9-6-2016

A Synthesis of Current Surveillance Planning Methods for the Sequential Monitoring of Drug and Vaccine Adverse Effects Using Electronic Health Care Data

Jennifer C. Nelson
Group Health Research Institute; University of Washington, nelson.jl@ghc.org

Robert Wellman
Group Health Research Institute

Onchee Yu
Group Health Research Institute

Andrea J. Cook
Group Health Research Institute; University of Washington

See next pages for additional authors

Recommended Citation

Nelson, Jennifer C.; Wellman, Robert; Yu, Onchee; Cook, Andrea J.; Maro, Judith C.; Ouellet-Hellstrom, Rita; Boudreau, Denise; Floyd, James S.; Heckbert, Susan R.; Pinheiro, Simone; Reichman, Marsha; and Shoaibi, Azadeh (2016) "A Synthesis of Current Surveillance Planning Methods for the Sequential Monitoring of Drug and Vaccine Adverse Effects Using Electronic Health Care Data," *eGEMs (Generating Evidence & Methods to improve patient outcomes)*: Vol. 4: Iss. 1, Article 17.
DOI: http://dx.doi.org/10.13063/2327-9214.1219
Available at: http://repository.edm-forum.org/egems/vol4/iss1/17

The Electronic Data Methods (EDM) Forum is supported by the Agency for Healthcare Research and Quality (AHRQ), Grant 1U18HS022789-01. eGEMs publications do not reflect the official views of AHRQ or the United States Department of Health and Human Services.

A Synthesis of Current Surveillance Planning Methods for the Sequential Monitoring of Drug and Vaccine Adverse Effects Using Electronic Health Care Data

Abstract

Introduction: The large-scale assembly of electronic health care data combined with the use of sequential monitoring has made proactive postmarket drug- and vaccine-safety surveillance possible. Although sequential designs have been used extensively in randomized trials, less attention has been given to methods for applying them in observational electronic health care database settings.

Existing Methods: We review current sequential-surveillance planning methods from randomized trials, and the Vaccine Safety Datalink (VSD) and Mini-Sentinel Pilot projects—two national observational electronic health care database safety monitoring programs.

Future Surveillance Planning: Based on this examination, we suggest three steps for future surveillance planning in health care databases: (1) prespecify the sequential design and analysis plan, using available feasibility data to reduce assumptions and minimize later changes to initial plans; (2) assess existing drug or vaccine uptake, to determine if there is adequate information to proceed with surveillance, before conducting more resource-intensive planning; and (3) statistically evaluate and clearly communicate the sequential design with all those designing and interpreting the safety-surveillance results prior to implementation. Plans should also be flexible enough to accommodate dynamic and often unpredictable changes to the database information made by the health plans for administrative purposes.

Conclusions: This paper is intended to encourage dialogue about establishing a more systematic, scalable, and transparent sequential design-planning process for medical-product safety-surveillance systems utilizing observational electronic health care databases. Creating such a framework could yield improvements over existing practices, such as designs with increased power to assess serious adverse events.

Acknowledgements

Author affiliations: Group Health Research Institute (Jennifer C. Nelson, Robert Wellman, Onchee Yu, Denise Boudreau, Andrea J. Cook); Department of Biostatistics, University of Washington, Seattle, Washington (Jennifer C. Nelson, Andrea J. Cook); Department of Population Medicine, Harvard Medical School and Harvard Pilgrim Health Care Institute, Boston, Massachusetts (Judith Maro) Food and Drug Administration, Silver Spring, MD (Rita Ouellet-Hellstrom, Simone Pinheiro, Marsha Reichman, Azadeh Shoaibi); Department of Epidemiology, University of Washington (Susan R. Heckbert); Department of Medicine, University of Washington (James S. Floyd) Funding: This work was supported by the Mini-Sentinel Contract, which is funded by the USFDA through the Department of Health and Human Services Contract Number HHSF223200910006I. The findings and conclusions in this paper are those of the authors and do not necessarily represent the views of the U.S. Food and Drug Administration. For their efforts related to study conception and oversight, data acquisition, and technical support, the authors would like to thank several additional scientists: Sophia Axtman, Aarthi Iyer, Ron Johnson, Jennifer Popovic and Andrew Petrone.

Keywords

adverse drug reaction reporting systems; drug-related side effects and adverse reactions; electronic health records; product surveillance, postmarketing; sequential analysis; vaccine/adverse effects

Disciplines
Biostatistics | Clinical Trials | Epidemiology | Health Services Research | Pharmacy Administration, Policy and Regulation | Statistical Methodology

Authors
Jennifer C Nelson, *Group Health Research Institute; University of Washington*; Robert Wellman, *Group Health Research Institute*; Onchee Yu, *Group Health Research Institute*; Andrea J Cook, *Group Health Research Institute; University of Washington*; Judith C Maro, *Harvard Medical School; Harvard Pilgrim Health Care Institute*; Rita Ouellet-Hellstrom, *FDA*; Denise Boudreau, *Group Health Research Institute; University of Washington*; James S Floyd, *University of Washington*; Susan R Heckbert, *University of Washington*; Simone Pinheiro, *FDA*; Marsha Reichman, *FDA*; Azadeh Shoaibi, *FDA*.

eGEMs
Generating Evidence & Methods to improve patient outcomes

A Synthesis of Current Surveillance Planning Methods for the Sequential Monitoring of Drug and Vaccine Adverse Effects Using Electronic Health Care Data

Jennifer C. Nelson, PhD;[i,ii] Robert Wellman, MS;[i] Onchee Yu, MS;[i] Andrea J. Cook PhD;[i,ii] Judith C. Maro, PhD;[iii,iv] Rita Ouellet-Hellstrom, PhD, MPH;[v] Denise Boudreau, RPh, PhD;[i,ii] James S. Floyd, MD, MS;[ii] Susan R. Heckbert, MD, PhD;[ii] Simone Pinheiro, ScD, MSc;[v] Marsha Reichman, PhD;[v] Azadeh Shoaibi, PhD, MHS, MS[v]

ABSTRACT

Introduction: The large-scale assembly of electronic health care data combined with the use of sequential monitoring has made proactive postmarket drug- and vaccine-safety surveillance possible. Although sequential designs have been used extensively in randomized trials, less attention has been given to methods for applying them in observational electronic health care database settings.

Existing Methods: We review current sequential-surveillance planning methods from randomized trials, and the Vaccine Safety Datalink (VSD) and Mini-Sentinel Pilot projects—two national observational electronic health care database safety monitoring programs.

Future Surveillance Planning: Based on this examination, we suggest three steps for future surveillance planning in health care databases: (1) prespecify the sequential design and analysis plan, using available feasibility data to reduce assumptions and minimize later changes to initial plans; (2) assess existing drug or vaccine uptake, to determine if there is adequate information to proceed with surveillance, before conducting more resource-intensive planning; and (3) statistically evaluate and clearly communicate the sequential design with all those designing and interpreting the safety-surveillance results prior to implementation. Plans should also be flexible enough to accommodate dynamic and often unpredictable changes to the database information made by the health plans for administrative purposes.

[i]Group Health Research Institute, [ii]University of Washington, [iii]Harvard Medical School, [iv]Harvard Pilgrim Health Care Institute, [v]Food and Drug Administration

CONTINUED

Conclusions: This paper is intended to encourage dialogue about establishing a more systematic, scalable, and transparent sequential design-planning process for medical-product safety-surveillance systems utilizing observational electronic health care databases. Creating such a framework could yield improvements over existing practices, such as designs with increased power to assess serious adverse events.

Introduction

New Safety Systems Using Electronic Data

Improving methods to monitor the safety of vaccines and drugs following United States Food and Drug Administration (FDA) approval is a crucial public health need.[1] Postmarket safety monitoring relies on passive surveillance from voluntary reports—from manufacturers, patients, and health care providers—of adverse effects suspected to be associated with a specific drug or vaccine.[2-3] Now, with the purposeful assembly of “big data” resources for public health research and surveillance, such as electronic health records and claims data maintained by health plans and insurers for administrative and clinical purposes, proactive safety surveillance is possible and can supplement passive surveillance. The Vaccine Safety Datalink (VSD)[4] and the Mini-Sentinel Pilot project to establish the Sentinel system[5-6] are two notable examples of national networks that are leveraging vast amounts of health care database information to conduct safety surveillance for marketed medical products. The VSD was created in 1990 by the Centers for Disease Control and Prevention (CDC) to study the adverse effects of vaccines and has involved collaboration with 10 health care systems. Sentinel is an electronic surveillance system that involves about 20 participating institutions and was initiated by the FDA in 2008 to monitor the safety of all FDA-regulated medical products.

