MINI-SENTINEL METHODS

FRAMEWORK FOR ASSESSMENT OF SIGNAL REFINEMENT POSITIVE RESULTS

Prepared by: David L McClure, PhD[1], Marsha A Raebel, PharmD, BCPS, FCCP[2,3], W Katherine Yih, PhD, MPH[4], Azadeh Shoaibi, MS, MHS[5], Jerry Mullersman, MD, PhD, MPH[6], Colin Anderson-Smits, MPH[7], Rita Ouellet-Hellstrom, PhD[5], Aloka Chakravarty, PhD[5], Clara Kim, PhD[5], Jason M Glanz, PhD[2]

Author Affiliations: 1. Marshfield Epidemiology Research Center, Marshfield Clinic Research Foundation, Marshfield WI. 2. Institute for Health Research, Kaiser Permanente Colorado, Denver CO. 3. University of Colorado Skaggs School of Pharmacy and Pharmaceutical Science, Aurora, CO. 4. Department of Population Medicine, Harvard Medical School and Harvard Pilgrim Health Care Institute, Boston MA. 5. Center for Drug Evaluation and Research, US Food and Drug Administration, Silver Spring, MD. 6. Center for Biologics Evaluation and Research, US Food and Drug Administration, Silver Spring, MD. 7. Center for Devices and Radiological Health, US Food and Drug Administration, Silver Spring, MD

December 20, 2012

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.

Mini-Sentinel Methods

Framework for Assessment of Signal Refinement Positive Results

Table of Contents

1. EXECUTIVE SUMMARY

This report describes a framework for assessment of signal refinement positive results. "Positive results" are defined to be when an association is detected between a medical product and an adverse outcome that exceeds a pre-specified threshold in the direction of increased risk.

The principal goal was to determine if the excess risk can be explained by something other than a cause and effect relationship, such as information or selection bias, confounding, or any errors associated with the signal refinement results.

This report addressed assessing *signal refinement*, which is the second of a three stage process (signal generation, signal refinement, signal evaluation) in medical product post-market safety surveillance.

Descriptions and general recommendation regarding sources of systematic error are presented. It begins with an assessment of data validity. This should be a regular activity before a safety signal occurs. The emphasis should be on ruling out errors in the data that contributed to the signal.

Sources of systematic error in medical product safety surveillance are information bias, selection bias, and confounding. Information bias is an error in measuring exposure, covariate, or outcome variables that result in different quality (accuracy) of information between comparison groups. Misclassification of categorical variables is a form of information bias.

For the Sentinel System, medical product exposures would be measured within electronic healthcare databases. Such data are an imperfect surrogate for actual biological exposures within individuals. Data sources that populate the electronic health databases are not originally created for research or public health purposes. The sensitivity and positive predictive values of electronic diagnostic codes for outcome assessment can be as low. It can be particularly challenging to identify outcomes when there are no specific diagnosis codes or the incident rate is very low and/or misdiagnosis occurs.

Selection bias is a distortion in an effect estimate due to the manner in which the study sample is selected from the source population. To avoid case selection bias, the cases (outcomes) that contributed to a safety signal must represent cases in the source population. For exposed groups, the exposure definition must be representative and realistic of the pertinent medical product users in the source population. Selection bias can occur for completely unexposed comparison groups, when they differ in important health related characteristics compared to exposed groups. To mitigate this, comparison groups can be based on exposures to other similar (but not identical) medical products, i.e. "active comparators".

Fundamentally, confounding is a mixing or confusion of effects. The apparent effect of the exposure is distorted from the true effect of exposure due to the effects of other factors outside the causal pathway of the medical product-outcome pair of interest. Medical products other than vaccines are generally given to treat existing health diseases and conditions. Usually, the patient has an indication for such treatment. Electronic medical databases usually do not have sufficient information to clearly define the indication. This "confounding-by-indication" is often a concern regarding the validity of a safety signal. Residual confounding can also occur when attempts to control known confounding are incomplete or other suspected confounders were unmeasured.

Sensitivity analysis, now known as quantitative bias analysis, is a general set of analytic methods that can assess the potential impact of confounding (measured, unmeasured, unknown), and information and selection biases. The goal of bias analysis is to quantify the influence of systematic error on a risk

estimate from an epidemiologic study. Quantitative bias analysis can establish that a result is invalidated by systematic errors or show that a result is robust to systematic errors and thus provides confidence to public health policy makers. For quantitative bias analysis to be feasible, a sufficiently precise estimate must have been generated. Quantitative bias analysis adjusts a conventional risk estimate based on bias parameters that determine the direction and magnitude of adjustment. The methods to generate and apply bias parameters vary from relatively simple analysis of a fixed value of a bias parameter to complex and comprehensive sets of several bias parameters and their probable ranges of values.

Simple bias analysis is most feasible when the bias parameters are known with certainty at one single value. Unfortunately this is almost never the situation and the more sophisticated methods described are recommended. Multidimensional bias analysis should be the minimum level of sophistication when analyzing systematic error. The more computationally intense probabilistic bias analysis or multiple biases modeling is recommended when the detail in the available data sources are sufficient. The relatively long history of active vaccine safety surveillance in VSD, with its large quantity of detailed longitudinal data, would allow probabilistic bias analysis or perhaps multiple biases modeling to be feasible. Since active safety surveillance is less mature for the vastly more variable drug exposures, the level of sophistication performed in quantitative bias analysis would be partly determined by the particular drug-adverse event pair under investigation. That is, on a case-by-case basis within the resources allocated. For the medical devices or blood products, unless there are detailed data from, for example, an established registry with unique product identifiers, quantitative bias analyses might not be feasible.

Overall, for *signal refinement*, quantitative bias analysis would be most useful in determining the maximum amount of systematic error that could be tolerated without obviating the signal. Subsequent signal evaluation might proceed differently if a relatively small amount of confounding, misclassification, or selection bias could overwhelm the signal versus if the signal is robust against a substantial amount of systematic error.

Each type of medical product has specific issues related to signal refinement. Cofounding-by-indication is a common issue for observational safety studies of drugs, medical devices, or blood products. Drug safety studies can be affected by channeling bias, concomitant drug exposures and uncertainties in exposure assessment. Vaccine safety studies must consider bias due to the healthy vaccine effect, confounding-by-contraindication, uncertainties in exposure time periods, seasonality, and multiple vaccines or combination vaccines. Medical device currently lack of a unique product identifier, leading to uncertainties in exposure assessments. Devices also have specific issues of product similarities/differences, human environment factors, and drug-device or device-device interactions.

A process for assessing signal refinement results should proceed in a series of steps. Once it has been established that a positive signal was detected, the most expedient next step is to review data validity, descriptive statistics, and analytic computer programs. If the positive signal still persists after this step, then additional secondary analyses should be considered. Patterns in time from the exposure to adverse event outcome should be examined. This could proceed via graphical methods and/or temporal scan statistics. If the graph or temporal scan statistical test indicates temporal clustering of outcomes after exposure, the risk window can be narrowed to include only this period and the analyses repeated, which would tend to result in higher relative risk estimates.

If the examination of descriptive statistics reveals possible confounders that might not have been adequately controlled for, these should be dealt with by applying standard pharmacoepidemiologic

methods such as multi-variable regression to the same dataset. Such analyses might include finer adjustment for age, adjustment for seasonal trends by month or by use of sinusoidal curves, adjustment for secular trends using a linear or other function, adjustment for chronic diseases or propensity to receive the drug or product of interest, and adjustment for one or more specific simultaneous exposures such as another drug or vaccine.

Another secondary analysis could be to use different comparison groups than in the original analysis, for example, historical comparison groups from different periods, concurrent controls that are selected or matched based on different criteria, and different control periods in self-controlled designs.

If after the previous steps, the positive signal still generated an estimate of reasonable precision, then quantitative bias analysis may be feasible. The objective is to determine the maximum amount of systematic error that could be tolerated without obviating the signal.

If quantitative bias analysis is properly performed, then the possible impact of misclassification, selection bias, and confounding will be assessed. The challenge is to accurately model the most relevant biases for the specific medical product and outcome pairs of interest. The particular medical product, adverse event outcome, surveillance database and the judgment of the investigators all influence the choice and performance of quantitative bias analysis methods.

The interpretation and reporting of the signal assessment activity should be based on the protocol and the findings. The report should generally reflect if and where the excess positive risk could be explained by something other than a cause and effect relationship, such as data or systematic errors (information or selection bias, confounding). The report should be written in the order of the assessment steps done as described in the previous sections. If the assessment included a quantitative bias analysis, then details of the analysis should be described.

2. INTRODUCTION

2.1 OBJECTIVE

The workgroup objective was to develop a framework for assessment of signal refinement positive results. "Positive results" are defined as when an association is detected between a medical product and an adverse outcome that exceeds a pre-specified threshold in the direction of increased risk.

2.2 SCOPE

The principal goal was to determine if the excess risk can be explained by something other than a cause and effect relationship, such as information or selection bias, confounding, or any errors associated with the signal refinement results. In general, the determination will be whether the exposure-outcome relationship appears to exist in its originating data source. The scope of this project did not extend to replication in other sources, to a determination of causality, dose-response relationships, or other aspects of the relationship which typically require more detailed epidemiologic analysis.

This report addressed assessing *signal refinement*, which is the second of a three stage process in medical product post-market safety surveillance (**Figure 1**). Signal refinement is the initial focus of Mini-Sentinel[1], and is defined as "the assessment of predefined exposure outcome pairs to determine whether there is evidence of association." [2]

