MINI-SENTINEL SURVEILLANCE PLAN

MINI-SENTINEL PROSPECTIVE ROUTINE OBSERVATIONAL MONITORING PROGRAM TOOLS (PROMPT): RIVAROXABAN SURVEILLANCE

Prepared by: Ryan Carnahan, PharmD, MS, BCPP,[1] Joshua J. Gagne, PharmD, ScD,[2] Jennifer Nelson, PhD,[3] Bruce Fireman, MA,[4] Shirley Wang, PhD, [2] Azadeh Shoaibi, MS, MHS,[5] Marsha Reichman, PhD,[6] Rongmei Zhang, PhD,[7] Mark Levenson, PhD,[7] David Graham, MD, MPH,[6] Ram Tiwari, PhD,[6] Mary Ross Southworth, PharmD,[6] Patrick Archdeacon, MD, MS,[5] Aloka Chakravarty, PhD,[7] Margie Goulding, PhD,[6] Rima Izem, PhD,[8] Jeffrey Brown, PhD,[8] Candace Fuller, PhD,[8] Catherine Rogers, MPH,[8] Darren Toh, ScD,[8] Elizabeth Chrischilles, PhD, MS[1]

Author Affiliations: 1. Department of Epidemiology, University of Iowa College of Public Health, Iowa City IA 2. Division of Pharmacoepidemiology and Pharmacoeconomics, Department of Medicine, Brigham and Women's Hospital, Boston, MA 3. Biostatistics Unit, Group Health Research Institute and Department of Biostatistics, University of Washington, Seattle, WA 4. Kaiser Permanente Northern California, Oakland, CA 5. Office of Medical Policy, Center for Drug Evaluation and Research, FDA, Silver Spring, MD 6. Office of Surveillance and Epidemiology, Center for Drug Evaluation and Research, FDA, Silver Spring, MD 7. Division of Biometric VII, Office of Biostatistics, Office of Translation Sciences, FDA, Silver Springs MD 8. Office of Biostatistics, Center for Drug Evaluation and Research, FDA, Silver Spring, MD 8. Department of Population Medicine, Harvard Medical School and Harvard Pilgrim Health Care Institute, Boston, MA

Version 3: October 29, 2015
Prior versions:
Version 1: January 13, 2014
Version 2: March 28, 2014

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

History of Modifications

Version	Date	Modification	By
V2	3/28/2014	• Modified cohort inclusion criteria to reflect when rivaroxaban was approved for atrial fibrillation • Expanded and modified list of potential covariates • Expanded description of analysis plan • Provided minor edits and clarifications in response to public comments	Mini-Sentinel Rivaroxaban Surveillance Team

V3	10/29/2015	• The workgroup updated the plan to reflect refinements made to the sequential monitoring tools that were made after the first interim analysis:
		  • Clarified the upper limit for variable ratio matching
		  • Added updates about the tools and associated new specifications
		  • Updated the anticoagulant list to include edoxaban, which was approved after the first interim analysis
		• The workgroup removed two CPT codes that were erroneously listed as exclusion criteria.
		• The workgroup prepared Addendum 1, which summarizes changes to the surveillance plan after the first interim analysis
			Mini-Sentinel Rivaroxaban Surveillance Team

This report is modified periodically to document all major changes made during surveillance plan implementation.

Mini-Sentinel Surveillance Plan

Mini-Sentinel Prospective Routine Observational Monitoring Program Tools (PROMPT): Rivaroxaban Surveillance

Table of Contents

I. FDA REQUEST

FDA has requested prospective routine observational monitoring of rivaroxaban for the safety outcomes of ischemic stroke, intracranial hemorrhage, and gastrointestinal (GI) bleeding in patients with atrial fibrillation (AF).

The Mini-Sentinel pilot has developed the capacity to perform prospective routine observational active surveillance of newly approved medical products as experience with these products accumulates in near real time. These Prospective Routine Observational Monitoring Program Tools (PROMPT) will enable FDA to assess the occurrence of a fixed number of pre-specified health outcomes of interest that may occur in association with use of newly approved medical products. The program's emphasis is on signaling of potential excess risks rather than formal assessment of causal relationships. As such, a small number of surveillance methods will be implemented through a library of programs that can be modified to accommodate specific agents, outcomes, populations, and time periods for evaluation. This approach requires that signals be carefully followed up to understand the explanation for the finding. The team has developed an overall users' guide detailing the process for utilizing the current system.

II. DESIGN DETERMINATION

This assessment will employ the PROMPT: Cohort Matching tool. This rapid assessment tool identifies new users of the product of interest, new users of one comparator product, and one outcome. At present, multiple outcomes and additional comparators are handled by running the program multiple times. Ratio and difference measures of effect can be estimated with this design. For three outcomes and one comparator, three runs of the tool will be required. The PROMPT: Cohort Matching tool is appropriate for the semi-automated surveillance question and its use is consistent with recommendations from the Taxonomy PROMPT Selection Tool. After the first interim analysis, the Mini-Sentinel suite of routine querying system was enhanced to include a Cohort Identification and Descriptive Analysis (CIDA) tool as well as the Propensity Score Matching (PSM) tool, both of which are used for this assessment.