Active Surveillance Using Sequential Monitoring

One approach used in these systems to assess safety is sequential monitoring, which permits repeated estimation and testing of associations between a new drug or vaccine and potential adverse events over time.[7-19] Compared to a traditional design with a single analysis or test at the study’s end, a sequential analysis computes the test statistic at periodic time intervals as data accumulate, compares this test statistic to a prespecified signaling threshold, and stops if the observed test statistic is more extreme than the threshold. In this way, sequential tests can facilitate earlier identification of safety signals as soon as sufficient information from the electronic health care database becomes available to detect elevated adverse event risks. While sequential methods have been used extensively in randomized trials,[20] using them to monitor safety in a multisite, observational electronic health care database setting raises new challenges:[21-22] (1) analyzing rare adverse events; (2) controlling for confounding and channeling; (3) accommodating dynamic database updating by health plans over time, and the unpredictable uptake of newly approved medical products. (Note: These challenges also apply more generally to traditional onetime safety assessments in this setting, but we focus here on sequential applications of these methods.) Consequently, many new sequential methods have emerged that (1) use exact test methods tailored for rare events, (2) offer a variety of confounder adjustment strategies, (3) can robustly handle real-time changes in the data, and (4) can accommodate unexpected patterns of new product uptake.[23] Approaches have included historically controlled or single arm designs,[16,24-25] self-controlled designs,[24,26] exposure matching,[24] stratification on categorical confounders,[27] as well as adjustment for confounding at the analysis phase through regression[28] or inverse probability weighting.[29]

Sequential Design Considerations

In addition to the standard design steps that are typically undertaken for a traditional epidemiological study with a single analysis at the study's end, planning for sequential safety surveillance involves additional considerations: (1) When should surveillance start and end? (2) How frequently should interim tests be performed? (3) What should the statistical threshold be for a safety signal, and should it change over time? Answers to these questions define the statistical properties of the study design (e.g., Type 1 error, power, and expected time until signal detection). Frameworks to address these questions in randomized trials are well established, and decisions are typically guided by the trial's scientific goals, ethical concerns, and practical circumstances.[30] Less consideration has been given to sequential design selection steps in an observational setting using electronic health care data sources, where the safety questions, consequences of confirming a signal, and costs of false positive and negative errors differ. In particular, a safety signal generated from observational surveillance using electronic data is a preliminary finding that requires considerable follow-up investigation. The level of evidence generated from a randomized trial is stronger and could more quickly lead to regulatory action.

In this paper, we describe current methods for planning sequential monitoring activities, including prevailing guidance for randomized trials. We also summarize examples of sequential surveillance planning steps from the VSD project and Mini-Sentinel pilot. We identify the strengths as well as opportunities to improve upon existing approaches, focusing on sequential design selection and sample size planning in these examples. Last, we provide suggestions for future sequential design planning and illustrate the proposed steps using an example of a drug with a known adverse effect: angiotensin-converting enzyme (ACE) inhibitors and the risk of angioedema.[31] The ultimate goal of this work is to further the dialogue about establishing a systematic sequential surveillance planning process for use within observational electronic health care database settings, both within government agencies that monitor the safety of regulated medical products (e.g., FDA and the CDC) and between those agencies and external scientists who conduct safety evaluations and research studies. Creating such a framework could yield important improvements over existing practices, such as designs with increased power to assess serious adverse events.

Existing Safety Surveillance Planning Practices

Guidance from Randomized Clinical Trials

The use of sequential designs to monitor randomized clinical trials is common practice and has been well described.[20] Thus, we do not provide a comprehensive review here but rather highlight selected statistical recommendations that reflect the current state-of-the-art in practice. For example, the FDA provides extensive guidance on statistical principles for clinical trials conducted by industry, many of which involve group sequential interim monitoring.[32] In addition, a set of minimum standards for adaptive randomized clinical trials has been recently developed for comparative effectiveness research conducted within the Patient-Centered Outcome Research Institute.[33] Table 1 summarizes key recommendations from both these sources and their potential relevance in an observational safety setting using electronic data like the VSD and Sentinel.

Experience from the Vaccine Safety Datalink (VSD) Collaboration

Continuous Sequential Testing Methods

Sequential designs, employing either continuous or group sequential testing, have also been developed for and implemented in the VSD's observational database surveillance setting. Table 2 summarizes the main features of these designs. After preliminary exploration with the original sequential probability ratio test (SPRT),[34] initial sequential safety-surveillance efforts within the VSD utilized the maximized sequential probability ratio test (MaxSPRT) method.[24] This approach involves near-continuous sequential monitoring and uses a one-sided likelihood ratio test (LRT) that rejects the null hypothesis of no difference in the risk of a prespecified adverse event between a vaccine of interest and comparator if the log likelihood ratio (LLR) exceeds a constant upper value. In other words, MaxSPRT uses a constant (or flat) signaling boundary over time on the scale of the LLR.

Surveillance using MaxSPRT has typically been conducted for a small number of prespecified outcomes (about 5 to 10) for a specific duration of calendar time, such as two or three years following introduction of a new vaccine[7-12] or, in the case of influenza vaccine monitoring, for the duration of influenza season.[13-14] This is in contrast to a monitoring approach that follows vaccine recipients until a specific sample-size requirement designed to achieve a desired level of statistical power is met (i.e., one that uses information time to determine the surveillance duration). In some instances, statistical power was computed post hoc after surveillance was completed.[35]

Continuous sequential testing (versus group sequential testing) is advantageous because, on average, it can identify true safety signals sooner. However, continuous testing may not be feasible within a large, multisite system if the data are not updated in a real-time, continuous fashion. In addition, continuous testing is inherently less powerful than designs with less frequent testing given a fixed sample size.[36] This is because more frequent testing increases the overall chances of a false signal or Type 1 error, and thus the signaling threshold must be increased to avoid this problem. A flat boundary can also enhance early identification of signals, as it imposes a lower, less conservative signal threshold at early tests. But, by not employing early conservatism, use of a flat boundary can lead to false positive signals based on relatively little information at early analyses. This problem was observed in several VSD studies[37] and led to the development of continuous methods that implement a "delayed start," postponing the first test until a specified minimum number of events has been observed.[38] Further technical details on the advantages and limitations of continuous, compared with group, sequential testing methods in a postlicensure safety setting are beyond the scope of this manuscript but have been described elsewhere.[21,36,39-40]

Group Sequential Testing Methods

Group sequential methods were first adapted from clinical trials for use in an observational safety setting using electronic data in a VSD study of a new pentavalent combination vaccine for infants (trade name: Pentacel).[16] Similar to prior VSD studies, the

Volume 4 (2016) Issue Number 1

Table 1. FDA[32] and PCORI[33] Recommendations on Sequential Testing in Clinical Trials and Their Relevance to Observational Electronic Health Care Database Safety Surveillance Settings like Sentinel

