the comparator (among saxagliptin users) are excluded from that analysis. PS are generated using a logistic regression model in which the dichotomous dependent variable indicates whether the individual is a new user of saxagliptin (=1) or of the comparator (=0). Independent covariates are shown in **Table 4**. The same covariates will be used by each participating Data Partner and standard programming for the PS models will be developed centrally and distributed to each Partner. All covariates will be included in each model at each Data Partner, regardless of whether the covariate appears to be statistically significantly associated with AMI. Additional diagnoses and procedures related to prior CVD will be included in the PS models for the subgroup with a history of CVD, but not for the subgroup without a history of CVD since, by definition, these covariates are not present in this subgroup.

Some Data Partners may have additional covariate information (e.g., blood pressure levels, body mass index, laboratory values, and race/ethnicity) that could contribute to a PS. Those sources will be asked to include such additional variables in a more detailed PS model that will be examined in secondary analyses to determine whether further adjustment for covariates may modify effect estimates. However, primary analyses will rely on a PS that is identically calculated for all Data Partners.

In the main analysis, PS calculation will not be performed until at least 300 new users of saxagliptin and 300 new users of a comparator (between August 1, 2009 and the date of the data pull) are available in at least one of the two CVD strata (prior history or no prior history) at the Data Partner. **Appendix B** shows the preliminary counts of saxagliptin initiators observed in the Data Partners during the first five months' availability in 2009. Anticipating gradual increases in its use during 2010, we expect that most Data Partners will have 300 new users of saxagliptin and a comparator within each CVD subgroup by the end of 2010. Requiring 300 new users in each drug group is intended to ensure that a PS can be reliably

estimated using a logistic regression model.[18] If fewer than 300 new users of either saxagliptin or a comparator are available in the initial pull for a Data Partner and a CVD subgroup, that source will not contribute to the main analyses for that subgroup until later when a more reliable PS can be calculated.

The multivariable adjusted predicted probability of being a saxagliptin user (i.e., a function of the PS) are output from the logistic regression model for each individual, ordered from lowest to highest and then used to match each saxagliptin user with the comparator having the closest available PS. However, before matching, saxagliptin and comparator users are stratified by quarter of cohort entry in order to match closely on time of entry and available follow-up times. Within these strata, the saxagliptin users are matched in a random order. Each saxagliptin user is matched to a comparator user with the nearest PS. If there are more than one eligible comparator users with the same PS, the match is chosen randomly among them. A comparator user is eligible to be matched to a given saxagliptin user if a) this comparator user has not already been matched, and b) the comparator user's PS differs from the saxagliptin user's PS by less than 0.01 (on the probability scale). If no comparator users are within this 0.01 "caliper" for a given saxagliptin user, then this hard-to-match saxagliptin user will not be included in analyses of saxagliptin versus this comparator. However, these individuals will be tallied and kept for later examination of the clinical characteristics that made them difficult to match.

The PS is recalculated quarterly after addition of new data on eligible new users of saxagliptin or comparators identified during the prior three months. These subsequent analyses are stratified on the quarter in which new users enter the cohort. (The third and fourth quarters of 2009 are combined into a single stratum because saxagliptin was newly licensed and the number of new users was low.) For each quarter, cumulative data for all available new users up to the time of calculation are included in calculating each PS, but the newly calculated PS will be used to match only the newly added users. Although those individuals matched previously are included in the new PS model, they will remain linked to their original PS matches. If a person could not be matched in the original search, they will not be subsequently matched to a comparator identified later in time. If covariate relationships with saxagliptin use appear to be changing substantially over time as new data accumulate, interaction terms of time (i.e., quarter) with covariates will be included in later PS models.

Heterogeneity in PS associations across Data Partners and particularly across sites within Data Partners, will be examined in models for the PS. PS models will be fit separately for each Data Partner, and matching will be done within Data Partners. However, for two of the Data Partners, Kaiser Permanente and the HMO Research Network, there may be a non-trivial amount of variation in predictors of drug use across sites within the Partner; that is across six regions within Kaiser Permanente or across the seven independent health plans in the HMO Research Network. Although all Kaiser Permanente regions use the same formulary and share clinical practice guidelines, local variations in practice or coding patterns may exist. We will first conduct PS calculations for each region. If these models suggest heterogeneity in associations, we will include indicators for region in a pooled model and test for interactions of site with other model covariates. If these analyses also suggest heterogeneity, we will include interaction terms in the final PS calculation and then perform all matching within region. This should not substantially impair our ability to find matches because we expect to have multiple users of each comparator for each eligible saxagliptin user.

For the seven sites of the HMO Research Network, we will use the same approach as for Kaiser Permanente. The HMO Research Network sites have agreed to share individual-level data with their coordinating center at the Group Health Research Institute. Preliminary data suggest that this pooling will be needed to satisfy the requirement for at least 300 new users saxagliptin and a comparator before

calculating a PS. Across the HMO Research Network sites, prescribing rates for sitagliptin appear to vary significantly, in part because formularies may differ. It is likely that AMI rate will also vary at least modestly. Therefore, the predictors of saxagliptin receipt may differ, as may predictors of AMI. We will include indicators for each participating health plans in calculating the PS and DRS and include interaction terms of site with other covariates. We would perform matching within sites. If or when data allow, we would calculate fully site-specific PS and use these for matching.

E. CALCULATION OF DISEASE RISK SCORES AND STRATIFICATION

Whereas the PS reduces covariate information on the basis of the multivariate relationship between covariates and the exposure (saxagliptin versus comparator), the DRS does so on the basis of the covariates' relationship to AMI (yes versus no). The DRS has several potential advantages over PS in the situation of outcome-based active surveillance. It can be calculated using data available on the large population of adults with diabetes in the MSDD prior to the start of surveillance in August 2009, so that a relatively reliable risk scoring algorithm is already available at that point. Covariate relationships with AMI should be relatively stable over time and across Data Partners, so that even though individual-level data cannot be pooled, DRS models at various sites should be quite similar. Importantly, a single DRS score is required to examine multiple comparators rather than the four distinct scores that must be calculated for the PS. A common DRS across all comparator choices also allows for comparisons of event rates between comparators and for pooling all exposure groups in a single analysis, whereas PS is calculated only for the pairwise analysis of a single comparator versus saxagliptin. Although not included in this protocol, a single DRS could be calculated for an entire cohort, allowing for assessments of AMI risk with a variety of drug-related exposures beyond anti-diabetic medications in persons with diabetes.

To calculate the DRS, a cohort is built by each Data Partner. Data from a baseline year (2007 for most Data Partners or the first year available at Data Partners whose data availability begins later) are used to identify members who have diabetes according to either or both criteria (a) and (b) below:

a) ≥ one dispensing of an anti-diabetic medication other than metformin during the baseline year (usually 2007)

b) ≥ one inpatient or outpatient diagnosis of diabetes plus ≥ one dispensing of metformin during the baseline year (usually 2007).

Eligible patients must also have continuous health plan membership, age >18 years and no pregnancy diagnosis during the baseline year (usually 2007). Although diabetic patients could also be identified solely by having diagnoses of diabetes, we are planning surveillance of diabetes patients who use pharmaceutical therapies. Those who do not require medications typically have milder diabetes, with lower risk for AMI and other macrovascular complications. Relationships of other covariates with AMI risk could also vary for this group of patients. Therefore, only diabetes subjects using at least one anti-diabetic medication are included. Those who receive only insulin during the entire 2-year period are not included in order to exclude possible type 1 diabetes patients.

For patients who meet these criteria for diabetes in the baseline year (usually 2007), follow-up begins on January 1 of the next year (usually 2008). Follow-up continues through the end of the following year (usually 2009). The cohort is immediately divided into two subgroups: those with versus those without a prior history of CVD during the 12-month baseline period. Persons with a principal hospital discharge diagnosis of AMI in the final 60 days of the baseline period (i.e., 2007) are excluded as they will be excluded in the actual surveillance analyses. As with the PS, all covariates in **Table 4** are measured and included in Cox proportional hazards model used to estimate the DRS. Separate DRS models are

estimated for the subgroups with and without prior CVD. Again, covariates are retained in the model regardless of whether they are statistically significantly associated with AMI. A centrally developed program for running the proportional hazards model will be distributed to all participating Data Partners. The dependent variable is the occurrence of an AMI at any time after the beginning of follow-up through the end of 2009. AMI is identified by the algorithm described in Section G below. Censoring of follow-up occurs if an enrollee dies or disenrolls before December 31, 2009. DRS model results are expected to be relatively similar across Data Partners because the covariates included are based on a well-understood epidemiology of AMI. We will consider use of meta-analysis to combine model results across sources, but we are proposing that Data Partner-specific model results be used in the primary surveillance analyses. As with the PS, Data Partners with additional covariates predictive of AMI will be asked to calculate a DRS based on this more detailed information, which will be used in secondary analyses.

Once the DRS model is run, the estimated model coefficients for each covariate are used to estimate the relative hazard of AMI for each new user in surveillance. It is important to note that coefficients from the fitted DRS model for either saxagliptin or a comparator cannot matter to the risk scores that we assign to new users of these drugs of interest, because risk is scored on the profile of morbidities and drugs from the pre-initiation period only – i.e., the DRSs of the new users in surveillance are based on the 12-month period *prior* to initiating saxagliptin or the comparator. These DRSs are then ordered from lowest to highest and divided into deciles. The same coefficients and decile boundaries are subsequently applied to calculate DRS and to stratify all new users identified after active surveillance begins (i.e., beginning on August 1, 2009). We do not expect there to be a need to re-run the proportional hazards model quarterly as we do the PS model, because we do not expect the predictors of AMI to vary over time as predictors of saxagliptin use likely will. However, we will periodically evaluate the possibility that risk factor relationships are changing over time as AMI events accumulate during surveillance.

Although we do not anticipate substantial heterogeneity across sites (within a Data Partner) in associations of the covariates with the AMI outcome, we will look for it and address it as appropriate in DRS stratification, either by including a term for site and interactions term for site and covariates in the DRS model or, if data permit, by calculating separate DRS within each site.

F. MEDICATION POSSESSION AND FOLLOW-UP TIME

For eligible new users identified after August 1, 2009, follow-up begins on the date of the initial dispensing and continues for as long as the patient continues under observation and uses the drug of interest (saxagliptin or the comparator). Follow-up ceases at the occurrence of a first AMI, or at the end of the surveillance period, or is censored if the patient: 1) disenrolls from the health plan and therefore from observation for more than 31 days – censoring in this case is on the date of disenrollment; 2) dies and the death is not identified as an AMI; 3) is being followed as a new user of any comparator and switches to or adds saxagliptin; or 4) ceases using the drug of interest. As noted above, a person may contribute follow-up simultaneously or sequentially to two or more comparator cohorts. A saxagliptin user is typically included in more than one comparison. If that user adds a new comparator, they are censored only in the analyses with that comparator. A person originally enrolled as a user of a comparator who subsequently adds saxagliptin is censored from analyses of that comparator, but they may be eligible to be new users of saxagliptin in comparisons with other drugs of interest they have never used.

Cessation of use is considered to occur when a person's days' supply appears to have been exhausted for a period equal to 1/3 of the days' supply of the most recent dispensing. Gaps in days' supply of less than 10 days are not considered cessation regardless of the prior dispensing quantity. Thus, if a person received a dispensing with a 30-days' supply or less, follow-up would be censored on the 11th day after the end of the days' supply. If a person received a 100-day dispensing, the allowable "gap" before censoring would be 33 days. Wide variation in the typical days' supply across Data Partners (from 30 to 100 days) precludes using a simpler uniform gap for all Data Partners. However, if patients within a single Data Partner appear to almost always receive the same days' supply, we will consider using a standard "gap" for that Data Partner to simplify programming. This extension of follow-up for the allowable gap makes the comparisons between persons who do not fill until the last day of the gap, those who appear to quit on the basis of going beyond the allowable gap, and persons who switch to another comparator. In the case of a switch from one comparator to another or to saxagliptin, it means that there will be a brief period of "overlap" in which a person contributes follow-up to two distinct exposures, in distinct comparison analyses.

In calculating days' supply of study drugs, "stockpiling" is accounted for by adding any remaining pills to the next dispensing's days' supply, up to a maximum of an additional 120 days' supply (accounting for stockpiling is done only for the study drugs; there is no accounting for stockpiling of covariate drugs). Periods during which a person is hospitalized for diagnoses that are ultimately specified as other than AMI (as the principal discharge diagnosis) are, by definition, periods in which a person could not have a primary discharge diagnosis of AMI. If they had an AMI, the attending physician and hospital coding staff obviously chose to list another diagnosis in the principal position, suggesting that an AMI may have a different etiology or epidemiology than AMI's which cause hospitalization. Importantly, in-hospital time represents a period when exposures are unclear. Patients are often switched off their usual medications. It can be argued that person-days spent in hospital for other diagnoses should be removed from follow-up of all cohort members. However, excluding these person-days based on information obtained later in the follow-up (i.e., at discharge) may introduce bias, and removal of these person-days is somewhat complex in the determination of risk sets. Risk sets are composed of all cohort members who are under observation on the date of the AMI that anchors the set. It will be necessary to identify and remove any persons who were in hospital for another cause on the day of the risk set formation. Therefore, hospitalized person-times during the follow-up will not be removed from the primary analyses.

Follow-up is updated and provided in aggregate with each new quarter's data (see Section I below). In order to calculate individual follow-up, it will be necessary to link dispensing information from the previous period in order to know days' supply accurately at the outset of the new quarter. Thereafter, ongoing medication possession can be determined and censoring dates established if medication supply is exhausted. Enrollment and death data, as well as dates of initiating new therapy with saxagliptin or a comparator will also be needed to identify censoring dates.

G. IDENTIFICATION OF ENDPOINTS

The primary endpoint for this surveillance activity is AMI occurring at any point during eligible follow-up time. An AMI is identified from hospital discharge diagnosis codes, emergency department diagnosis codes and any available death records using either of the two algorithms in **Table 5**. Discharge codes of 410.x2 are not included as AMI because these specifically refer to "prior" rather than acute events, and because one validation activity suggested that inclusion of these codes lowers the positive predictive value for AMI.[19] There is no length of stay requirement for hospitalizations. Hospitalization episodes that appear to have two or more principal discharge diagnoses are included if any one meets the criteria for AMI. Admission dates, length of stay, and discharge status (alive, dead) are captured. The second criterion for AMI (death following an emergency department visit for ischemic heart disease) is anticipated to be infrequent, but intended primarily to capture deaths from AMI in the emergency department that may be missed by hospital discharge claims. It is similar to the definition used by Graham et al. in a recent study of CVD endpoints in users of rosiglitazone versus pioglitazone.[20]

Table 5. Definition of Acute Myocardial Infarction

a) ICD-9-CM hospital discharge codes (principal) of 410.x0 and 410.x1.

b) Deaths occurring on the day or the day after an emergency room encounter associated with an acute ischemic heart disease (ICD-9-CM code: 410.x0, 410.x1, 411.1, 411.8x, 413.x).*

* If there are multiple discharge diagnoses associated with a single hospital episode, a single discharge diagnosis of AMI is sufficient.

A potential need for a second endpoint had been suggested by the finding that the positive predictive value for an AMI diagnosis varies between health systems, being lower in systems dominated by fee-for-service care.[21] This suggests that hospitals in predominantly fee-for-service environments may code AMI more aggressively than hospitals in other settings. In preliminary data (**Appendix B**) from four Data Partners, AMI occurrence did appear to be greater in two large fee-for-service health plans relative to two HMO-based networks. The possibility was considered that in these latter, less aggressive settings, patients could be more likely to receive discharge diagnoses consistent with unstable angina (411.1), occlusion without infarction (411.8), or coronary insufficiency (411.89). However, the preliminary data did not suggest such a compensatory excess of non-AMI acute coronary syndrome events in the HMO-based Data Partners. On the basis of this finding, and because of the lack of validation studies of acute coronary syndrome, we will not examine acute coronary syndrome as an endpoint in this surveillance activity.

Addendum (Version 4): Upon reviewing results from the initial looks based on the original AMI definition above, the workgroup observed a higher-than-expected variation in AMI incidence rate across Data Partners. After consulting with the Data Core, the workgroup concluded that the variation was likely due to Data Partners populating their principal discharge diagnosis variable differently at the early stage of

the Mini-Sentinel program. After deliberation, we included an alternative, separate outcome algorithm that considered either primary or non-secondary discharge diagnosis of AMI (i.e., AMI diagnosis not known to be secondary) in our surveillance. We will assess the robustness of the findings based on these two slightly different outcome algorithms.

H. ANALYSIS PLAN

Several approaches are proposed for managing confounding and for modeling the association of saxagliptin use with risk for AMI in the context of sequential testing. With each approach, the relative risk for AMI in saxagliptin users versus users of the comparator is the primary target for estimation. Relative risks will be examined separately for persons with and without a prior history of CVD, but the primary test will combine these two subgroups unless marked differences in subgroup-specific relative risks are observed. After combining data across all participating Data Partners, we will "signal" concern about the safety of saxagliptin if we can reject the null hypothesis according to our alpha spending plan (described in Section H.4.). If there is no safety signal, then we will assess how much of a relative risk can be confidently ruled out and then how much reassurance is appropriate. In this section, we first describe plans for estimating the relative risk. Then we describe our alpha spending plan and how we follow it. In addition to relative risks, we also describe estimation of absolute risks and risk differences between users of saxagliptin and comparators.

H.1. Preliminary Analysis

We will begin by describing the analytic cohort at each of the Data Partners and over time. We will describe the rates at which new users of each of the drugs of interest enter the analytic cohort, accumulate follow-up, and are censored for various reasons (stopped using the drug of interest, left the health plan, etc.). We will look across Data Partners to characterize patterns of days' supply and adherence. The users of each of the drugs will be profiled with respect to age, sex, concomitant anti-diabetic medications, CVD and other comorbidities noted during the baseline period. Numbers of AMI events, and unadjusted incidence rates, will be summarized. Survival curves, where time extends from start of follow-up until AMI, will be plotted by the Kaplan-Meier methods for each of the drugs of interest.

All Data Partners will participate from the first day of surveillance in these activities, including creation of the new-user cohort, calculation of the DRS, and in producing the descriptive data described above. None of these activities requires a minimum number of prospectively identified new users of saxagliptin or any comparator. As described previously, initial calculation of the PS at each Data Partner will require that 300 new users of saxagliptin and of a comparator be identified in at least one of the two CVD strata (prior history or no prior history). The retrospective simulated sitagliptin surveillance (**Appendix C**) will provide an opportunity to evaluate the stability of the PS when there are relatively low numbers of new users during the early stages of active surveillance. If models are unstable, matches will not be robust. This would become apparent when we re-assess the closeness of original matches using the improved PS scores from subsequent periods. A second problem is that finding acceptable matches could prove difficult in the early data pulls, when there are relatively fewer numbers of comparator new users. Either of these problems, if seen in the simulated sitagliptin surveillance could lead us to modify the required number of new users for the initial PS calculations.

Heterogeneity in AMI rates across the Data Partners will be examined and implications considered. For example, an unusually high or low incidence of AMI at a Data Partner may permit us to identify and address errors in a database or unusual coding or diagnostic practices. Differences in DRS and PS models

may indicate either different prescribing patterns (for PS) or possible differences in coding or diagnostic practice (for either score).