III. COHORT IDENTIFICATION

The surveillance activity will assess one primary cohort—patients with atrial fibrillation (AF) (patients with one or more outpatient or inpatient diagnosis of atrial fibrillation or atrial flutter [ICD-9-CM codes 427.31 or 427.32] in the 183 days prior to initiation of rivaroxaban or comparators). Codes for atrial flutter are included because a large fraction of these patients have both AF along with atrial flutter, and per national clinical practice guidelines they should be treated similarly. Both codes will be added to the propensity score pre-specified variables (see below) to help ensure an equal balance. Eligible patients must meet the atrial fibrillation/flutter definition based on data in the 183 days before initiation of rivaroxaban or the comparator. Exclusions for mitral stenosis or mechanical heart valve, joint replacement, renal dialysis, or a history of renal transplant will be made (**Table 1**). Rivaroxaban is not indicated for valvular AF and should not be used for patients with poor renal function. Patients with joint replacement are excluded because this evaluation is only for the AF indication.

Table 1. Diagnosis and procedure codes for inclusion and exclusion conditions of the cohort.

DESCRIPTION	CODE TYPE AND CODE	INCLUDE/EXCLUDE
Atrial Fibrillation	ICD9(D): 427.31	Include
Atrial Flutter	ICD9(D):427.32	Include
Codes Suggestive of Chronic Dialysis	ICD9(D): 792.5x, V56.2x ICD9(P): 39.95, 54.98 CPT: 90935, 90937, 90945, 90947, 99512	Exclude
Kidney replaced by transplant	ICD9(D): V42.0x, 996.81 ICD9(P): 55.6x CPT: 50340, 50360, 50365, 50370, 50380	Exclude
Mitral stenosis or mechanical heart valve	ICD9(D): 394.0x, 394.2x, 396.0x, 396.1x,746.5x, 996.02, 996.71 ICD9(P): 35.20, 35.22, 35.23, 35.24, 35.97 CPT4: 33405, 33430, 33420, 33422, 33425-33427, 92987	Exclude
Recent joint replacement/ arthroplasty surgery	CPT4: 01214, 01215, 01402, 24363, 27130, 27132, 27134, 27137, 27138, 27447, 27486, 27487 ICD9(P): 00.70 - 00.77, 00.80 - 00.87, 81.51 - 81.57, 81.59, 81.80, 81.81, 81.84, 81.88, 81.97	Exclude
x= All codes beginning with the values noted will not be included		

IV. EXPOSURES

Rivaroxaban was approved for AF on November 4, 2011. The cohort will include new users of rivaroxaban (15 mg or 20 mg) or the comparator, warfarin, whose first exposure after at least 183 days without anticoagulants exposure occurred on or after November 1, 2011.

While there is interest in comparison of rivaroxaban among switchers from warfarin, the reason for switching may predict poor outcomes and therefore will not be included in the surveillance activity.[1]

The recommended dose of rivaroxaban for AF is higher than for orthopedic surgery-related DVT prophylaxis (20 mg vs. 10 mg, respectively, and 15 mg for AF patients with renal impairment). This surveillance will only include patients who receive the 15 mg or 20 mg formulations. This restriction does not exclude those with an AF diagnosis who might have received rivaroxaban for acute venous thromboembolism, as they receive 15 mg and 20 mg products.

Dabigatran was considered as an additional active comparator. Its advantage over warfarin is that the treatment episode gap would be comparable (for warfarin, dose adjustment could result in misclassification of warfarin exposure). However there would likely be fewer dabigatran new users than warfarin new users available for matching.[2] A Mini-Sentinel one-time protocol-based assessment is currently underway for dabigatran (with warfarin comparator group) which could provide a platform for a future dabigatran-rivaroxaban comparison.

[1] Discussed by the Surveillance Team. It was agreed that the concern for confounding was too great for this to be a primary interest.
[2] Discussed by the Surveillance Team. It may best be addressed in a one-time look at the end of monitoring or a protocol-based assessment.

Oral rivaroxaban and warfarin are readily identified from National Drug Codes (NDCs). Effectiveness and safety of warfarin is highly dependent on the level of anticoagulation intensity, but this information is not currently available for most patients in the Mini-Sentinel Distributed Database (MSDD). However, comparison with real-world use of warfarin is of interest clinically.

The CIDA and PSM tools will create new treatment episodes of rivaroxaban or the comparator (warfarin). From the treatment episode, the module will create "days at risk", during which events of interest are assessed. Treatment episodes and days at risk are created using several different parameters and assumptions, described here.

A treatment episode of rivaroxaban (or warfarin) will be identified if three criteria are all met: (1) the member has been enrolled in medical and drug coverage for the prior 183 days; (2) the member has not been exposed to any anticoagulant (rivaroxaban, warfarin, dabigatran, apixaban, and edoxaban[3]) in the prior 183 days; and (3) this is the first exposure to rivaroxaban (or warfarin) for the member in the entire query period. Only one treatment episode per member will be identified.