RECOMMENDATION	DESCRIPTION	RELEVANT FOR OBSERVATIONAL SURVEILLANCE?
Prespecify statistical design and primary analysis and document changes.	All statistical methods should be prespecified prior to obtaining information on treatment outcomes, including the schedule of interim analyses, stopping rules and their properties, primary hypotheses, underlying statistical model, use of one- versus two-sided tests, and designation of primary versus exploratory analyses. It is important to document protocol deviations as changes made to the original plans can weaken and even invalidate the results.	Yes. It is equally important in observational settings to prespecify analytic plans to the extent possible. However, observational surveillance is subject to many more unknowns and may need to flexibly accommodate some changes when plans cannot be implemented as initially expected. Such changes should be documented and explained so that appropriate interpretations may be made.
Evaluate statistical properties of the design in advance.	The statistical properties of the design should be evaluated a priori so that they are understood prior to implementation and in the context of the research question (e.g., adequate power for several assumed treatment effects). For complex designs, this might include evaluating properties over a range of assumptions relating to size of treatment effect, missing data, dropout rates, etc. Technical details should be included in an appendix (e.g., statistical models and significance thresholds for the primary analyses along with calculation details or software used, operating characteristics for the design along with methods and assumptions for computing them).	Yes. But it may not be as desirable or practical to conduct an extensive performance evaluation for surveillance applications because of the following: (1) Surveillance may be done for many exposure-outcome pairs at once, making it less feasible to conduct an extensive evaluation for each design, and (2) many unknowns can lead to changes in the actual versus designed implementation, which may downweight the need to understand the planned design's performance in depth. It also may be helpful to use relatively simple designs that are well understood, can be reused, and can be scaled up.
Communicate and vet the design in advance.	The sequential design and analyses should be clearly communicated and vetted with those designing and interpreting the safety surveillance activity to assess acceptability to address the primary aims.	Yes. It is important that those designing and interpreting the safety surveillance activity (e.g., FDA) understand how the design will work in practice so any potential actions taken based on a safety signal are suitable.
Account for multiple testing.	The chance of making a Type 1 error will increase due to testing multiple outcomes, treatment comparisons, subgroups, or repeated analyses over time and should be addressed, potentially using frequentist Type 1 error adjustment methods.	Yes. However, the importance of strict accounting for random variation via multiple testing may be less in an observational surveillance setting since systematic variation will be (relatively) larger and sample sizes relatively larger. It is likely worth adjusting for sequential tests across multiple analysis time points, but it may be less necessary to adjust across multiple outcomes (since very few outcome are targeted for surveillance) or subgroups (since this is already designated as exploratory).
Interpret exploratory analyses with caution.	Exploratory analyses (e.g., in subgroups) should be interpreted with caution and should generally not be used to make definitive conclusions regarding treatment effects.	Yes. In general, surveillance results are more exploratory than results from trials. However, when prespecified, surveillance may reasonably test specific hypotheses. Results of surveillance analyses that are *not* prespecified should be considered as hypotheses for further evaluation.
Ensure proper oversight and reporting.	Proper statistical oversight of trial conduct should be in place, and reporting of the results should be done in a consistent fashion.	Yes. Statistical oversight and reliable reporting are key components for surveillance, given the data and analysis complexities and the desire for transparent presentation.

Table 2. Key Features of the Planned Sequential Designs Used in the VSD Collaboration and MS Pilot

SEQUENTIAL DESIGN FEATURES	CONTINUOUS TESTING: SEVERAL VSD STUDIES	GROUP SEQUENTIAL TESTING: VSD PENTACEL SAFETY STUDY[16] (SEPT. 2008–JAN. 2011)	GROUP SEQUENTIAL TESTING: VSD PCV13 SAFETY STUDY[15] (APR. 2010–JAN. 2012)	GROUP SEQUENTIAL TESTING: SAXAGLIPTIN EVALUATION IN MS[17] (AUG. 2009–JAN. 2014)	GROUP SEQUENTIAL TESTING: RIVAROXABAN EVALUATION IN MS[18] (NOV. 2011–APR. 2015)
Surveillance start	As soon as uptake begins or delayed until a preset # of events occur	Delayed start until 1 year of uptake (for early conservatism)	Specified in doses (information time) based on power for specific RRs	Specified in new users (information time) based on power for specific HRs	Specified in new users (information time) based on power for specific HRs
Surveillance end	Specified in calendar time ~2–3 years after the first dose	Specified in doses (information time) based on power for specific RRs; varied by event prevalence (N=72,000 doses if common, 150,000 if rare)	Specified in information time based on power to detect specific RRs; varied by adverse event prevalence	Specified in information time and based on power to detect specific HRs; resulted in last analysis ~6 years after licensure	Specified in information time and based on power to detect specific HRs
Frequency of testing	Specified in calendar time as weekly	12 total tests based on doses (information time); spacing between analyses depended on event prevalence: 3,500 or 10,500 doses	12 total tests based on information time; spacing depended on event prevalence	7 total tests, planned to be equally spaced based on information time	5 total tests, planned based on information time to occur at 35, 47, 62, 80, and 100% of the total person-time
Duration of surveillance	Specified in calendar time as 2–3 years	Specified in information time; resulted in ~2.5 years	Specified in information time; resulted in ~2 years	Specified in information time; resulted in ~6 years	Specified in information time
Shape of signaling threshold over time	Constant (flat) threshold on the scale of the LRT statistic	Constant (flat) threshold on the scale of the LRT statistic	O'Brien-Fleming threshold on the LRT scale, which is higher at earlier analyses	Constant (flat) threshold on the scale of the Wald statistic	Constant (flat) threshold on the scale of the Wald statistic
Test statistic	LRT	LRT	LRT	Wald	Wald
Test type	one-sided	one-sided	one-sided	one-sided	two-sided
Adjust thresholds?	No	Yes	No	No	No
Apply data lag so data are more complete?	2–3 months	2–3 months	2–3 months	Varied by Data Partner (some lag by 6–9 months, others do not lag)	Varied by Data Partner (some lag by 6–9 months, others do not lag)
Freeze prior data?	Freeze results from prior analyses and add only new information.	*Primary:* Cumulatively refresh all data since start of surveillance at each interim analysis. *Secondary:* Freeze results from prior analyses and add only new data.	Cumulatively refresh all data since start of surveillance at each new interim analysis	Cumulatively refresh all data since start of surveillance but preserve matches from prior analyses whenever feasible.	Cumulatively refresh data since start of surveillance.

Volume 4 (2016) Issue Number 1

Pentacel safety study used a one-sided LRT with a flat signaling boundary to test whether the risk of several targeted adverse events was elevated among Pentacel recipients versus comparators. Instead of continuous testing, however, 12 group sequential interim tests were planned. The first test, which occurred after one year of Pentacel uptake (N=33,308 doses), was purposely delayed to apply early conservatism and minimize early false positive signaling. Subsequent tests were planned to be equally spaced, based on the number of newly accruing Pentacel vaccine recipients needed to achieve specific statistical power goals. In other words, the spacing between interim analyses was based on the number of new Pentacel doses (i.e., information time) as opposed to a preset number of weeks or months (i.e., calendar time).

Given this sequential design and the expected adverse event rate among comparators, the maximum total sample size required to achieve at least 80 percent power to detect a specific minimum relative risk of interest for each outcome was computed. For more common events, this resulted in tests being performed after each additional batch of 3,500 doses of Pentacel was observed, up to a maximum sample size of about 72,000 doses. For rarer events, tests were planned to occur after each new 10,500 doses accrued among VSD enrollees, with a maximum sample size of about 150,000 doses. In addition to prespecified adverse events, a nonspecific severe outcome (e.g., any-cause hospitalization) and several control outcomes were analyzed as end-of-study, nonsequential endpoints.

In settings like the VSD and Sentinel, where data are captured and dynamically updated over time by health care organizations for administrative and clinical purposes, many unanticipated changes to the data can occur for newly approved products during the surveillance period. These unpredictable factors can constrain the ability to conduct sequential analyses exactly according to a prespecified plan. Complications that arose in the Pentacel study included the following:

1. There was unanticipated differential uptake of vaccine by age and by data partner.
2. Each planned interim analysis could not be performed at exactly the number of doses that was prespecified because data were not refreshed continuously but rather in discrete weekly batches. For instance, the second analysis was planned to occur at 36,808 doses. However, it was conducted at 37,851 doses in week 59 of surveillance because fewer than the required 36,808 doses were available at week 58 and more than 36,808 had accrued by week 59.
3. Due to an unforeseen data quality issue that was identified and later corrected, an unexpectedly large amount of previously missing Pentacel vaccine data was updated at a single time point from one data partner.

This lack of experimental control affects the adverse event variability and, in turn, the probability of committing a Type 1 error that investigators want to control. To account for these unpredictable changes in the data and still maintain proper error control in the Pentacel analysis, the planned sequential thresholds were modestly adjusted at each analysis to reflect the actual (versus planned) way in which the data were analyzed.[16] Since actual departures from the planned analyses were small, threshold adjustments were correspondingly small.