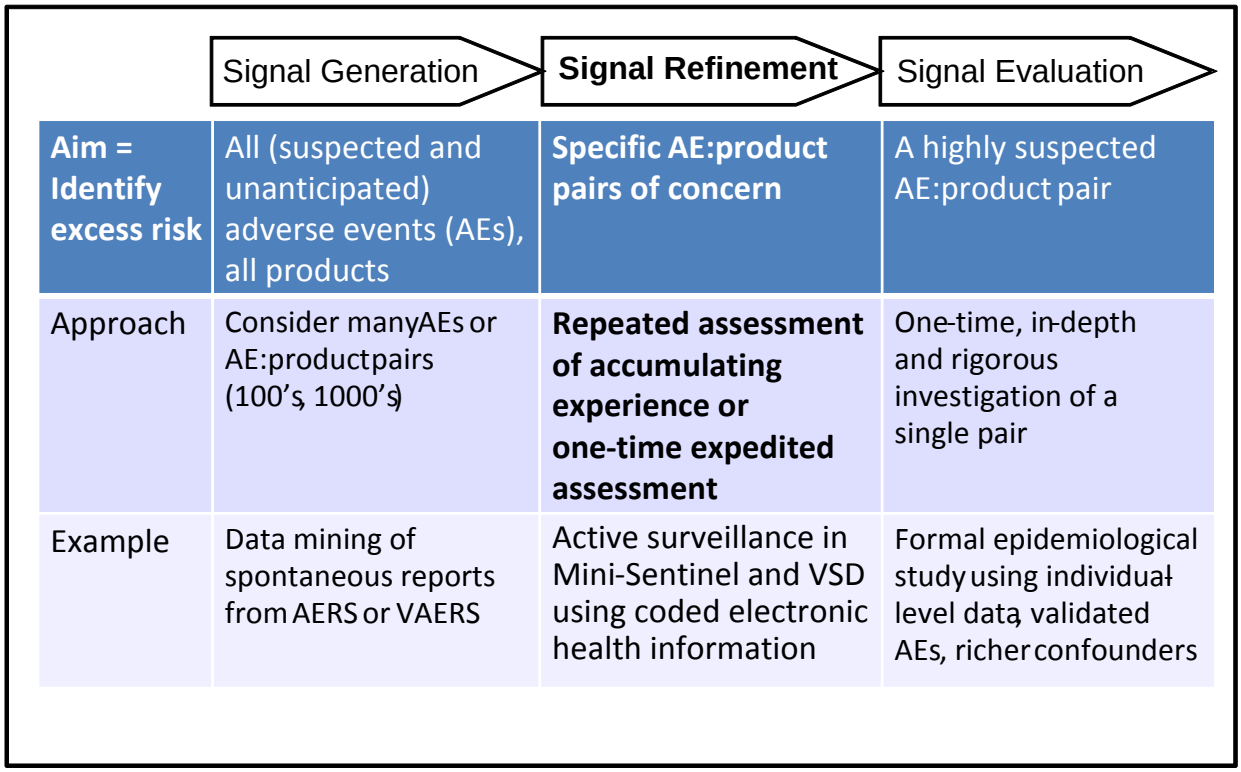

Figure 1. Stages of Medical Product Post-market Safety Surveillance

Previous reports from the Mini-Sentinel Methods Core, focusing on various aspects of signal refinement, were implicitly (or sometimes explicitly) based on good accepted principles of statistics and epidemiology and state-of-the-art techniques.[3-7] Thus, signal refinement, especially as conducted via active surveillance, should be considered an epidemiologic and statistical activity. The Centers for Disease Control and Prevention (CDC) sponsored Vaccine Safety Datalink (VSD) has several years' experience in conducting active surveillance of vaccines, and considers it an essential *epidemiologic* activity.[8,9]

Previous work regarding signal refinement in the context of post-market surveillance has been most thoroughly assessed for vaccines[8-10] and somewhat less so for drugs.[11-13] It is relatively unknown for medical devices where prospective safety monitoring has been limited.[14]

3. SOURCES OF SYSTEMATIC ERROR IN MEDICAL PRODUCT SAFETY SURVEILLANCE

3.1 INFORMATION BIAS – MISCLASSIFICATION

In the context of epidemiology, bias has been defined as "the systematic deviation of results or inferences from truth or processes leading to such deviation".[15] Bias is fixed with respect to sample size and is synonymous with systematic error. In contrast, the net effect of random errors approaches zero as the sample size increases. Bias can arise from measurement errors in the information gathered (information bias) or in errors of selection of the study sample from the source population (selection bias).

Information bias is an error in measuring exposure, covariate, or outcome variables that result in different quality (accuracy) of information between comparison groups.[15] It is not necessarily independent of selection bias. Misclassification of categorical variables is a form of information bias.

3.1.1 Exposure misclassification

For the Sentinel System, medical product exposures would be measured within electronic healthcare databases. Such data are an imperfect surrogate for actual biological exposures within individuals.[16] How exposures are practically assessed and the impact of misclassification depends on the particular medical product type, for example, a chronic intermittent drug exposure, an acute vaccine exposure, or a chronic continuous exposure from an implanted medical device. Details on the uncertainties of exposure assessment are in section 5, Specific Product Considerations.

3.1.2 Non-differential vs. differential misclassification and the direction of bias

If the exposure misclassification does not depend on the values of other variables (including case status), then it is non-differential. Only when the exposure is accurately modeled as dichotomous and the misclassification is independent of other errors, among other assumptions, is the bias of effect only in the direction of the null, i.e. the measured risk is diminished from the real risk. When the exposure misclassification is likely to differ by the value of other variables (mainly the outcome variable), it is differential, and the bias can affect the risk estimates either toward or away from the null.

Recall bias can lead to differential misclassification of exposures. For example, consider a case-control study of over-the-counter nonsteroidal anti-inflammatory drugs (NSAID) and gastrointestinal bleeding. Individuals identified as cases often will contemplate possible reasons for their symptoms. They may recall and be less likely to forget use of NSAIDs than controls. More cases than controls would be classified as exposed due to the effects of recall bias, resulting in differential misclassification of exposure status.

3.1.3 Outcome misclassification

Data sources that populate the electronic health databases are not originally created for research or public health purposes. The sensitivity and positive predictive values of electronic diagnostic codes for outcome assessment can be as low as 20%.[17,18] It can be particularly challenging to identify outcomes when there are no specific diagnosis codes or the incident rate is very low and/or misdiagnosis occurs. Examples include Stevens-Johnson syndrome,[18] newborn pulmonary hypertension[19] or Churg-Strauss syndrome.[20]

In general, many positive signals with risk ratios greater than 1.0 (i.e., an increased risk following exposure) are negated by medical record review because a large percentage of the electronically-identified events do not represent incident disease.[17,18] Instead, these electronically recorded events often represent follow-up visits for chronic (prevalent) conditions, miscoded diseases, or rule-out diagnoses. In other instances, the electronic diagnostic dates imprecisely measure the date of disease onset, which can shift an event from being classified as exposed to unexposed or vice versa. For these reasons, studies relying on automated data alone are associated with a high rate of false positive signals.

The misclassification of automated data highlights additional limitations of safety screening and near-real time surveillance. The data are initially analyzed as if the outcomes are perfectly measured, and it is

standard practice to validate only the positive associations with a medical record review. However, given the considerable misclassification of the automated data, it is possible that true positive associations are being overlooked. Statistical power calculations should have been performed to establish the lowest level of risk that could be detected for the particular safety study that is re-examined for validity.

3.1.4 Methods to assess misclassification

The statistical literature on methods to correct for misclassification and measurement error is vast. Such methods differ by theoretical approach, which can be parametric, semi-parametric or nonparametric.[21] Although differential error has been addressed in the literature,[22] most methods have focused on correcting for non-differential error.[21] Both Bayesian and frequentist approaches have been applied. The particular approach is often predicated on whether or not external validation data is available.

A majority of the published methods for assessing misclassification bias assume no error in the response variable (outcome); they instead tend to focus on the impact of mismeasured exposures, confounders and other predictors of risk. Methods that do explicitly address misclassification of response variables are in the context of traditional hypothesis confirmation analyses, where the risk of a single outcome is modeled within a case-control or cohort framework.

3.2 SELECTION BIAS

Selection bias is a distortion in an effect estimate due to the manner in which the study sample is selected from the source population.[23] Selection bias can be related to information bias,[24] and can also have an impact on confounding.[23]

3.2.1 Selection of cases

To avoid case selection bias, the cases (outcomes) that contributed to a safety signal must represent cases in the source population. Cases should meet a predefined case definition that may include electronic diagnostic codes, chronicity (acute or chronic) and medical setting (outpatient, inpatient). Exposure to a particular medical product is not necessarily in the case definition. Unexposed cases are needed in appropriate comparison groups, and ascertaining cases “blinded to exposure” helps avoid introducing information bias.

3.2.2 Selection of exposed groups

For exposed groups, the exposure definition must be representative and realistic of the pertinent medical product users in the source population. Exposures to medical products should be defined as precisely as possible with respect to (but not limited to) duration of exposures, continuous or intermittent exposure periods, and whether medical products of similar therapeutic class are included.[25]

3.2.3 Selection of comparison groups

Even if cases or exposed groups are selected without bias, the inappropriate selection of controls or unexposed comparison groups introduces selection bias. Examples are provided below.

3.2.3.1 Exposure based comparison groups (unexposed to main exposure of interest)

For exposure based analyses, as in cohort or self-control study designs, comparison groups should provide a valid outcome estimate free of exposures due to the medical product of interest. Selection bias can be introduced if there were differential losses to follow-up between exposed and comparison groups.

3.2.3.1.1 Completely unexposed comparison groups

Selection bias can occur for completely unexposed comparison groups, when they differ in important health related characteristics compared to exposed groups. For example, in vaccine safety studies, completely unvaccinated individuals often have different follow-up times among other fundamental differences in morbidity, health behaviors, or other characteristics not otherwise captured in the available data. They may not be representative of members from the source population that could have been exposed to the specific medical product of interest but were otherwise not. Moreover, unexposed groups may not be "completely unexposed" if the "look back" time for prior exposures is too short relative to an index date.

For medical device studies, it is sometimes difficult to have completely unexposed comparison groups from a similar source population. Many devices are implanted among patients who have severe disease progression where other medical treatments have failed.[26] The device may be the only option or in some cases such as cardiac catheterization the only choice and it would be unethical to have a completely unexposed group.

3.2.3.1.2 Active comparator groups

To mitigate this selection bias, comparison groups can be based on exposures to other similar (but not identical) medical products, i.e. "active comparators". This is often used in drug or device efficacy or safety studies, where the adverse event risk from users of a new drug is compared to the risk from users of an established "standard-of-care" drug. Both drugs (and devices) should have been prescribed for the same indication within the population under study.

Active comparator controlled studies can have specific challenges. Often MCO formularies and treatment guidelines determine which drugs are considered for primary and secondary use. These are generally drugs with a longer history of use and better known efficacy and safety profile and often are drugs available as lower cost generic formulations. Drugs newer to the market are often only available as a single branded product and are generally more expensive. They may be relegated to tertiary use as a covered drug in the pharmacy benefit package only when primary and secondary agents fail to deliver effective treatment. For some clinical indications, only expensive products are available. In both these situations, economic factors can force patients who cannot afford particular drugs to either be prescribed less than ideal drugs, cut pills to extend drug supply, or avoid treatment altogether by leaving prescriptions unfilled.[27] These factors can be difficult to assess in claims data and can make valid comparisons particularly challenging.

3.2.3.1.3 Temporal comparison groups

The relative timing between members of a comparison group to members of the exposed group can also be a factor in selection bias.

In concurrent controls, members can be matched to exposed members based on calendar time and other covariates deemed important for the particular medical product-outcome pair under surveillance. This is efficient for near-real time safety surveillance, where a relative risk can be rapidly estimated without needing statistical adjustment from regression modeling. However, if the outcomes are particularly rare, a stable rate in a concurrent comparison group might not be possible.

The use of historical comparison groups is an alternative. Instead of individual level matching, estimated background rates are stratified by predetermined covariate groupings. Rates may have to be generated from data that predates by several years the occurrence of the exposures of interest. A selection bias could be introduced if the historical rates are not representative of what a true unexposed rate would be during the time of the exposures.