We will also look for evidence of possible bias in the accumulating data due to receipt of partial data at some sites (with additions or corrections to the late-arriving data then made during subsequent quarterly data pulls). A data stability assessment is now underway in Mini-Sentinel to quantify the occurrence of these additions and corrections across Data Partners. In these prospective surveillance analyses, we will compare risk ratio and risk difference estimates obtained using all available data at each look with similar estimates obtained after waiting for the period found in the data stability report to insure data completeness at each Data Partner. We will produce tables for both risk ratios and risk differences showing results with confidence intervals for each delay. If we see evidence of bias in risk estimates in the early looks of the saxagliptin surveillance, we will examine associations of missingness and corrections with prescribing patterns, and we will consider whether to adopt a standard delay for future looks, from some or all Data Partners.

Once PSs have been calculated at one or more Data Partners, we will consider conducting the first PS-matched exposure–outcome analyses. These analyses will require at least some AMI events. Preliminary data gathered from four Data Partners suggested annual AMI event rates of approximately 9 per 1,000 diabetic patients per year. With at least 300 new users of saxagliptin and an equal number of each comparator at each contributing site accumulated over a period of 18 months, and an estimated average of approximately 6 months (0.5 years) of exposed follow-up per person, we expect that a Data Partner would contribute at least 2.7 events to each PS-matched analysis (600 new users x 0.5 years per new user x 0.009 AMIs per person-year). The smallest number of events across all Data Partners that could yield a significant signal is seven. With seven events and a 7-0 split with all events occurring either in saxagliptin users or in comparator users, a signal would have a nominal p-value = 0.0078, which is below our 0.0144 threshold for signaling (see Section H.4.). Any smaller number of events could not yield a significant signal. Because it is desirable to identify an elevated relative risk as soon as possible, we propose to conduct the initial analysis provided that at least seven events have been observed across all contributing Data Partners in any 1:1 matched comparison. Since the separate analyses of saxagliptin versus each comparator are complementary (i.e., each analysis is helpful in interpreting the others), we believe it is worthwhile to conduct all comparisons even if only one meets criteria specified for a “look.”

In contrast to the PS, a DRS requires no new users of saxagliptin because it is performed on pre-existing data from the period just before saxagliptin was released. We will learn from the simulated sitagliptin surveillance how comparable the DRS-stratified and PS-matched analyses are with each other and with the conventional multivariable analytic approach. Again, however, we would require that at least seven endpoints be observed across all Data Partners. In these analyses, Data Partners could contribute aggregate data on very small numbers of news users of either saxagliptin or only of a comparator. However, only those strata with at least one AMI event would contribute to pooled analyses.

H.2. Analytic Strategies

After considerable discussion, the Methods Working Group concluded that there was merit in employing two approaches to adjustment for confounding in sequential analyses, and further value in examining results of a third approach at the end of the surveillance period. Both sequential analytic approaches rely on construction of summary confounder scores, the PS in one case and a DRS in the other. By using both the PS and the DRS score, we will gain experience with each in the sequential testing environment.

Neither is specified as the primary approach. If substantial differences in relative risk estimates are observed using the two approaches, we will attempt to determine the source of these differences.

We will use 1:1 matching with the PS and we will stratify by the DRS. Matching provides the closest possible adjustment for differences in covariates at baseline, assuming that there is substantial overlap of scores between users of saxagliptin and the comparator. For those matched, baseline covariate distributions are essentially identical and saxagliptin users and their comparators can be compared using transparent, simpler unmatched methods. Comparisons using Kaplan-Meier survival curves can also be generated more easily. We will have to monitor for differential dropout over time to ensure that the unmatched analysis remains appropriate. Stratification (on deciles of the DRS) has the advantage over 1:1 PS-matching of including larger numbers of patients on comparator agents (assuming that there will generally be many more new users of the comparator than of saxagliptin), thereby augmenting power somewhat in a context where power is critically important.

In both matching and stratification, saxagliptin users may be excluded if comparable comparators cannot be found. The numbers of saxagliptin users that cannot be matched, and their clinical characteristics, will be maintained by each Data Partner. In the stratified analysis, it is also possible that an entire stratum might be eliminated from analysis if either saxagliptin users or users of the comparator were extremely skewed within the stratum. In this instance, "restriction" has been recommended to guard against persistent unmeasured confounding.[22]

The third proposed approach uses individual covariate information in typical proportional hazards models rather than constructing either score. This approach has the advantage of familiarity and also provides an opportunity to examine individually the potential confounding effects and interactions of individual covariates. At present, there is no plan to share individual-level data, even de-identified data across Data Partners. Thus, these multivariable models would be run by each Data Partner and model results pooled using meta-analytic techniques.

H.3. Choices of Models

Cox regression will be the primary approach for estimating relative risks for all analyses. Proportional hazards models are familiar to a broad audience and are designed to accommodate differential length of follow-up across individuals. We specify a stratified Cox proportional hazards regression model in which the hazard ratio for each stratum, at each point in time (i.e., at the time of each AMI event), is a function of a binary indicator of saxagliptin use, an unspecified baseline hazard, and no additional covariates. Covariate adjustment is accomplished completely either by the PS matching or by the DRS stratification. Time is measured in days from the first saxagliptin dispensing until the person has an AMI or is censored. With PS matching, each risk set includes every person in a given Data Partner who is uncensored on a day when there was at least one AMI in the new-user cohort; the risk sets for the Cox regression are stratified only by Data Partner, quarter of cohort entry, and presence/absence of CVD. With DRS stratification, the risk sets are additionally stratified by DRS level (in deciles).

In such Cox models, the hazard, at time t in the k^{th} stratum, is equal to $(\lambda_{kt})e^{\beta x}$, where β is the estimate of the log of the hazard ratio and is assumed to be constant across the strata and over time. X is a saxagliptin indicator: 1 for saxagliptin users and 0 for users of the comparator. λ_{kt} is the baseline hazard of AMI at time t, in stratum k, among users of the comparator. Our target for estimation is e^{β}, which is the hazard ratio for AMI in saxagliptin users versus users of the comparator. We can find e^{β} without estimating the nuisance parameters – which represent the baseline hazards in the strata over time –

because these "cancel out" in the partial likelihood, assuming that our target hazard ratio is constant across the strata and over time.

This stratified Cox model is especially simple to fit because it includes only one binary covariate. In the usual case when individual patient-level data are available and can be pooled, we would use the PHREG procedure in SAS to estimate the hazard ratio. In our case, we will not be able to pool the patient-level data, at least not in the active surveillance and sequential analysis portion of this project. Nevertheless, we can obtain identical results by fitting a logistic regression model to a small dataset that contains all the relevant information required to fit the usual Cox model and is built from data aggregated at the Data Partner and then pooled across Data Partners (shown in the tables in Section I below). This small dataset includes only one record per risk set.[23] The outcome variable is whether or not the AMI, which anchors the risk set, happened to a saxagliptin user; the sole predictor is the proportion of people in the risk set who are saxagliptin users. The logit of this sole predictor is specified as an offset (because it indicates the expected log odds, under the null, that the AMI occurred in a saxagliptin user).

If users of saxagliptin and comparators are lost to follow-up at equal rates over time in the 1:1 PS-matched samples, then the proportion of saxagliptin users will be 0.5 in all of the risk sets. If so, the ratio of AMIs in saxagliptin users to AMIs in users of the comparator will be our estimate of the relative risk. Furthermore, we could do an exact binomial test of the null hypothesis because the number of AMIs in saxagliptin users, conditional on the total AMIs, would have a simple binomial distribution (like coin flips). In the DRS-stratified analysis, the ratio of saxagliptin users to users of comparators will vary across the risk sets; so the relative risk estimator – and its distribution under the null hypothesis – are less transparent but nonetheless possible to calculate exactly. (For convenience and transparency, we will use familiar SAS software – the LOGISTIC procedure – for statistical tests of the null hypothesis, but we note here that it would be equally possible to use exact tests, which were also used in the simulation described below to find the threshold p-value for signaling that would control alpha-spending across planned sequential analyses).

If it turns out that saxagliptin users and comparator users are censored at equal rates, and the balance achieved by 1:1 PS-matching is sustained, then the strata (the rows of **Table 6** below) will be collapsible, and the relative risk would be simply estimated by the ratio of AMI incidence in saxagliptin users divided by AMI incidence in comparator users. With DRS stratification, the relationship of saxagliptin person-times to comparator person-times will vary across the strata and so the strata would not be collapsible.

In all analyses, we will describe whether the relative risks vary over time or across Data Partners (i.e., we will examine the "proportionality" assumption and "interactions"). If we find meaningful amounts of such variation, then this finding will be considered and reported.

After fitting the regression models, AMI incidence in saxagliptin users and relative risk estimates will be used to calculate risk differences in absolute terms between users of saxagliptin and comparators. Specifically, we will estimate the risk difference by subtracting the expected incidence of AMI in saxagliptin users from the observed incidence, where the "expected" amounts to the observed incidence divided by the relative risk estimate. Confidence intervals (CIs) will be reported for these risk differences as well as for the relative risks. Interpretation of these CIs – given our plan for multiple quarterly analyses – is discussed below under "sequential analysis." If significant risk differences are found, corresponding "numbers needed to harm" will be presented.

H.4. Sequential Analysis

Our sequential analysis plan is designed to have 80% statistical power to detect a relative risk of 1.33 over the surveillance period. In other words, the probability of ever signaling during planned surveillance is 0.80 if saxagliptin really increases the risk of AMI by 33%. We plan 10 sequential analyses. The first analysis will be done as soon as is feasible, which we expect to be near the end of 2010 or early 2011. Then the nine subsequent analyses will follow on a quarterly basis. Our proposed alpha-spending plan would keep the chances of a Type I error (i.e., a false signal) to 5%, spread across the maximum of 10 sequential analyses. The alpha spending plan described below was obtained using simulations based on the observed AMI event rates for patients with diabetes in preliminary data from four Data Partners (**Appendix B**) and on the following assumptions: 1) that by the end of the first surveillance period (which covers the first five quarters after licensure of saxagliptin), 0.25% of the diabetes population will have become eligible for the analysis as a new user of saxagliptin; 2) that in the 5th quarter of that first period the quarterly accrual rate as new users of saxagliptin would be equivalent to 0.1% of the diabetes population; 3) that thereafter, the total number of new users of saxagliptin would increase by approximately 15% per quarter over the next nine quarters (2.25 years); and 4) that the average follow-up time on saxagliptin and the comparators (and eventually available for the primary analyses) will be six months per new user. These assumptions regarding saxagliptin uptake were needed because we considered the preliminary 5-month uptake data available to us and shown in **Appendix B** to be too early to make estimates for a full 3-year period. Nevertheless, the preliminary data do raise the possibility that uptake could be slower than we have assumed. We believe that this surveillance project would prove useful even if saxagliptin use remains limited during the surveillance period because it will allow us to pilot methods for active population-based surveillance and afford rich opportunities to examine the relative safety of the four comparators.

We assumed that saxagliptin use would increase by an average of 15% per quarter over the next two years. On this basis, we expect 205 cumulative AMIs in saxagliptin users. Specifically, we assumed that there will be:

a) 46,000 new users of saxagliptin entering the PS-matched cohort (and slightly more entering the DRS-stratified cohort) and followed for an average of six months each (until an AMI or censoring event). If the overall diabetes population in all participating Data Partners amounts to about 1.6 million at the outset of surveillance, as expected, then 46,000 new users of saxagliptin would amount to just under 2.3% of this diabetes population.

b) 23,000 person-years of follow-up in saxagliptin users and the same amount or more for users of each comparator.

c) 9 AMIs per 1,000 person-years of follow-up in new users of saxagliptin and comparator, under the null hypothesis that saxagliptin and all comparators have the same event rate.

At each of the 10 planned looks at the accumulating data, according to this alpha-spending plan, we will analyze all of the informative data that will have been accrued cumulatively since saxagliptin licensure. To maintain an alpha of 0.05 across all 10 looks, we require a one-tailed p-value of 0.0144 to signal at any look. We plan to keep the signaling threshold fixed – in terms of the nominal p-value – at all looks, following the general approach of Lan and Demets.[24-26] In order to find the specific threshold level of 0.0144 that we are proposing, we performed a simulation in which we tailored this alpha-spending plan to the AMIs expected at each look in new users of saxagliptin (under the null hypothesis that saxagliptin is as safe as the comparators), and assuming that the number of comparator users in all informative risk

sets is always the same as the number of saxagliptin users (under PS matching) or greater (under DRS stratification).

The threshold p-value for signaling is straightforward to calculate because the proposed stratified Cox regression model yields test statistics that are distributed (in relatively large samples) approximately the same as the test statistics from Bernoulli trials ("coin flips"). In other words, the cumulative AMIs available for analysis at each look can be examined as if they were coin flips that land "heads" (when the AMI is in a saxagliptin user) or "tails" (when the AMI is in a user of the comparator drug). The conditional probability of our test statistic (conditional on not having signaled at a prior look) can also be calculated exactly. Our test statistic is the number of AMIs in comparator users, given the total number of AMIs ascertained thus far and the proportions of comparator users in the risk sets; and equivalently we can say that our test statistic is the corresponding nominal p-value.

The overall power of the proposed design to detect relative risks of 1.25, 1.33, 1.4, and 1.5 by the final analysis is 0.61, 0.81, 0.91, and 0.98, respectively. The chance of detecting these relative risks by the fifth quarterly analysis is 0.33, 0.49, 0.62, and 0.78, respectively. If the average person has a longer period of follow-up than our assumption (above) of six months, then power is increased. Conversely, if we accumulate only half as much follow-up time, then we will have 80% power to detect a relative risk of 1.50 by the final analysis.

Nominal CIs will be reported to facilitate interpretation and – in the absence of a signal – to facilitate evaluation of the amount of reassurance that is warranted. It should be kept in mind that with multiple sequential analyses, there is somewhat more than a 5% chance that the true relative risk will be outside the nominal 95% CI at one or another analysis. Because we used a nominal p-value of 0.0144, rather than 0.05 at each look, if there is truly no safety problem for saxagliptin or the comparator, there is a 8.1% chance that the lower bound of the 95% CI will have been above 1.0 at least once. CIs will also be reported for the risk difference and for AMI events per 1,000 person-years of saxagliptin use, along with information on the performance of these statistics in the context of sequential surveillance.

It is important to note that the proposed threshold level of the p-value, required for signaling, is not adjusted for the multiple comparisons at each periodic look. At each look, we are planning separate pairwise comparisons of saxagliptin with each of the four comparator drugs. Furthermore, we plan to conduct each comparison separately in a) the subgroup of patients with a history of CVD, b) the subgroup of patients with no history of CVD, and c) the entire new-user cohort including both types of subgroups. We further plan analyses that are adjusted by PS matching, and others adjusted by DRS-stratification. Thus, each quarterly look will include 4 x 3 x 2 = 24 analyses, yielding 24 estimates of the relative risk for AMI in saxagliptin users versus users of a comparator. Any one of these 24 analyses can yield a "signal." Although the results of these 24 analyses seem likely to be closely correlated, we need to acknowledge that if saxagliptin is really entirely safe, the chance of at least one false signal, among all planned comparisons at all planned looks, is above 0.05. If the relative risk estimates differ by subgroup, or over time, or by method of adjusting for confounding, then such differences will be examined further to facilitate interpretation.

We will monitor the accumulation of saxagliptin use, events, and the average period of continuous follow-up among new users across all Data Partners quarterly. If trends in saxagliptin uptake in Mini-Sentinel or in average follow-up fall substantially short of expectations, or if event rates prove to be less than we have assumed, the FDA will be notified immediately of the consequences for our power to detect relative risks of various magnitudes within the planned time frame. Initial calculations of power

and alpha spending were based on the estimates of both the numbers of new users and the numbers of events described above. If accrual of new users or event rates is substantially lower than expected, the complex question arises of whether the next look should be delayed. Each look uses a portion of alpha. We could spend alpha inefficiently and thereby attenuate our power to detect drug-event associations.

Because each subsequent look adds follow-up time as well as new users, new information could be obtained even if no new users were added. We therefore propose to defer any quarterly look only if fewer than four additional events have become available for all matched comparisons. If any scheduled look is deferred, we will consider re-computing the threshold level (of the p-value) that is required for a subsequent signal. By raising the threshold level slightly to reflect the lower total number of looks during follow-up, we can reduce (though not eliminate) the loss of power from lower accrual. Regardless of the ultimate number of looks, a final analysis will be conducted after 30 months, as planned.

Alternatively, as we examine the accrual of saxagliptin users and events from quarter to quarter, revised projections of the statistical power anticipated by 30 months could lead to a joint decision (with the FDA) to extend the duration of surveillance beyond 30 months. In that case, we would then retain the 10-look design without revising the threshold for signaling. By this alternative approach, we would retain the power of the original design but it would not be achieved until later (when we expect that sufficient AMIs will be ascertained).

Addendum (Version 4): As of November 2013, we have completed 5 sequential looks. Upon reviewing the uptake of saxagliptin in the MSDD and the timeline, the workgroup revised the number of looks from 10 to 7. The signaling threshold – calculated based on 10 looks – remained unchanged. Keeping the signaling threshold while reducing the number of looks would result in a more conservative analysis, i.e., we may fail to "signal", but we believe the revision would not otherwise have any substantial impact on the scientific validity of the surveillance.

H.5. "End-of-Surveillance" Analysis

At the end of the surveillance period, we propose to also fit a Cox proportional hazards model using individual covariate data rather than the two confounder scores within each Data Partner. This will allow us to compare and confirm findings from the somewhat less familiar PS and DRS approaches, to understand the role of various covariates as they may affect associations of exposure and AMI, and to examine possible treatment-covariate interactions that would not be appreciated using the summary covariate scores. These end-of-surveillance analyses would remain stratified by subgroup defined by Data Partner and prior CVD as in the sequential analyses. We would also perform separate analyses for those with versus without a prior history of CVD.

These models would first be run at each Data Partner using all available data and a centrally developed, standard program. After examination of results for possible differences across Data Partners, results would be combined using meta-analytic techniques. At present there are no plans to pool de-identified individual-level data across Data Partners. The model would use the same algorithms for determining follow-up as described in Section F above. However, the availability of individual-level data would also allow for some more complex approaches to measuring exposure. Thus, in end-of-surveillance analyses we will relax criteria for censoring patients promptly at cessation of initial, continuous use of the drug of interest. For example, resumption of use of the drug of interest after an interval of non-use would contribute additional follow-up time to cumulative exposure. This is likely to increase our available follow-up time and possibly provide a proportionally greater opportunity to examine longer term cumulative exposure. In these analyses, follow-up during which patients are not taking the drugs of

interest would be attributed to either "recent" or "former" use. These analyses will be particularly useful if the average duration of initial use of the drugs of interest is brief, as they were in the study of rosiglitazone versus pioglitazone in U.S. Medicare recipients.[20] These analyses would carefully quantify the duration or cumulative exposure and also account for time-varying exposure to other therapies.