Tseng et al. also used a group sequential approach to monitor 13-valent pneumococcal conjugate vaccine (PCV13) safety in children in the VSD.[15] As in the Pentacel vaccine study, actual conduct of the PCV13 safety study was modestly different than initially planned.[15] In particular, investigators planned to finish surveillance for all prespecified outcomes within two years. However, accrual of information for the rarest events did not occur quickly enough to meet this goal. Thus, some testing plans needed modification. Table 2 provides more detail on the selected sequential features of this study. Note that the final two rows of Table 2 address two technical data-related questions that sequential surveillance plans have faced in the VSD. First, should investigators impose a data lag to improve completeness? In other words, instead of including all data captured in the databases up to the day before each interim analysis is conducted, should investigators wait several weeks or months before including a patient's data in an analysis? This would increase the probability that all relevant information (i.e., on vaccine exposure, adverse events, and confounders) has been correctly and completely captured in the database. Second, at each interim analysis when data are cumulatively examined since the surveillance start, how should prior data be treated? Should the previously analyzed data be frozen and only new data be appended that have been captured since the prior analysis? Or, should all the information observed since the beginning of the study be cumulatively refreshed?

With regard to data lagging, the standard practice within the VSD has been to simply lag the incoming data for analysis by about two to three months. For instance, if an analysis were conducted on March 1, the most recent health encounter data included would be those observed through January 1. This lag period has been instituted because some relevant vaccine and adverse event information is known not to be captured in the databases instantaneously, for example, due to relatively slower-arriving claims data when enrollees are seen at hospitals outside the integrated health system data partner. The rationale for waiting two to three months is that VSD data have been shown to stabilize and become much more complete after this period, which improves the validity of the results.[35] With respect to freezing prior data, the approach has varied by VSD study, depending on specific design and method considerations. In some cases, multiple approaches were used to assess the impact of different strategies on the final results.[16]

Ongoing Safety Assessments in Mini-Sentinel

A small number of sequential safety evaluations for drugs[17-18] and vaccines[19] have been conducted within Mini-Sentinel. Many of the lessons learned from sequential safety studies conducted within the VSD have been applied when planning these surveillance activities. Table 2 summarizes the key features of these designs. Since Mini-Sentinel data are updated on an approximately quarterly schedule, rather than near-continuously as in the VSD, group sequential designs have been the primary method utilized within Mini-Sentinel thus far.

As described for the VSD studies in the previous section, the actual sequential conduct of pilot Mini-Sentinel evaluations was not always the same as specified in initial plans, particularly for new products. For instance, in the rivaroxaban surveillance activity,[18] sample sizes were estimated for various potential scenarios of interest that varied the minimum hazard ratios (HRs) of interest detectable with 80 percent power. Calculations assumed that five group sequential analyses would be conducted based on information time when 35, 47, 62, 80, and 100 percent of new users were observed. Based on these calculations, the maximum sample size required to achieve 80 percent power to detect the smallest desired HR of 1.5 for the least common outcome of intracranial hemorrhage was estimated to be about 16,000 new rivaroxaban users. (Note: The maximum sample size is defined to be the sample size at the fifth and final planned analysis if no safety signal is detected.) In practice, largely because this was a first-of-its-kind pilot activity for drug surveillance within the Mini-Sentinel

Volume 4 (2016) Issue Number 1

environment, the actual timing of sequential tests was conducted when it was feasible based on operational factors. Specifically, the first test was conducted as soon as possible in calendar time after the surveillance plan was finalized, after about 15,000 new rivaroxaban users had been observed. Thus, by the time of the first analysis, the sample size was already almost as large as the estimated maximum sample size. In other situations, slower-than-expected new drug uptake may occur, yielding the opposite situation. Both circumstances highlight the challenge of aligning sequential design plans (which may be based on information time spacing between interim analyses so that power considerations are well understood) and the actual implementation of these analyses (which may be driven by practical calendar time and logistical constraints).

Summarizing the Lessons Learned from Prior Sequential Evaluations

Many of the established planning practices for randomized trials can help increase the integrity of a sequential safety evaluation in an observational, electronic health care–database surveillance setting. However, the extent to which each recommendation applies may vary due to practical and scientific differences from the clinical trial setting and population. Table 1 highlights the relevance of recommendations from clinical trials to safety surveillance settings. The sequential vaccine-safety surveillance experience within the VSD and the pilot surveillance activities conducted within Mini-Sentinel offer further lessons that should be considered when planning future surveillance activities. Key among these are the following:

1. *Collect and use preliminary data to inform planning*. This can reduce the number of assumptions that need to be made at the planning phase and, in turn, can minimize downstream changes to initial sequential plans. For instance, assessing the amount of existing new drug or vaccine use prior to developing the surveillance plan can better facilitate sample-size estimation and provide insight into how quickly in calendar time sample-size needs may be achieved. Examining the distribution of key potential confounders in the population of interest and computing background rates of adverse events among the likely comparator group can also help refine sample-size calculations.
2. *Provide an opportunity for preliminary discussion* with those designing and interpreting the safety surveillance activity. Clear communication in advance of a sequential design's operating characteristics and joint selection of the final design with those designing and interpreting the safety surveillance activity is essential. Then, the definition of a safety signal, which depends on the selected sequential design's signaling thresholds over time, will be well understood and will be better aligned with the follow-up actions that may be taken should a signal occur.
3. *Employ early conservatism at the surveillance start*. Using a design with a delayed start (as in the VSD Pentacel safety study[16]) or with a higher boundary at early versus later analyses (as in the VSD PCV13 safety study[15]) can help reduce false positive signals based on relatively little data at early analyses.
4. *Conduct a traditional sample-size calculation*. This can help facilitate an understanding of how much information is needed to address a particular safety question, ensure that there is an adequate amount of new data between interim analyses to warrant performing a new data analysis, and better estimate how long it will be necessary to conduct surveillance.
5. *Prepare to accommodate dynamically changing health plan data and prescribing patterns*. Implementing sequential analyses in an unpredictable, observational electronic health

care–data setting needs flexibility and caution. Even with informed and well-vetted planning steps in place, the precise rate of new drug or vaccine uptake, the population composition of new users, and the timing of database updates by health plans are not known in advance. Thus, investigators need to be prepared to adjust initial plans based on actual uptake, acceptance, and other constraints.

Since post hoc changes to initial plans can potentially introduce bias, any resulting modifications to initial surveillance plans should be justified and well documented. In addition, implementing a time lag between when data are first captured by a data partner and when they are included in an analysis is important to increase data accuracy and completeness and to reduce instability that may be caused by health plan data updates. The ability to make these informed adjustments when unexpected changes occur in the data, and the ability to successfully implement data lagging strategies to reduce bias, inherently require an in-depth, local understanding of the data from each contributing data partner. The value of having and utilizing this local data knowledge in this way cannot be overstated.

Applying Prior Lessons to Future Surveillance Planning

Potential Improvements to Future Safety Evaluations Using Observational Electronic Health Data

In this section, we translate these lessons learned from prior studies into concrete, sequential design-planning steps that could be used to improve future safety evaluations in observational, electronic health care–database settings, either for a onetime analysis or multiple sequential analyses over time. We illustrate these steps using an example of a product-outcome pair where there is a known adverse effect: ACE inhibitors and risk of angioedema.[31] The goal is to design a set of steps that meet the following criteria: (1) simple, so planning can be rapid, efficient, and scalable; (2) interpretable, so the steps are easy to understand and repeatable, (3) transparent, so planning decisions can be easily shared with those designing and interpreting the safety surveillance activity; and (4) scientifically sound, to ensure rigorous surveillance that leads to maximal public health benefit. The proposed steps are as follows:

1. *Use available data (or existing literature) to conduct a feasibility assessment and prespecify the surveillance plan*. This can provide a rough estimate of the overall sample size needed to address the designated safety question of interest in the target population.
2. *Describe uptake for the product of interest* to determine whether or not there is adequate uptake to meet these sample-size needs and thus to move forward with additional surveillance planning activities for either a onetime or a sequential analysis.
3. *Statistically evaluate, jointly select, and clearly communicate the final sequential design* with those designing and interpreting the safety surveillance activity. To conserve resources, this more time-intensive planning step, which includes finalizing the sample-size requirements, should occur only after enough product use has been observed. To cope with the dynamically changing data, investigators should plan for some flexibility in implementing the design and documenting any changes to initial plans.