3.2.3.2 Self-controls

In self-control study designs, cases serve as their own controls. The comparison is based on whether the case occurred during an exposed time period versus an unexposed time. By definition there cannot be a selection bias due to "control group selection". However, studies that use self-controls work best for acute exposures and acute outcomes with prompt onset such as in most vaccine safety studies and where time-varying confounders do not exist. Another limitation is an issue of information bias if there is misclassification as to how exposed and/or unexposed time periods were assigned.

3.2.3.3 Outcome based comparison groups (non-case controls in case-control studies)

Controls based on lack of the outcome of interest (non-cases) imply a case-control study setting. Controls as non-cases must be as representative of the underlying study population as cases. However, case-control studies can be particularly vulnerable to selection biases and confounding. In this era of large longitudinal electronic health care databases, case-control studies are not necessarily an efficient and economical alternative to cohort studies.[28]

3.2.3.4 Methods to assess selection bias

Since the origin of selection bias is essentially in the study design phase, it cannot be corrected by statistical analysis alone.[17] Selection bias can also be described graphically in the context of causal diagrams. [29] These diagrams show that it is sometimes possible to disentangle the effects of selection bias if the determinants of participation are accurately known. A relatively simple example is for an odds ratio as calculated from a 2x2 table of case/noncase versus exposed/unexposed. If the selection probabilities are known for each of the 4 cells, then the odds ratio is multiplied by a ratio defined by the 4 selection probabilities.[30] Unfortunately it is almost never known what the true selection probabilities were, as partitioned among all states of outcome status, exposures, and other important covariates.

Probabilistic bias analysis, described in the subsequent Quantitative Bias Analysis section, can address the extent of how selection bias might impact findings from a study.[30,31] This method is based on Monte Carlo sampling of plausible distributions of selection probability bias parameters.

3.3 CONFOUNDING

Fundamentally, confounding is a mixing or confusion of effects. The apparent effect of the exposure is distorted from the true effect of exposure due to the effects of other factors outside the causal pathway of the medical product-outcome pair of interest.[32]

3.3.1 Confounding by indication

Medical products other than vaccines are generally given to treat existing health diseases and conditions. Usually, the patient has an indication for such treatment. Electronic medical databases usually do not have sufficient information to clearly define the indication. This "confounding-by-indication" is often a concern regarding the validity of a safety signal. Confounding-by-indication arises where the treatment is a marker for a medical condition that serves as the "indication" for using the treatment, and at the same time that medical condition increases (or possibly decreases) the risk of the outcome of interest. In contrast, vaccines are given to a generally healthy population and confounding by contra-indication can be present (for some drug safety analyses too).

3.3.2 Measured confounders

Other factors than indication can confound the true association of a medical product-outcome pair of interest. Some of these factors such as demographics and coded comorbidities are captured in electronic health care databases. Others such as socio-economic factors and health related behaviors may not be. Often there are many covariates in a database that could be constructed into a possible measured confounder. Examples are disease risk scores or exposure propensity scores.[18]

3.3.3 Residual, unmeasured, and unknown confounding

Residual confounding can occur when attempts to control known confounding are incomplete or other suspected confounders were unmeasured.[33] Known confounders may have been improperly constructed from covariates in the data, especially if categorized from continuous measures. The categories may have been too broad. Another reason is that a covariate may poorly represent a possible confounding factor. For example, census tracts may or may not adequately control for possible confounding from economic factors. Covariates that represent confounders may also have measurement errors. Levels of confounding factors may be misclassified into the wrong exposure or outcome groups. Even if important confounders were well measured, but they were not included in the final multivariate risk models, they will strongly contribute to residual confounding.

Another kind of residual confounding is unmeasured confounding. This originates from known confounding factors that were not captured in the data. For example, frailty in the elderly was a known confounder in some drug safety studies but measures of frailty are not usually captured in health claims data.[34] Other possible confounders not always measured well in such data are exercise level, smoking or alcohol use status, diet or family health history.

Perhaps the most problematic is when a possible confounding factor is unknown to not only the investigator, but to the larger scientific community.[35] Possible unknown confounding is suspected when a new observed association of an exposure and outcome is considered counter-intuitive. An interesting example is the continuing scientific debate about the association of moderate alcohol consumption and lower risk of coronary heart disease and the possible effects of unknown confounders.[36]

3.3.4 Methods to assess and control confounding

If confounders are known and measured, then there are several possible methods that can be considered. The first and simplest is restriction. [37]Individuals are only included in a study if they have a single category or level of a confounding characteristic. There can be no confounding if there is no variation in the confounding covariate in the study data. If sex is an important confounder in an association of exposure and outcome risk, then a study would only include either males or females, but not both.

The second method is stratification on levels of the confounding covariate.[38] Matching on confounding covariates between exposed and unexposed is a special case of stratification. Matching in cohort studies can be a way to reduce confounding effects but also can reduce statistical efficiency (wider confidence intervals).

Multivariate adjustment in regression modeling is a third method of confounding assessment and control. When regression modeling is performed correctly and important confounding covariates are included (and measured without error), then the risk estimate of the exposure of interest is "adjusted" and free of confounding.

A fourth method for assessment and control of confounding is the use of propensity scores either by matching or stratification, or inverse probability weighting.[39] Propensity scores are probabilities for exposure conditioned on various covariates (which are often important measured confounders). Each individual in a study has a specific nonzero propensity score, regardless of the actual exposed/unexposed status. By matching or stratifying on the exposure propensity score, the effects of measured confounders are eliminated or at least reduced.

Finally, sensitivity analysis, now known as quantitative bias analysis, is a general set of analytic methods that can assess the potential impact of confounding (measured, unmeasured, unknown), and information and selection biases.

4. QUANTITATIVE BIAS ANALYSIS

The goal of bias analysis is to quantify the influence of systematic error on a risk estimate from an epidemiologic study. Quantitative bias analysis can establish that a result is invalidated by systematic errors or show that a result is robust to systematic errors and thus provides confidence to public health policy makers. The following described methods are based on the chapter "Bias Analysis" in Modern Epidemiology, 3rd edition[42] and the textbook by Lash, et al., "Applying Quantitative Bias Analysis to Epidemiologic Data".[31]

For quantitative bias analysis to be feasible, a sufficiently precise estimate must have been generated. Note that this is not determined by "statistical significance", rather, that the confidence interval of the estimate is narrow enough such that bias and confounding analyses are feasible. Consider two example risk estimates.[40] The first example is a relative risk (RR) and 95% confidence interval (CI)of 1.4 (0.80-2.4) with p=0.2. The width of the CI, based on the ratio of the upper and lower limits, is 3. The second example is a "significant" RR of 4.1 (1.2-14), p=0.02. Here the width of CI is 11.7. The first example is more feasible for quantitative bias analysis than the second because of its more precise CI. Ultimately, it is the investigators responsibility to judge whether quantitative bias analysis is necessary and feasible.

4.1 BIAS PARAMETERS

Quantitative bias analysis adjusts a conventional risk estimate based on bias parameters that determine the direction and magnitude of adjustment. Table 1 shows examples of bias parameters for outcome misclassification, selection bias, and unmeasured confounding.

Table 1. Examples of bias parameters for quantitative bias analysis. RRtrue is the true relative risk, RRobs is the observed biased relative risk, and M is a function of the bias parameters that expresses magnitude and direction of the bias.

Misclassification of the Outcome
Se1 = sensitivity among exposed, Sp0 = specificity among exposed,
Se2 = sensitivity among unexposed, Sp0 = specificity among unexposed
Mmc = function of (Se1, Sp1, Se0, Sp0), RRobs=Mmc x RRtrue
If all the sensitivities and specificities equal 1 then there is no bias from misclassification. RRobs=RRtrue if there are no other biases present.

Selection Bias
Sc1 =probability of selecting exposed study cases from the population of all exposed cases
Sc0= probability of selecting unexposed study cases from the population of all unexposed cases
Sn1 =probability of selecting exposed study noncases from the population of all exposed noncases
Sn0= probability of selecting unexposed study noncases from the population of all unexposed noncases
Msb = function of (Sc1, Sc0, Sn1, Sn0), RRobs=Msb x RRtrue
If all the selection probabilities are equal then there is no selection bias. RRobs=RRtrue if there are no other biases present.

Unmeasured Confounder
RRco =relative risk associated with the confounder and outcome
Pc1=prevalence of the confounder among exposed
Pc0= prevalence of the confounder among unexposed
Muc = function of (RRco, Pc1, Pc0), RRobs=Muc x RRtrue
If Pc1=Pc0 or RRco=1 then there is no unmeasured confounding. RRobs= RRtrue, if there are no other biases present.

The methods to generate and apply bias parameters vary from relatively simple analysis of a fixed value of a bias parameter to complex and comprehensive sets of several bias parameters and their probable ranges of values.

4.2 SIMPLE BIAS ANALYSIS

In simple bias analysis, one fixed value is assigned to each bias parameter. The bias parameters are analyzed one at a time and a single revised estimate of association of risk is presented. Simple bias analysis is computationally simple and straightforward; however it does not incorporate the joint effect of random error.

For example, for an initially electronically assessed outcome in a case-control study, misclassification of the odds ratio can be accomplished by generating positive (negative) predictive values from secondary data such as medical chart review. The correction factor for disease misclassification is 1/[PPV+NPV-1), here PPV (NPV) are the positive (negative) predictive values of the outcome.[41]

4.3 MULTIDIMENSIONAL BIAS ANALYSIS

Multidimensional bias analysis is similar to simple bias analysis, except that more than one value is assigned to each bias parameter. Bias parameters are also analyzed one at a time. Multidimensional bias analysis yields a set of corrected estimates, but not a real frequency distribution. As in simple bias analysis, this method is computationally simple and also does not include the effect of random error.

4.4 PROBABILISTIC BIAS ANALYSIS

Probabilistic bias analysis is a further progression in sophistication compared to multidimensional bias analysis. It begins with assigning probability distributions to the bias parameters; however bias parameters are still analyzed one at a time. Monte Carlo methods are used to repeatedly sample from the probability distributions and to correct for the bias. The main advantage of probabilistic bias analysis is that a frequency distribution of revised risk estimates are produced that also incorporate random error. The mean of the distribution and its 95% simulation interval can be interpreted similarly to a conventional point estimate and its confidence interval. While probabilistic bias analysis has these desirable properties over the previous simpler methods, it is computationally more intense. The general procedure for conducting probabilistic bias analysis is shown in Table 2. Examples of probabilistic bias analysis are also presented in Appendix A.