When creating treatment episodes, a stockpiling algorithm is automatically applied. The stockpiling algorithm accounts for the fact that members may refill their drug prescriptions before the end of days supply of the prior prescription. For example, if a member receives a 30-day dispensing for rivaroxaban on January 1st and then receives a second 30-day dispensing for rivaroxaban on January 20th, the stockpiling algorithm will adjust the second dispensing so that it starts on January 31st, after the first dispensing has been used in full. The treatment episode will thus be 60 days in total, through March 1st (assuming February has 28 days).

The analysis will also make use of a 7 day episode gap when creating treatment episodes. The episode gap is the maximum number of days of interrupted days supply that can be found between two claims of the same query group to be considered part of the same treatment episode. If a gap of treatment between two claims of the same treatment is smaller than or equal to the allowable gap, the algorithm "bridges" these two claim periods to build a continuous treatment episode. If, however, the allowable gap is exceeded between the same two claims, the treatment episode ends at the end of the first claim. The allowable gap is assessed after claim service dates are adjusted by the stockpiling algorithm.

The analysis will also make use of a 7 day exposure extension period when examining days at risk. This parameter extends the treatment episode by a set number of days (in this case 7). For example, if a treatment episode ends on December 31st, 2010 and an episode extension of 7 days is allowed, an event occurring between January 1st, 2011 and January 7th, 2011 will be considered valid (i.e., exposed event).

For each treatment, the number of days "at risk" for an exposed event are reported by the tool. Days at risk are based on the length of treatment episode; days at-risk end with the length of the treatment episode or upon identification of an exposed event of interest. By including NDCs for the anticoagulant drugs this will censor patients' follow-up at the date of dispensation of a non-index anticoagulant. This is specified using the CIDA Cohort Codes file.

[3] The FDA approved edoxaban on January 8, 2015.

V. OUTCOMES OF INTEREST

The primary outcomes of interest are (a) ischemic stroke; (b) intracranial hemorrhage; and (c) GI bleeding (**Table 2**).[4]

Positive predictive values for stroke and GI bleeding for any position inpatient discharge diagnosis codes are generally acceptable (>80%), though lower than when a first position discharge diagnosis is required.[5,6,7] The position designation for discharge diagnosis codes in the Mini-Sentinel Common Data Model can be coded as primary, secondary, unable to classify, or missing. After several discussions the surveillance planning team decided to include primary and non-secondary codes.[8]

Table 2. Inpatient Diagnosis codes for serious thromboembolic events and major hemorrhage

OUTCOME	ICD-9 CODES
Ischemic Stroke	433.x1, 434.x1, 436
Intracranial hemorrhage	Hemorrhagic stroke: 430 (subarachnoid), 431 (intracerebral) Other ICH (per Mini-Sentinel dabigatran protocol): 432.0 (nontraumatic extradural), 432.1 (subdural), 432.9 (unspecified ICH), 852.0x (subarachnoid after injury*), 852.2x (subdural after injury*), 852.4x (extradural after injury*), 853.0 (other and unspecified ICH after injury*) *="without mention of open intracranial wound"
GI bleeding	Gastroduodenal site: 530.21, 531.0x, 531.1x, 531.2x, 531.4x, 531.6x, 532.0x, 532.1x, 532.2x, 532.4x, 532.6x, 533.1x, 533.2x, 533.4x, 533.6x, 534.0x, 534.1x, 534.2x, 534.4x, 534.6x, 535.01, 535.11, 535.21, 535.31, 535.41, 535.51, 535.61, 537.83, 537.84, 562.02, 562.03, 562.12, 562.13, 569.86 Esophageal site: 456.0, 456.20, 530.21, 530.7, 530.82 Upper GI Unspecified: 534.0x, 534.1x, 534.2x, 534.4x, 534.6x, 562.02, 562.03, 578.0 Lower GI Site: 455.2, 455.5, 455.8, 562.02, 562.03, 562.12, 562.13, 568.81, 569.3x Unspecified GI Site: 533.0x,533.1x, 533.2x, 533.4x, 533.6x, 568.81 (add to Mini-Sentinel HOI algorithm), 578.x, 569.85, 569.86

VI. PRE-DEFINED COVARIATES

The CIDA and PSM tools allow the requestor to pre-define covariates and return estimates from cohorts matched on propensity scores including three sets of covariates: (1) pre-defined covariates only; (2) pre-defined covariates and empirically selected covariates; and (3) empirically selected covariates only. This assessment will use both pre-defined and empirically selected covariates. All pre-defined variables in **Table 3** will be forced into the propensity score models. In each new monitoring period, the

[4] Discussed by the Surveillance Team. There was clear consensus that serious events were of primary interest, that thrombosis be separated from bleeding events, and that intracranial hemorrhage be examined separately from other major bleeding. Whether to include several serious types of major bleeding events in one outcome definition was discussed. The large majority of major bleeding events will be GI bleeds. It may be best to examine intraocular, intraspinal, retroperitoneal, and perhaps other serious but uncommon bleeding events in a one-time look at the end of monitoring.

[5] Andrade et al, Pharmacoepidemiol Drug Saf 2012;21(Suppl 1):100-28.

[6] Schelleman et al, Am J Med 2010;123(2):151-7.