In a final check for possible delayed effects of exposure to saxagliptin, we will use an approach in which persistent exposure is not required after a minimum duration of exposure is reached. If an agent of interest such as saxagliptin should influence aspects of atherosclerosis progression that predispose individuals to increased risk later in life, this may be missed in analyses that terminate follow-up as soon as a gap is noted. Even the relaxation of these censoring rules to allow resumption of use, may yield a dataset that is weighted heavily toward early months or years of exposure. In the final check, we will continue follow-up for all persons who had been exposed cumulatively for at least one year, regardless of their subsequent exposures. The only grounds for censoring will be addition of the comparator agent.

I. DESCRIPTION OF DATA TO BE PROVIDED BY EACH DATA PARTNER

In this section, we describe the requirements for the aggregate data that will be needed from each Data Partner at the initial and at subsequent quarterly data pulls. No individual-level data will be requested during active surveillance. Privacy is further enhanced by use of the two confounder scores which summarize (and obscure) individual covariate information.

As shown in **Table 6,** for each of the four distinct comparisons, aggregated data are cumulated, quarter by quarter, in strata defined by the quarter of entry into the cohort, the DRS decile (for the DRS stratified analyses) and the presence/absence of prior CVD. The data are hypothetical data representing expected numbers from one of the participating Data Partners for three surveillance periods. These expected numbers were used in the simulations described above to assess power. At the time the third sequential analysis is done, there will be three cohorts that have entered the surveillance (one for each period of data pull) and from one to three periods of follow-up per stratum. However, the numbers shown in the table represent the numbers after combining all strata of the DRS and both strata of CVD history.

Table 6. Aggregated Data Required for Each Comparator for the DRS Stratified Analysis

Prior CVD	DRS Decile	Period of Follow-up	Period of New Use	Persons		Person-years of follow-up		Number of AMI Events	
				Saxa-gliptin	All Comptr	Saxa-gliptin	All Comptr	Saxa-gliptin	All Comptr
0	1	1	1	1,092	10,995	364	3,647	3	32
0	1	2	1	819	8,207	171	1,710	1	14
0	1	2	2	483	4,744	55	544	0	6
0	1	3	1	546	5,349	123	1,204	1	14
0	1	3	2	362	3,601	75	750	2	9
0	1	3	3	555	5,601	64	642	1	6
.....									
0	1	10	1	256	2,593	62	632	1	10
0	1	10	2	119	1,182	29	288	1	5
.....									
0	1	10	9	964	4,821	201	1,005	1	12
0	1	10	10	1,478	7,231	169	826	1	6

As shown in the table, the initial numbers of comparators within each stratum is expected to be much larger than the numbers of saxagliptin users. For each cohort defined by a period of new use, the numbers of users of saxagliptin and comparators in follow-up declines over time, reflecting censoring. For our simulation, these declines were comparable between saxagliptin and comparator cohorts. A data display such as this could help visualize whether this assumption is met.

A similar table could be drawn to represent the data that will be provided for the PS-matched analysis. Here there would be no column for DRS decile; the number of saxagliptin users may be a little smaller if sources are unable to match some to comparator users. The number of comparators would be identical to the number of saxagliptin users in the initial period of follow-up. These numbers would remain similar during follow-up, unless there is differential cessation of use or switching.

The aggregate data, such as in **Table 6**, will be sufficient for the Cox proportional hazards model if we are willing to discretize time into 3-month time points. However, it may be desirable to have a finer view of the order of occurrence of events than the quarterly aggregate. To permit the finest ordering of outcome event in time, each Data Partner would prepare and send a table such as **Table 7**, with one row for each AMI (i.e., for each risk set).

Table 7. Risk-set Level Information Required of One Data Partner.						
Prior CVD	DRS Decile	Period of Follow-up	Period of New Use	Number of AMI Events	Drug used by person who had AMI	Proportion in risk set who used saxagliptin
0	1	1	1	1	Comparator	.0903
0	1	1	1	2	Comparator	.0898
0	1	1	1	3	Saxagliptin	.0901
0	1	1	1	4	Comparator	.0889
0	1	1	1	5	Comparator	.0890
0	1	1	1	6	Comparator	.0904
0	1	1	1	7	Comparator	.0893
0	1	1	1	5	Comparator	.0901
0	1	1	1	6	Comparator	.0899
0	1	1	1	7	Comparator	.0892

J. RECOMMENDATIONS FOR THE MINI-SENTINEL AMI VALIDATION WORKGROUP

The validity and high positive predictive value (PPV) of the AMI diagnosis have been demonstrated repeatedly in hospital discharge data from various databases.[19, 21, 27-32] However, it has also been suggested that the PPV for this diagnosis may vary between health care systems. This AMI active surveillance project is novel in combining data from multiple large health care systems, representing distinct hospitals, financing systems and geographic areas. Although all analyses will be stratified by Data Partner, the risk remains that the validity of the AMI endpoints may vary somewhat by system. If this proves to be the case, apparent associations of saxagliptin use and AMI would also vary across systems. Differential use of saxagliptin by system is anticipated. Thus, despite stratification, final safety estimates will rely more heavily on some sources than others. Thus, it will be important to know whether the PPV for this diagnosis varies by system. The planned Mini-Sentinel AMI validation activity, if it purposefully samples from each Data Partner, will be able to address concerns about differences in validity across sources.

A second concern is that at least one Mini-Sentinel Data Partner does not appear to be able to identify a principal discharge diagnosis in its hospital discharge data. Instead, it has relied on the "first listed" diagnosis in previous studies. In one such study,[32] the PPV for AMI using the first listed discharge diagnosis was relatively low at 88%. There is an ongoing study that is validating the first listed diagnosis of AMI in a pediatric population,[33] but there are no AMI validation studies in adults that address this directly. This difference could contribute to apparent differences in endpoint validity across systems and should be addressed if the Data Partner cannot find a way to reliably obtain the principal discharge diagnosis. Again, the planned Mini-Sentinel AMI validation activity, if sampled in a balanced fashion across sources, will be able to address possible differences in validity between the first position code and a principal discharge diagnosis.

A third need relates to possible differences in validity of the ICD-9-CM discharge diagnosis code of 410.x0 ("episode of care unspecified") and code 410.x1 (initial episode of care). A third code is 410.x2, which is clearly designated as an episode of care following an AMI. That code is rarely included in AMI studies. In one study that included the 410.x2 codes variably, their inclusion lowered PPVs of the 410 code for AMI modestly.[19] We recommend that the planned Mini-Sentinel AMI validity activity stratify

the sample in equal numbers between those with a discharge code of 410.x0 and 410.x1. A look at the preliminary data suggests that the 410.x0 constitute approximately 30% of the endpoints identified.

The workgroup acknowledges the careful work contained in the systematic reviews[34, 35] prepared by the Observational Medical Outcomes Partnership (OMOP). Our recommendations are made in light of the findings of these two reports and are consistent with their conclusions.

K. TIMELINE

The proposed timeline below distinguishes work done by the analytic center at Kaiser Permanente Northern California (white rows) and that done by Data Partners (shaded rows).

Table 8a. Original Timeline For Surveillance of Saxagliptin Use and AMI

	2Q 2011	3Q 2011	4Q 2011	1Q 2012	2Q 2012	3Q 2012	4Q 2012	1Q 2013	2Q 2013
Develop and test initial programs for creating the new-user cohort	X								
Distribute initial program to Data Partners	X								
Identify and describe analytic cohort, Aug 2009 – December 2010 or later (Data Partners)	X								
Develop and test programs for calculating PS and doing 1:1 matching – using Kaiser Permanente and HealthCore data	X								
Distribute program for PS calculation and 1:1 matching	X								
Initial quarterly update with PS program (Data Partners)	X								
Prepare and test programs for calculating DRS	X								
Distribute program for calculating DRS	X								
Run the DRS program (Data Partners)	X								
Quarterly update, including updates of PS and other analyses (Data Partners)		X	X	X	X	X	X	X	X
Develop and test programs for creating aggregate data		X							
Distribute program for creating aggregate data		X							
Work with Data Partners to create and check aggregate data		X							
Receive and analyze aggregate data			X	X	X	X	X	X	
Final look; conduct “end-of-surveillance” analysis									X
Prepare final report									X

Addendum (Version 4): Revised timeline.

Table 8b. Revised Timeline For Surveillance of Saxagliptin Use and AMI												
	2Q 2011	3Q 2011	4Q 2011	1Q 2012	2Q 2012	3Q 2012	4Q 2012	1Q 2013	2Q 2013	3Q 2013	4Q 2013	1Q 2014
Develop and test initial programs for creating the new-user cohort	X											
Distribute initial program to Data Partners	X											
Identify and describe analytic cohort, Aug 2009 – December 2010 or later (Data Partners)	X											
Develop and test programs for calculating PS and doing 1:1 matching – using Kaiser Permanente and HealthCore data	X											
Distribute program for PS calculation and 1:1 matching	X											
Initial quarterly update with PS program (Data Partners)	X											
Prepare and test programs for calculating DRS	X											
Distribute program for calculating DRS	X											
Run the DRS program (Data Partners)	X											
Quarterly update, including updates of PS and other analyses (Data Partners)		X		X		X		X		X		
Develop and test programs for creating aggregate data		X										
Distribute program for creating aggregate data		X										
Work with Data Partners to create and check aggregate data		X										
Receive and analyze aggregate data		X		X		X		X		X		
Final look; conduct “end-of-surveillance” analysis												X
Prepare final report												X

V. APPENDIX A

DELIBERATIONS RELATED TO THE DEVELOPMENT OF A PROTOCOL FOR ACTIVE SURVEILLANCE OF ACUTE MYOCARDIAL INFARCTION IN ASSOCIATION WITH USE OF ANTI-DIABETIC AGENTS

A. OBJECTIVES

To document the decisions, alternatives and reasoning that contributed to the development of "A Protocol for Active Surveillance of Acute Myocardial Infarction in Association with Use of Anti-Diabetic Agents."

B. DELIBERATION DEVELOPMENT

This report documents the deliberations of a three-month protocol development process. Deliberations were conducted in bi-weekly meetings for all investigators, in frequent meetings and email communications of the Methods, Data, Diabetes, and Endpoints working groups. Questions raised at the Investigators' meeting or at any working group were directed to the appropriate working group and answers and proposed strategies were brought back to the Investigators' meetings. Responses to these questions guided development of the first draft of the protocol. After circulating early drafts of the proposal, more granular suggestions and comments focused on aspects of measuring exposures, follow-up time, and outcomes. The most complex decisions were related to aspects of data analyses. Final decisions on the protocol have considered all input, represent consensus among the large group of scientists involved, and take into account limitations of the data, epidemiologic and clinical trials standard practice, current uncertainties regarding optimal methods for active surveillance, and the need to recognize a safety signal as quickly as possible while preserving validity. We also considered the design of the FDA mandated post-market randomized controlled trial of saxagliptin safety.

C. SURVEILLANCE DESIGN

We considered, at least briefly, three designs often used in observational studies of drug safety:

a) Mixed prevalent and new user cohort

b) New-user cohort

c) Self-controlled case series

A mixed-user design includes both incident drug users (i.e., those starting on a saxagliptin or a comparator during a fixed period of observation) and prevalent users (i.e. those taking saxagliptin or a comparator for an indeterminate period at (prior to) the start of the observation period). Although this design could theoretically identify slightly more users of saxagliptin and many more users of comparators, it has several drawbacks compared to the new-user design and was therefore quickly discarded. The disadvantages are listed below (as advantages of the new-user design).

In a new-user design the analytic cohort is defined only by new users of the drugs of interest during a defined period of observation. One compelling argument for adopting the new-user design rather than the mixed-user design is that it will be possible to fix the beginning of the observation period precisely on the date that saxagliptin was approved for use. That is, essentially all saxagliptin users can be treated as new users. Thus, the major potential advantage of the mixed cohort design is nullified – few

additional saxagliptin users would be found. The only exceptions will be people who appear to have joined the health plan less than a year prior to starting saxagliptin or, who appear to already be taking saxagliptin when they join. Because we expect saxagliptin users to be much less numerous than users of any comparator, especially early on in active surveillance, there would be little gain in power with the mixed-user design. Other advantages of the new-user design have been described by Ray et al[6] and Schneeweiss et al[7] and include the following:

a) All covariates in each group can be ascertained prior to the start of therapy. Therefore at the start of follow-up, no covariate is a consequence of the therapy choice (and in the causal pathway) rather than a determinant of the therapy choice.

b) No events occurring during early weeks of treatment will be missed in either saxagliptin or comparator cohorts. Missing these events differentially in one group could conceivably bias estimates of drug-associated risk if risk is either increased early or delayed.

c) Because no persons who experience an early event or who stop therapy early on because of adverse events are overlooked in either group, the cohort is not selected for persons who survive or who persist and are adherent with therapy. These characteristics and behaviors could reflect "depletion of susceptibles" or a healthy-user bias and lead to an underestimation of acute myocardial infarction (AMI) risk.

d) Duration of use is known in users of saxagliptin and each comparator. By contrast, in a mixed-user design, duration is often unknown for the prevalent users and may vary widely between the study agents.

The self-controlled case series was considered only briefly. This design can be especially useful for exposures that are intermittent, and that are not affected by earlier outcome events or changes in risk for the outcome. It has been used most often to examine brief "windows" of risk that occur relatively promptly after exposure. In the case of saxagliptin and other anti-diabetic drugs, exposure is intended to persist after initiation, i.e., to not be intermittent. Informative cases would be comprised of patients who experienced an AMI and were exposed to saxagliptin at some point. For each patient, the "unexposed" period would necessarily be a period after saxagliptin use stopped (at least once). That is, one could not sample on AMI and then examine a period before saxagliptin use – that could introduce bias. Thus, those patients who took the drug as intended, without stopping it would be uninformative. Those who would be informative could have stopped or restarted use at times when risk is changing due to unmeasured within-person confounders such as a worsening in the severity of diabetes. Decisions to start or stop saxagliptin may be associated with risk for AMI.

D. CHOICE OF COMPARATORS

The simplest analysis would compare saxagliptin with a single comparator agent. However, the group agreed to evaluate multiple comparators immediately because there are so many treatment options for type 2 diabetes mellitus, with differing cardiovascular disease (CVD) risk profiles. Further, the actual risks and differences in risk between these options are not well defined, so no obvious primary or "standard" comparator that could provide the benchmark for safety occurred to the group. Thus, there was no rationale for designating one comparator as primary. It was felt that evaluating multiple comparators would add value in that saxagliptin's level of risk could be appreciated from multiple perspectives.

The agents used to treat type 2 diabetes mellitus include sulfonylureas, thiazolidinediones (TZD), glucagon like peptide-1 (GLP-1) analogues, long-acting or basal insulin preparations, alpha glucosidase

inhibitors, amylin analogs, and meglitinides. The Diabetes working group felt the choice of comparators should be limited to agents used with some frequency as alternatives to saxagliptin. Some members thought saxagliptin would be used primarily as a third-line agent, being added to metformin and a sulfonylurea. For this application, drugs used at a similar stage in therapy would be the other DPP-4 inhibitor, sitagliptin, a TZD, and basal insulin. Others felt it may often be used as second-line therapy, being added to metformin. In this case, sulfonylureas would be the principal additional alternative. These differing roles, as a second-line and third-line therapy align with the views of the American Diabetes Association[1] and American Association of Clinical Endocrinologists/American College of Endocrinology[8] with regards to the place of DPP-4 inhibitors.

Four comparators were chosen: sitagliptin, pioglitazone, long-acting or basal insulin, and second-generation sulfonylureas (glyburide, glipizide, and glimepiride). Sitagliptin was chosen because it is the only other DPP-4 inhibitor marketed at the time and would be the most obvious treatment alternative to saxagliptin.

The TZDs were chosen because this class would be a likely alternative to DPP-4 inhibitors. We proposed to consider only pioglitazone as a comparator because new use of rosiglitazone would likely be low, given recent published evidence linking it with increased risk of CVD.[20] Given uncertainty regarding the magnitude of rosiglitazone's effect on AMI risk, a saxagliptin-rosiglitazone comparison would be especially difficult to interpret. Further, what use there is of rosiglitazone would also be potentially more selected on the basis of perceived risk for CVD.

Long-acting insulin was chosen as a comparator because it is also a potential alternative to saxagliptin initiation, especially in patients who are already using two or more oral agents. The Diabetes working group recognized that the potential risk of an AMI would likely be higher in patients initiating insulin, because the appearance of other complications of diabetes (e.g., diabetic nephropathy) often triggers a recommendation for insulin therapy by the clinician and may also be a marker of increased risk for AMI and other CVD events. The working group also recognized that earlier initiation of insulin therapy is being encouraged in some quarters, so that it may increasingly be seen as a true alternative to saxagliptin initiation.

Sulfonylureas were not initially identified as a comparator for saxagliptin, because they have historically been used as a first- or second-line therapy. However, prescribing patterns are changing. DPP-4 inhibitors are listed among second-line agents in recent guidelines[1] and concurrently, sulfonylureas may more often be reserved as a third-line therapy, after a DPP-4 inhibitor or a TZD. The Diabetes working group felt sulfonylureas would be competing alternatives and should be studied. A specific agent was not chosen. Rather comparators will include users of any of the three second-generation agents, although we will preserve our ability to distinguish among these exposures.

Metformin was not chosen as a comparator because metformin is now widely recommended and used as first-line therapy. It was considered likely most patients' initial diabetes treatment would be with metformin and the Diabetes working group felt that so few patients would use saxagliptin as initial therapy that a comparison would not be reasonable.

GLP-1 analogues were not recommended because the Diabetes working group felt that rates of use of these agents would be relatively low and that persistence could be an additional problem. The other anti-diabetes agents, including alpha glucosidase inhibitors, amylin analogs, and meglitinides, were also

not recommended, again because they are not prescribed broadly and do not have a distinct, recommended place in therapy which is comparable to saxagliptin.

E. COHORT IDENTIFICATION

Several issues surrounding cohort selection were considered by one or more working groups. These included strategies for identifying persons with type 2 diabetes mellitus, the definition of a new user, exclusion criteria for specific comparisons, and the need to identify a previous history of CVD.

E.1. Identifying Patients with Type 2 Diabetes Mellitus

The age distribution for the cohort is all otherwise eligible patients age 18 years and older. Our rationale is that an advantage of observational assessments is the ability to examine all ages, to the extent that patients in an age group can have the condition and be users of the agent. We exclude patients < 18 years of age because of the low probability that these are type 2 diabetes mellitus patients. To further reduce the chances of including persons with type 1 diabetes, we require evidence of at least one prior or concomitant dispensing of an anti-diabetic agent (excluding short-acting insulin) for persons identified as new users of long-acting insulin. We did not require concurrent use of an oral medication because the Diabetes working group felt that a substantial portion of patients with type 2 diabetes mellitus who newly initiate long-acting insulin would not be using another oral medication at that time. Note that the propensity score (PS) and disease risk score (DRS) will account for concomitant and prior medication use, largely as a means of adjusting for severity of diabetes. To the extent that concomitant use of multiple agents is driven down by insulin initiation, this could result in comparing sicker insulin-using patients with a subset of saxagliptin users who may be earlier in the course of their diabetes. However, we will be adjusting for medications used in the year prior to initiation of insulin. This fact and our planned adjustment for prior complications and comorbidities should help reduce this potential bias. Nevertheless, the decision to initiate insulin may be associated with additional aspects of disease severity not captured in prior diagnoses. Thus, we anticipate that risk should be lower in users of saxagliptin.