Table 2. Probabilistic bias analysis procedure

1. Identify likely sources of important bias and bias parameters
2. Assign probability distributions to each bias parameter
3. Randomly sample from each bias parameter distribution
4. Use simple bias analysis to correct for the bias
5. Save corrected estimates and repeat steps 3 and 4
6. Summarize the corrected estimates with a frequency distribution that yields a central tendency and simulation interval

Adapted from Lash TL, Fox MP, Fink AK. Applying Quantitative Bias Analysis to Epidemiologic Data. New York: Springer. 2009, p 132

4.5 SUMMARY DATA VS. RECORD LEVEL CORRECTIONS

Simple and multidimensional bias analyses are performed on estimates of association and summary data—typical of what is reported in text and tables in published manuscripts. Probabilistic bias analysis was first developed via simulating corrections to record level data.[42,43] This allows direct regression

modeling (with appropriate multiple imputation techniques) and subsequent estimation of confidence intervals that incorporate both random and systematic error. The major limitation of record level corrections is the computer programming expertise and resources required. Recently, methods have been developed to perform probabilistic bias analyses on summary data that combines random and systematic errors.[44]

4.6 MULTIPLE BIASES MODELING

The most sophisticated and comprehensive quantitative bias analysis is multiple biases modeling.[45] Probability distributions are assigned to bias parameters and several bias parameters analyzed at once. Information bias, selection bias, and confounding are usually modeled in order with additional random error. A frequency distribution of revised risk estimates is produced and its mean and its 95% simulation interval are interpreted as for conventional estimates. Multiple biases modeling is the most computationally intense method of bias analysis and is performed at the record level of data.

4.7 GENERAL RECOMMENDATIONS FOR QUANTITATIVE BIAS ANALYSIS

Simple bias analysis is most feasible when the bias parameters are known with certainty at one single value. Unfortunately this is almost never the situation and the more sophisticated methods described are recommended. Multidimensional bias analysis should be the minimum level of sophistication when analyzing systematic error. The more computationally intense probabilistic bias analysis or multiple biases modeling is recommended when the detail in the available data sources are sufficient. The relatively long history of active vaccine safety surveillance in VSD, with its large quantity of detailed longitudinal data, would allow probabilistic bias analysis or perhaps multiple biases modeling to be feasible. Since active safety surveillance is less mature for the vastly more variable drug exposures, the level of sophistication performed in quantitative bias analysis would be partly determined by the particular drug-adverse event pair under investigation. That is, on a case-by-case basis within the resources allocated. For the medical devices or blood products, unless there are detailed data from, for example, an established registry with unique product identifiers, quantitative bias analyses might not be feasible.

Overall, for *signal refinement*, quantitative bias analysis would be most useful in determining the maximum amount of systematic error that could be tolerated without obviating the signal. Subsequent signal evaluation might proceed differently if a relatively small amount of confounding, misclassification, or selection bias could overwhelm the signal versus if the signal is robust against a substantial amount of systematic error.

5. SPECIFIC PRODUCT CONSIDERATIONS

5.1 DRUGS

5.1.1 Post-approval drug safety surveillance

Post-marketing drug safety surveillance is a core component of how the US Food and Drug Administration (FDA) protects and promotes public health.[1] In the past the FDA has had to rely on spontaneous reporting systems such as the Adverse Event Reporting System (AERS)[46], as one surveillance component, to monitor drug product safety post-marketing. Spontaneous reporting

systems depend on voluntary reports of adverse events and quality problems from healthcare practitioners and patients. Voluntary reporting is subject to both underreporting and incomplete reporting and, while spontaneous reports provide numerator data (number of outcomes), no denominator data (number of exposures) is provided. Thus, from spontaneous reporting, the rate of an adverse outcome cannot be determined within the context of an overall population of exposed and unexposed individuals. Recently, as a result of provisions within the FDA Amendments Act of 2007, FDA has embarked on the Sentinel Initiative. Once fully operational, the Sentinel Initiative will provide active surveillance capabilities that complement and supplement spontaneous reporting and provide population-based denominator (as well as numerator) data to assess outcomes. This Initiative is currently under development (between 2009 and 2014), with this developmental, pilot phase known as Mini-Sentinel. The infrastructure necessary to provide population-based numerator and denominator data is being implemented using health care information collected in the process of patient care as well as insurance claims data from data partners that are voluntarily participating. These administrative and clinical health care data are being extracted, transformed, and loaded into a distributed data system common model (the Mini-Sentinel Common Data Model or MSCDM). The MSCDM will enable the FDA to monitor the safety of drugs (and other regulated medical products) under conditions of actual use in the US. Post-approval drug product safety surveillance within the Sentinel Initiative program is initially focusing on signal refinement (further investigating an identified safety signal) to determine whether evidence exists to support a relationship between the exposure and the outcome.

5.1.2 Issues specific to drugs

5.1.2.1 Confounding by indication

Drug safety studies are often "confounded by indication." As mentioned previously, drugs are generally given to prevent disease or to treat existing conditions, and electronic medical databases usually do not have sufficient information to clearly define the indication. Numerous strategies have been explored to control for confounding in observational pharmacoepidemiology. Strategies to control for confounding-by-indication can be thought of as two types, those that adjust for known confounders and those that adjust for unknown confounders. The latter also often adjust for known confounders, but usually include additional assumptions or restrictions to generalizability.[47] Strategies to handle measured confounders include restriction and matching in the design phase, and standardization, stratification, multivariate regression (including propensity scores and disease risk scores) and marginal structural models in the analytic phase.[47] Strategies to handle unmeasured confounders include two-stage sampling and external adjustment for confounders that might be measurable in a separate study. For truly unmeasurable confounders, case only or crossover study designs, instrumental variables, and sensitivity analysis may be useful.[47] Exposure based propensity scores or instrumental variables (if they exist) have shown some utility in at least partially controlling for confounding by indication. Restricting the study population to those with a clearer indication for the specific drugs under investigation works well, but at the expense of generalizing the interpretation of results.

5.1.2.2 Channeling bias

Channeling occurs when drugs with similar therapeutic indications are prescribed to groups of patients with prognostic differences.[48] Channeling is sometimes thought of as a subtype of confounding by indication in that claimed advantages of a new drug may direct (channel) the drug towards use in highly selected patient subgroups. The consequence can be that higher adverse event or lower efficacy rates

than expected are observed and are incorrectly attributed to the drug. For example, a newly-marketed drug for rheumatoid arthritis may be initially prescribed only to the subgroup of patients who have the most severe disease, who did not achieve remission or disease modification with previously-marketed drugs, or who experienced adverse reactions to older drugs. Channeling bias can at times be identified when large healthcare databases provide information on prior and concurrent dispensings of other medications and on patient comorbidity or disease duration.

5.1.2.3 Sources of uncertainty in drug exposure

Electronic pharmacy dispensing records are the prime data source for post-market active safety surveillance of drugs.[49] For example, the Mini-Sentinel Common Data Model (MSCDM) Dispensing file contains information typically found in electronic pharmacy records sourced from managed care organizations (MCOs), large health insurance carriers, and government insurance programs such as Medicare. These variables include person-level identifiers, drug dispense date, amount dispensed, days supply, and the National Drug Code (NDC) which identifies the particular drug, specific strength, dosage and package forms and sizes.

However drug dispensing data have limitations that can have an impact on drug exposure misclassification. Pharmacy databases usually contain outpatient drug dispensing information, obtained either from pharmacy insurance claims or from actual drugs dispensed at pharmacies within an integrated health care delivery system. Moreover, the actual prescribed dose regimen is usually not captured and must be inferred from other variables. In addition, drugs administered during an inpatient stay are usually not available from pharmacy insurance claims databases and are only sometimes available from MCOs electronic health records databases. Therefore, drugs administered during an inpatient hospitalization often represent missing data and consequently the potential for misclassification as unexposed. Another important limitation is that drug dispensing information available from these databases only indicates that the patient had the medication in his/her possession, but not whether the patient actually took the drug or took it at the recommended dosage and frequency. Thus, patient adherence to the prescribed regimen can only be inferred from the variables available within these databases.[50] Finally, the pharmacodynamics and pharmacokinetics of the drug must always be considered when estimating drug exposure and the potential for misclassification as when the risk period begins and ends is in large part determined by these pharmacological considerations. Drug exposure patterns can be acute or chronic with short- or longer-term, sporadic, intermittent, or continuous characteristics. Thus, with drugs, both exposure and outcome can be time-varying.[51,52] As Faich and Stadel observed over two decades ago, “unless you observe the patient swallowing the drug and monitor the body drug concentrations, drug exposure must always be considered as measured approximately.” [53]

Uncertainties in measuring and understanding patterns of drug exposure can broadly be categorized as associated with patient, provider, system, and drug characteristics as well as drug exposure time patterns. Some of these were discussed above. Patient-related characteristics that can contribute uncertainty include demographics and socioeconomic status (e.g., younger aged adults often have poorer medication adherence than do older adults), health status/co-morbidities (e.g., smoking can be a marker for risky health behaviors), severity of disease, and adherence to or persistence with drug therapy (e.g., if a patient’s adherence is considered to be 70%, it cannot be confirmed whether the non-adherent days were evenly spaced throughout the refill period, all occurred within the initial portion of the refill period, or all occurred at the end of the period). Provider/prescriber characteristics that

contribute to lack of clarity in drug exposure can include for example, interviewing skills, completeness of documentation in the medical record, and even prescribing preferences (e.g., preferring to prescribe new drugs as they become available versus preferring to prescribe older established drugs with known safety profiles). System characteristics that may foster uncertainty include the use of paper versus electronic medical records (e.g., paper records or radiology films may be destroyed after a certain date), drug benefit and health care utilization co-payment structures (e.g., high drug copayments can be associated with poorer medication adherence), and poor access to pharmacies (e.g., no transportation to obtain refills of medication). Finally, drug characteristics and exposure patterns that are sources of uncertainty include prescribing (e.g., paper prescriptions versus electronic prescription order entry), dispensing (e.g., no electronic documentation of provision of medication samples), medication adherence, cumulative dose, the pharmacodynamics (duration and pattern or adverse event risk window) and pharmacokinetics of the drug, concomitant medications including both the number of unique medications and the types of unique medications, and drug switching patterns.

5.1.2.4 Defining the appropriate risk window for assessing risk of the outcome

Choosing the appropriate risk window implies that the correct period of increased risk of outcome following exposure has been identified. Ideally, this time window includes the full period of potential excess risk, that is, the duration of both drug intake and of the pathogenic process that results in development of clinically manifest events. In reality, the usual process for identifying the "risk" window is more aligned with identifying dispensings of the drug of interest, the duration of "exposure" based on days of drug supplied to the patient across one or more refill intervals, the dosage and dosing interval, and a surrogate measure of medication adherence (often identified from the information available within the pharmacy dispensing or claims database). Consequences of incorrect risk window measurement can be dire, in that cohort studies where lifetime exposure history includes pertinent exposure and irrelevant time can fail to discover effects with short or transient risk periods and case-control studies with too wide time window include time outside the true risk window yielding study results that differ greatly from study results obtained using narrowly-defined time windows. Over- or underestimation of duration of risk often results in bias towards the null. Conversely, the best estimate of risk duration is the one that maximizes the relative risk estimate. These points are exemplified in published studies that have conflicting results.