Step 1: Feasibility Assessment

Step 1 can occur as soon as a product has been identified as being a priority for surveillance. This feasibility assessment should be informed by existing data (e.g., data from the same sources or a subset of the same sources that will be used in the actual surveillance activity) and should roughly estimate the sample size needed to address the prespecified safety questions based on background rates

Volume 4 (2016) Issue Number 1

estimated in the comparator group. Specifically, one can estimate required sample sizes to detect a minimum relative risk or risk difference of interest both for a onetime analysis and for a very basic sequential design (e.g., with four or eight total tests equally spaced based in information-time, a flat signaling threshold over time, one-sided test, 90 percent power, and 5 percent Type 1 error), varying the prevalence of exposure over a plausible range.

Table 3 (see top half) displays this type of preliminary data for a logistic regression analysis of the association between ACE inhibitors and the risk of angioedema within 30 days of exposure. For example, if 25 percent of the study population uses ACE inhibitors and a relative risk of 2 is of interest to detect, then them a study cohort of 308,745 total users (ACE inhibitors and comparators combined) is needed for a onetime assessment with 90 percent power, assuming an estimated outcome rate of 3.08 events per 10,000 person-months. Larger sample sizes are needed if multiple analyses are performed, but the increment in sample size required decreases as the number of additional sequential tests increases (371,041 for 4 analyses, 394,857 for 8 analyses, and 415,189 for 16 analyses). Table 3 (see bottom half) presents this same information for a linear regression analysis designed to estimate a risk difference. Considerably smaller sample sizes are needed to detect comparable signals on the risk difference scale since the risk difference is more stable than a relative difference measure when events are rare. Selection of the relative risk (or risk difference) that surveillance should aim to detect should be based on the risk-benefit profile of the new drug and, in particular, what safety signal threshold value is meaningful to regulators and should thus raise an alert.

Step 2: Uptake of the Medical Product

Once the approximate sample size is estimated, those requirements should be compared with the actual product uptake observed in the database (Step 2). This descriptive uptake assessment can help guide decisions about whether a well-powered onetime analysis can address the safety surveillance question, whether there are *not* enough users for a onetime analysis but there is adequate uptake to initiate routine sequential surveillance, or whether continued uptake monitoring is needed. For an existing medical product that has been on the market for many years, there may already be an adequate number of users to allow a single, well-powered analysis. For newer products, there may be too few users for a onetime analysis, but there may be enough to support the initiation of routine sequential surveillance. For other new products, uptake may be very slow, and continued uptake monitoring may be needed before any further planning is worthwhile. Although continued product uptake is generally expected, this may not always be the case.

Step 3: Performance Characteristics of the Final Design

As soon as the observed uptake numbers (from Step 2) reach the estimated preliminary sample-size needs (from Step 1) either for a onetime analysis or for initiation of a sequential evaluation, one can finalize the surveillance plan (Step 3). Because this third step involves more extensive planning, it should occur only once it is evident that uptake is adequate to conduct an evaluation. The goal of this step is to examine the properties of several potential designs in more detail so that they are fully understood prior to implementation. This process should involve clear communication and collaborative vetting of several potentially suitable sequential-surveillance designs with those designing and interpreting the safety surveillance activity in order to assess acceptability of that design in addressing the primary safety aims. Choices include the number and timing of analyses as well as the shape of the signaling boundary over time. Once the final design is selected, then final estimated sample-size requirements can be computed for that

Table 3. Maximum Sample Size for Logistic Regression Analysis By Number Of Analyses

LOGISTIC REGRESSION TO ESTIMATE A RELATIVE RISK (RR)					
		MAXIMUM SAMPLE SIZES			
% of total sample who are ACE users	RR	1-TIME	4-TIMES	8-TIMES	16-TIMES
25%	1.5	902,285	1,084,340	1,153,941	1,213,358
	2	**308,745**	**371,041**	**394,857**	**415,189**
	3	122,903	147,701	157,182	165,275
50%	1.5	676,714	813,255	865,456	910,019
	2	231,559	278,281	296,143	311,392
	3	92,178	110,776	117,887	123,957

LINEAR REGRESSION TO ESTIMATE A RISK DIFFERENCE (RD)					
		MAXIMUM SAMPLE SIZES			
% of total sample who are ACE users	RD	1-TIME	4-TIMES	8-TIMES	16-TIMES
25%	1.5	625,032	751,145	799,360	840,519
	3	**156,258**	**187,787**	**199,840**	**210,130**
	6	39,065	46,947	49,960	52,533
50%	1.5	468,774	563,359	599,520	630,389
	3	117,194	140,840	149,880	157,598
	6	29,299	35,210	37,470	39,400

Notes: *Assumptions: Binary outcome: Angioedema in 30 days after exposure; Comparator group: Beta blockers; Estimated rate of outcome among comparator group: 3.08/10,000 person-months; Boundary shape: Flat on standardized Z-statistic scale; Power: 90% to detect a given relative risk or risk difference. **Maximum sample size is defined as the number of new ACE inhibitor users that are required to achieve 90% power to detect a specified minimum RR or RD of interest if no signal is detected during the course of a sequential evaluation.

Volume 4 (2016) Issue Number 1

design, which will ultimately determine how long surveillance needs to be conducted.

Table 4 displays the type of detailed data that are useful for making sequential design decisions. Specifically, sample-size estimates are shown for a wide variety of potential sequential designs that implement a logistic regression analysis to estimate the relationship between ACE inhibitors and the occurrence of angioedema within 30 days of exposure. Both the number of total planned analyses and the shape of the signaling threshold over time are varied across a range of values for exposure uptake and minimum detectable relative risks (RRs) of interest. Several common sequential-threshold options are shown, including a Pocock signaling threshold that is constant over time on the scale of the test statistic (a Z-score),[41] a curvilinear O'Brien-Fleming threshold that is highest at the first few analyses to achieve early conservatism,[42] and a power family threshold that lies "in between" these two extremes.[43-44] Figures 1–2 show the magnitude of the signaling thresholds on the test statistic (Z-score) scale as well as the more interpretable RR scale for selected designs.

For instance, based on Figure 1, a sequential design with only 4 analyses would result in the first analysis not being conducted until about 80,000-90,000 new users have been observed for any design. This may be viewed as waiting too long if there truly is increased harm. Focus might then turn to designs with more frequent analyses, such as those with 8 or 16 total planned analyses presented in Table 4 and Figure 2. At the first analysis, all designs with 16 total tests could signal after about 5 adverse events are observed in the comparator group (see Analysis 1, bottom of Figure 2). This number of events may be deemed too small a number upon which to base a preliminary safety signal, and that may direct further attention to the designs with 8 analyses. The designs with 8 analyses require about 10 events in each group before a signal would be raised (see Analysis 1, top of Figure 2). Among those designs, the O'Brien-Fleming boundary threshold may be considered too conservative at the first analyses, requiring an extremely high RR of 27 or more (Figure 2, data point not shown) to generate a signal. This might lead a surveillance team to choose an 8-analysis plan with either the Pocock threshold (which would signal if the RR is about 4 at the first analysis) or a threshold in between these two extremes (which would require a RR of about 8 to signal at the first analysis).

This ACE inhibitors and angioedema example illustrates the type of statistical information that could be used to communicate the operating characteristics of different sequential designs to those designing and interpreting the safety surveillance activity prior to surveillance implementation. And, in an oversimplified way, it shows how such information could be used to compare the performance of competing designs, and to facilitate a dialogue among those designing and interpreting the safety surveillance activity about their design preferences. And the use of such information could lead to more informed final decisions about the choice of appropriate signaling thresholds. Clearly, though, the factors that influence the choice of sequential design selection are more complicated than this illustration conveys. Numerous scientific, ethical, and practical considerations (e.g., the magnitude of the vaccine or drug's benefit, the prevalence and severity of the adverse event of interest, etc.) should bear on this choice, and the relative importance of each factor may depend on the specific safety question of interest. Our intent here is not to comprehensively discuss these factors but rather to describe a high-level framework for how statistical information can be used by those designing and interpreting the safety surveillance activity—to better weigh these factors when making sequential design decisions.