The Diabetes working group also recommended that we not include non-diabetic persons, and was not convinced that a new dispensing of saxagliptin or any comparator except insulin would be sufficient to confidently identify a patient as having type 2 diabetes mellitus. We therefore aimed to reduce the chances of including patients with other diagnoses (e.g., polycystic ovary syndrome or metabolic syndrome) by requiring, in addition to the newly initiated therapy, either a diagnosis of diabetes or dispensing of another anti-diabetic agent (excluding short-term insulin) at any point during the prior year.

E.2. Identifying "New Use"

In identifying new use, our preference would have been to require a full 2-year period prior to initiation of the agent of interest (saxagliptin or comparator) to rule out prior use. However, we were advised by Data Partners that turnover in enrollment is sufficient that this would substantially reduce the cohort size. We therefore require at least one year's continuous enrollment prior to the first dispensing to identify a new user. Because saxagliptin was newly marketed in August 2009, this is an issue primarily for identifying new use of comparators. The year's prior enrollment is also essential for identifying baseline comorbidities uniformly. There is a small risk that the shorter pre-observation period will allow for inclusion of a few people who are not true new users of the comparator, having used the same agent earlier – at some point between one and two years previously. We will attempt to quantify the extent to

which this could be a problem by looking for earlier use of the same medication in persons who meet the criteria for new use but have more than the required 12 months of prior enrollment.

The protocol recognizes that patients can be true new users of more than one medication of interest during the course of the surveillance activity. At any time during the surveillance period (from licensure in August 2009 until our final planned look at the end of 2012 or beyond), a patient with type 2 diabetes mellitus can enter into a saxagliptin versus comparator comparison, as a new user of saxagliptin or a particular comparator, if they are not known to have previously used the agent earlier. There are four parallel comparisons. We will permit a patient to start contributing to one of our four comparisons even if they have previously (or simultaneously) contributed follow-up in another comparison, provided they meet all eligibility criteria. This ability to identify new use of second medications of interest will be particularly useful for maximizing the numbers of eligible saxagliptin initiators, where numbers are likely to be smallest, and use likely to increase over time.

Patients may switch during follow-up from one comparator to another, or from a comparator to saxagliptin. Switching from saxagliptin to a comparator would typically disqualify one from being in any comparison. However, if they remained off the saxagliptin for more than 12 months, they would meet eligibility criteria if they became a new user of a comparator. Inclusion of these patients is no different from inclusion of a patient who switched or added just prior to their initial cohort entry (i.e., the first observed new use). A patient could contribute follow-up to two comparator cohorts at the same time if they start new use of two agents on the same day or if a second comparator is added to therapy at some time during follow-up. If a person adds new use of saxagliptin to use of a comparator, that person is censored immediately from comparisons with that comparator, but they may join the saxagliptin user cohorts for other comparisons. Addition of a comparator to saxagliptin use results in censoring from analyses with that comparator, but the person remains in analyses for other comparisons. When a person becomes eligible to enter a new comparison, covariates must be re-assessed for the prior 12 months followed by calculation of PS and DRS, and matching or stratification with other new users from the same time period of entry.

E.3. Exclusion Criteria

There are few exclusion criteria in this surveillance activity. Patients less than 18 years of age are excluded because of the high probability that they are patients with type 1 diabetes. Patients are also excluded if their age is missing or if their sex is neither “F” nor “M”.

A key area of discussion was the patient with a recent AMI and others with a prior diagnosis of CVD. Persons with a recent AMI (within past 60 days of saxagliptin or comparator initiation) are at extremely high risk for a subsequent AMI, much greater than those with a more distant AMI or other CVD. One major concern was that enrolling patients in randomized trials after a very recent cardiovascular event could add "noise" and bias towards showing no effect (if one existed) because these patients may have another event soon after randomization that is unrelated to recent treatment choices but purely to the fact that they had had a very recent event.

Patients with a recent AMI are also more likely to get treatment intensification with a new agent and among these agents, they are more likely to start insulin either in hospital or just after discharge. A second concern, therefore, is that confounding by indication may be quantitatively greater shortly after discharge. The FDA-mandated post-market randomized trials excluded patients who had been discharged within the prior 60 days for an AMI.[9] We will do the same. This is a very small group of persons, but they could contribute a larger number of events and distort findings. Information on these

patients will be kept by each Data Partner, and if sufficient numbers accumulate and if there is some variation in treatment choices, it may be feasible to examine the comparisons in this very high risk subgroup.

Patients' medication history will impact the comparisons they are eligible for. Thus, in analyses of each comparator, prior users of saxagliptin or that comparator (within the year preceding the start of follow-up) are excluded. For this reason, the composition of saxagliptin users in the four comparisons will differ somewhat.

One additional exclusion criterion is applied when saxagliptin is being compared to pioglitazone. Because congestive heart failure is a contraindication to pioglitazone and also a predictor of AMI, we exclude patients with any history of congestive heart failure from both the saxagliptin group and the pioglitazone group for this comparison only. The excluded saxagliptin initiators with congestive heart failure can be included in the other comparisons. There were no other contraindications or precautions strong enough to warrant exclusion.

E.4. Stratification of Population for Prior History of CVD

Concerns remained about other patients with known CVD. Regardless of duration, its presence is also suspected of causing confounding by indication. This is true across a broad range of vascular disease diagnoses because these are all considered to be complications of diabetes. Not only may the presence of these complications be a direct confounder, increasing risk for AMI and the likelihood of certain treatment choices, but relationships of other diagnoses and medications with diabetes treatment choices and with outcomes could vary in the presence of CVD. In other words, those with prevalent versus no prevalent CVD were considered too different with respect to outcome risk, factors impacting use of exposures, and confounding mechanisms to combine and be assured that common confounder adjustment would be adequate. Rather than excluding all patients with CVD, and making this a pure incidence study, the workgroup felt it was critically important to include and be able to generalize inferences from safety data to these patients as well.

We therefore propose that separate analyses be done in the subgroups defined by prior CVD, and that analyses combining the two subgroups will be stratified by subgroup (as well as by site, etc.) This approach was selected, in lieu of simply relying on adjustment for CVD comorbidities, because of the concerns that numerous confounders may operate differently in the presence of CVD. The stratification is then carried through to the analytic phase. The stratified approach should not diminish the power of the stratified analyses on the combined analytic cohort (including both subgroups defined by prior CVD), assuming that substantial interaction is not observed in the safety signal between the subgroups. In that case, it would be inappropriate to combine the subgroups.

There is also a concern about under-recognition of prior CVD, given the limitation of 12 months of follow-back prior to cohort entry. We will certainly misclassify some persons whose CVD was diagnosed or treated more than 12 months earlier. This can be investigated by looking for prevalent disease noted only earlier in those with a full two years follow-back. We will ask each site to report the prevalence of additional CVD noted during this period. Despite this misclassification, the two subgroups will differ dramatically in the prevalence of CVD. Because the missed disease is more distant, with less evidence of recent activity, the undetected CVD in the "no CVD stratum" is likely to reflect somewhat lower risk. Further, adjustment will still be made for the same set of covariates, so it seems highly unlikely that this misclassification could distort our effect estimates substantially.

F. CONTROLLING FOR CONFOUNDING

We propose limiting model covariates in all primary analyses to variables that are consistently available in the Mini-Sentinel Distributed Database (MSDD). Using the same variables across Data Partners allows for a single, centralized writing of the code for regression models that produce the PS and DRS described below. This will reduce the data programming burden and the chance for programming errors across sites and having the same variables at all sites will also ease data interpretation. At least two of the Data Partners could produce a number of additional risk factor covariates from electronic medical records (blood pressure, LDL-cholesterol, hemoglobin A1c levels, and body mass index). The effect of further adjusting for these covariates will be examined in secondary analyses at the Data Partners that possess them if sufficient resources are present.

Most covariates are either demographic traits (age, sex), diagnoses and procedures recorded, or medications dispensed during the prior 12 months. Several variables quantifying utilization – of hospital, outpatient visits and pharmacy – are included as surrogates for comorbidity level. The absence of a diagnosis will be taken as absence of the condition.

We also considered the possibility of using a more empirical approach to evaluating a longer list of potential covariates, and selecting only those that met defined criteria for confounding (e.g., result in a 10% change in the exposure-outcome odds ratio when added to a regression model), but discarded it because of its complexity and because of the likely result that there would be hard-to-justify variation across Data Partners in the potential confounders that we would be adjusting for. We opted instead for a lengthy but uniform, *a priori,* list of covariates, selected because of their likely associations with the outcome of AMI, their availability in the MSDD, and their likely prevalence of at least 1% in the population. Each comorbid condition listed has been reported to be associated with increased AMI risk. We are concerned that inclusion of other covariates that predict exposure choice but not outcomes (i.e., that are instrumental variables) could lead to reduced precision and possibly even bias.[15,16] Use of indices such as the Charlson comorbidity index was rejected in lieu of using large numbers of diagnoses relevant specifically to AMI in PS and DRS that should capture the comorbidities of such an index at least as well without over-adjusting.

There was a desire to determine the severity of diabetes because patients with more severe diabetes have a higher risk of CVD and may be more likely to use certain medications, especially insulin. Limitations in the data would not allow one to assess either the duration of diabetes or the presence of certain comorbidities such as microalbuminuria. Current anti-diabetic medications are a helpful proxy for severity as will be the measures of utilization.

We have proposed conducting several analyses that will address the issues related to covariate data and confounding (**Table A1**).

Table A1. Proposed Analyses to Examine Covariate Adjustment Issues

a) We will examine and report differences in observed comorbidity prevalence between Data Partners and between new users of saxagliptin and each comparator.

b) We will examine, compare and report the results of PS and DRS models across Data Partners.

c) In each Data Partner, we will examine the additional follow-back prior to 12 months before cohort entry for all patients who have additional follow-back in order to estimate how often we miss specific diagnoses or procedures using just 12 months data.

Table A1. Proposed Analyses to Examine Covariate Adjustment Issues
d) Data Partners with additional covariate data will be asked to collect that information. At the end of the surveillance, these variables will be included in the various adjustment procedures to see whether further adjustment affects estimates of the associations of saxagliptin with AMI.

F.1. Adjusting for Confounding

The Methods working group considered a variety of strategies that could be used to adjust for measured confounders, the rationale for each of the three strategies ultimately selected, and the manner in which each will be implemented are described in the following sections. The strategies selected included use of PS, DRS, and individual covariate adjustment. Although the PS was initially considered to be the primary approach for confounder adjustment, further deliberation led the group see a number of potential advantages of the DRS and determine that there was no rationale for making the PS the preferred or primary approach. The Methods working group also discussed the relative merits of 1:1 or 1:N matching versus stratification on the two scores extensively. We recognized our particular uncertainty and lack of previous experience applying either score to active surveillance. Our recommendations reflect the hope that this first active surveillance project will provide insight into the advantages and disadvantages of each approach.

F.1.1. Propensity Score (PS)

A summary score approach to adjusting for confounding was considered first by the Methods working group because these scores obscure individual-level covariate values, reducing any concerns about data sharing and because summary scores are more practical than including large numbers of covariates in surveillance analyses, and can be more efficient during early surveillance when the number of endpoints is small relative to the numbers of covariates. In this case, the summary score has shown greater power than individual covariate adjustment.[15] The PS was considered first among summary scores because of its relatively greater use and familiarity in recent years (versus DRS), One further advantage of the PS is that, were there to be multiple endpoints of interest, the single PS suffices for more than one comparison.

Conversely, one disadvantage of the PS is that a separate PS must be calculated for each pairwise comparison of exposures. In this case, four separate PSs must be calculated by each Data Partner, each derived from an analysis of all saxagliptin new users and all new users of one comparator. In fact, because we propose to stratify the analyses on history of CVD, eight PSs must be calculated. Although a single multinomial PS model could theoretically be used,[10] such models have not been widely used or evaluated, especially in the case of non-ordinal (i.e., categorical) exposures.[14] Further, we have proposed that the pairwise safety comparisons be conducted so that earlier and concurrent users of the specific comparator could be excluded, while earlier and concurrent users of other comparators are included. This pattern of inclusion and exclusion would be impossible to replicate in a single model that pools all new users – for either the PS or the outcomes models. Thus, the preferred PS, like the final analyses for safety, would be based on comparisons of saxagliptin new users with users solely of the other agent.

Although the PS will contain the same predictor covariates at each Data Partner, the PS will be generated locally by Data Partner staff using a program developed centrally by the lead team. This choice was made in part because it requires much less sharing of individual covariate information, but it was also recognized that the predictors of treatment choice are likely to differ greatly across Data Partners (e.g., in relationship to clinician practice styles and formulary differences). Coefficients in the model as well as the distribution of the PS are very likely to differ by data source. In this case, the idea

that a single Mini-Sentinel-wide PS could explain as much of the variation in treatment choice as locally generated scores seemed unlikely. Moreover, matching and stratification on the summary scores also occur within strata defined by Data Partners. The better the fit of the PS models to available data, the better matching and stratification will balance important covariates.

One downside of local computing of PS scores is that the number of exposed persons will accumulate more slowly, possibly delaying the initial PS model construction (due to expected lack of reliability of PS estimation) and slowing the initiation of the first and possibly subsequent looks. If little variation in the major predictors is seen and if relatively small numbers of exposures at some sites is precluding their use, the application of a broader PS could be reconsidered.

Another concern with PS is that the predictors of saxagliptin use will change over time as this newly introduced agent becomes more familiar and possibly as more is learned of its effectiveness. We address this in two ways. First, PSs are re-calculated as each new data pull is accomplished by pooling all previous baseline information with the newly acquired data. PS score calculation becomes increasingly robust over time. The newly calculated PSs are used to match newly identified saxagliptin users with new users of comparators who enter during the same quarter (or other period). Once matched, new users remain paired with their original match, even though their baseline data are used again the following quarter for re-calculating the PS. By matching closely on time of entry, we also tend to match on available follow-up time. The second step will be to include interactions of time (i.e., quarter) with other predictors in PS models if they appear to improve model fit.

F.1.2. Disease Risk Score (DRS)

The concept of a summary confounder score based on identifying those covariates that predict the outcome of interest (in this case AMI) rather than exposure, was proposed long before the PS.[36] To date, however, the statistical properties and value of what has come to be called a DRS have been studied less thoroughly.[13, 14] In the course of our deliberations, several potentially attractive characteristics of the DRS came into clear view, and we therefore propose that it be implemented and that its performance and findings be compared with those of the PS. These advantages (**Table A2**) increased the workgroup's interests in the DRS during our deliberations.

Table A2. Potential Advantages of the DRS versus the PS for Active Surveillance

a) A single DRS can accommodate all exposure comparisons with respect to the outcome of AMI. It is not necessary to calculate separate DRSs for users of various comparators.

b) We expect that the DRS can probably be calculated once, prior to the start of surveillance. There should be enough events during the baseline period that stable model coefficients can be obtained, within or at least by pooling across Data Partners, even for the earliest surveillance looks. By contrast, the PS will require at least 300 users of saxagliptin and a comparator at each Data Partner before a model can be estimated.

c) The single DRS will also allow comparisons of event rates between any two comparators, which may prove to be as or more valuable than comparisons with saxagliptin, especially if saxagliptin uptake is slow.

d) The major predictors of AMI are well established and should be relatively similar across Data Partners, increasing comfort with pooling data across these sources.

e) The major predictors of AMI are unlikely to change over time, increasing comfort with pooling data across time.

f) The meaning of the DRS is clearer than that of the PS. Thus, comparisons of DRS distributions, between agents of interest as well as across Data Partners, will be more useful for understanding potential confounding than those for the PS.

g) Because the DRS reflects CVD risk status, interactions of treatment choice with the DRS are also of interest and readily interpretable.

Miettinen,[36] Arbogast[13] and others recommend that the DRS be calculated in an "unexposed" cohort. In the present comparative situation, it is difficult to apply such a rule, because everyone is exposed to one agent or the other. Moreover, the intent of this advice was primarily to guard against estimating scores in cohorts artificially enriched with exposed patients, which is not the case in our proposed cohort. Because the DRS is generated prior to baseline, there will in fact be extremely little exposure to saxagliptin during follow-up. However, we have recommended that the cohort be limited to those patients with diabetes who are exposed to at least one anti-diabetic medication.

The DRS will be calculated within each Data Partner using a program written centrally. Separate scores will be derived for persons with and without prior CVD. We will examine and compare model results across Data Partners. If one or more Data Partners appear to have unstable coefficient estimates due to relatively small numbers of events, we will carefully consider the possibility of combining data across Data Partners to create a Mini-Sentinel-wide DRS model. This is more feasible than for the PS, which we expect to produce heterogeneous results across Data Partners. Pooling of data could be accomplished by meta-analysis of the results of identical models from the Data Partners.

F.1.3. Conventional Multivariable Regression Modeling

Inadequate numbers of saxagliptin users or AMI events, as well as possible issues related to sharing of individual-level data, will make it difficult or impossible to use conventional multivariable modeling with individual-level covariate and outcome data during each stage of active surveillance. However, we believe that a conventional model conducted at the end of the surveillance, whether or not a signal is found, will be important for corroborating findings from the surveillance analyses. All the variables used to create the PS and DRS will be entered individually into a multivariable Cox regression. Separate

models would be fit by each Data Partner within CVD history strata, all using the same methods to examine time-to-AMI in relation to the drugs of interest, after adjustment for the potential confounders. We will use meta-analytic methods to combine these identical models, test for heterogeneity and estimate pooled effects.

F.2. Matching versus Stratification

Matching (either 1:1 or 1:N) on either the PS or the DRS was considered to produce the closest adjustment for the large number of confounders. In fact, matching was expected to yield such similar covariate distributions for the cohorts of saxagliptin and comparators at baseline that some Methods working group members felt that, after matching, cohorts of saxagliptin and comparator users could be treated as if they had been randomized. If initial comparisons confirmed nearly identical baseline covariate distributions, then cohorts could be followed and compared during surveillance using simple, unmatched statistics. Relative risks could be estimated by simply dividing AMI incidence in saxagliptin users by AMI incidence in comparator users, where each incidence rate is simply a count of AMIs divided by the person-times of follow-up. Kaplan-Meier survival curves could be shown, and log-rank tests used to test whether differences between two curves might be due to chance alone.

Using unmatched analyses (i.e., ignoring the matched pairs) during surveillance would require that the number of matched comparators be fixed (either 1:1 or 1:N) to create identical covariate distributions. If we want to preserve the simplicity of analysis, N must be constant across matched sets. Consequently, the number of comparators would likely be either one or two, simply because the requirement for a larger fixed number would result in excluding some saxagliptin users for lack of close matches. Exclusion of any saxagliptin users would result in loss of power, especially earlier in the surveillance. Using an unmatched approach to follow-up would also require close monitoring for differential dropout which could lead to a loss of balance in covariates with continued surveillance. If the unconditional comparisons become unbalanced, and matched analyses are called for, then additional power would be lost because an entire pair would be lost to follow-up when either the saxagliptin user or the comparator user is censored.

The primary advantage of stratification was the ability afforded to use a larger proportion (or all) of comparator patients in analyses. Although we do not know how quickly saxagliptin use will increase in Mini-Sentinel populations, it is likely that the number of users will be small enough early in surveillance to make limited statistical power a significant concern. In this case, the availability of multiple comparator users per saxagliptin user will improve power somewhat. A second advantage of stratification, which pertains only to the DRS, is the ability to examine possible heterogeneity of effects across the range of risk for CVD. A stratified Cox model is also quite flexible in that it allows a different baseline hazard for each stratum rather than assuming a single common baseline hazard.