Transient or time-varying adverse event risk is the most common type of outcome window with drug exposures. Transient risk windows are characterized by a period of normal (background) risk and periods of excess risk and reduced risk. In fact, with drug exposures, adverse event risk that is constant over time is rare. A classic example of time-varying risks associated with drugs include diethylstilbestrol (DES) and adenocarcinoma of the vagina[54] where the overall risk across the lifetime among women exposed in utero approximates 1 per 10 million person-years, but the risk for these women between ages 20 to 30 years is 1 per 10 thousand person-years. Other examples include anaphylaxis associated with penicillin (risk window is minutes after exposure), hospitalization for peptic ulcer disease associated with NSAIDs (risk is highest in first 30 days after exposure),[55] and angioedema with angiotensin converting enzyme inhibitors [56](risk is logarithmically higher in the first week of exposure). Other types of risk windows that occasionally are important include delayed onset where the risk begins after some defined period of drug exposure, and carry over windows where there is a period of persisting risk (i.e., delayed recovery from risk) after discontinuing the medication. The length of time the adverse event risk persists can range from days (e.g., NSAIDs and gastrointestinal hemorrhage; acetaminophen and drug-induced liver

disease) to weeks (e.g., phenytoin and drug-induced liver disease) to months or even years (e.g., isotretinoin and birth defects).

5.2 VACCINES

5.2.1 Post-approval vaccine safety surveillance

In 1990, two federally funded programs, the Vaccine Adverse Event Reporting System (VAERS) and the Vaccine Safety Datalink (VSD) were established to support vaccine safety surveillance. The Centers for Disease Control and Prevention (CDC) and Food and Drug Administration (FDA) jointly administer VAERS.[11] VAERS is a passive surveillance system that accepts reports of possible adverse events following immunization (AEFI) from the vaccine manufacturing industry, health professionals and the public. While there are no restrictions on reporting possible AEFIs, reporting bias and lack of denominator data limit VAERS from being used to establish causation.

VSD is a collaborative project between the CDC and several managed care organizations (MCOs) across the United States.[8] The participating MCOs comprise a population of over 10 million members annually (~3% US population). VSD conducts both retrospective and near-real time population–based vaccine safety studies. Retrospective observational studies test hypothesis from the medical literature, VAERS, and public concern, or generate hypotheses from large cohorts of vaccinated individuals that are screened for several medically-attended events following vaccination. For near-real time safety surveillance of newly approved vaccines, a Rapid Cycle Analysis (RCA) team creates weekly analytic data sets, creates study cohorts, and analyzes them with appropriate statistical methods.[9]

The federally funded Post-Licensure Rapid Immunization Safety Monitoring (PRISM) program began in 2009 for monitoring the safety of the H1N1 influenza vaccine.[57] PRISM now monitors the largest US general population cohort (~ 10% of US) for active safety surveillance of several vaccines. PRISM is based on data from national health insurance plans and state and city immunization registries and is now part of the Mini-Sentinel program. PRISM data are updated quarterly and medical record review can be accomplished.

5.2.2 Issues specific to vaccines

5.2.2.1 Vaccine exposure misclassification

Vaccines are acute exposures, usually administered during medical office visits, on a predetermined schedule. Some vaccine administrations occur in retail pharmacy (e.g., influenza and zoster vaccines) or emergency department (e.g., tetanus vaccine) settings. The data in electronic vaccine registry databases are generally considered of high quality.[58] The main source of possible exposure misclassification is uncertainties in the risk window length. If the risk window following vaccination is too long, then a portion of unexposed time may be classified as exposed, if too short, then the opposite may occur. Usually a risk window length, typically in the range of 3 days to 2 months is specified from what is known from previous safety studies and/or biologic plausibility. Data driven approaches currently being investigated show some utility in quantitative risk window determinations.[59]

5.2.2.1 Healthy vaccinee effect

The healthy vaccinee effect is a postulated selection bias in which vaccinated individuals are less likely to have had a visit to a medical provider due to a serious illness in the days preceding vaccination.[60] The

corollary is that vaccination is deferred among those who are acutely ill.[61] The time of vaccination is in one of the healthier periods, as measured by the use of health services data.

5.2.2.2 Confounding by contra-indication

Confounding by contra-indication can occur if unvaccinated individuals are used as a comparison group. Contraindicated individuals often have medical conditions that result in higher morbidity (or ultimately mortality) than in the general population. Fine and Chen[62] in 1992 showed that earlier studies of diphtheria, tetanus, and (whole-cell) pertussis (DTP) vaccine and sudden infant death syndrome or encephalopathy had likely underestimated the risk due to confounding by indication. However, established contraindications to a specific vaccine are relatively rare in pediatric vaccine safety studies, as they are usually related to rare congenital conditions.[63]

5.2.2.3 Seasonality

Disentangling the effects of a seasonally occurring adverse event from a vaccine administered one during specific seasons is particularly problematic. For example, over 85% of influenza vaccines are given in October or November in any particular year.[64] Many of the adverse events, such as febrile seizures, also have high incidence during these months. No analytic methods are completely effective in controlling for this seasonal cofounding in some vaccine safety studies.[65]

5.2.2.4 Multiple vaccines and combination vaccines

In the US, pediatric vaccinations are administered to most children according to recommended schedules.[66] There are now up to 11 vaccines given in the first year of life to protect against 15 vaccine preventable diseases. Increasingly, vaccines are being combined into one administered shot, for example, diphtheria, tetanus, acellular pertussis, inactivated poliovirus, and Haemophilus influenzae type B combination vaccine (DTaP-IPV-Hib).

When multiple and combination vaccinations are administered simultaneously, determining whether events are attributable to particular components or one of several combinations is frequently difficult or impossible.[64,65]

5.3 DEVICES

Devices are a very broad and diverse group of medical products that primarily function by means other than chemical or being metabolized in humans or animals.[14] A detailed legal definition of “device” is in section 201(h) of the Federal Food Drug & Cosmetic (FD&C) Act:

> “an instrument, apparatus, implement, machine, contrivance, implant, in vitro reagent, or other similar or related article, including any component, part, or accessory, which is—(1) recognized in the official National Formulary, or the United States Pharmacopeia, or any supplement to them, (2) intended for use in the diagnosis of disease or other conditions, or in the cure, mitigation, treatment, or prevention of disease, in man or other animals, or (3) intended to affect the structure or any function of the body of man or other animals, and which does not achieve its primary intended purposes through chemical action within or on the body of man or other animals and which is not dependent upon being metabolized for the achievement of its primary intended purposes. “

The FDA classifies medical devices into 3 categories according to the controls necessary to establish safety and effectiveness.[67] Examples of Class I devices are disposable gloves, tongue depressors, and dental floss. Class I devices are considered to be low risk and have the least regulatory control. Class II devices, such as infusion pumps, surgical or acupuncture needles, have a greater risk and thus have a higher level of regulatory control. Class III devices are the most complex, the most invasive, have the highest risk and the most control. Examples are artificial heart valves, implanted joints, and implanted defibrillators.

5.3.1 Post-approval medical device safety surveillance

Unlike other medical products, most medical devices are chronic implants and designed to provide a life-time worth of benefits after they have failed more conservative treatments. Although many devices are approved to be marketed after thorough evaluation of clinical pre-market studies (PMAs), the safety and effectiveness of the device may deviate substantially from what was observed under the typically very controlled study conditions of short duration. Therefore, most information about the safety and effectiveness of a particular medical device is gained in the postmarket setting. The FDA Center for Devices and Radiological Health (CDRH) has implemented several surveillance systems to monitor device safety post-marketing:

The Manufacturer and User Facility Device Experience (MAUDE) database is the FDA's mechanism for collecting all adverse events involving medical devices; it is a passive system which relies on spontaneous reports from patients, healthcare practitioners, distributors and manufactures.[68] Much like AERS previously described, MAUDE is subject to underreporting, incomplete reporting and lack of denominator data. Dissimilar to AERS, however, MAUDE reports lack unique identifiers for the devices for which the event is being reported, making it difficult to identify the specific exposure.

The Medical Product Safety Network (MedSun) is an adverse event reporting program launched in 2002 which consists of a voluntary sample of over 350 hospitals and health care facilities.[69] On site representatives report adverse medical events directly to the FDA. This system is still under development and currently functions most similar to a passive reporting system such as the MAUDE, along with the similar limitations.

Apart from these passive systems, under 21 CFR 814.82 the FDA has the authority to impose post-approval studies of Class III devices. This is to continue evaluation and reporting on the safety, effectiveness and reliability of the device for its intended use. In addition, under Section 522, the FDA has the authority to order postmarket surveillance studies for Class II or Class III medical devices that meet the statutory criteria. These mandated studies aim to evaluate specific safety and/or effectiveness concerns that were identified during the pre-market review cycle, through review of the literature, or through the passive surveillance systems described above. Many device manufactures and professional societies are beginning to develop patient registries either as part of required post-market studies or for research and surveillance purposes.

5.3.2 Issues specific to devices

5.3.2.1 Source of exposure data (registry, health insurance claims)

Device exposure can be measured through various sources such as patient registries and health insurance system administrative claims. Examples of registries are the Kaiser Permanente National Total

Joint Replacement Registry,[70,71] or the Interagency Registry for Mechanically Assisted Circulatory Support (INTERMACS).[72] Within registries, there is a large depth of information on the actual procedure and baseline demographics of the patients receiving a medical device with minimal selection bias. However, there is often limited information on longer term outcomes, and these registries are not structured to follow outcomes associated with a specific device. Moreover, most multicenter registries are relatively small with less than 5000 patients needed to detect and quantify a 1-in-200 adverse event frequency with adequate power.[73] They further suffer from limited external validity as the data is most often generated at large referral centers with highly experienced operators.[74]

Most health insurance data systems only capture procedure codes for medical device installation yet have substantial follow-up medical encounter data. For example using this approach to evaluate the impact of a medical device on subacute stent thrombosis rates requires the use of a surrogate for stent thrombosis such as rehospitalization for acute myocardial infarction, repeated catheterization, and/or intervention. Such surrogates are intrinsically limited because acute myocardial infarction or repeated coronary intervention may occur for reasons other than subacute stent thrombosis. Furthermore, using claims-based data does not provide adequate data from which to discern operator/procedural behavior and selection bias, for example, in identifying limitations in device availability based on individual anatomical composition. This issue is not unique to medical devices but is unique to claims-based data for research/surveillance purposes. The claims data also typically do not distinguish between essentially the same products made by different manufacturers, thus making desirable comparisons, very difficult. Finally, gaps in coverage must be taken into account when longer-term outcomes are assessed.