Table 4. Maximum Sample Size for Regression Analyses by Boundary Shape

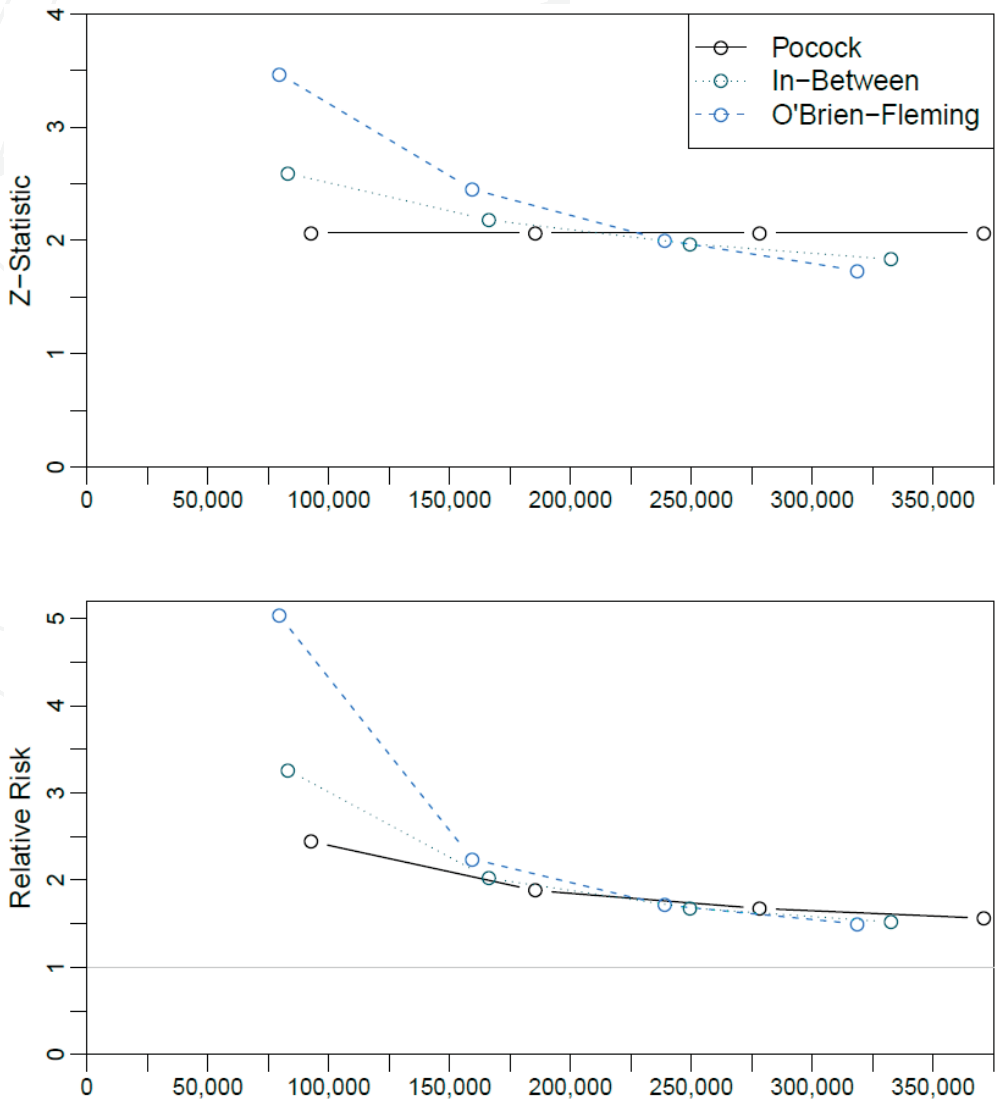
LOGISTIC REGRESSION (TO ESTIMATE A RELATIVE RISK)					
			MAXIMUM SAMPLE SIZES		
# of Analyses	% of total sample who are ACE users	RR	POCOCK	IN-BETWEEN	O'BRIEN-FLEMING
8	25%	1.5	1,153,941	990,736	943,715
		2	**394,857**	**339,012**	**322,922**
		3	157,182	134,951	128,546
	50%	1.5	865,456	743,052	707,786
		2	296,143	254,259	242,191
		3	117,887	101,214	96,410
16	25%	1.5	1,213,358	1,003,258	951,930
		2	**415,189**	**343,296**	**325,733**
		3	165,275	136,657	129,666
	50%	1.5	910,019	752,444	713,948
		2	311,392	257,472	244,300
		3	123,957	102,493	97,249

LINEAR REGRESSION (TO ESTIMATE A RISK DIFFERENCE)					
			MAXIMUM SAMPLE SIZES		
# of Analyses	% of total sample who are ACE users	RD (per 10k person-months)	POCOCK	IN-BETWEEN	O'BRIEN-FLEMING
8	25%	1.5	799,360	686,304	653,731
		3	**199,840**	**171,576**	**163,433**
		6	49,960	42,894	40,859
	50%	1.5	599,520	514,728	490,298
		3	149,880	128,682	122,575
		6	37,470	32,171	30,644
16	25%	1.5	840,519	694,978	659,422
		3	**210,130**	**173,745**	**164,856**
		6	52,533	43,437	41,214
	50%	1.5	630,389	521,234	494,567
		3	157,598	130,309	123,642
		6	39,400	32,578	30,911

Notes: *Assumptions: Binary outcome: Angioedema in 30 days after exposure; Comparator group: Beta blockers; Estimated rate of outcome among comparator group: 3.08/10,000 person-months; Power: 90% to detect a given relative risk or risk difference.

Volume 4 (2016) Issue Number 1

Figure 1. Signaling Thresholds for a Design with Four Analyses

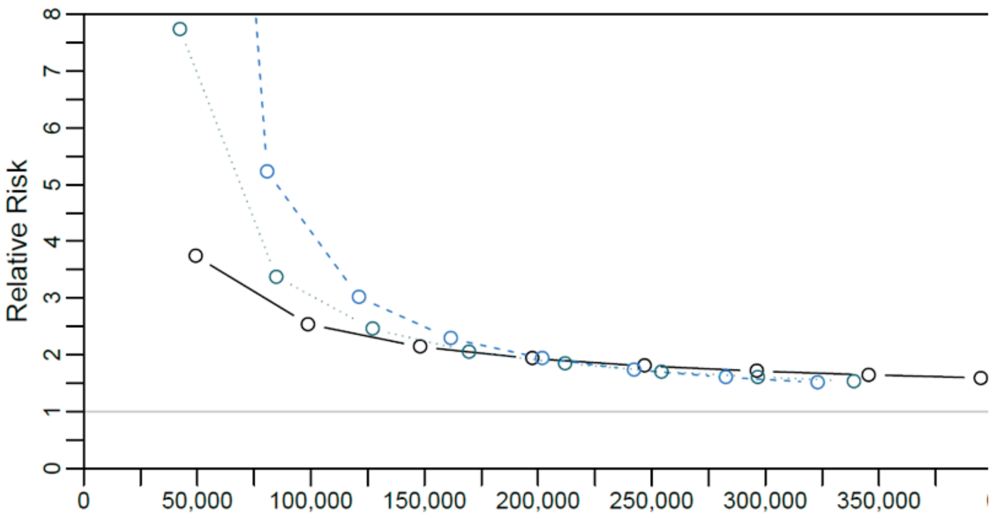


Analysis Times: Total Sample Size (Outcomes in Unexposed, Exposed)

	Analysis 1	Analysis 2	Analysis 3	Analysis 4
Pocock:	92760(21, 17)	185520(43, 27)	278281(64, 36)	371041(86, 45)
In−Between:	83140(19, 21)	166279(38, 26)	249418(58, 32)	332558(77, 39)
O'Brien−Fleming:	79664(18, 31)	159328(37, 28)	238991(55, 32)	318655(74, 37)

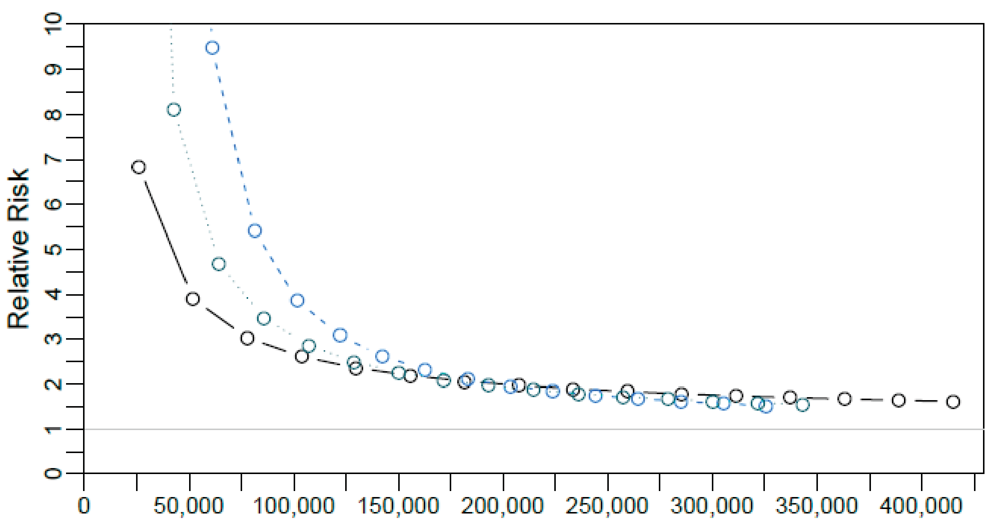
Notes: Assumptions: Binary outcome: Angioedema in 30 days after exposure; Proportion using ACE inhibitors (versus a comparator like beta blockers): 25%; Estimated rate of outcome among comparator group: 3.08/10,000 person-months; Power: 90% to detect a RR=2.