[7] Go et al, JAMA. 2003;290(20):2685-2692.

[8] Mini-Sentinel and FDA have explored this issue. This decision is based on those investigations.

propensity score is estimated on all eligible new users of rivaroxaban and warfarin from all prior monitoring periods up to and including the current period. Only new users identified in the current monitoring period are matched using the most recent propensity score model and all prior propensity score matches remain in the analysis. The CIDA and PSM tools will retain matches for the duration of the surveillance activity. The follow-up for outcome events will be updated so that we can use late-arriving data and corrected data that pertains to outcome events. Information about outcome events is expected to be stable, but if changes occur it is important to the validity of our findings that we use the corrected data.

Table 3. List of potential confounders for inclusion as pre-defined covariates in cohort matching program.

	BROAD CATEGORIES OF CONFOUNDERS	SPECIFICS
1	Risk factors for bleeding	• Prior intracranial bleed without open intracranial wound (same as outcome but any setting) • Prior gastrointestinal bleed • Other GI ulcer disease • Other intracranial bleed • Prior other bleed • Alcoholism • Advanced liver disease • Coagulation defects
2	Risk factors for ischemic stroke	• Atrial fibrillation • Atrial flutter • Prior ischemic stroke • Prior transient ischemic attack • Other ischemic cerebrovascular disease • Non-specific cerebrovascular symptoms • Other arterial embolism • Prior VTE or phlebitis • VTE risk NOS indicators • Prior central venous thrombosis • Heart failure/cardiomyopathy • Hypertension • Diabetes • Hyperlipidemia • Prior myocardial infarction • Prior acute coronary syndrome • Prior percutaneous coronary intervention • Prior coronary artery bypass graft • Peripheral vascular disease • Cancer (non-metastatic) • Metastatic cancer • Central venous catheter • Other venous catheter • Major trauma potentially causing prolonged immobilization • Selected surgeries* • End stage renal disease • Tobacco use**

	BROAD CATEGORIES OF CONFOUNDERS	SPECIFICS
3	Measures of overall health status, including of frailty	• # of distinct medications • # of prior hospitalizations • # of prior outpatient visits • Combined comorbidity score • Renal disease (other than end-stage) • Use of home oxygen • Wheelchair use • Walker use • Cane use • Commode chair use • Osteoporotic fracture • Falls

4	Medications (from outpatient pharmacy claims)	• Cardiovascular and antidiabetic agents
		  • Statins
		  • Non-statin lipid lowering agents
		  • ACE inhibitors
		  • Angiotensin receptor blockers
		  • Aldosterone receptor antagonists
		  • Beta blockers
		  • Calcium channel blockers
		  • Diuretics
		  • Other antihypertensives
		  • Antianginal vasodilators
		  • Anti-arrhythmic agents
		  • Oral antidiabetic agents
		  • Insulin
		  • Estrogens
		  • Progestins
		• Medications that increase bleeding risk
		  • Aspirin (to the extent captured)
		  • Antiplatelet agents
		  • Prescription NSAIDs
		  • COX-2 inhibitors
		  • SSRIs or SNRIs
		  • Heparin, low molecular weight heparin, or fondaparinux
		• Medications that may protect from bleeding
		  • H2 antagonists
		  • Proton pump inhibitors
		  • PPI/antibiotic combination products for H. Pylori
		• Interacting prescription medications described in section 5.6 of the label
		  • CYP3A4 and P-glycoprotein inhibitors: ketoconazole, itraconazole, lopinavir/ritonavir, ritonavir, indinavir, and conivaptan
		  • CYP3A4 and P-glycoprotein inducers:

	BROAD CATEGORIES OF CONFOUNDERS	SPECIFICS
		carbamazepine, phenytoin, and rifampin
	* Open gynecologic or urologic surgery, bariatric surgery, intracranial neurosurgery, hip or knee surgery, and spinal cord surgery ** De novo broad algorithm recommendations for smoking or other tobacco use are from an ongoing working group effort to identify algorithms for cohorts within Mini-Sentinel.	

VII. STATISTICAL ANALYSES

A. OVERVIEW

To detect potential safety signals more rapidly than would be possible with a single retrospective evaluation, sequential monitoring will be conducted. A group sequential design will be used to repeatedly compare the risk of each adverse event between new users of rivaroxaban and new users of warfarin over time as new users and additional follow-up data accumulate within the MSDD. Repeated analyses were planned to be approximately quarterly, which would have yielded about five analyses during the surveillance period. As of the writing of the third version of this surveillance plan, the Surveillance Team deliberated and decided to skip three planned interim looks given that more rivaroxaban use was observed than originally expected. The original choice of testing frequency was informed by the prior experience of the saxagliptin customized surveillance activity and was designed to balance key safety, practical, and statistical criteria[9]. From a safety perspective, it is often desirable to conduct more frequent tests in order to either identify potential signals as rapidly as possible or provide reassurance that there is no evidence for a major safety concern. However, each time an analysis is conducted, resources (which are not unlimited) must be devoted to oversee and manage the receipt of the data, and to review, troubleshoot, interpret, and act on the results. Thus, for this evaluation, quarterly testing (5 times) was originally selected as the most frequent rate of testing that would both provide potentially valuable new information at each analysis and also be practically feasible with available resources.