5.3.2.2 Assessing medical device exposures

An important current limitation of medical devices is the lack of a unique device identifier that facilitates tracking of patient level exposures.[75] Although a proposed rule that most medical devices distributed in the US carry a unique device identifier (UDI) was recently published in the Federal Registrar, it has yet to be fully implemented.[76] At present, devices lack unique identifiers, making exact source exposure considerably difficult. This further limits the potential for identifying a single poor performer within a device class. Once a UDI is implemented, it will allow for more accurate exposure data and provide the ability of linking data on specific patients across registries, claims data, electronic health records and other databases for more comprehensive device surveillance, and potential validation methods.

While uncertainty in device exposure is not as great as with drugs, there are circumstances where it is not clear cut. Most exposures to implantable devices are chronic (with exception of temporary implantables, such as inferior vena cava filters), with the onset of exposure clearly defined at implantation. However, in some cases such as breast implants even after exposure ends with implant removal there could be silicone leakage of a ruptured breast, or as with dermal fillers absorbed over time, making the duration of exposure less obvious.[14]

5.3.2.3 Confounding by indication

As for drugs, the principles of confounding by indication apply to device safety studies. Similar analytical techniques described previously can be used to help to control for confounding-by-indication.

5.3.2.4 Product similarities/differences

Unlike drugs and vaccines in which a single change in the molecular structure of the compound will constitute a new product, the structural components of devices can be changed and modified and so long as the modifications are considered "substantially equivalent" to the predicate device iterative versions of the device are considered the same. This allows for the potential of unforeseen differences in performance and risk of events, based on only subtle differences in the composition of the device. To account for potentials in the difference in study outcomes by the device implanted, a standard poolability analysis can be carried out.

5.3.2.5 Learning curve bias for device use/implantation

When new devices are being studied in an early phase, most often elite investigators are selected to participate. Once the device becomes more saturated in clinical practice, new or less experienced uses of a medical device can undergo a substantial learning curve, which is defined as a constant proportional improvement in performance such as clinical outcomes of medical procedure, with each doubling of cumulative experience.[77] This learning curve can have a substantial impact on event patterns observed, and may be difficult to account for without having data on a provider's or an institution's experience in terms of volume with the device in question.

5.3.2.6 Variable device product life

There are many factors that play a role in the total product life cycle of a medical device. Not only could outcomes be affected by underlying patient factors and device factors (such as biomaterials) but also importantly by the settings and use of the device by the patient and provider. For example the failure modes of many types of neurostimulators which function by providing electrical stimulation to various parts of the central or peripheral nervous system through the use of an implantable battery pack are dependent on the settings programmed by the provider and/or patient. Further compounding this obstacle is the fact that there can be a large amount of fluctuation in the settings of device for the patient over the course of therapy. Rates of adverse events such as loss of therapy or overstimulation for the same device can therefore vary from patient to patient, and could only be accounted through the use of a multivariate analysis by having detailed data collection on device settings. This level of data capture is often times very difficult leading to the potential for a large amount of unmeasured confounding beyond that accounted for in the collected information, leading to the possibility of having to rely on proxy adjustments or instrumental variable-based methods.[78]

5.3.2.7 Drug-Device, Device-Device, Interaction

Devices are frequently approved or used as systems treating an overall pathological condition involving several components including drugs and/or other devices; device components can also be used in combination to deliver a drug such as drug eluting stents. Devices can also be approved to be used in the presence of another device such as performing NovaSure ® endometrial ablation in the presence of Essure ® microinserts for tubal sterilization. In this case an observed rate of fallopian tubal perforation may be difficult to ascertain as to whether it was the result of either the NovaSure ® procedure alone, the Essure microinserts alone or the result of thermal heating of the Essure device caused by the NovaSure procedure.

A pharmacological treatment used in conjunction with a medical device to provide treatment for the same indication is an effect modifier for the device. Similarly, when devices are used in combination such as NovaSure® in the presence of Essure ® there is device-device interaction.

Effect modification can be assessed by examining the exposure-adverse event associations in subgroups (calculating rates of events among patients that only have one device or not receiving concomitant pharmacological treatment versus those that have both devices and/or receiving concomitant pharmacological treatment) and by statistical modeling, for example, use of interaction terms in regression analysis.

5.3.2.8 Environmental factors

Environmental factors on a medical device may pose significant effects on additional hazards. The engineering components of medical devices are subject to many environmental factors such as electromagnetic interference (EM) originating from sources as diverse as MRI machines, anti-theft detectors, power lines etc. These sources may cause device malfunctions leading to unintended events such as burns, loss of effect, or overstimulation. Trauma to the device may also cause a device to malfunction leading to unexpected adverse events. Environmental factors such as those listed above are all potential confounders.

5.4 BLOOD COMPONENTS AND BLOOD DERIVATIVES

5.4.1 Safety surveillance of blood components and blood derivatives

The medicinal products manufactured from human donor blood generally fall into two broad categories: 1) whole blood and blood components that are prepared as units from a single donation of an individual donor and 2) blood derivatives that are prepared from pools of material (typically plasma) obtained from many donors. The circumstances related to the production and to the use of these two categories of products are sufficiently different that the FDA requirements for the reporting of adverse events are distinct for these two groups of products. Except in the case of fatal adverse events, the collecting or transfusing facility is not required to report to FDA any adverse events related to administration of blood or blood components but must analyze and record these events locally (21CFR606.170). In contrast, a licensed manufacturer is required to report adverse events related to the administration of blood derivatives to FDA's Adverse Event Reporting System (AERS) according to the provisions set out for other biological products (21CFR600.80).

This dichotomy in the reporting requirements for blood-derived products reflects some fundamental differences in the production and use of these two related groups of products. Whole blood and blood components are collected, subjected to quality control testing, and processed in myriad facilities and then administered by transfusion services, often after additional processing and laboratory testing of the donor and the blood component. Each unit of whole blood or blood component is effectively a lot and can vary substantially from other units, depending upon the genetic background and the environmental exposures of the donor and also depending upon the specific methods and materials used in the processing, packaging, and laboratory testing of the collected material. As a consequence, the safety characteristics of each individual blood component unit may differ to a degree that is substantially higher than what can be attained for synthetic drugs. In the case of blood derivatives, the

pooling of material from large numbers of donors can mitigate this variability by averaging the composition and diluting potentially noxious substances (e.g., certain antibodies). Additionally, many blood derivatives are subjected to purification and decontamination processes that further increase the product's homogeneity and reduce the risk from impurities like infectious disease agents.

The special characteristics of the production and distribution of blood components have recently spurred the development of hemovigilance programs in many parts of the world.[79] In the U.S., the CDC has implemented a Hemovigilance Module for its National Healthcare Safety Network.[80] Participating transfusion services report blood component-related adverse events to CDC according to detailed case definitions on standardized forms. It is anticipated that the accumulated data will be useful in identifying informative patterns in the occurrence of adverse events associated with transfusion of blood components. In addition to the passive reporting activities noted above, efforts to obtain information regarding the occurrence of adverse events in conjunction with the administration of blood components and blood derivatives using active surveillance approaches have been recently reported. Menis and associates have examined the occurrence of serious noninfectious transfusion-related complications among U.S. elderly for the years 2007 and 2008 by analyzing administrative claims codes submitted to Medicare.[81] Also, Daniel and colleagues have recently published their findings from a study of thrombotic adverse events associated with administration of immunoglobulin preparations as detected in insurance claims found in the HealthCore Integrated Research Database for 2008 through September of 2010.[82] Thus, safety surveillance activities for blood components and blood derivatives have been evolving rapidly in recent years and include claims-based active surveillance activities that can help to inform the efforts of Mini-Sentinel with regard to surveillance for these product categories.

5.4.2 Issues specific to blood components and blood derivatives

In most respects, blood derivatives and blood components resemble conventional drugs quite closely in terms of the potential sources of confounding. However, variability in the properties of blood components could introduce some additional complexities. Certain types of variability (such as changes in product performance due to long periods of storage) would be expected to be distributed fairly homogeneously across healthcare settings represented within the data provided by Mini-Sentinel data partners. This more homogeneous type of effect would yield a safety signal that would be expected to be reliably detected by Mini-Sentinel. However, some potentially deleterious characteristics (such as those that might result from a particular set of donor policies and/or processing steps) could occur in only a restricted set of healthcare facilities. This more clustered type of product heterogeneity might then be detected in a less reliable fashion. Depending upon the distribution of this characteristic, a safety signal deriving from such a clustered effect might be detected by nationwide passive surveillance activities but not within the sample of information provided by the Mini-Sentinel data partners. Conversely, a safety signal that was marginally detected by passive surveillance efforts might yield a very strong signal from Mini-Sentinel, should the data partner coverage of the effected healthcare facilities be fortuitously favorable. Consequently, clustering of variability in certain qualities of blood components could conceivably lead to misperceptions of certain safety signals for blood components in either direction and might result in false-negative results or overly positive assessment of signals.

6. A PROCESS FOR ASSESSING SIGNAL REFINEMENT RESULTS

Assessment of a safety signal from a signal refinement activity should proceed in a series of steps. If the signal can be largely explained in a particular step, then some of the subsequent steps may not be necessary.

6.1 DEVELOP A PRODUCT SPECIFIC ASSESSMENT PLAN

The first step should be to develop a product specific protocol for assessing a positive signal. A protocol for assessing a positive signal should be developed during the study planning phase. This protocol should be completed before actual safety surveillance begins for the specific medical product of interest. It should first state what explicitly defines a positive safety signal. It is assumed to be based on the association of a medical product exposure and adverse event outcome that exceeded a pre-specified threshold in the direction of increased risk. The pre-specified threshold would be determined by the statistical methods and/or expert judgment. A statistical threshold should account for the precision (random error) of the signal estimate, in addition to other considerations such as multiple comparisons in time (for longitudinal sequential analysis). The details of possible statistical methods are beyond the scope of this framework, but other Mini-Sentinel Reports have addressed[3-7] or are expected to address this. Expert judgment on establishing a detection threshold may be particularly important for rare and serious outcomes. It is conceivable that just 2 exposed cases compared to none unexposed could be worrisome, yet it would not necessarily be statistically significant.

Second, the protocol should include plans for data error checking among study sites that contributed data that generated the signal. Third, the protocol should describe possible threats to validity in consideration of specific medical product issues such as likely sources of systematic error. Fourth, investigators need to describe possible secondary analyses such as examining patterns in time from exposure to outcome, adjusting for additional confounders, and using other comparison groups. Finally, if quantitative bias analysis is planned, it needs to be described based on the threats to validity anticipated and the resources available.

6.2 REVIEW DATA VALIDITY, DESCRIPTIVE STATISTICS, AND ANALYTIC COMPUTER PROGRAMS

Once it has been established that a positive signal was detected, the most expedient next step is to review the data. The emphasis should be on ruling out errors in the data that contribute to the signal such as event counts (number of cases) and other important case information. Basic data verification steps would involve checking for missing values, invalid character values, outliers among numeric values, and variable (covariate) coding and programming errors. Observed and expected counts and rates should be scrutinized and compared with each other and with incidence rates from the literature. Tables and graphs of descriptive statistics should be assembled to allow examination of rates by age, gender, data-providing organization, calendar year, possibly season, and simultaneous exposure to other drugs/products. These examinations can turn up explanations for the signal such as errors in background rates used, changes in coding practices at a participating organization, confounding due to secular trends, or inadequate adjustment for other confounders.