Figure 2. Signaling Thresholds for Designs with 8 (Top) or 16 (Bottom) Analyses



Analysis Times: Total Sample Size (Outcomes in Unexposed, Exposed)

	Analysis 1	Analysis 2	...	Analysis 8
Pocock:	49357(11, 14)	98714(23, 19)		394857(91, 48)
In−Between:	42376(10, 12)	84753(20, 17)		339012(78, 42)
O'Brien−Fleming:	40365(9, 12)	80730(19, 16)		322922(75, 40)



Analysis Times: Total Sample Size (Outcomes in Unexposed, Exposed)

	Analysis 1	Analysis 2	...	Analysis 16
Pocock:	25949(6, 14)	51899(12, 16)		415189(96, 52)
In−Between:	21456(5, 11)	42912(10, 13)		343296(79, 43)
O'Brien−Fleming:	20358(5, 11)	40717(9, 12)		325733(75, 41)

Notes: Assumptions: Binary outcome: Angioedema in 30 days after exposure; Proportion using ACE inhibitors (versus a comparator like beta blockers): 25%; Estimated rate of outcome among comparator group: 3.08/10,000 person-months; Power: 90% to detect a RR=2.

Conclusions

Existing methods used for sequential design planning in randomized trials and observational safety surveillance assessments within the VSD and Mini-Sentinel provide a strong foundation upon which to build a more formal framework to plan future routine safety evaluations using electronic health care databases. We have provided recommendations on how practices from randomized trials can be adapted to accommodate the unique challenges of conducting safety surveillance activities in the observational setting of electronic health care databases, which contributes to an emerging literature on this topic.[45-46] We have also illustrated ways in which existing methods from observational settings like the VSD and Mini-Sentinel could be improved—by further leveraging well-established best practices from trial settings and tailoring them to meet the challenges posed by an electronic data environment.

This review points to three important sequential design steps that should be addressed during the planning phase for safety surveillance activities utilizing observational electronic health care databases:

1. Prespecification of the surveillance design and analytic plan is critical.
2. Use of existing data to inform surveillance planning can reduce the number of assumptions that need to be made at the planning phase and, in turn, minimize downstream changes to initial sequential plans.
3. Selection of a sequential design should include statistical evaluation and clear communication of the sequential design and analysis with all those designing and interpreting the safety surveillance activity so that the operating characteristics are well understood in advance of implementation.

In addition, due to the dynamic nature of the health care data sources, it is important that selected methods offer the ability to be flexible in their implementation and that investigators document any resulting changes to initial plans that are caused by unpredictable data. We hope that this work can spark further dialogue among regulatory scientists about more systematic sequential-design planning processes and, ultimately, that it will lead to formal guidance with recommended best practices that can be used in future safety evaluations that are conducted using health care database information.

Acknowledgements

This work was supported by the Mini-Sentinel Contract, which is funded by the USFDA through the Department of Health and Human Services Contract Number HHSF223200910006I. The findings and conclusions in this paper are those of the authors and do not necessarily represent the views of the U.S. Food and Drug Administration. For their efforts related to study conception and oversight, data acquisition, and technical support, the authors would like to thank several additional scientists: Sophia Axtman, Aarthi Iyer, Ron Johnson, Jennifer Popovic and Andrew Petrone.

References

1. Nelson JC, Cook AJ, Yu O, Zhao S, Jackson LA, Psaty BM. Methods for observational post-licensure medical product safety surveillance. *Statistical Methods in Medical Research*. 2015; 24(2):177-193. Epub 2011 Dec 2.
2. Chen RT, Rastogi SC, Mullen JR, Hayes SW, Cochi SL, Donlon JA, et al. The Vaccine Adverse Event Reporting System (VAERS). *Vaccine*. 1994 May;12(6):542-50.
3. Rodriguez EM, Staffa JA, Graham DJ. The role of databases in drug postmarketing surveillance. *Pharmacoepidemiool Drug Saf*. 2001 Aug-Sep;10(5):407-10.
4. Baggs J, Gee J, Lewis E, Fowler G, Benson P, Lieu T, et al. The Vaccine Safety Datalink: a model for monitoring immunization safety. *Pediatrics*. 2011;127:S45-S53.
5. Behrman RE, Benner JS, Brown JS, McClellan M, Woodcock J, Platt R. Developing the Sentinel System — A National Resource for Evidence Development. *N Engl J Med*. 2011. DOI: 10.1056/NEJMp1014427.