It is possible that stratification (e.g., on deciles of DRS) could create extreme subgroups in which use of either saxagliptin or a comparator is very rare. In this case, matching within the accepted caliper would also be difficult. It is also argued that such extreme strata are so highly unrepresentative of the users of one agent or the other that they are not of great interest and should simply be excluded (i.e., "restriction").[22] There is no wish to be able to generalize to such unlikely individuals. A further concern is that the rare users who are found within these extreme subgroups may be there by virtue of some very strong unmeasured confounders. We agree that elimination of such deciles or subgroups would be appropriate, but believe that this would be more likely to occur with the PS, which aims directly at predicting use of one agent versus another, than with the DRS. It seems unlikely that the association of treatment with AMI risk would be sufficiently strong to create DRS strata with severe imbalance in

treatment choice. The fact that we are stratifying on prior history of CVD before calculating DRS should make this even less likely because the strongest confounder of treatment choice with AMI risk is very likely a known prior history of CVD.

F.3. Proposed Use of PS, DRS, Matching and Stratification

The Methods working group ultimately did not feel that either matching or stratification was a clearly superior approach in the active surveillance setting, in large part because neither has been used extensively in the active surveillance setting. Advantages and disadvantages of each have been cited above. In the context of the first Mini-Sentinel surveillance project, it seemed more reasonable to use and evaluate both approaches without selecting one as primary. However, we do not propose to conduct both matched and stratified analyses for both the PS and DRS. Rather, we proposed to use matching in association with the PS analyses and stratification in association with the DRS analyses. Conducting all four combinations was felt to be too cumbersome. We further felt that the relative merits of each (PS versus DRS; matching versus stratification) could be isolated.

Our reasoning in linking stratification with the DRS was explained above. Namely, strata of the DRS have epidemiologic meaning, being associated with risk status for the outcome. Tests of interaction with treatment choice are interpretable, whereas “interactions with the PS” would be less likely and more difficult to interpret. Secondly, the chance of having extreme deciles of the PS that would have to be eliminated is greater than for the DRS.

G. OBSERVATIONAL PERIOD

The observation period is broken into two sections, the time before the index dispensing of new use of saxagliptin or a comparator and the follow up period after the medication is prescribed.

G.1. Pre-Follow Up Period

The pre-observation period is the 12-month time frame to assess cohort eligibility and capture of patient characteristics at baseline. The information captured during this period is used to generate the PS and DRS. Two decisions were needed to define the pre-observation period: 1) Should the pre-period be equal across all sites; and 2) If the pre-period is equal, how long should it be?

The rationale for a fixed, 12-month pre-baseline assessment period was given in the section above. Requiring a longer, fixed period would improve classification for prior CVD and other comorbidities, but would lead to exclusion of a substantial number of otherwise eligible patients who have been enrolled for less than two years. Conversely, a fixed period of less than one year may increase sample size somewhat but at the expense of making appropriate classification difficult or impossible. Using all available follow-back for each patient risks systematic bias. New joiners, who are typically healthier than longer term members, would look even healthier by virtue of having less follow-back for capturing comorbidities. Thus, adjustment would be biased.

G.2. Follow Up

When a person enters the cohort by virtue of a new dispensing, follow-up begins on the dispensing date of the index dispensing. A second dispensing is not required, because doing so could prevent the capture of acute effects that occurred after a single dispensing. Follow-up then continues until the first of any of the following events:

a) The first occurrence of AMI;

b) Known death from any other cause (ICD-10 code other than I21.09, I21.19, I21,11, I21.29, I21.3, I21.4);

c) A gap in continuous enrollment;

d) Discontinuation, or a gap in use, of saxagliptin or the comparator;

e) Switching or augmenting therapy to saxagliptin (for any user of a comparator); or switching to a comparator (if on saxagliptin);

f) The end of the surveillance period.

Although some patients may have multiple AMI events during observed follow-up, their inclusion would require more complex analytic methods. The predictors of repeat AMI, including drug exposures, may differ from those of an initial AMI. Moreover, treatments are likely to change following an AMI. For all of these reasons, only the first AMI during surveillance will be counted and follow-up will be terminated at that point.

Death from any other cause. Mortality data will not be routinely available. However, some deaths will be identified from hospital discharge summaries and occasionally from membership or enrollment data. When identified, non-AMI death is always a cause for censoring. Although the availability of death data may differ by Data Partner, matching and stratification both occur within Data Partner.

Disenrollment. While the initial thought was to require continuous enrollment, short allowable gaps were later considered because these typically are due to administrative errors rather than genuine loss of coverage. Two options for allowable enrollment gaps were proposed: 30 or 60 days. The majority of participating Data Partners reported that any gap longer than 31 days likely meant a true absence of coverage within their health system, during which events or new dispensing could be missed. However, there is a difference in the coding of monthly enrollment across Data Partners; some but not all use the first day of the month to indicate the start of enrollment. Therefore, 30-day was selected as the maximum allowable gap in enrollment allowable across all Data Partners. This gap applies to the requirement for continuous enrollment during the 12 months of the pre-baseline period as well as to the period of follow-up.

Discontinuation of medication of interest. Initially, a 30-day gap in medication possession (after exhaustion of the apparent day's supply and before filling a next dispensing) was proposed for defining the discontinuation of a medication and the end of follow-up. Such gaps recognize that many individuals on chronic medications occasionally fail to take the medication every day due to forgetfulness, acute illnesses, or lost medication. Subsequent discussion determined that the average size of a dispensing varied across the participating Data Partners from 30 to 100 days' supply. Under these circumstances, a uniform gap of 30 days seemed excessive for those with only 30 days supply at the last dispensing. One recent study of the association of rosiglitazone versus pioglitazone with CVD[20] used a uniform gap of seven days, which resulted in an average follow-up time to censoring of only six months. However, when

the authors extended the gap to 30 days, person-time increased only modestly. (D Graham, personal communication) Thus, it is not entirely clear how much difference in total available follow-up would result from varying the allowable gap. The final decision was to define the allowable gap in medications as a proportion (1/3) of the days' supply of the latest dispensing. A minimum gap of 10 days was also specified for those whose most recent dispensing was for < 30 days' supply. If a Data Partner proves to have a very typical dispensing size (and most do), we will allow the Data Partner to simply fix the allowable gap at one third of the typical days' supply. This will be computationally simpler.

Switching and augmentation. As discussed in Section E.2., switching medications during surveillance can lead to censoring from some or all analyses, or in some instances may lead to entry into a new cohort. Briefly, if a new user of saxagliptin subsequently adds a comparator, that user is censored from analyses involving the added comparator, but not from other comparisons. If a saxagliptin user switches to a comparator, that user is censored from all analyses because saxagliptin use has ceased and they cannot enter any analysis as a new user of a comparator because they are a prior user of saxagliptin. If a user of a comparator adds or switches to saxagliptin, they are censored from the analysis involving that comparator, but can enter other analyses as a new user of saxagliptin. Finally, if a user of a comparator adds another comparator, they continue in the analyses for the first comparator, but become eligible to enter analyses for the second comparator.

There was discussion about whether to resume follow-up for people who discontinued treatment based on our definition, but later restarted the same treatment. This is done in many drug safety studies by following subjects during such periods and attributing the follow-up to an exposure category called "former user." Several concerns led the group away from this approach for the active surveillance analyses. The first was simply the complexities of data management and analyses in having to monitor all users on a quarterly basis for resumption of use while doing active surveillance. Person-time would have to be tracked in the new category of former user. Complexity would arise with our plan to stratify analyses on the quarter (or period) of follow-up. Whereas this stratification was intended to group patients with similar length of follow-up, the follow-up stratum would no longer be equivalent to duration of use for some users. A second concern was the possibility that those users who interrupted and then resumed treatment with the same medication were different from continuous users in ways that were unmeasured but that could bias comparisons. For similar reasons, it was determined that a person could not meet eligibility criteria for new use of the same drug twice, even if more than 12 months elapsed between episodes of use. The group did feel, however, that at the end of surveillance, final analyses should try to examine the more complete exposure experience, incorporating exposure categories for former use and allowing resumption of medication use.

Follow-up during hospitalizations Periods during which a person is hospitalized for diagnoses that are ultimately specified as other than AMI (as the principal discharge diagnosis) are, by definition, periods in which a person could not have a primary discharge diagnosis of AMI. If they had an AMI, the attending physician and hospital coding staff obviously chose to list another diagnosis in the principal position, suggesting that an AMI may have a different etiology or epidemiology than AMIs which cause hospitalization. Importantly, in-hospital time represents a period when exposures are unclear. Patients are often switched off their usual medications. It can be argued that person-times spent in hospital for other diagnoses should be removed from follow-up of all cohort members. However, excluding these person-days times based on information obtained later in the follow-up (i.e., at discharge) may introduce bias, and removal of these person-days times is somewhat complex in the determination of risk sets. Risk sets are composed of all cohort members who are under observation on the date of the AMI that anchors the set. It will be necessary to identify and remove any persons who were in hospital

for another cause on the day of the risk set formation. Therefore, hospitalized person-times during the follow-up will not be removed from the primary analyses.

H. OUTCOMES

Discussion by the Endpoints working group and the larger Investigator group touched on the selection of appropriate hospital discharge diagnosis codes for identifying AMI, whether to impose a length of stay requirement, on possibilities for identifying out-of-hospital AMIs ending in death, and on the advisability of expanding the endpoint definition to include other events with similar pathophysiology to AMI. Descriptive data provided by four Mini-Sentinel Data Partners on the occurrence of AMI, by age and by presence of diabetes are given in **Appendix B**. The data suggest strong, expected associations with both increasing age and with presence of diabetes. Some differences in rates are apparent across four Data Partners, with two Partners having consistently lower rates than the other two.

H.1. Selection of ICD-9-CM Codes

Previous studies have used various sets of ICD-9-CM discharge codes to identify AMI outcomes. Many have used any ICD-9-CM code of 410 code, regardless of 4th or 5th digits (i.e., 410.xx).[34, 35] However, the fifth digit of "2" refers specifically to a "subsequent episode of care following an AMI." Although it suggests that an earlier AMI episode probably occurred, the absence of a claim for that hospitalization raises concern. One validation study observed that the positive predictive value (PPV) of an AMI diagnosis declined only slightly (from 94.1% to 92.3%) when hospitalizations with a discharge code of 410.x2 were included.[19] Such hospitalizations were relatively rare, however, amounting to only 7% of discharges, suggesting that the PPV for these cases was substantially lower. Therefore, we have specifically excluded hospitalizations with a discharge code of 410.x2, leaving all discharges with a principal diagnosis of 410.x0 (unspecified episode) or 410.x1 (initial episode) as the primary AMI endpoint.

This same criterion has been used in other recent studies.[19, 37-39] all of which reported PPV's >90%. We are not aware of any validation study that specifically compares the validity of a 410.x0 with that of a 410.x1. However, a preliminary look at the AMI's identified by the Data Partners found that as many as one third of all episodes were identified by a 410.x0 code. We know from Kiyota et al[19] and from Yeh et al[39] that this combination yields a PPV of 95% or above. Therefore, we can deduce that the PPV for 410.x0 must be quite high too, and we do not wish to exclude such a high fraction of highly likely cases. However, we do recommend that the planned AMI validation activity specifically stratify its sample on the basis of the fifth digit (0 versus 1) to assure that this is not a source of substantial misclassification at any Data Partner.

In addition to hospital discharges of AMI, the protocol also specifies that any deaths known to have occurred within one day of an emergency department visit for acute ischemic heart disease (ICD-9-CM code: 410.x0, 410.x1, 411.1, 411.8X, 413.x) also be defined as AMI. It is unclear how many deaths Data Partners will identify, but we expect the co-occurrence of an emergency department visit for ischemic heart disease followed by an out-of-hospital death to be rare. However, it is included because of the possibility that some AMI events may result in death in the emergency department without a hospital admission, and therefore without a hospital discharge. This strategy is intended to identify as many of such episodes as possible, and to minimize the possibility that bias would arise from the possible effect of a drug on the chance of surviving an AMI.

ICD-9-CM codes other than 410 (e.g., 411) have been shown to have a low PPV for AMI and therefore were not considered for inclusion in the assessment of risk for AMI.[34, 35]

H.2. Diagnosis Position for Identifying the AMI ICD-9-CM Codes

Many previous database studies of AMI have included only those cases with the ICD-9-CM code in the primary or principal position. Approximately the same number have accepted a diagnosis in either the primary or the second position.[34, 35] A somewhat larger number of validation studies have been reported for studies using the primary discharge diagnosis only. However, Kiyota et al found only a minor difference in the PPV: 95.1% for a primary position diagnosis versus 94.1% when either a primary or secondary position was considered.[19] There is some uncertainty about the meaning and use of AMI in the second position. In some cases, the AMI may truly be secondary to another condition (e.g., hypotension or hypovolemia during surgery; pneumonia). In such cases, the AMI may a have different underlying pathophysiology and may be unrelated to drug exposures. If so, these events amount to "noise" in the analysis, which is undesirable because it tends to bias risk estimates toward 1.0 and slow detection of a true signal. In some instances, AMI in the second position may be accompanied by another ischemic heart disease code (ICD-9-CM codes 411-414) in the primary position. And in others, hospitals may opt to put another diagnosis in the primary position if it results in a Diagnosis-related Group that is better reimbursed. There is uncertainty as to which of these mechanisms may be predominant in our data and whether this varies by Data Partner. For these reasons, the Endpoints working group has recommended that we define AMI as a primary or principal discharge diagnosis of 410.x0 or 410.x1. We do recommend, however, that the Data Core pursue additional analyses of secondary position AMI diagnoses early in surveillance to determine its frequency, variation in occurrence across Data Partners, and accompanying primary discharge diagnoses.

It also became apparent in discussions with Data Partners that one Data Partner does not receive diagnoses with designations of primary or principal. Rather, diagnoses are simply listed in order. Researchers at this Partner have often used the diagnosis in the "first position" as the equivalent of the primary discharge diagnosis. However, this practice has not been carefully validated. The protocol recommends that this Data Partner continue the practice of selecting only diagnoses in the first position for identifying endpoints and that the Mini-Sentinel AMI validation workgroup specifically addresses the PPV of this diagnosis.

H.3. Length of Hospital Stay Requirement

Some outcomes studies have required a length of stay restriction of two or three days to avoid misclassifying patients being admitted for diagnostic evaluation for suspected AMI; however, Kiyota et al found that eliminating a requirement for any minimum length of stay increased the number of AMI's identified by approximately 6% with no reduction in PPV for an AMI.[19] Shorter lengths of stay (without indication of in-hospital death) typically indicate a transfer, either to or from another hospital. On the basis of the Kiyota study findings, we recommend that length of stay not be a requirement in identifying AMI.

H.4. Other Possible Sources of Hospital-Related Diagnosis Codes

Data Partners pointed out that some sources of data contain diagnostic codes submitted by clinicians for professional care delivered during a hospitalization. However, these codes have never been validated and are not typically used for research by the Data Partners. A decision was made to include only hospital facility codes and emergency department codes. Professional codes will not be used to identify AMI because there could more error associated with them, not all Data Partners could identify professional claims, and use of codes from professional claims have not be used in any previous validation study.

H.5. Possible Additional Acute Coronary Syndrome Endpoints

The Endpoints working group considered the potential value of examining additional ischemic heart disease outcomes (i.e., acute coronary syndrome without infarction) in secondary analyses. Assuming that the underlying pathophysiology is similar, a broader endpoint definition could provide an earlier warning by increasing event rates. A second concern was that differential coding practices for AMI across Data Partners could lead to apparent variation in AMI rates.[21] Such "upcoding" could be manifest by apparent "compensation" or substitution, such that Data Partners with lower AMI rates would have higher rates for other acute coronary syndrome diagnoses. That is, some sites may code more aggressively for AMI more frequently, while others coded more conservatively.

However, preliminary data from four Data Partners (**Appendix B**) suggests that although the occurrence of AMI does appear to vary in frequency across four Data Partners, with rates being lower in integrated HMOs than in network model plans, there was no evidence of compensatory coding of non-AMI acute coronary syndrome in sites with lower apparent rates of AMI. Thus, concerns about possible missed AMI's in the lower rate plans are reduced. This surveillance activity will not include acute coronary syndrome as a secondary endpoint.

Sudden cardiac death events were also considered as potential outcomes. However, identification of such events would be delayed in most cases for 18 to 24 months, until mortality data from state or national death files could be obtained, matched and validated. This process would be infeasible for an active surveillance activity.

Most sites will not be able to conduct analyses that include inpatient biomarker data (e.g., cardiac troponin I, CK-MB), but the workgroup highlighted that sensitivity analyses for sites that have these data available could be very important for assessing both the presence of a confirmed AMI as well as possibly determining AMI severity. When the MSDD expands to include clinical data, these markers may be considered in future surveillance assessing AMI.

H.6. Recommendations for the Mini-Sentinel AMI Validation Workgroup and AMI Other Validation Activities

We have several recommendations for the validation activity that will be conducted by the Mini-Sentinel AMI validation workgroup or others. The first is to examine the validity of identifying AMI using the first listed diagnosis in the Data Partner that cannot identify primary discharge diagnosis. This can be accomplished by modestly oversampling records from that Data Partner. While there is an ongoing AMI validation in a pediatric population using the first diagnosis designation,[33] there are no AMI validation studies in adults from data sources which do not have primary discharge diagnosis. A second recommendation is to stratify the sample on patients with ICD-9-CM code 410.x0 versus 410.x1 to see if the PPV for 410.x0 is comparably high. A preliminary look at data from several Data Partners indicated

that the code 410.x0 was found with sufficient frequency to make 50:50 stratification of the proposed samples feasible.

I. DATA ANALYSIS

I.1. Overview of Analysis Plan

The protocol specifies several approaches to adjusting for confounding and estimating exposure-outcome associations. Associations are expressed in primary analyses as relative risks. We recognize that risk differences are important in decision-making, providing a scale on which risks can be weighed against benefits. However, risk differences tend to vary much more across subgroups than the relative risk and therefore most methods of adjustment for confounding have targeted the relative risk. Once adjusted relative risks are estimated, adjusted risk differences (in subgroups and overall) can be derived from estimates of the relative risk in conjunction with incidence information. Thus, in addition to reporting relative risks and confidence intervals (CIs), we will describe the incidence of AMI in users of saxagliptin and comparators and provide estimates of risk differences. If we find heterogeneity in the relative risks across subgroups, then subgroup specific relative risk estimates can be applied to the specific incidence rates to estimate risk differences. It should be noted that in surveillance for rare outcomes (rarer than AMI) the relative risk estimate tends to be unstable, and it can be preferable to model the risk difference directly, rather than derive it from the relative risk estimate.