At each pre-specified analysis time point, new rivaroxaban users will be variable ratio matched to new warfarin users based on a propensity score that is estimated within each Data Partner. The upper bound of the variable ratio matching will be set to 1:10, which is a pre-defined upper limit of the PSM tool.

A Cox regression model with time-since-drug-initiation as the time scale, stratified by Data Partner and matched set will be used to estimate hazard ratios (HRs) that compare rivaroxaban and warfarin users. A two-sided test based on the standardized Wald statistic from the Cox regression analysis (i.e., log(HR)/sqrt(var(log(HR))) will be computed using model-based standard errors. This observed statistic will be compared to a preset signaling threshold that was initially planned to be 2.37 (on the scale of the standardized test statistic – the log hazard ratio divided by its standard error) at all analyses (i.e., a Pocock boundary specified using the unifying family), designed to hold the overall Type 1 error rate (i.e., alpha) across all tests at 0.05, and computed using large sample assumptions (using PROC SEQDESIGN in SAS).

[9] Mini-sentinel Medical Product Assessment: A Protocol For Active Surveillance Of Acute Myocardial Infarction In Association With Use Of Anti-Diabetic Agents. 2015. Available at http://www.mini-sentinel.org/work_products/Assessments/Mini-Sentinel_AMI-and-Anti-Diabetic-Agents_Protocol.pdf

If the statistic exceeds the threshold at any analysis, the null hypothesis is rejected, a signal is generated, and signal follow-up activities will be initiated for the outcome that has had a signal generated. For the remaining outcomes where a signal has not been generated, the planned sequential testing will continue. For as long as surveillance continues, planned analyses will be conducted for all 3 outcomes at all interim and final stages of analysis, yielding comparable analyses of all 3 outcomes (even though the threshold for sequential testing would lose its rationale after a signal for the outcome that signaled).

Analysis will not account for multiple testing across the three separate outcomes of interest. A variety of diagnostic analyses will also be conducted to help assess the validity of each analysis, for example, to examine overlap in the propensity score distributions of rivaroxaban and warfarin users.

B. POWER AND SAMPLE SIZE

Given the desired overall Type 1 error rate for each outcome, minimum detectable HR of interest, the planned sequential testing schedule, and the background incidence of each outcome among the comparator group (i.e., warfarin new users), the total number of rivaroxaban users required to achieve a specified level of power can be computed. Preliminary sample size calculations for each outcome are shown in **Table 4** using the following inputs: 2-sided test, Type 1 error = 0.05; Power = 0.80; a range of minimum HRs of interest to detect [1.5-3.0]; five tests conducted with an expected cumulative proportion of total information at each analysis of 0.35, 0.47, 0.62, 0.80, and 1.0, respectively; and background incidence rates of 17, 7, and 10 per 1,000 person-years for ischemic stroke, intracranial hemorrhage and GI bleed, respectively, which were estimated from warfarin initiators in the MSDD. Additional details of sample size and power estimations are provided in Appendix A. Power calculations also assume that the average expected follow-up time is six months after drug initiation and that an average of three warfarin comparators will remain matched (i.e., not lost to follow-up) to each rivaroxaban user for analyses. HRs indicative of potential safety problems with rivaroxaban are our foremost interest so we considered the sample sizes required to detect several levels of relative risk above 1.0 for each outcome, as shown in **Table 4**.

Table 4. Number of rivaroxaban new-users that are needed at final look to detect relative risks of 1.5, 2.0, or 3.0 for each of the 3 safety outcomes, (Assuming 5-looks (Pocock), 1:3 matching, 2-sided alpha = .05, power=.8)

MINIMUM DETECTABLE RELATIVE RISK	INTRACRANIAL HEMORRHAGE	GASTROINTESTINAL BLEEDING	ISCHEMIC STROKE
1.5	17,073	11,951	7,030
2.0	4,961	3,473	2,043
3.0	1,536	1,075	632

C. PRACTICAL CONSIDERATIONS

There are several practical constraints that may influence the implementation of this surveillance plan. First, the speed of uptake and the pattern of diffusion of rivaroxaban in the population of interest is uncertain. Second, information on new users is only available for analysis when each Data Partner refreshes their data, which occurs on a quarterly basis that is not synchronized across all Data Partners. As a result, the actual amount of information (i.e., number of new users) available for analysis each quarter may be more or less than assumed. It also is uncertain what overall duration of calendar time will be necessary to accrue the ideal total number of rivaroxaban users required to achieve power to detect HRs of regulatory interest. To address these issues, feasibility data will be pulled directly prior to

the start of surveillance to refine the expected uptake assumptions and minimize discrepancies between the amounts of assumed versus actual information that is analyzed each quarter. If observed uptake does not occur according to our expectations, then we will adjust our planned signaling thresholds to reflect the actual uptake pattern and adjust the schedule of interim analyses so that we maintain the integrity of our Type 1 error spending plan. If uptake is especially slow and it becomes apparent that the calendar time required to accrue the needed doses will be impractically long (e.g., beyond the planned funding period for the Mini-Sentinel pilot), surveillance may end early due to limited use of rivaroxaban and the need to prioritize surveillance resources to other more pressing questions.