6. Nguyen M, Ball R, Midthun K, Lieu TA. The Food and Drug Administration's Post-Licensure Rapid Immunization Safety Monitoring program: strengthening the federal vaccine safety enterprise. *Pharmacoepidemiol Drug Saf*. 2012 Jan;21 Suppl 1:291-7. doi: 10.1002/pds.2323.
7. Lieu TA, Kulldorff M, Davis RL, Lewis EM, Weintraub E, Yih K et al. Real-time vaccine safety surveillance for the early detection of adverse events. *Med Care*. 2007;45(10 Suppl 2):S89-95.
8. Klein NP, Fireman B, Yih W, Lewis E, Kulldorff M, Ray P, et al. Measles-mumps-rubella-varicella combination vaccine and the risk of febrile seizures. *Pediatrics*. 2010;126:e1-e8. (doi:10.1542/peds.2010-0665).
9. Yih WK, Nordin JD, Kulldorff M, Lewis E, Lieu TA, Shi P, et al. An assessment of the safety of adolescent and adult tetanus-diphtheria-acellular pertussis (Tdap) vaccine, using near real-time surveillance for adverse events in the Vaccine Safety Datalink. *Vaccine*. 2009;27:4257-4262. (doi:10.1016/j.vaccine.2009.05.036).
10. Belongia EA, Irving S, Shui IM, Kulldorff M, Lewis E, Yin R, et al. on behalf of the Vaccine Safety Datalink Investigation Group. Real-time surveillance to assess risk of intussusception and other adverse events after pentavalent, bovine-derived rotavirus vaccine. *Pediatr Infect Dis J*. 2010;29(1):1-5.
11. Gee J, Naleway A, Shui I, Baggs J, Yin R, Li R, et al. Monitoring the safety of quadrivalent human papillomavirus vaccine: Findings from the Vaccine Safety Datalink. *Vaccine*. 2011;29(46):8279-84. Epub 2011 Sep 9.
12. Daley MF, Yih WK, Glanz JM, Hambidge SJ, Narwaney KJ, Yin R, et. al. Safety of diphtheria, tetanus, acellular pertussis and inactivated poliovirus (DRaP-IPV) vaccine. *Vaccine*. 2014; 32(25):3019-24. Epub 2014 Mar 31.
13. Lee GM, Greene SK, Weintraub ES, Baggs J, Kulldorff M, Fireman BH et al. on behalf of the Vaccine Safety Datalink Project. H1N1 and seasonal influenza vaccine safety using prospective weekly surveillance in the Vaccine Safety Datalink Project. *Am J Prev Med*. 2011;41(2):121-128.
14. Tse A, Tseng HF, Greene SK, Vellozzi C, Lee GM on behalf of the VSD Rapid cycle Analysis Influenza Working Group. Signal identification and evaluation for risk of febrile seizures in children following trivalent inactivated influenza vaccine in the Vaccine Safety Datalink Project, 2010-2011. *Vaccine*. 2012 Mar 2;30(11):2024-31.
15. Tseng HF, Sy LS, Liu IL, Qian L, Marcy SM, Weintraub E, et. al. Postlicensure surveillance for pre-specified adverse events following the 13-valent pneumococcal conjugate vaccine in children. *Vaccine*. 2013 May 24;31(22):2578-83. Epub 2013 Apr 8.
16. Nelson JC, Yu O, Dominguez C, Cook AJ, Peterson D, Greene SK, et al. Adapting group sequential methods to observational post-licensure safety surveillance: Results of a pentavalent combination DTaP-IPV-Hib (Pentacel) vaccine safety study. *Am J Epidemiol*. 2013; Epub 2012 Dec 12.
17. Fireman B, Toh S, Butler MG, Go AS, Joffe H, Graham D et al. A protocol for surveillance of acute myocardial infarction in association with use of a new anti-diabetic pharmaceutical agent. *Pharmacoepidemiology and Drug Safety*. 2012; 21 Suppl 1:282-90. doi: 10.1002/pds.2337.
18. Carnahan R, Gagne J, Nelson JC, Fireman B, Wang S, Shoabi et al. Mini-Sentinel Prospective Routine Observational Monitoring Program Tools (PROMPT): Rivaroxaban Surveillance. 2014. Available from http://www.mini-sentinel.org/work_products/Assessments/Mini-Sentinel_PROMPT_Rivaroxaban-Surveillance-Plan.pdf. Accessed August 14, 2015.
19. Yih WK, Zichittella L, Sandhu SK, et al. Accessing the Freshest Feasible Data for Conducting Active Influenza Vaccine Safety Surveillance [Internet]. 2015 Available from: http://www.mini-sentinel.org/work_products/PRISM/Mini-Sentinel_PRISM_Active-Influenza-Vaccine-Safety-Surveillance-Report.pdf. Accessed May 12, 2015.
20. Whitehead J. ***The design and analysis of sequential clinical trials***, revised ***2nd ed***. Chichester, West Sussex: John Wiley & Sons Ltd; 2000:135-181.
21. Nelson JC, Cook AJ, Yu O, Zhao S, Dominguez C, Fireman B, et al. Challenges in the design and analysis of sequentially-monitored post-licensure safety surveillance studies using observational health care utilization data. *Pharmacoepidemiology and Drug Safety*. 2012; 21 Suppl 1:62-71. doi: 10.1002/pds.2324.
22. Brown JS, Kulldorff M, Petronis KR, Reynolds R, Chan KA, Davis RL, et. al. Early adverse drug event signal detection within population-based health networks using sequential methods: key methodologic considerations. *Pharmacoepidemiol Drug Saf*. 2009 Mar;18(3):226-34.
23. Cook AJ, Wellman RD, Marsh T, Li L, Heckbert S, Heagerty P, Tiwari RC, Nelson JC. Statistical approaches to group sequential monitoring of post-marketing safety surveillance data: Current state of the art for use in the Mini-Sentinel surveillance project. *Pharmacoepidemiology and Drug Safety*. 2012; 21 Suppl 1:72-81. doi: 10.1002/pds.2320
24. Kulldorff M, Davis RL, Kolczak M, et al. A maximized sequential probability ratio test for drug and vaccine safety surveillance. *Seq Anal*. 2011;30:58-78.
25. Shih M-C, Lai TL, Heyse JF, et al. Sequential generalized likelihood ratio tests for vaccine safety evaluation. *Statistics in Medicine*. 2010;29:2698-2708.
26. Hocine MN, Musonda P, Andrews NJ and Farrington CP. Sequential case series for pharmacovigilance. *Journal of the Royal Statistical Society series A*, 2009, 172: 213-236.
27. Li L, Kulldorff M. A conditional maximized sequential probability ratio test for pharmacovigilance. *Stat Med*. 2010 Jan 30;29(2):284-95.
28. Cook AJ, Wellman RD, Nelson JC, Jackson LA and Tiwari RC. (2014), Group sequential method for observational data by using generalized estimating equations: application to Vaccine Safety Datalink. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 2014 Sep 23.doi: 10.1111/rssc.12075 [Epub ahead of print]
29. Cook AJ, Wellman RD, Marsh T, Tiwari RC, Nguyen MD, Russek-Cohen E, et al. (2012) Statistical methods for estimating causal risk differences in the distributed data setting for postmarket safety outcomes. Available from http://www.mini-sentinel.org/work_products/Statistical_Methods/Mini-Sentinel_PRISM_Statistical-Methods-for-Estimating-Causal-Risk-Differences.pdf. Accessed July 20, 2015.

30. Emerson S, Kittelson J, Gillen D. Tutorial in biostatistics: Frequentist evaluation of group sequential clinical trial designs. *Statistics in Medicine*. 2007;26:5047–5080.
31. Israili ZH, Hall WD. Cough and angioneurotic edema associated with angiotensin-converting enzyme inhibitor therapy. A review of the literature and pathophysiology. Ann Intern Med 1992;117(3):234-42.
32. Food and Drug Administration. Guidance for Industry: E9 Statistical Principles for Clinical Trials. Available from http://www.fda.gov/downloads/drugs/guidancecomplianceregulatoryinformation/guidances/ucm073137.pdf. Accessed July 20, 2015.
33. Detry MA, Lewis RJ, Broglio KR, Connor JT, Berry SM, Berry DA. Standards for the design, conduct, and evaluation of adaptive randomized clinical trials. Available at http://www.pcori.org/assets/Standards-for-the-Design-Conduct-and-Evaluation-of-Adaptive-Randomized-Clinical-Trials.pdf. Accessed July 20, 2015.
34. Davis RL, Kolczak M, Lewis E, Nordin J, Goodman M, Shay DK. Active surveillance of vaccine safety: a system to detect early signs of adverse events. *Epidemiology*. 2005 May;16(3):336-41.
35. Greene SK, Kulldorff M, Yin R, et al. Near real-time vaccine safety surveillance with partially accrued data. *Pharmacoepidemiol Drug Saf*. 2011;20(6):583-590.
36. Silva I and Kulldorff M. Continuous versus group sequential analysis for post-market drug and vaccine safety surveillance. *Biometrics*. 2015 May 22. doi: 10.1111/biom.12324. [Epub ahead of print]
37. Yih WK, Kulldorff M, Fireman BH, et al. Active surveillance for adverse events: the experience of the Vaccine Safety Datalink project. *Pediatrics*. 2011;127;S54-S64.
38. Kulldorff M, Silva IR. (2015). Continuous post-market sequential safety surveillance with minimum events to signal. Available at: http://arxiv.org/abs/1503.01978.
39. Zhao S, Cook AJ, Jackson LA, Nelson JC. Statistical performance of group sequential testing methods for post-licensure medical product safety surveillance: a simulation study. *Statistics and Its Interface*. 2012; 5:381-390.
40. Kulldorff M. Sequential statistical methods for prospective post-marketing safety surveillance. In: Strom BL, Kimmel SE, Hennessy S, eds. *Pharmacoepidemiology*. 5th ed. Sussex, England: Wiley; 2011.
41. Pocock SJ. Group sequential methods in the design and analysis of clinical trials. *Biometrika*. 1977; 64(2):191-199.
42. O'Brien P, Fleming T. A multiple testing procedure for clinical trials. *Biometrics*. 1979;35:549-556.
43. Kittleson JM, Emerson SS. A unifying family of group sequential test designs. Biometrics. 1999;55(3):874-882.
44. Wang SK, Tsiatis AA. Approximately optimal one-parameter boundaries for group sequential trials. *Biometrics*. 1987; 43(1):193-199.
45. Maro JC, Brown JS, Kulldorff M. Medical Product Safety Surveillance: How Many Databases to Use? *Epidemiology*. 2013 Sep;24(5):692-9.
46. Maro JC, Brown JS, Dal Pan GJ, Kulldorff M. Minimizing signal detection time in postmarket sequential analysis: balancing positive predictive value and sensitivity. *Pharmacoepidemiol Drug Saf*. 2014 Apr 3. doi: 10.1002/pds.3618