I.2. Analytic Strategies

We believe that each of the three proposed approaches has important advantages. The PS and DRS are alternative ways to address "the curse of dimensionality" – the problems that arise when we adjust for a large number of individual covariates (some of them collinear, and some of them indicator variables flagging relatively rare risk factors) with only a limited number of outcome events. This is typical in surveillance studies of newly approved medical products, especially early in surveillance when the number of events is lowest, but the number of covariates is large. We can simplify and stabilize the analysis by combining the large number of covariates into a single PS or DRS, which can then be used to balance comparisons of saxagliptin users versus comparator users. Additionally, there remain concerns among Data Partners about pooling individual-level data, PS-matching and DRS-stratification permit stratified analyses using only aggregate data.

PS matching has the intuitive advantage of balancing the analytic cohort in a way that mimics a randomized trial of saxagliptin versus comparator. DRS stratification has the intuitive advantage of facilitating comparison of saxagliptin versus comparator in subgroups defined by level of risk for AMI. PS is easier to use when comparing multiple outcomes in relation to a binary exposure; DRS is easier to use when comparing multiple types or levels of exposure with respect to a single type of outcome. PS is advantageous when there is relatively more available data and externally-derived knowledge for modeling the exposures of interest; DRS is advantageous when there is much available data, and externally-derived knowledge, to model the outcomes of interest. Although the PS has been used more often than the DRS in recent years, the proposed surveillance has multiple comparators, and more data available early on for modeling the outcome of AMI. Predicting exposure (i.e., saxagliptin use) will require some waiting at each Data Partner until sufficient saxagliptin users accumulate. Additionally, the outcome of AMI has been modeled in multiple studies, is well understood, and is likely to be fairly consistent across Data Partners. These factors each render the DRS advantageous – and worthy of use – in this surveillance activity.

The protocol calls for use of both 1:1 matching (with the PS) and stratification (with the DRS). Matching (1:1) can minimize bias by restricting analyses to the single best-matching comparator users for each saxagliptin user. But 1:1 matching discards relevant information from all potential matches that are not used. Stratification can maximize efficiency by making use of relevant information from all of the comparable users of comparator drugs. 1:1 matching is more intuitive because it facilitates simple transparent analyses as could be done with data from a randomized trial (with 1:1 allocation). Stratification is less burdensome to implement in multi-site, sequential surveillance. One-to-one matching could be augmented to 1:N matching, increasing statistical efficiency. However, N must be constant across matched sets if the simplicity of analyses is to be preserved. In reality N would have to be small (likely 1 or 2) because some saxagliptin users will only have one acceptable match and we would be reluctant to exclude any saxagliptin users.

Our proposal to use both PS-matching and DRS-stratification, complemented by individual covariate adjustment at the final stage of analysis, will permit us to 1) gain extra insight from these complementary methods into the association of the anti-diabetic drugs with AMI, and 2) gain experience and understanding for Mini-Sentinel – in this initial surveillance effort – with the implementation, strengths and possible limitations of each of these approaches.

I.3. Choice of Models

The protocol specifies that Cox proportional hazards models will be used in all primary analyses. Cox regression is powerful, flexible, and widely used in cohort studies such as ours, given 1) follow-up that will be long for some users and will be censored for most users (due to dropout, non-adherence, switching, etc.), and 2) an outcome that is binary and non-repeatable (first AMI). PS-matching and DRS-stratification both permit the use of stratified Cox models in which there is only a single binary covariate: saxagliptin versus comparator. Such models can be easily fit to pool data across all Data Partners with only one record per risk set.[23] These models also facilitate the design of an alpha-spending plan for sequential analysis because the risk sets can be easily simulated as Bernoulli trials.

Poisson regression is expected to yield very similar results. Results would be virtually identical if the ratios of follow-up time (in saxagliptin users versus comparator users) in the Data Partners, subgroups, and time periods specified by the Poisson model are the same as the ratios of saxagliptin users to comparator users in the risk sets of the Cox model. If Cox and Poisson yield estimates that differ nontrivially, then the Cox model is less vulnerable to bias because its risk sets are anchored to specific time points, which would presumably make them more homogeneous in level of risk than the stratified time periods used in Poisson regression. However, Poisson regression is more intuitive to some researchers, and yields explicit estimates, not only of the relative risk, but also of AMI incidence rate – in subgroups and time periods as well as overall.

I.4. Sequential Testing Plan

We will follow the general alpha-spending approach of Lan and DeMets.[24-26] The test statistic for each of our planned periodic analyses will be the nominal p-value associated with the estimated saxagliptin effect on the hazard of AMI. This estimate will be obtained by stratified Cox regression. Lan and DeMets showed that in large samples the overall chance of a Type I error, across multiple looks, can be limited to a desired level of alpha by using (at every look) a threshold level of the nominal p-value that can be derived from the following simple formula:

$$\text{Threshold} = \text{alpha} / \log(1 + N*(e - 1)),$$

where N is the number of planned looks, and alpha is the overall chance of a Type I error.

Given our interest in limiting the chance of a Type I error to 0.05 across 10 planned looks, the threshold level of the nominal p-value in large samples would then be:

Threshold = 0.05 / log(1 + 10*(2.718 – 1)) = .05 / log(18.18) = 0.017

Although our Mini-Sentinel-wide diabetes population is very large, the number of additional AMIs that will be ascertained at each quarterly look for each pairwise comparison of saxagliptin users versus comparator users (PS-matched 1:1) is expected to be in the range of 20 to 60. We conducted a simulation to ascertain whether these expected numbers are large enough for us to rely on the formula that Lan and Demets derived for large samples. We found that our expected numbers of outcome events would not be quite large enough: a threshold slightly more stringent than 0.017 would be required, specifically a threshold level of 0.0144 for the nominal p-value would be required for us to limit Type I errors to 0.05, given the expected accumulation of outcome events in our analytic cohort. The simulation we conducted was similar to that proposed by Li and Kulldorff for a conditional sequential sampling procedure.[40]

This alpha-spending plan, like other "flat-boundary" sequential designs (see Kulldorff et al[41] and Pocock[42]) "spends" more alpha early in surveillance rather than "conserving" it for the final analysis. We also considered an intermediate strategy, which would spend alpha evenly (such that we would have a 0.005 chance of a Type I error at each of the 10 planned looks), but decided that our proposed plan is more intuitive and transparent (with a signaling threshold that is constant across looks in terms of the nominal p-value) and more consistent with the priority for timeliness in safety surveillance.

We also considered timing analyses according to the "information fraction": the proportion of needed information (needed for statistical power) that has come in, rather than the amount of months that have gone by. The power of our planned analyses at any point will be driven by information regarding AMI incidence and the numbers of new users of saxagliptin that have been identified to that point. In order for the planned 10-look sequential design to have 80% power to detect a relative risk of 1.33, the number of AMI events expected in saxagliptin users – under the null hypothesis – has to be about 207 (This number is derived from our power calculation that a relative risk of 1.33 is detectable by surveillance of 46,000 saxagliptin users for 23,000 persons-years with AMI incidence rate of 9 per 1,000 person-years: 23,000 * 0.009 = 207).

Therefore, 10 sequential analyses could be conducted whenever the expected number of AMIs in saxagliptin users (approximately 10% of 207) has accumulated (given the saxagliptin uptake and AMI incidence observed during surveillance) rather than every three months. The initial look would be conducted as soon as the initial data can be collected (to assure ourselves that we are not allowing a very large effect go undetected) and the final, 10th look performed when we've reached at least 207 AMIs. If saxagliptin uptake or AMI incidence turns out to be lower than we now expect, then our analyses will be either less timely or less powerful than we now expect. We could adapt by either reducing the number of planned looks (and thereby permit a slight increase in the nominal p-value required for signaling at each look – to retain as much power as possible while controlling Type I errors) or else we could increase the intervals between looks and thereby increase the expected duration of surveillance (in order to keep power the same as is now planned). We are inclined to look at 3-month intervals rather than intervals based on information fraction (numbers of users and event rates) for several reasons:

a) After doing the work to manage the data and ascertain the drug use, follow-up, and AMI incidence on a quarterly basis, then deferring statistical analyses amounts to "shutting our eyes" to the information that has become readily available. Opening our eyes – by looking at the data that we are making available – would sacrifice very little power unless the amount of new information is trivial (or much less than expected). "Shutting our eyes" to available data is harder to justify in lengthy public-sponsored safety surveillance than in many randomized trials. This is clearly our rationale for taking the first look as soon as possible rather than waiting for a certain number of events.

b) The timeliness that is prioritized in safety analyses is meaningful in terms of the calendar rather than information. Even if only a modest about of information is accumulated, so that only a very big safety problem is detectable, active surveillance provides some worthwhile reassurance: it will be reassuring that Mini-Sentinel is frequently looking at whatever data is available. Again, this is more critical for the first look than for subsequent looks, assuming that the first look suggested no important association.

c) We are interested in permitting a rather comprehensive assessment of saxagliptin safety at each look, with analyses planned of several comparator drugs, two subgroups (with and without prior CVD), and two important measures of safety (the relative risk and risk difference, reassurance and safety). Even if there is little new information for one of the planned analyses, there may be enough information for another planned analysis to make a comprehensive set of analyses worthwhile. Given the multiple analyses planned at each look, it is hard to specify specific transparent criteria in advance that justify delays in surveillance. For example, if reassurance matters and is based on the upper limit of the 95% CI of the risk difference, then lower-than-expected AMI incidence would make it easier for us to rule out a big risk difference despite making it harder for us to "rule out" a big relative risk.

We have also considered and are prepared to implement a compromise approach whereby, after an initial look, we would defer subsequently planned quarterly analyses for up to three additional months – but for no more than three months – if the AMI events accrued in the new-user cohort are far less than expected.

Our deliberations included consideration of the special problems posed by sequential analysis of survival data, especially if the timeline is time-from-entry to the cohort rather than calendar time. The risk set anchored to an AMI that contributes to an early analysis can grow by the time of a later analysis, and the proportion in the risk set who are saxagliptin users can change. However, in simulation a modest degree of such change mattered very little to the threshold required for alpha control. The Methods working group recommends that as the AMI surveillance activity is launched, the work on sequential methods workgroup sponsored by Mini-Sentinel will be very helpful for refining analyses. We expect to coordinate surveillance plans with the efforts of this workgroup.

The workgroup also considered the advantages and disadvantages of the proposed alpha-spending plan, in which the same level of the nominal p-value is required for signaling at each analysis, versus a plan that would incorporate a lower, more stringent threshold in early analyses that would "save alpha", permitting a less stringent threshold – closer to the conventional 0.05 – at the end of planned surveillance. With the proposed plan, more of our chances (for a Type I error) are spent early, but we have more power to detect a safety problem early. If instead we conserve alpha during early surveillance then, on average, we sacrifice timeliness for greater power later on.

Many randomized trials conserve alpha at early interim analyses so that the final look will be able to use a threshold for signaling that is fairly close to 0.05. However, the constraints and priorities are different in Mini-Sentinel's sequential safety surveillance than in most sequential randomized trials in a number of relevant ways: 1) timeliness may be a higher priority in Mini-Sentinel – a larger alpha early on enhances ability to detect a true signal if it is emerging; 2) in Mini-Sentinel, signaling is less closely linked to a decision to stop – use of the drug is not stopped immediately in the U.S. and there is no requirement that follow-up cease; thus, if lower power at the end of planned surveillance appears to preclude detecting a possible effect, the surveillance activity can be extended (albeit with additional chance of Type I error); 3) if more than 2-4 interim looks are planned, it is intuitive and convenient to use the same threshold at all looks. It is worth noting, however, that the primary analytic method we are proposing – stratified Cox regression – can be used with any alpha-spending plan, if a consensus emerges that an alternative alpha spending plan would be preferable.

Although we have proposed to adjust the threshold p-values in these sequential analyses, we are not proposing to adjust the CIs or point estimates of relative risks or risk differences. Multiple sequential analyses yield multiple opportunities for a false positive signal (a Type I error). The possible harms of Type I errors for decisions make it important to limit the chance of a Type I error to a specified amount across all planned analyses. However, it does not necessarily follow that CIs should be similarly widened. The primary role of the CI is not to replace the hypothesis test. That is, we do not focus on its lower bound in deciding whether to signal. The CI provides richer added information and its upper bound can serve to provide a measure of reassurance during surveillance when there is no signal. Widening CIs would inappropriately distort the utility of the upper limit for reassurance.

Alternatives to the nominal CI (such as post-hoc statements about power or the "repeated CIs" that merely invert sequential tests) are less useful than the nominal CI for interpreting uncertainty about effect size. Post-hoc power ignores the observed data, and repeated CIs incorporate a type of uncertainty (about the furthest extent of any CI at any analysis) that should usually be less relevant to reassurance (which is not usually linked to costly-to-reverse decisions) than to signaling. We expect that point and CI estimates will be assessed routinely, as part of surveillance, and will not be assessed only at the time of a signal. For policymakers who would only consider the CI at the time of a signal, we plan to accompany the CI with guidance about interpretation in the context of sequential surveillance: we will caution the audience regarding the extent to which – at the time of a signal – the point and CI estimates are more likely to be high than low (relative to the truth).

I.5. Multiple comparisons; Primary versus secondary methods of analysis; Consistency of results across comparator drugs, across subgroups defined by CVD history or site, and across time periods defined by time-on-drug

At each quarterly (or periodic) analysis, we are planning to conduct multiple hypothesis tests. Saxagliptin will be separately compared to four comparator drugs. Separate comparisons will be done in patients with and without a history of CVD, and in these two subgroups combined. And each hypothesis test will be conducted once with PS-matching to adjust for confounders and once with DRS-stratification. Thus, there will be 4 x 3 x 2 = 24 hypothesis tests done at each quarterly analysis, and 24 corresponding relative risk estimates will be obtained. Each of these 24 "primary" analyses will be accompanied by examination and testing of the heterogeneity of the relative risk estimate across Data Partners and over time. With stratified Cox regression, subgroup-by-saxagliptin interaction terms, and time-by-saxagliptin interaction terms will be used to examine heterogeneity (and to test the hypothesis that there is none).

Throughout, we will be explicit and transparent about the extent to which there is an elevated chance (> 0.05) that a Type I error will occur at least once. The chance of at least one Type I error is driven not only by the number of tests conducted, but also by the pattern of correlations among the multiple tests (which is now unknown). At the end of surveillance, the observed data will have defined this pattern of correlations, and we can then ascertain the chance of a Type I error by conducting a randomization test. For such a test, we simulate random re-assignment of the drugs of interest to the observed analytic cohort with its observed follow-up and outcomes, and then we re-analyze the data. By repeatedly randomizing and re-analyzing of the observed data, we ascertain the rate of Type I errors in a large number of simulated assessments, each of which has the observed patterns of follow-up and outcomes. Although we plan to be explicit and transparent in this way about the chance of a Type I error, we do not believe that it would be appropriate to impose a corresponding formal adjustment of the nominal p-value required for a signal.

While we acknowledge that heterogeneity in results (by comparator, method, subgroup, or time period) can arise due to chance alone, we plan to examine heterogeneity and consider possible implications for confounding, for the strengths and weaknesses of alternative methods, and for possible variation in real drug safety. The goal is a comprehensive assessment of the safety of saxagliptin with respect to the risk of AMI.

VI. APPENDIX B

RESULTS FROM THE DATA REQUEST FOR THE DATA CORE

A list of queries and criteria were submitted to the Mini-Sentinel Data Core to capture the utilization of the anti-diabetic agents of interest as well as the prevalence of acute myocardial infarction (AMI) and acute coronary syndrome from four Data Partners. Summary of the criteria and results of the queries are listed below.

A. FREQUENCY OF NEW USE OF VARIOUS ANTI-DIABETIC AGENTS IN 2009

New use is defined as no earlier use of the same agent in the prior 12 months:

1) Saxagliptin – count the use by month over the last six months of 2009
2) Sitagliptin
3) Exenatide
4) Long-acting insulin – alone or as mixed preparation with short-acting insulin
5) Thiazolidinedione – pioglitazone or rosiglitazone

Inclusion criteria:

1) Identify all recipients of each medication at their first dispensing during 2009.
2) Then identify the subset of recipients who have no prior dispensing in the 12 months before the first dispensing in 2009.
3) Present counts overall and by age group (e.g., 20-44, 45-65, 65-74, 75+ years) and keep separate by Data Partner so that differences across sites can be appreciated.
4) For this table, include all members covered by each Data Partner, even those for whom claims are not available (i.e., the Medicare Part D Plan only enrollees).
5) For each medication, present a descriptive table of concurrent anti-diabetic medications (defined as any medication prescribed in the prior six months).

Table B1. New Users of Saxagliptin and Each Proposed Comparator during 2009, by Data Partner

Drug	Data Partner 1	Data Partner 2	Data Partner 3	Data Partner 4	Totals
Saxagliptin[1]	260	89	8	0	357[1]
Sitagliptin	15,530	6,232	908	500	23,170
Exenatide	4,187	1,432	475	163	6,257
Pioglitazone	14,417	20,679	1,202	5,489	41,787
Long-acting insulin[2]	20,026	37,959	4,205	15,450	82,915

[1] Based on data from August 2009 to the end of the year
[2] Prescribed alone or in combination with short-acting insulin

B. FREQUENCY OF PERSONS WITH DIABETES AGED 40 YEARS OR OLDER DURING 2007-2008

Classify as diabetic all members aged 40 years and above who have one or more of the following during the two year period 2007-2008:

1) A dispensing of an anti-diabetic medication (excluding metformin)
2) An inpatient principal diagnosis of diabetes
3) Two outpatient diagnoses of diabetes on separate days
4) A dispensing of metformin + a single outpatient diagnosis of diabetes

Present counts by Data Partner and by 10-year age group beginning at age 40 years through 80+.

Table B2. Numbers of Diabetic Members during 2007-2008, by Data Partner and Age

	a. Data Partner 1 (N=4,612,356[1])	Data Partner 2 (N=1,956,049)	Data Partner 3 (N= 828,241)	Data Partner 4 (N=2,270,346)
All Diabetics	404,983 (8.8%)	368,428 (18.8%)	86,038 (10.4%)	261,719 (11.5%)
40-49	57,181	17,022	9,738	31,379
50-59	110,048	42,680	21,993	63,749
60-69	113,563	107,097	24,809	75,870
70-79	70,934	131,703	17,128	57,794
80+	45,004	67,424	12,370	32,927

[1]N is based on all members on January 1, 2009

C. ESTIMATES OF THE CRUDE INCIDENCE OF AMI AND ACUTE CORONARY SYNDROME HOSPITALIZATIONS IN ADULTS AGED 40 YEARS OR OLDER IN 2009

Estimate the "crude incidence" of hospitalized AMI, by Data Partner, during 2009 for persons age 40 years and above, stratified by whether or not they have diabetes (as discerned in 2007-2008).

1) Identify all first hospitalizations with a principal diagnosis of 410.x0 and 410.x1 in 2009, these are the numerator.
2) The denominators are all persons who were members on January 1, 2009, stratified by age and by whether or not diabetes was identified in 2007-2008.
3) Create event rate tables stratified by Data Partner, by diabetes status (identified in 2007-2008), and by 10-year age group beginning at age 40 years through 80+.
4) Identify the first occurrence of non-AMI acute coronary syndrome: unstable angina (ICD-9-CM codes 411.1 or 411.8) diagnosis at any position, or a primary diagnosis of 414.x with a secondary diagnosis of 411.1 or 411.8.
5) For both diabetic and non-diabetic enrollees, look at the crude incidence separately for subendocardial MI (410.7x or 410.9x) versus STEMI (410.0-410.6, 410.8).