A third issue is that estimation and testing will be inherently less stable, uncertainty will be greater, and power will be limited at early analysis time points that are based upon relatively little information (i.e., few adverse events). This problem diminishes as more new rivaroxaban users accrue since later analyses are based on the cumulatively available information. To address these issues, the first test will not be conducted until at least 7 total events (both groups combined) are observed and descriptive analyses show that balance is achieved with the propensity score matching procedure. If these criteria are not met, then analyses will wait and criteria will be re-checked at the next quarter. Also, a Data Partner will not be included in the overall analysis until the propensity score model successfully converges at that Data Partner. Before data are aggregated and statistical analyses are performed, propensity score distributions for rivaroxaban and warfarin initiators will be examined from each Data Partner.

VIII. DRAFT INPUT SPECIFICATIONS

PROMPT: Cohort Matching tool will be used for this activity. See the input specifications for this tool below (**Table 5**).

Table 5. Input Specifications for Rivaroxaban Cohort Matching Analysis.

INPUT NEEDED FOR PROMPT COHORT MATCHING PROGRAM	SPECIFICATION
Eligibility Information	
Maximum enrollment gap	45 days
Inclusion/exclusion conditions	Include: Age 21+; inpatient or outpatient AF/flutter during baseline(ICD-9 427.31, 427.32); Exclude: on dialysis or history of kidney transplant; valvular disease during baseline; joint replacement during baseline. (See for codes) Medical coverage: Yes Drug coverage: Yes
Exposure Information	
Medical product of interest	Dataset of NDCs for Rivaroxaban limited to 15 mg and 20 mg. Update with each sequential monitoring to capture new doses or bottle sizes.
Comparator of interest	Dataset of NDCs for warfarin. Update

INPUT NEEDED FOR PROMPT COHORT MATCHING PROGRAM	SPECIFICATION
	with each sequential monitoring.
Matching ratio	Variable, upper limit 10
Matching caliper	0.05 on the propensity score scale
New user definition:	
Duration/Wash-out	183 days
Products to define new use	Dataset of all anticoagulant agents (warfarin, dabigatran, apixaban, rivaroxaban, edoxaban)
How incident use will be defined (washout type)	Single (one new use; WASHTYP=SING)
Exposure definition during follow-up:	
Induction period	1 day?; Begin follow-up the day after the index date
Treatment episode gap	Consider consecutive claims with up to 7 day gap to be part of same episode
Episode extension period	7 days
Minimum episode duration	0
Minimum days supply	0
As-treated (default) or intention-to-treat analysis	As-treated
Incident query file	Censor follow-up at the date of dispensation of a different (i.e. non-index) anticoagulant
Covariate Information	
Length of covariate assessment period	183 days preceding index date
Pre-specified covariates:	
Procedures	Dataset with groupings of procedure codes: risk factors for bleeding, ischemic stroke. See **Table 3**.
Conditions	Dataset with groupings of diagnosis codes: risk factors for bleeding, ischemic stroke. See **Table 3**.
Medications	Dataset with NDCs for oral cardiovascular agents, medications that increase bleeding risk, interacting medications. See **Table 3**.
Combined comorbidity score	Include
Health service utilization variables	Include
Subgroups	Prior occurrence of outcome of each corresponding analysis, i.e.: • Prior ischemic stroke • Prior GI bleed • Prior intracranial hemorrhage Age groups (< 65; 65+)
Age groups for propensity score calculation	5-year bands

INPUT NEEDED FOR PROMPT COHORT MATCHING PROGRAM	SPECIFICATION
High-dimensional propensity score options:	
Ranking algorithm	Exposure only (small number of outcome events anticipated)
Covariates considered	100 (default)
Covariates selected	Smaller of 200 or number of initiators (default)
Zero cell correction for association with exposure	Yes (default)
Outcome information	
GI Bleeding:	
Outcome of interest	See **Table 2**. Inpatient claim, primary or non-secondary position
Outcome washout	Allow[1] prior GI bleed
Outcome incidence type	Multiple (WASHTYP=MULT; alternative is MIN which would define incidence using all available data rather than only the washout period)
Intracranial Hemorrhage(ICH)	
Outcome of interest	See **Table 2**. Inpatient claims, primary or non-secondary position.
Outcome washout	Allow[1] prior ICH
Outcome incidence type	MULT
Ischemic stroke:	
Outcome of interest	See **Table 2**. Inpatient claims, primary or non-secondary position
Outcome washout	Allow[1] prior ischemic stroke
Outcome incidence type	MULT
Sequential analysis information	
One vs. two-sided	2-sided
Type 1 error[2]	.05
Desired power[2]	.8
Minimum relative risk of interest	range: 1.5, 2.0, 3.0
Preferred boundary shape over time[2]	Flat
Maximum sample size requirement (total number of events in both exposure groups combined)	See **Table** 4
Number of cases needed to be accrued prior to first look[2]	7
Frequency of testing[2]	Every 3 months