The assessment of data validity should be a regular activity before a safety signal occurs. Ideally, a signal would arise from a validated population-based cohort designed for pharmacovigilance.[83] To ensure this, a data coordinating center would enforce uniform data quality of all site data files via a common data model built for medical research purposes. However, electronic health care data systems are designed

for clinical care or billing and administrative functions and not necessarily for medical research. As such, data not missing-at-random can be an issue. Missing data can arise from data collected during a clinical encounter but not entered into a searchable database (but possibly entered in an individual patient's medical record) or data that were not obtained (but possibly should have been) and not present in any data system, including a patient's medical record. The distinction is that in the first instance, the data missing in an electronic database may be recovered by laborious medical record review. In the second instance, the data are completely missing and unrecoverable.

Regular data validity checks are particularly challenging under (near) real-time active surveillance for newly introduced medical products. In this context, data validity reviews should ideally occur at the same frequency as data are acquired (e.g. weekly). Completeness and timeliness of data are competing factors in real-time active surveillance, however, and the lag may vary by data source and by type of medical encounter. For instance, pharmacy data are commonly available in real-time, outpatient encounters can take a few days to process, and inpatient hospital encounters can take several months to process if originated from health insurance claim data. From the Mini-Sentinel Distributed Database (MSDD)experience, the average time lag is 6 months. The lag time may be even longer when using other data sources.

The possibility of error in the analysis computer programs should also be considered and the code checked. An equivalent or very similar analysis should be carried out using the same data as the data generating the signal. For example, if the signal emerged during sequential analysis, an equivalent non-sequential analysis should be done, which should produce similar risk estimates as the original analysis if both were done correctly.

6.3 CONDUCT SECONDARY ANALYSES

6.3.1 Examine patterns in time from exposure to outcome

The number of cases by day after the exposure (or after the start of the exposure period) can be visually graphed. If there is no relationship between exposure and outcome, the cases should be more or less uniformly distributed in time. Temporal scan statistics can also be used for detecting possible temporal clusters of outcomes events (cases) following exposure events.[84,85] This is usually performed after a positive signal is detected to further refine the risk window length in subsequent secondary analysis. The temporal scan statistic can be used for formal statistical inference; it adjusts for the multiple testing of many possible risk windows. If the graph or temporal scan statistical test indicates temporal clustering of outcomes after exposure, the risk window can be narrowed to include only this period and the analyses repeated, which would tend to result in higher relative risk estimates.

6.3.2 Adjust for additional confounders

If the examination of descriptive statistics reveals possible confounders that might not have been adequately controlled for, these should be dealt with by applying standard pharmacoepidemiologic methods such as multi-variable regression to the same dataset. Such analyses might include finer adjustment for age, adjustment for seasonal trends by month or by use of sinusoidal curves, adjustment for secular trends using a linear or other function, adjustment for chronic diseases or propensity to receive the drug or product of interest, and adjustment for one or more specific simultaneous exposures such as another drug or vaccine.

6.3.3 Use other comparison groups

Another step in the investigation could be to use different comparison groups than in the original analysis, for example, historical comparison groups from different periods, concurrent controls that are selected or matched based on different criteria, and different control periods in self-controlled designs.

6.4 CONSIDER QUANTITATIVE BIAS ANALYSIS

If after the previous steps, the positive signal still generated an estimate of reasonable precision, then quantitative bias analysis may be feasible. The objective is to determine the maximum amount of systematic error that could be tolerated without obviating the signal.

Sources of systematic error most likely to influence the results from signal refinement need to be considered first: measurement error (misclassification for dichotomous measures), selection bias, and/or unmeasured confounding. Next, appropriate bias parameters and their values need to be determined for each likely source of systematic error. Enough information on bias parameters might be gathered from the research literature. For example, if obesity was an unmeasured confounder, then the research literature might provide estimates of the likely prevalence of obesity in study population among users and nonusers of the medical product under investigation, along with an estimate of the relative risk of obesity to the outcome of interest. The estimates of prevalence and relative risk are bias parameters that would be applied in a quantitative bias analysis.

In terms of choosing a method of quantitative bias analysis, it is important to note that estimates of bias parameters are themselves measured with error. Probabilistic bias analysis can account for the measurement error in bias parameters and the systematic error they induce. Probabilistic bias analysis begins with assigning probability distributions to the bias parameters; however bias parameters are still analyzed one at a time. Monte Carlo methods are used to repeatedly sample from the probability distributions and to correct for the bias. The main advantage of probabilistic bias analysis is that a frequency distribution of revised risk estimates are produced that also incorporate random error. The mean of the distribution and its 95% simulation interval can be interpreted similarly to a conventional point estimate and its confidence interval.

If quantitative bias analysis is properly performed, then the possible impact of misclassification, selection bias, and confounding will be assessed. The challenge is to accurately model the most relevant biases for the specific medical product and outcome pairs of interest. The particular medical product, adverse event outcome, surveillance database and the judgment of the investigators all influence the choice and performance of quantitative bias analysis methods.

Electronic resources for performing quantitative bias analysis include user developed spreadsheets or statistical program user defined macros and functions. As a companion to the Lash, et al text,[43] Microsoft® Excel spreadsheets were set up for probabilistic bias analysis. A SAS® macro for performing multiple biases modeling these spreadsheets, including a SAS® macro for performing multiple biases modeling, are currently available via the Internet.[86]. STATA® also has certain functions that are useful for quantitative bias analysis. A description of these along with a user defined macro command that performs both simple and probabilistic bias analysis is described in the STATA journal.[87] Other epidemiologic software also have simple sensitivity analysis functions such as WINPEPI.[88]

When reporting the results of a quantitative bias analysis, the following steps should be followed. First, a clear rationale should be stated for assigning values and distributions (if used) to bias parameters.

Second, the sources of data for the bias parameters should be described, whether from an internal source within the overall study or externally, using published data from other studies. Third, the method of bias analysis should also be described so that the reader of the report will see the connection from the conventional results to the bias adjusted results via the mathematical application of the bias parameters. Fourth, if multiple biases are analyzed, then a description should be completed for each bias and the rationale for the order of corrections should be clearly explained.

6.5 INTERPRET AND REPORT ASSESSMENT

The interpretation and reporting of the signal assessment activity should be based on the protocol and the findings. The report should generally reflect if and where the excess positive risk could be explained by something other than a cause and effect relationship, such as data or systematic errors (information or selection bias, confounding). The report should be written in the order of the assessment steps done as described in the previous sections. If the assessment included a quantitative bias analysis, then details of the analysis should be described.

A checklist for assessing signal refinement results is presented in Appendix B.

7. ACKNOWLEDGEMENTS

We thank Elizabeth Cavagnaro and Madhavi Vajani for their expert project management, Ryan Carnahan, Robert Greenlee, Sean Hennessy, and Richard Platt for their helpful comments and suggestions. We especially thank Richard Platt for his overall leadership of Mini-Sentinel.

8. APPENDIX A: WORKED EXAMPLES OF QUANTITATIVE BIAS ANALYSIS

8.1 THE IMPACT OF AN UNMEASURED CONFOUNDER

Rofecoxib, a non-steroidal anti-inflammatory (NSAID) selective Cox-2 inhibitor, was withdrawn from the US market in 2004. Several population based observational studies have shown that rofecoxib at daily doses of greater than 25 mg was associated with an approximate two-fold risk of AMI. [89] While these studies had adjusted for potential confounders such as age, sex, vascular risk factors, comorbidities and prescribed medications, other important confounders were generally not controlled: non-prescription aspirin or NSAIDs, alcohol use, body mass index (BMI), or tobacco use.

Consider a hypothetical case-control surveillance study of AMI in 1000 rofecoxib users compared to 4000 naproxen users. The data in Table 3 generated an odds ratio of 2.026 with 95% confidence interval of (1.247, 3.291). However this odds ratio is not adjusted for the important unmeasured confounder of obesity. If we have other information of the odds ratio of obesity (assessed as a binary covariate) and AMI, and the prevalence of obesity among both exposure groups, then a corrected odds ratio can be generated by probabilistic bias analysis.

Table 3. Hypothetical case-control study

	AMI	no AMI	
rofecoxib	25	975	
naproxen	50	3950	
odds ratio : 2.026 (1.247, 3.291) p=0.0043			

From previous published data[90], the prevalence of obesity (BMI≥30) was 19% among rofecoxib users, 20% among naproxen users., and the odds ratio of obesity (BMI≥30 vs. BMI<30) to AMI was 1.7. Triangular probability distributions can be defined for these bias parameters. For the prevalence of obesity, the minimum = 0.1, the mode=0.2 (peak of the triangle) and the maximum = 0.3. This distribution will be used for both exposure groups. For the odds ratio of obesity and AMI , the minimum =1.4, the mode=1.7, and the maximum=2.0. An internet available Microsoft Excel application, available as a companion to the Lash, et al text[43], will be used for the simulation.[86] Using these bias parameters with 10000 simulations, shows that the impact of the unmeasured confounding of obesity was minimal. The corrected odds ratio and 95% interval for the risk of AMI of rofecoxib users compared to naproxen 2.02 (1.23-3.31). This is very close to the original odds ratio of 2.03 (1.25-3.29).

Using these spreadsheet tools, the maximum amount of unmeasured confounding from obesity that would push the original result to non-significance can also be estimated. As shown in **Figure 2**, the prevalence of obesity in the rofecoxib users would have to be a mode of 70% to obviate the original association of AMI and rofecoxib use, corrected odds ratio: 1.51 (0.96-2.54). This level of obesity relative to the comparison group is unlikely.

PROBABILISTIC SENSITIVITY ANALYSIS UNMEASURED CONFOUNDING

This spreadsheet can be used to conduct a probabalistic sensitivity analysis to correct for unknown of unmeasured counfounding and random error simultaneously.

Input Bias Parameters

	Min	Mode	Mode2	Max
p_1	0.60	0.70	0.70	0.80
p_0	0.10	0.20	0.20	0.30
RR_{DC}	1.40	1.70	1.70	2.00
Sims	100			

Error	
No error	
Corr p	0.80

Instructions

Enter distributions for the bias parameters in the blue cells to the left and the crude data in the blue cells below. Cells in green give the results after adjusting for the unmeasured confounder. Note that white cells are expected values and therefore do not have to be integers.