Table B3. Occurrence of AMI in 2009, by Data Partner, Age, Diabetes Status								
	Data Partner 1 N (nondiab/diab) 4,207,673/404,983		**Data Partner 2** N (nondiab/diab) 1,587,621/368,428		**Data Partner 3** N (nondiab/diab) 742,203/86,038		**Data Partner 4** N (nondiab/diab) 2,270,236/261,719	
	N	%	N	%	N	%	N	%
All non-Diabetics	10,648	.003	6151	.004	1,587	.002	4,252	.002
40-49	1169	.0007	226	.0008	149	.0006	389	.0006
50-59	2413	.002	595	.002	425	.002	959	.002
60-69	2315	.003	1489	.004	377	.003	980	.002
70-79	1913	.007	1922	.005	264	.005	858	.004
80+	2838	.015	1919	.009	372	.009	1066	.008
All Diabetics	3,926	.01	3,315	.009	684	.008	1,868	.007
40-49	197	.003	61	.004	31	.003	83	.003
50-59	580	.005	236	.005	103	.005	228	.004
60-69	901	.008	848	.008	208	.008	517	.007
70-79	1,076	.015	1,253	.009	157	.009	556	.010
80+	1,172	.026	917	.013	185	.014	484	.015
[1]N is based on all members on January 1, 2009								

Table B4. Occurrence of non-AMI Acute Coronary Syndrome in 2009, by Data Partner, Age, Diabetes Status								
	Data Partner 1 N (nondiab/diab) 4,207,673/404,983		**Data Partner 2** N (nondiab/diab) 1,587,621/368,428		**Data Partner 3** N (nondiab/diab) 742,203/86,038		**Data Partner 4** N (nondiab/diab) 2,270,236/261,719	
	N	%	N	%	N	%	N	%
All non-Diabetics	8719	.002	3384	.002	886	.001	2,033	.0009
40-49	975	.0006	111	.0004	91	.0004	183	.0003
50-59	2180	.002	384	.001	244	.001	482	.0008
60-69	2231	.003	1081	.003	263	.002	588	.001
70-79	1860	.007	1214	.003	160	.003	458	.002
80+	1478	.008	595	.003	128	.003	322	.002
All Diabetics	3415	.008	2089	.006	437	.005	835	.003
40-49	245	.004	59	.003	25	.003	48	.002
50-59	655	.006	270	.006	91	.004	158	.002
60-69	974	.008	716	.007	146	.006	275	.004
70-79	965	.013	734	.006	107	.006	243	.004
80+	573	.013	310	.005	68	.005	111	.003
[1]N is based on all members on January 1, 2009								

Table B5. Percent of STEMI Among all AMIs in 2009, by Data Partner, Age, Diabetes Status				
	Data Partner 1 N (nondiab/diab) 4,207,673/404,983	**Data Partner 2** N (nondiab/diab) 1,587,621/368,428	**Data Partner 3** N (nondiab/diab) 742,203/86,038	**Data Partner 4** N (nondiab/diab) 2,270,236/261,719
	% STEMI	% STEMI	% STEMI	% STEMI
All non-Diabetics	30	30	38	36
40-49	38	44	40	38
50-59	36	42	40	37
60-69	34	36	36	32
70-79	25	29	40	25
80+	21	21	36	20
All Diabetics	22	19	29	25
40-49	24	26	35	40
50-59	25	25	36	26
60-69	25	21	20	21
70-79	21	18	30	16
80+	17	15	32	14
[1]N is based on all members on January 1, 2009				

VII. APPENDIX C

PROTOCOL ADDENDUM: SIMULATED ACTIVE SURVEILLANCE FOR SITAGLIPTIN EXPOSURE, 2006-2010

A. INTRODUCTION

It is difficult to predict how quickly saxagliptin use will increase in the health systems affiliated with the Mini-Sentinel Data Partners. The availability at this time of four years or more of previously collected longitudinal data across multiple Data Partners provides an opportunity to learn about the active surveillance methods proposed in the saxagliptin surveillance protocol. We will simulate active, sequential surveillance using data on the early years of use of sitagliptin (from October 2006 through the first quarter of 2010). In doing so, we will follow the saxagliptin protocol as closely as possible. However, because all data needed for the entire sequence of analyses are available immediately, several aspects of the protocol differ. In the following paragraphs, we point out aspects of the plan that differ from the prospective surveillance proposed in the original protocol for saxagliptin.

In this simulation, we will conduct 10 scheduled "looks" to parallel the saxagliptin protocol. Each of these 10 looks is comprised of a set of analyses on the cumulative data that is available (from the time when sitagliptin was licensed [October 2006] through one of the 10 dates when we are scheduling simulated surveillance). The first look will use data from October 2006 through the end of 2007, simulating saxagliptin surveillance that would have begun as soon as data through 2010 became available. Nine subsequent looks will each incorporate an additional quarter's data. The statistical criterion for simulating a safety signal at each look will follow the alpha-spending plan of the saxagliptin protocol. However, no real safety signal will be based on the data used in simulated surveillance. The only formal hypothesis-testing analyses of sitagliptin will use all available data through the end of 2010, i.e., it will incorporate more data than are used in the simulation exercise, which only uses data through the first quarter of 2010. This analysis will be done very near in time to the simulated surveillance analyses, since all data are available before the simulations begin.

B. IDENTIFICATION OF NEW USERS OF SITAGLIPTIN AND COMPARATORS

We will identify new users of sitagliptin, long-acting insulin, pioglitazone, second-generation sulfonylureas, beginning in October 2006, when sitagliptin was approved. New users will be identified using the same algorithm proposed in the saxagliptin protocol. Identification begins once the individual has 12 full months of enrollment. These initial 12 months of data are needed so that new use can be distinguished from repeat fills in newly enrolled individuals and so that the earliest identified new users will have a 12-month period for ascertaining baseline comorbidities. The date at which identification becomes feasible may vary by Data Partner, depending on how far back the data is available, but will be no earlier than October 2006.

C. EXTRACTION OF OTHER NEEDED DATA

Using centrally written programs, each Data Partner will extract all needed information on all new users identified above between October 1 2006 and December 31, 2010, in a single data pull. This includes subsequent enrollment and pharmacy benefit data; subsequent dispensing and day's supply information to calculate persistence and the end of exposure, switching and addition of new drug(s) of interest, censoring and disenrollment dates, and incidence of acute myocardial infarction (AMI) through the end of 2010 in all new users. Data extraction programs will be written so that new use of a second comparator can be identified at each look.

D. DISEASE RISK SCORE ANALYSIS

The disease risk score (DRS) will again be calculated based on a single Cox proportional hazards model analysis conducted onsite at each Data Partner. Cohort inclusion and exclusion criteria will be identical to those used in calculating the DRS in the saxagliptin protocol. In most Data Partners, the cohort will be ascertained during 2004-2005 and followed through the end of 2006. Thus, no member would be on sitagliptin when first ascertained and essentially none would start it in the final three months of follow-up. There would be essentially no overlap of event ascertainment between the DRS cohort and the surveillance cohort. It is possible however that a Data Partner may be unable to go back as far as 2004 to build a DRS cohort. In that case, the first two years of available data would be used to create the DRS and surveillance for sitagliptin would begin as soon as data were available for at least 12 months (to distinguish new users). In this case, there could be some overlap in membership of the DRS cohort and the sitagliptin surveillance cohort. The concern with such overlap is that when the same patients and events contribute to both calculating the DRS and to surveillance analyses for AMI that use the DRS, then the surveillance models might be overfit and the precision of associations in the surveillance analyses might be overstated. However, we expect that the extent of overlap would be small, even if the time frames for DRS estimation and surveillance were fully concurrent, because each surveillance analysis focuses on new users of sitagliptin versus a single comparator, whereas the DRS is derived from a cohort including new and prevalent users of many anti-diabetic drugs. In any ascertainment window, prevalent users will likely greatly outnumber new users.

Follow-up will begin as soon as each individual has a year of total observation (i.e., continuous enrollment). Predictors in the DRS model will be the same as those in the saxagliptin analysis (**Table 4** of the saxagliptin protocol) and will include dummy variables for baseline use of each class of anti-diabetic medications, including all comparator drugs. Their inclusion is necessary to obtain unbiased (i.e., un-confounded) estimates of the other model coefficients. However, in calculating the DRS for the new users who will be in surveillance, the coefficients for sitagliptin and the comparator will be inconsequential (as if they were set at "0") because new users cannot have used either of these drugs during the 12-month pre-initiation period used for risk profiling. Thus the effects of drugs of interest can be examined without having their effects obscured in the DRS.

We will also examine for heterogeneity of DRS model findings across Data Partners and across sites within two of the Partners, Kaiser Permanente and the HMO Research Network. However, we expect heterogeneity to be much less than for the propensity score (see below), because the major predictors of AMI are well known and unlikely to vary substantially (unless coding differences alter the meaning of specific diagnoses across sites).

E. PROPENSITY SCORE ANALYSIS

The propensity score (PS) will also be calculated separately by each Data Partner. Per the saxagliptin protocol, the first PS score will be calculated as soon as 300 new users of sitagliptin and of the comparator have been accumulated at a Data Partner in either the stratum with or without prior cardiovascular disease (CVD). The same set of covariates used in the saxagliptin protocol will be used for both the PS and the DRS. The length of the period needed to accumulate 300 new users in each drug group will vary somewhat across Data Partners. In preliminary data collected from several Data Partners, some appear to have sufficient numbers by the first quarter of 2007. Others either lack any preliminary data until 2008 or had small enough numbers that several quarters of data would be required. We anticipate that all Data Partners will have sufficient numbers of sitagliptin new users by the first surveillance analysis (with data through end of 2007).

For Data Partners that can begin PS calculations at the end of 2007, there are four years of follow-up in this simulation, compared to approximately three years in the saxagliptin protocol. However, to preserve the similarities in design and statistical issues to the saxagliptin protocol, we will take only 10 quarterly looks. The first will use data through the end of 2007. The final quarterly look will be at the end of the first quarter of 2010.

As in the saxagliptin protocol, re-calculations of the PS for each interval will build cumulatively on the models from the previous period by adding new users identified in the most recent period, but matching will only take place among the new users in that quarter. We will carefully examine the PS model re-calculations over time to evaluate possible changes in the predictive value of various confounders (as prescribing patterns may change). If covariate relationships appear to be changing substantially over time as new data accumulate, interaction terms of time (i.e., quarter) with covariates will be evaluated in later PS models.

Heterogeneity in PS associations across Data Partners and particularly across sites within Data Partners will be examined in models for the PS. As in the saxagliptin protocol, PS models will be fit separately for each Data Partner, and matching will be done within Data Partners. However, for two of the Data Partners, Kaiser Permanente and the HMO Research Network, there may be a nontrivial amount of variation in predictors of drug use across sites within the Partner; that is across six regions within Kaiser Permanente or across the seven independent health plans in the HMO Research Network. Although all Kaiser Permanente regions use the same formulary and share clinical practice guidelines, local variations in practice or coding patterns or patient demographics may exist. We will first conduct PS calculations for each region. If these models suggest heterogeneity in associations, we will include indicators for region in a pooled model and test for interactions of site with other model covariates. If these analyses also suggest heterogeneity, we will include interaction terms in the final PS calculation and then perform all matching within region. This should not substantially impair our ability to find matches because we expect to have multiple users of each comparator for each eligible sitagliptin user.

For the seven sites of the HMO Research Network, we will use the same approach as for Kaiser Permanente. HMO Research Network sites have agreed to share individual-level data with their coordinating center at the Group Health Research Institute. Preliminary data suggest that this pooling will be needed even in the sitagliptin surveillance project to satisfy the requirement for at least 300 new users each of sitagliptin and of a comparator before calculating a PS. Across the HMO Research Network sites, prescribing rates for sitagliptin appear to vary significantly. It is likely that AMI rate will also vary at least modestly. Therefore, the predictors of sitagliptin use may differ, as may predictors of AMI. We will include indicators for each participating health plan in calculating the PS and DRS and include interaction terms of site with other covariates. We would perform matching within sites. If or when data allow, we would calculate fully site-specific PS and use these for matching.

The sitagliptin simulation will allow us to examine the stability of early PS estimates (based on at least 300 new users each of sitagliptin and a comparator) by comparing PS rankings of individuals and the quality of matches obtained from the first samples with those obtained at PS re-calculations in the second and third periods. If rankings and matches appear to change substantially, we will consider delaying the initial PS calculations and matchings until more data (e.g., 600 new users each of sitagliptin and of a comparator) become available.

F. CONVENTIONAL MULTIVARIABLE REGRESSION ANALYSIS

The saxagliptin protocol specifies that a conventional multivariable-adjusted regression analysis, using the entire set of predictors that is included in the PS and DRS, will be conducted at the end of the surveillance period using meta-analytic techniques. Specifically, a conventional multivariable-adjusted regression analysis will be done at each Data Partner; then the effect estimates will be pooled. The pooled estimate of the effect of sitagliptin on AMI risk will be calculated as a weighted average of the Data Partner-specific estimates, using precision weights – also known as inverse variance weights. The 95% confidence interval (CI) and hypothesis test will be calculated, using the inverse of the sum of the precision weights to estimate the variance. Similarly, heterogeneity of the Data Partner-specific estimates will be assessed by a chi-square test (based on the weighted sum of squared differences of the site-specific estimates from the pooled estimate, again using precision weights, and where the degrees of freedom equal the number of Data Partners minus one). The amount and possible sources of heterogeneity will be considered when we evaluate the pooled effect estimate and the hypothesis test, even if the heterogeneity could be deemed "statistically insignificant" (i.e., somewhat above 0.05). These methods for meta-analysis are described by Greenland and O'Rourke in Modern Epidemiology, 3rd ed., p 668-672.

The FDA has requested that the workgroup also consider conducting sequential multivariable-adjusted regression analyses at individual Data Partners during the active surveillance period. This simulation activity provides an opportunity to evaluate that strategy. We will provide a centrally written program to each Data Partner to run Cox proportional hazards analyses predicting time to AMI incidence with instructions to attempt the models at the same time the PS are calculated, using the same samples. Separate models would compare sitagliptin to each comparator in unmatched analyses using all new users of sitagliptin and the comparator.

Because of the relatively small numbers of observations per Data Partner initially, the large number of covariates, and the very small numbers of endpoints expected with short follow-up, we expect that many early Data Partner-specific multivariable-adjusted models will not converge. Data Partners will be instructed to repeat the process sequentially after adding the next period of data. When models first converge for a Data Partner, model results (coefficients and their standard errors) will be transmitted to the analytic center at Kaiser Permanente Northern California. Once results are available from two or more Data Partners, they will be combined meta-analytically. Trajectories of serial results (hazard ratios for sitagliptin use versus each comparator with CIs) will then be compared with the same estimates obtained using PS matching and DRS stratification. We will be able to evaluate the relative precision of the estimates obtained and draw some conclusions about the earliest point in the saxagliptin surveillance at which this approach could reasonably be recommended.

G. EVALUATING POTENTIAL BIAS OR INSTABILITY DUE TO RETROSPECTIVE UPDATING OF CLAIMS AND ADMINISTRATIVE DATA

In the active surveillance context, it is desirable to maximize available information at each look by using all data as quickly as it becomes available to each Data Partner. However, in most systems, the most recent data is typically incomplete and subject to updating in the months that follow. Updates are of two types. First, late-arriving pharmacy or hospital discharge claims may simply be added at a later date. Thus, with respect to complete data, earlier looks would suffer from some "missingness." A second type of update is the correction of earlier entries. In this case, earlier looks would have some "misclassification" of outcomes and exposures. Most Data Partners know the time required for data in their system to "settle;" that is, for most additions and corrections to have occurred. We suspect that

misclassification (due to corrections) is a much smaller issue than missingness when using data before it is finalized.

Missingness and misclassification in data sampled early may be entirely random with respect to drug exposures or AMI events. If so, missingness simply reduces the statistical power when compared with waiting for full data, but the advantage of having partial data earlier remains. If misclassification is random, it would be expected to create a bias toward the null if indeed there is a drug-outcome association. Whether such non-differential misclassification undoes the advantage of taking earlier looks at data depends on the volume of corrections to endpoints (and to a much less extent drug exposures) after initial entry. Its impact would be greater earlier in surveillance when recent events comprise a larger fraction of total endpoints.

However, either missingness or misclassification could also be non-random or differential with respect to the drugs of interest. For instance, some providers may routinely submit claims later than others and also be more (or less) likely to prescribe specific anti-diabetic agents. The presence and magnitude of such differential bias could vary by Data Partner and also by the exposure–outcome pair being studied. Determining whether and how severely early or incomplete data may be biased compared with looking later at complete data is a critical question for Mini-Sentinel. A related question is how long it is necessary to wait in order to ensure that data are essentially complete for each Data Partner. Both of these questions can be addressed empirically. Mini-Sentinel is currently conducting a data stability assessment. This will provide valuable information on the volume of late arriving and corrected data and the interval needed to insure that data are complete or stable for each Data Partner.

It is not possible to examine the question of bias in this retrospective surveillance simulation activity of sitagliptin, because it requires being able to identify the late additions and corrections. Data Partners and their parent health plans do not preserve earlier versions of datasets. Only a complete dataset, after all additions and corrections have been made, will be available. However, we can and will examine the effects of non-differential missingness of data on patterns of risk ratios and risk differences at each look by randomly removing fractions of the most recent data. If the data stability assessment activity suggests that missingness varies by Data Partner, we will modify the random removal exercise to reflect observed differences in missingness across sites. Similarly, if there is evidence that delays in data entry are more common for some drugs (e.g., newer drugs requiring prior authorization in some settings), we can simulate greater missingness for those drugs. We will produce two tables, one for risk ratios and one for risk differences and their CIs. Each table will present the estimates, at each look, for complete data and for random exclusions of 5, 10, and 20% of data from the most recent interval. We expect that point estimates will not change materially, but that CIs will be wider, especially at the earlier looks. It would also be possible to simulate the random misclassification that could occur due to late corrections. If the data stability report suggests that such corrections are numerous enough to cause concern, we will consult with the FDA whether such simulation would be useful.

H. DECIDING WHETHER TO END SITAGLIPTIN SURVEILLANCE OR CONTINUE IT PROSPECTIVELY

Based on power considerations for the saxagliptin analyses, we believe it unlikely that we would need to continue follow-up for sitagliptin beyond the 10th look, after the first quarter of 2010. Even if a weak though causal association is present, the available data should yield a precise estimate. Preliminary looks at new-user data suggest that there will be nearly three times as many sitagliptin users identified (~125,000) between 2007–2010 as the projected number of saxagliptin users (~46,000). We have further specified that the formal hypothesis tests for AMI risk with sitagliptin will be performed on the full dataset through the end of 2010, conducted at the same time as the 10 sequential analyses but containing the additional data from the last three quarters of 2010.

There is no plan or expectation to continue follow-up beyond the end of 2010. However, it is nevertheless possible that we could see a suggestion of an association, possibly in one stratum (e.g., in those with a prior CVD), and that the precision for this subgroup estimate is insufficient. In that case, it would be straightforward to continue assessment of sitagliptin into the prospective saxagliptin surveillance project. The DRS-stratified analysis in particular affords the opportunity to readily compare risks with sitagliptin to that of users of each other comparator. Depending on what we find regarding the comparability of PS- and DRS-derived estimates, we may elect to continue calculating PS for matching sitagliptin users to other comparators too or we may decide that the DRS stratified analyses are sufficient.