[1]If different incident washouts for the three outcomes were to be used this would result in different cohorts due to excluding different people. In theory, if all inputs are the same across outcomes, the program will obtain the same propensity-matched cohort each time.
[2] Same for each outcome event

IX. PLAN FOR FOLLOW-UP OF SIGNALS

Data will be evaluated descriptively with each look to identify anomalies suggestive of data quality problems. The PROMPT: Cohort Matching tool provides extensive control for confounders. Signals that are detected will be further investigated by verifying data quality and analytic code. Adjustment for additional confounders and testing against additional comparators as well as a variety of pattern evaluation descriptive analyses would next be considered. These will be conducted using the cumulative data. Where applicable, a temporal scan will be conducted, to assess whether the observed adverse outcomes cluster within the specified at-risk time window. Subgroup analyses (e.g., by site, by age group, by risk factor, and by specific diagnosis or procedure code within a given outcome group) will be performed to assess the robustness of the signal. If a signal persists, medical record review will be considered, depending on the nature of the adverse event and the results of the above analyses. Decisions regarding the follow-up of a particular positive or negative signal will be made by FDA in consultation with the Surveillance Team.

X. APPENDIX A. MINI-SENTINEL PROMPT RIVAROXABAN SAMPLE SIZE AND SEQUENTIAL BOUNDARY CALCULATION PLAN

Table 6. Sample size and power for sequential Rivaroxaban Surveillance

Health outcome of interest: Ischemic stroke (expected incidence: 17 per 1,000 person-years, 8.5 per 1,000 users)

5 Looks, approximately quarterly (cumulative proportion of information expected for looks 1, 2, 3, 4, and 5, respectively: .35, .47, .62, .80, and 1.0)

3 warfarin-users per rivaroxaban-user (on average), 2 sided alpha=.05, flat boundary

ROW	MINIMUM DETECTABLE	POWER	THRESHOLD WALD Z (B/SE) FOR SIGNAL	THRESHOLD FOR SIGNAL ON P-VALUE SCALE	RIVAROX USERS ASSUMED AT 1ST LOOK	RIVAROX AES EXPECTED 1ST LOOK UNDER H1	RIVAROX USERS AT LAST LOOK	RIVAROX AES EXPECTED LAST LOOK UNDER H1
1	2.0	0.8	2.37	0.0089	715	12.2	2,043	34.7
2	1.5	0.8	2.37	0.0089	2,460	31.4	7,030	89.6
3	3.0	0.8	2.37	0.0089	221	5.6	632	16.1
4	2.0	0.9	2.37	0.0089	957	16.3	2,735	46.5
5	1.5	0.9	2.37	0.0089	3,294	42.0	9,411	120.0
6	3.0	0.9	2.37	0.0089	296	7.6	846	21.6

Table 7. Sample size and power for sequential Rivaroxaban Surveillance

Health outcome of interest: Intracranial hemorrhage (expected incidence: 7 per 1,000 person-years, 3.5 per 1,000 users)

5 Looks, approximately quarterly (cumulative proportion of information expected for looks 1, 2, 3, 4, and 5, respectively: .35, .47, .62, .80, and 1.0)

3 warfarin-users per rivaroxaban-user (on average), 2 sided alpha=.05, flat boundary

ROW	MINIMUM DETECTABLE	POWER	THRESHOLD WALD Z (B/SE) FOR SIGNAL	THRESHOLD FOR SIGNAL ON P-VALUE SCALE	RIVAROX USERS ASSUMED AT 1ST LOOK	RIVAROX AES EXPECTED 1ST LOOK UNDER H1	RIVAROX USERS AT LAST LOOK	RIVAROX AES EXPECTED LAST LOOK UNDER H1
7	2.0	0.8	2.37	0.0089	1,736	12.2	4,961	34.7
8	1.5	0.8	2.37	0.0089	5,975	31.4	17,073	89.6
9	3.0	0.8	2.37	0.0089	537	5.6	1,536	16.1
10	2.0	0.9	2.37	0.0089	2,324	16.3	6,641	46.5
11	1.5	0.9	2.37	0.0089	7,999	42.0	22,855	120.0
12	3.0	0.9	2.37	0.0089	719	7.6	2,056	21.6

Table 8. Sample size and power for sequential Rivaroxaban Surveillance

Health outcome of interest: GI Bleed (expected incidence: 10 per 1,000 person-years, 5 per 1,000 users)

5 Looks, approximately quarterly (cumulative proportion of information for looks 1, 2, 3, 4, and 5, respectively: .35, .47, .62, .80, 1.0)

3 warfarin-users per rivaroxaban-user (on average), 2 sided alpha=.05, flat boundary

ROW	MINIMUM DETECTABLE	POWER	THRESHOLD WALD Z (B/SE) FOR SIGNAL	THRESHOLD FOR SIGNAL ON P-VALUE SCALE	RIVAROX USERS ASSUMED AT 1ST LOOK	RIVAROX AES EXPECTED 1ST LOOK UNDER H1	RIVAROX USERS AT LAST LOOK	RIVAROX AES EXPECTED LAST LOOK UNDER H1
13	2.0	0.8	2.37	0.0089	1,215	12.2	3,473	34.7
14	1.5	0.8	2.37	0.0089	4,183	31.4	11,951	89.6
15	3.0	0.8	2.37	0.0089	376	5.6	1,075	16.1
16	2.0	0.9	2.37	0.0089	1,627	16.3	4,649	46.5
17	1.5	0.9	2.37	0.0089	5,600	42.0	15,999	120.0
18	3.0	0.9	2.37	0.0089	504	7.6	1,439	21.6