I. TIMELINE FOR THE SIMULATED ACTIVE SURVEILLANCE ANALYSIS OF SITAGLIPTIN

We propose a timeline of nine months to complete these analyses using data through the end of 2010. The workplan, by month, is given in the table below. Programs will be written at the analytic center at Kaiser Permanente Northern California and tested in both Kaiser Permanente data and with Data Partner collaborators at HealthCore. Once distributing programs to all Partners, we will work closely with each Data Partner to insure that the programs work well and that the data generated appear to be reasonable. The programs will include the steps needed to randomly delete small fractions of the most recent data at each Data Partner in order to look at the effects of random missingness on patterns of associations.

Timeline									
2011→	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Develop and test initial programs for creating the new-user cohort	x								
Distribute initial program to Data Partners		x							
Develop and test programs for calculating the PS and doing 1:1 matching		x							
Distribute program for PS calculation and 1:1 matching			X						
Prepare and test programs for calculating DRS			X						
Distribute program for calculating DRS				x					
Develop and test programs for creating aggregate data				x					
Distribute program for creating aggregate data					x				
Work with Data Partners to create and check aggregated data						x			
Receive and analyze data, prepare final report							x	x	x

Addendum (Version 4): As we revised the timeline for the saxagliptin surveillance, we decided to continue the simulated sitagliptin surveillance beyond 2010. The cost to continue the surveillance is marginal because a single data request provides data needed for both saxagliptin and sitagliptin surveillance.

VIII. ADDENDUM: ASSESSMENT OF HOSPITALIZED HEART FAILURE OUTCOME

Prepared by: Darren Toh, ScD[1]; Christian Hampp, PhD[2]; Marsha E. Reichman, PhD[2]; David J. Graham, MD, MPH[2]; Suchitra Balakrishnan, MD, PhD[3]; Frank Pucino, PharmD, MPH[3]; Jack Hamilton, AB[4], Samuel Landle, PhD[4]; Aarthi Iyer JD, MPH[1]; Malcolm Rucker, MS[1]; Madelyn Pimentel, BA[1]; Neesha Nathwani, BS[1]; Marie R. Griffin, MD, MPH[5]; Nancy J. Brown, MD[6]; Bruce H. Fireman, MA[4]

Affiliations: [1]Department of Population Medicine, Harvard Medical School and Harvard Pilgrim Health Care Institute, Boston, MA, [2]Office of Surveillance and Epidemiology, Center for Drug Evaluation and Research, Food and Drug Administration, Silver Spring, MD, [3]Office of New Drugs, Center for Drug Evaluation and Research, Food and Drug Administration, Silver Spring, MD, [4]Division of Research, Kaiser Permanente Northern California, Oakland, CA, [5]Departments of Health Policy and Medicine, Vanderbilt University and the Mid-South Geriatric Research Education and Clinical Center, VA TN Valley Health Care System, Nashville, TN, [6]Vanderbilt University School of Medicine, Nashville, TN

Background. A large placebo-controlled randomized trial (Saxagliptin Assessment of Vascular Outcomes Recorded in Patients with Diabetes Mellitus–Thrombolysis in Myocardial Infarction 53 [SAVOR-TIMI 53]) found that saxagliptin did not increase or decrease the rate of the primary combined outcome—nonfatal myocardial infarction, nonfatal ischemic stroke, or cardiovascular death—in patients with type 2 diabetes (T2D) who were at risk for cardiovascular events.[43] However more patients in the saxagliptin group than in the placebo group were hospitalized for heart failure (HF): 3.5% vs. 2.8%, hazard ratio: 1.27 (95% CI, 1.07 to 1.51, p=.0007), causing concern that DPP-4 inhibitors, such as saxagliptin or sitagliptin, may increase HF risk. Additionally, hospitalized HF (hHF) events may be associated with increased morbidity and mortality in patients with T2D.

Objective. Examine the risk of hHF in new users of saxagliptin and sitagliptin versus new users of second generation sulfonylureas, pioglitazone, and long-acting insulin, using the infrastructure of the Mini-Sentinel AMI surveillance of these antihyperglycemic drugs.

New-User Cohort Design. The protocol for AMI surveillance specifies most features of the study design, including identification of eligible T2D patients who are new users of the study drugs, ascertainment of baseline covariates, assessment of follow-up time and censoring, stratification by prior cardiovascular disease, and analysis by stratified Cox regression. As in AMI surveillance, there will be two approaches to covariate adjustment: one set of analyses adjusting for covariates by matching on a propensity score (PS), and another set of analyses adjusting for covariates by stratifying on a disease risk score (DRS). New users of the same antihyperglycemic drugs as have been monitored in AMI surveillance will be followed from the initial filled prescription until an outcome event, which will be a hHF event rather than AMI. The risk of a first HF hospitalization (after initiating a study drug) in new users of a DPP-4 inhibitor will be compared with risk in new users of a comparator drug using a stratified Cox regression model. Unlike AMI surveillance, there will not be multiple "sequential" analyses; instead we propose a single analysis in the spring of 2014 of all information available on new users since licensure of sitagliptin in 2006 (and licensure of saxagliptin in 2009). Key features of the design include:

- hHF will be identified by ICD-9-CM codes 402.x1, 404.x1, 404.x3, and 428 recorded as principal diagnosis in an inpatient encounter in the MSDD. Previous validation studies of hospital

discharge diagnosis codes 402.x1, 404.x3, and 428 in the first position showed a PPV of 85% to 96%.[44]

- Seven separate pairwise comparisons, as in AMI surveillance: (a) saxagliptin will be compared to each of three other types of antihyperglycemic drugs: pioglitazone, second generation sulfonylureas, and long-acting insulin, (b) sitagliptin will be compared to each of the same 3 comparators, and (c) the two DPP-4 inhibitors, saxagliptin and sitagliptin, will be compared to each other.
- Patients hospitalized during the 60-day period before the index dispensing date will be excluded from the analysis if the principal discharge diagnosis was AMI or HF (due to the high potential for residual confounding during follow-up that would start <60 days after such an event).
- Each pairwise comparison will be balanced by 1:1 PS matching, using the same baseline covariates as are used to adjust for potential confounding in AMI surveillance.
- Each pairwise comparison will also be balanced by a DRS, in a separate analysis from the PS-matched analysis. Whereas the AMI surveillance calibrated the DRS by risk of AMI in the underlying study population (at each Data Partner), the HF analysis will calibrate the DRS by risk of hHF in the underlying study population (at each Data Partner).
- Subgroup analyses will focus on the patients who are at relatively high risk of hHF, as indicated by decile of the DRS. This subgroup will be comprised of the decile or deciles that are predicted to yield an average incidence of hHF in new users of the study drugs that is nearest to the incidence observed in the placebo arm of the SAVOR-TIMI trial: 13.6 per 1,000 person-years.
- We will also assess whether relative risk estimates differ across Data Partners and over time (time-on-study-drug and calendar time).

Sample size considerations and power. In pairwise comparisons of saxagliptin with the comparator drugs that were conducted in June 2013 (for Look 5) the new users of saxagliptin contributed the following person-years of follow-up to the comparison with: sulfonylurea (~11,000), pioglitazone (~14,000), long-acting insulin (~18,000), and sitagliptin (~18,000). Simulated sitagliptin surveillance included many more new users and a longer study period: we had about 6.6 times more person-years in sitagliptin new users than in saxagliptin new users eligible for the pairwise comparisons. In 2014, more new users and follow-up are available: at least 20% more person-years can be expected for the saxagliptin comparisons and 10% more for the sitagliptin comparisons. Using 1:1 PS matching we expect each saxagliptin comparison to have totals (counting the comparator-users as well as the saxagliptin users) of ~24,000 to ~44,000 person-years, and the sitagliptin comparisons to have totals of ~152,000 to ~262,000 person-years of follow-up.

The incidence of hHF in SAVOR-TIMI 53 was 3.5% in saxagliptin group versus 2.8% in the placebo group, amounting to 17.1 and 13.6 events per 1,000 person-years, respectively. The patients in the trial were older than Mini-Sentinel users of DPP-4 inhibitors: mean age 65 years versus 57 (saxagliptin) and 58 (sitagliptin); and a higher percentage of patients in the trial had a history of HF at baseline: 12.8% versus 4.7% (saxagliptin) and 6.5% (sitagliptin). Thus, it may be reasonable to expect outcome events in the Mini-Sentinel cohorts to occur at lower rates. The incidence of AMI in Mini-Sentinel saxagliptin users is about 6 per 1,000 person-years compared to 15 per 1,000 in the trial. A preliminary estimate is that rates of hHF events in users of the study drugs may range from roughly 6 to 9 per 1,000 person-years.

Given baseline incidence of 6 per 1,000 person-years, the proposed design would provide 80% power to detect relative risks of 1.15 to 1.20 in the 3 pairwise sitagliptin comparisons, and 1.38 to 1.53 in the 3 saxagliptin comparisons, using 2-sided tests with alpha=0.05 or 1-sided alpha=0.025. If our threshold for a signal is 1-sided alpha=0.05 rather than 1-sided alpha=0.025, then our power to detect these relative risks would be 88% rather than 80%.

If baseline incidence of hHF outcomes in new users of the study drugs is 9 per 1,000 person-years rather than 6, then the proposed design would provide 80% power to detect relative risks of 1.12 to 1.16 in the 3 pairwise sitagliptin comparisons (rather than 1.15 to 1.20 if baseline incidence is 6 per 1,000 person-years), and the least detectable relative risks for the saxagliptin comparisons would be 1.30 to 1.42 (rather than 1.38 to 1.53).

In subgroup analyses, focusing on highest quintile of the DRS among patients in the stratum with prior CVD, and assuming that hHF risk in this high-risk quintile will be 25 per 1,000 person years (which may be expected if prior cardiovascular disease and the hHF DRS predict the HF outcome as well as prior cardiovascular disease and the AMI DRS predict the AMI outcome): we would have 80% power to detect a relative risk of about 1.35, 1.31 and 1.26 for sitagliptin versus the sulfonylureas, pioglitazone and long-acting insulin, respectively. Similarly, relative risks of about 1.98, 1.84, and 1.70 would be detectable in analyses of saxagliptin versus these 3 comparators, respectively.

It should be noted that patients with prior HF are excluded from comparisons of the DPP-4 inhibitors with pioglitazone (because prior HF can be a contraindication for pioglitazone); this reduces the expected incidence of hHF, and consequently reduces the power of the pioglitazone comparisons (below the estimates suggested above).

IX. REFERENCES

1. Nathan DM, Buse JB, Davidson MB, et al. Medical management of hyperglycemia in type 2 diabetes: a consensus algorithm for the initiation and adjustment of therapy: a consensus statement of the American Diabetes Association and the European Association for the Study of Diabetes. Diabetes care 2009;32:193-203.

2. Onglyza. In: package insert. In. Princeton, NJ: Bristol-Myers Squibb Company; 2009.

3. Frederich R, Alexander JH, Fiedorek FT, et al. A systematic assessment of cardiovascular outcomes in the saxagliptin drug development program for type 2 diabetes. Postgraduate medicine 2010;122:16-27.

4. U.S. Food and Drug Administration. Guidance for Industry: Diabetes Mellitus — Evaluating Cardiovascular Risk in New Antidiabetic Therapies to Treat Type 2 Diabetes. In.

5. Scirica BM, Bhatt DL, Braunwald E, et al. The design and rationale of the saxagliptin assessment of vascular outcomes recorded in patients with diabetes mellitus-thrombolysis in myocardial infarction (SAVOR-TIMI) 53 study. Am Heart J 2011;162:818-25 e6.

6. Ray WA. Evaluating medication effects outside of clinical trials: new-user designs. American journal of epidemiology 2003;158:915-20.

7. Schneeweiss S. A basic study design for expedited safety signal evaluation based on electronic healthcare data. Pharmacoepidemiol Drug Saf 2010;19:858-68.

8. Rodbard HW, Jellinger PS, Davidson JA, et al. Statement by an American Association of Clinical Endocrinologists/American College of Endocrinology consensus panel on type 2 diabetes mellitus: an algorithm for glycemic control. Endocr Pract 2009;15:540-59.

9. Ray WA. Population-based studies of adverse drug effects. N Engl J Med 2003;349:1592-4.

10. Glynn RJ, Schneeweiss S, Sturmer T. Indications for propensity scores and review of their use in pharmacoepidemiology. Basic & clinical pharmacology & toxicology 2006;98:253-9.

11. Joffee MM, Rosenbaum PR. Invited commentary: propensity scores. American journal of epidemiology 1999;150:327-33.

12. Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. Biometrika 1983;70:41-55.

13. Arbogast PG, Ray WA. Use of disease risk scores in pharmacoepidemiologic studies. Stat Methods Med Res 2009;18:67-80.

14. Cadarette SM, Gagne JJ, Solomon DH, Katz JN, Sturmer T. Confounder summary scores when comparing the effects of multiple drug exposures. Pharmacoepidemiol Drug Saf 2010;19:2-9.

15. Cepeda MS, Boston R, Farrar JT, Strom BL. Comparison of logistic regression versus propensity score when the number of events is low and there are multiple confounders. American journal of epidemiology 2003;158:280-7.

16. Brookhart MA, Schneeweiss S, Rothman KJ, Glynn RJ, Avorn J, Sturmer T. Variable selection for propensity score models. American journal of epidemiology 2006;163:1149-56.

17. Brookhart MA, Sturmer T, Glynn RJ, Rassen J, Schneeweiss S. Confounding control in healthcare database research: challenges and potential approaches. Medical care 2010;48:S114-20.

18. Peduzzi P, Concato J, Kemper E, Holford TR, Feinstein AR. A simulation study of the number of events per variable in logistic regression analysis. Journal of clinical epidemiology 1996;49:1373-9.

19. Kiyota Y, Schneeweiss S, Glynn RJ, Cannuscio CC, Avorn J, Solomon DH. Accuracy of Medicare claims-based diagnosis of acute myocardial infarction: estimating positive predictive value on the basis of review of hospital records. Am Heart J 2004;148:99-104.

20. Graham DJ, Ouellet-Hellstrom R, MaCurdy TE, et al. Risk of acute myocardial infarction, stroke, heart failure, and death in elderly Medicare patients treated with rosiglitazone or pioglitazone. JAMA : the journal of the American Medical Association 2010;304:411-8.

21. Lanes SF, de Luise C. Bias due to false-positive diagnoses in an automated health insurance claims database. Drug Saf 2006;29:1069-75.

22. Schneeweiss S, Patrick AR, Sturmer T, et al. Increasing levels of restriction in pharmacoepidemiologic database studies of elderly and comparison with randomized trial results. Medical care 2007;45:S131-42.

23. Fireman B, Lee J, Lewis N, Bembom O, van der Laan M, Baxter R. Influenza vaccination and mortality: differentiating vaccine effects from bias. American journal of epidemiology 2009;170:650-6.

24. DeMets DL, Lan KK. Interim analysis: the alpha spending function approach. Statistics in medicine 1994;13:1341-52; discussion 53-6.

25. Lan KK, DeMets DL. Changing frequency of interim analysis in sequential monitoring. Biometrics 1989;45:1017-20.

26. Lan KKG, DeMets DL. Discrete sequential boundries for clinical trials. Biometrika 1983;70:659-63.

27. Choma NN, Griffin MR, Huang RL, et al. An algorithm to identify incident myocardial infarction using Medicaid data. Pharmacoepidemiol Drug Saf 2009;18:1064-71.

28. Petersen LA, Wright S, Normand SL, Daley J. Positive predictive value of the diagnosis of acute myocardial infarction in an administrative database. Journal of general internal medicine 1999;14:555-8.

29. Rosamond WD, Chambless LE, Sorlie PD, et al. Trends in the sensitivity, positive predictive value, false-positive rate, and comparability ratio of hospital discharge diagnosis codes for acute myocardial infarction in four US communities, 1987-2000. American journal of epidemiology 2004;160:1137-46.

30. Solomon DH, Schneeweiss S, Glynn RJ, et al. Relationship between selective cyclooxygenase-2 inhibitors and acute myocardial infarction in older adults. Circulation 2004;109:2068-73.

31. Varas-Lorenzo C, Castellsague J, Stang MR, Tomas L, Aguado J, Perez-Gutthann S. Positive predictive value of ICD-9 codes 410 and 411 in the identification of cases of acute coronary syndromes in the Saskatchewan Hospital automated database. Pharmacoepidemiol Drug Saf 2008;17:842-52.

32. Wahl PM, Rodgers K, Schneeweiss S, et al. Validation of claims-based diagnostic and procedure codes for cardiovascular and gastrointestinal serious adverse events in a commercially-insured population. Pharmacoepidemiol Drug Saf 2010;19:596-603.

33. Hennessy S, Schelleman H, Daniel GW, et al. Cardiovascular safety of ADHD medications: rationale for and design of an investigator-initiated observational study. Pharmacoepidemiol Drug Saf 2010;19:934-41.

34. Kachroo S, Jones N, Reynolds MW. Systematic literature review for evaluation of myocardial infarction. Final report prepared for the Foundation for the National Institutes of Health via the Observational Medical Outcomes Partnership (OMOP). Lexington, MA: United Biosource Corporation; 2009.

35. Lux LJ, West S. Systematic Evaluation of Health Outcome of Interest Definitions in Observational Studies and Clinical Definitions for the Observational Medical Outcomes Partnership: Myocardial Infarction Report. Final Report prepared for the Foundation for the National Institutes of Health via the Observational Medical Outcomes Partnership (OMOP). Research Triangle Park, NC: Research Triangle Institute; 2009.

36. Miettinen OS. Stratification by a multivariate confounder score. American journal of epidemiology 1976;104:609-20.

37. Ishani A, Herzog CA, Collins AJ, Foley RN. Cardiac medications and their association with cardiovascular events in incident dialysis patients: cause or effect? Kidney Int 2004;56:1017-25.

38. Romano PS, Chan BK. Risk-adjusting acute myocardial infarction mortality: are APR-DRGs the right tool? Health services research 2000;34:1469-89.

39. Yeh RW, Sidney S, Chandra M, Sorel M, Selby JV, Go AS. Population trends in the incidence and outcomes of acute myocardial infarction. The New England journal of medicine 2010;362:2155-65.

40. Li L, Kulldorff M. A conditional maximized sequential probability ratio test for pharmacovigilance. Statistics in medicine 2010;29:284-95.

41. Kulldorff M, Davis RL, Kolczak M, Lewis E, Lieu T, Platt R. A maximized sequential probability ratio test for drug and vaccine safety surveillance. Sequential Analysis 2011;30:58-78.

42. Pocock S. Group sequential methods in the design and analysis of clinical trials. Biometrika 1977;64:191-9.

43. Scirica BM, Braunwald E, Bhatt DL. Saxagliptin, alogliptin, and cardiovascular outcomes. N Engl J Med 2014;370:483-4.

44. Saczynski JS, Andrade SE, Harrold LR, et al. A systematic review of validated methods for identifying heart failure using administrative data. Pharmacoepidemiol Drug Saf 2012;21 Suppl 1:129-40.