Table 9. Minimum detectable relative risks, by N, of rivaroxaban new-users that are needed at final look to detect relative risks of 1.5, 2.0, or 3.0 for each of the 3 health outcomes of interest (Assuming 5-looks (Pocock), 1:3 matching, 2-sided alpha = .05, power=.8)

ROW	IF FINAL N OF RIVAROXABAN NEW-USERS (AT LAST LOOK) IS:	DETECTABLE RELATIVE RISK, INTRACRANIAL HEMORRHAGE IS:	DETECTABLE RELATIVE RISK FOR GI BLEED IS:	DETECTABLE RELATIVE RISK FOR ISCHEMIC STROKE IS:
1	632	4.72	3.86	3.00
2	1,075	3.54	3.00	2.43
3	1,536	3.00	2.59	2.16
4	2,043	2.71	2.37	2.00
5	3,473	2.23	2.00	1.74
6	4,961	2.00	1.82	1.61
7	7,030	1.81	1.67	1.50
8	11,951	1.61	1.50	1.38
9	17,073	1.50	1.41	1.31

XI. ADDENDUM 1. CHANGES TO SURVEILLANCE PLAN AFTER THE FIRST INTERIM ANALYSIS

A. OVERVIEW

The purpose of this addendum is to list and provide rationale for three changes that were made to the Rivaroxaban surveillance plan between Looks 1 and 2.

1. Reducing the number of planned looks

Rationale: Repeated analyses were planned to be approximately quarterly, which would have yielded about five analyses during the surveillance period. We expected to observe 35%, 47%, 62%, 80%, and 100% of the total required information at each analysis. As of the writing of the third version of the surveillance plan, considerably more rivaroxaban users had been observed over time compared to initial expectations. In fact, at our first analysis, the accrued sample size (n=~15,000 rivaroxaban users) was already almost as large as the estimated maximum sample size required for all five planned analyses (N=~17,000) to achieve desired levels of power for the least common adverse event. Thus, it was clear that only one additional analysis – a final analysis – would be needed to meet the surveillance goals, and so the number of planned interim analyses will be reduced. In other words, three planned interim sequential analyses will be skipped and only two formal sequential tests will be performed.

2. Adjusting the signaling threshold to preserve the alpha spending plan

Rationale: To reflect both the reduction in the total number of interim analyses as well as the change in the amount of actual information available for each analysis, the original signaling threshold for the final analysis was adjusted. The original surveillance plan based on 5 sequential analyses specified a signaling threshold of 2.37 on the scale of the Wald Z-score. The available sample size at the second (and now final) analysis will be larger than that planned for the 5-analyses plan, yielding additional power. To correctly maintain the desired overall Type 1 error rate, we recomputed the signaling threshold to be used for the final test for each outcome that has not already signaled. Without such adjustment, the overall Type 1 error rate would have been smaller than desired. Given our expectation that our final analysis will include twice the amount of information as was available at the first analysis, the revised signal threshold is 2.06 on the scale of the Wald Z-score (the logHR estimate divided by its standard error), or 0.04 on the scale of the 2-sided p-value. For GI bleeding this threshold corresponds to a hazard ratio estimate of 1.15 or higher (indicating a relative risk that is unfavorable to rivaroxaban) or 0.87 or lower (indicating a relative risk that is favorable to rivaroxaban). For intracranial hemorrhage the threshold corresponds to an unfavorable hazard ratio estimate of 1.43 or higher and a favorable hazard ratio estimate of 0.70 or lower. On the risk difference scale, an unfavorable hazard ratio estimate of 1.15 (for GI bleeding) would amount to a harm of 8.6 GI bleeds per 1000 person years; a favorable hazard ratio estimate of 0.70 (for intracranial hemorrhage) would amount to a benefit on the risk difference scale of 4.0 events per 1000 person years.

3. Using predefined covariates to estimate propensity scores

Rationale: The CIDA and PSM tools allow the requestor to include pre-defined covariates and/or empirically select covariates to be included in the propensity scores used for matching. Previous versions of the plan specified that we would conduct one set of analyses using pre-defined covariates and a second set of analyses using empirically selected covariates (i.e., high dimensional propensity

scores). Because there are ongoing methodological investigations within Mini-Sentinel on the performance of the high dimensional propensity score analysis tool and the results of these investigations have not been completed, it was decided that rivaroxaban surveillance will not include analyses using empirically selected covariates.

It is important to note that most key analysis parameters will not change, including:

- Exposure identification algorithm
- Outcome identification algorithm
- Inclusion and exclusion conditions
- Definition of new use
- Exposure definition during follow-up
- Variable ratio matching
- Matching caliper
- Subgroups
- Cohort washout period