Run Simulation

Variable Names			
Confounder	Obesity	Exposure	rofecoxib
		Outcome	AMI

Data (Enter Crude rofecoxib-AMI Data in Blue Cells)

	Total		Obesity +		Obesity -	
	rofecoxib	rofecoxib -	rofecoxib +	rofecoxib -	rofecoxib +	rofecoxib -
AMI +	25 a	50 b	18.4 A_1	11.8 B_1	6.6 A_0	38.2 B_0
AMI -	975 c	3950 d	627.8 C_1	667.7 D_1	347.2 C_0	3282.3 D_0
Total	1000 m	4000 n	646.2 M_1	679.5 N_1	353.8 M_0	3320.5 N_0

Crude and Adjusted Measures of rofecoxib-AMI Relationship

Crude	Measure (95% CI)
RR (rofecoxib	2 (1.24 - 3.22)
OR (rofecoxib	2.03 (1.25 - 3.29)

Adjusted SMR	
RR (rofecoxib-AMI)	1.63
OR (rofecoxib-AMI)	1.65

Chosen Values	
p_1	64.6%
p_0	17.0%
RR_{DC}	1.51

rofecoxib-AMI Relationship Adjusted for Obesity

RR Simulation Results (N=101)	
Analysis	Median (2.5^{th}-97.5^{th} percentile)
Conventional	2 (1.24 - 3.22)
Systematic	1.52 (1.38 - 1.65)
Total Error	1.56 (1.01 - 2.65)

OR Simulation Results (N=101)	
Analysis	Median (2.5^{th}-97.5^{th} percentile)
Conventional	2.03 (1.25 - 3.29)
Systematic	1.54 (1.39 - 1.67)
Total Error	1.51 (0.96 - 2.54)

Illegal Values
0

Figure 2. Probabilistic bias analysis of the impact of obesity as an unmeasured confounder on the risk of acute myocardial infarction (AMI) following use of rofecoxib. [89]

8.2 THE IMPACT OF OUTCOME MISCLASSIFICATION

Using the previous example, consider a hypothetical differential misclassification of the AMI outcome. For illustrative purposes, assume that AMI would be more likely to be diagnosed following rofecoxib exposure compared to naproxen. For the rofecoxib group the triangular distributions of sensitivity was assigned minimum=0.9, mode=0.95, maximum=1.0, and for the naproxen group, minimum=0.85, mode=0.9, maximum=1.0. Specificity was assigned minimum 0.95, mode=0.97, maximum=1.0, for both groups. As shown in Figure 3, impact of outcome misclassification, given these bias parameters was severe. The calculated odds ratio interval due to systematic error was wide: 2.49 (0.1 - 346), which completely overwhelms the variation due to random error.

PROBABILISTIC OUTCOME MISCLASSIFICATION

This spreadsheet can be used to conduct a probabalistic sensitivity analysis to correct for outcome misclassification and random error simultaneously.

Input Bias Parameters

	Min	Mode	Mode2	Max	Error
Se (rofecoxib+)	0.90	0.95	0.95	1.00	No error
Se (rofecoxib-)	0.85	0.90	0.90	1.00	Corr Se
Sp (rofecoxib+)	0.95	0.97	0.97	1.00	0.80
Sp (rofecoxib-)	0.95	0.97	0.97	1.00	Corr Sp
					0.80

Misclassification type: D Sims: 100

Instructions

Enter the bias parameter distributions in the blue cells the left and the observed data in the blue cells below Cells in green give the results after correcting for outco misclassification. Note: green cells are expected value and therefore do not have to be integers.

Run Simulation

Variable Names

Exposure	rofecoxib	Outcome	AMI

Data (Enter rofecoxib-AMI Data in Blue Cells)

Observed Data

	rofecoxib	rofecoxib -	Total
AMI +	25 a	50 b	75
AMI -	975 c	3950 d	4925
Total	1000 m	4000 n	

Single Simulated Corrected Data

Corrected Data

	rofecoxib +		rofecoxib -		Total
AMI +	-15.3	A_1	-142.1	B_1	75
AMI -	1015.3	C_1	4142.1	D_1	4925
Total	1000.0	M_1	4000.0	N_1	

Chosen values	
Se(E+)	95.6%
Se(E-)	90.5%
Sp(E+)	96.1%
Sp(E-)	95.7%

Observed	Measure (95% CI)
RR (rofecoxib-AMI)	2 (1.24 - 3.22)
OR (rofecoxib-AMI)	2.03 (1.25 - 3.29)

Single Corrected Estimate	Measure
RR (rofecoxib-AMI)	0.43
OR (rofecoxib-AMI)	0.44

RR Simulation Results (N=101)

Analysis	Median (2.5th-97.5th percentile)
Conventional	2 (1.24 - 3.22)
Systematic	2.45 (0.1 - 343.31)
Total Error	2.45 (0.09 - 422.18)

OR Simulation Results (N=101)

Analysis	Median (2.5th-97.5th percentile)
Conventional	2.03 (1.25 - 3.29)
Systematic	2.49 (0.1 - 346.13)
Total Error	2.66 (0.1 - 320.9)

Illegal Values
1744

Figure 3. Probabilistic bias analysis of the impact of outcome misclassification on the risk of acute myocardial infarction (AMI) following use of rofecoxib. [89]

9. APPENDIX B: CHECKLIST FOR ASSESSING SIGNAL REFINEMENT RESULTS

□ 1. Develop a product specific assessment plan

- Explicitly define what would be a positive safety signal.
- Include plans for data error checking.
- Describe possible threats to validity.
- Describe possible secondary analyses.
- Plan for quantitative bias analysis based on the threats to validity anticipated and the resources available.

☐ 2. Review data validity, descriptive statistics ,and analytic computer programs

- Rule out errors in the data.
- Examine descriptive statistics.
- Check analytic computer program code

☐ 3. Conduct secondary analyses

- Perform an equivalent or similar analysis using the same data.
- Examine patterns in time from exposure to outcome.
- Adjust for additional confounders.
- Use other comparison groups.

☐ 4. Consider quantitative bias analysis

- Assess feasibility of quantitative bias analysis (positive signal generated an estimate of reasonable precision).
- Consider sources of systematic error most likely to influence the results.
- Determine appropriate bias parameters, their values, and the quantitative bias analytic method.
- Perform quantitative bias analysis to determine the maximum amount of systematic error that could be tolerated without obviating the signal.

☐ 5. Interpret and report assessment

- Based on the protocol and the findings of the assessment.
- Should generally reflect if and where the excess positive risk could be explained by something other than a cause and effect relationship.
- If the assessment included a quantitative bias analysis, then details of the analysis should be described.

10. REFERENCES

[1] Robb MA, Racoosin JA, Sherman RE, Gross TP, Ball R, Reichman ME, Midthun K, Woodcock J. The US Food and Drug Administration's Sentinel Initiative: expanding the horizons of medical product safety. Pharmacoepidemiol Drug Saf. 2012 Jan;21 Suppl 1:9-11.

[2] Platt R, Carnahan RM, Brown JS, et al. The U.S. Food and Drug Administration's Mini-Sentinel program: status and direction. Pharmacoepidemiol Drug Saf. 2012 Jan;21 Suppl 1:1-8.

[3] Gagne JJ, Fireman B, Ryan P. Taxonomy for monitoring methods within a medical product safety surveillance system: Report of the Mini-Sentinel Taxonomy Project Work Group. September 22, 2010.

[4] Cook AJ, Tiwari RC, Wellman RD, et al. Mini-Sentinel methods development Sequential Testing Working Group report. May 12, 2011.

[5] Jeremy A. Rassen JA, Myers J. Mini-Sentinel methods development signal evaluation working group report. August 12, 2011.

[6] Gagne JJ, Nelson JC, Fireman B. Taxonomy for monitoring methods within a medical product safety surveillance system: year two report of the Mini-Sentinel Taxonomy Project Workgroup. January 31, 2012.

[7] Cook AJ, Wellman RD, Marsh TL. Mini-Sentinel methods development: Statistical methods for estimating causal risk differences in the distributed data setting for postmarket safety outcomes. March 1, 2012.

[8] Baggs J, Gee J, Lewis E, et al. The Vaccine Safety Datalink: a model for monitoring immunization safety. Pediatrics. 2011 May;127 Suppl 1:S45-53.

[9] Yih WK, Kulldorff M, Fireman BH, et al. Active surveillance for adverse events: the experience of the Vaccine Safety Datalink Project. Pediatrics. 2011; 127 Suppl 1:S54-64.

[10] 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.

[11] Varricchio F, Iskander J, DeStefano F, et al. Understanding vaccine safety information from the Vaccine Adverse Event Reporting System. Pediatr Infect Dis J. 2004;23(4):287–294.

[12] Kessler DA. Introducing MEDWatch. A new approach to reporting medication and device adverse effects and product problems. JAMA. 1993 Jun 2;269(21):2765-8.

[13] Ahmad SR. Adverse Drug Event Monitoring at the Food and Drug Administration. J Gen Intern Med 2003; 18: 57-60.

[14] Marinac-Dabic D, Normand SL, SedrakyanA, Gross T. Epidemiologic studies of medical devices: methodologic considerations for implantable devices. In: Strom BL, Kimmel SE, Hennessy S, eds. Pharmacoepidemiology. 5th ed. Sussex, England: Wiley; 2011.

[15] Porta M, editor. A Dictionary of Epidemiology.5th edn. New York: Oxford University Press; 2008.

[16] Gagne JJ, Fireman B, Ryan PB, Maclure M, Gerhard T, Toh S, Rassen JA, Nelson JC, Schneeweiss S. Design considerations in an active medical product safety monitoring system. Pharmacoepidemiol Drug Saf. 2012 Jan;21 Suppl 1:32-40.

[17] Mullooly JP, Donahue JG, DeStefano F, et al. Predictive value of ICD-9-CM codes used in vaccine safety research. Methods Inf Med. 2008;47(4):328-35.

[18] Eisenberg DF, Daniel GW, Jones JK, et al.. Validation of a claims-based diagnostic code for Stevens-Johnson syndrome in a commercially insured population.Pharmacoepidemiol Drug Saf. 2012 Jul;21(7):760-764.

[19] Andrade SE, McPhillips H, Loren et al. Antidepressant medication use and risk of persistent pulmonary hypertension of the newborn. Pharmacepidemiol Drug Saf 2009;18(3):246-52.

[20] Harrold LR, Patterson MK, Andrade SE, et al. Asthma drug use and the development of Churg-Strauss syndrome (CSS). Pharmacoepidemiol Drug Saf. 2007 Jun;16(6):620-6.

[21] Guolo A. Robust techniques for measurement error correction: a review. Stat Methods Med Res. 2008 Dec;17(6):555-80.

[22] Roy S, Banerjee T, Maiti T. Measurement error model for misclassified binary responses. Statist Med 2005;24:269-283.

[23] Rothman KJ, Greenland S, Lash TL. Selection bias. In: Rothman KJ, Greenland S, Lash TL, eds.,Modern Epidemiology. 3rd ed. Philadelphia: Lippincott Williams and Wilkins, 2008; 134-137.

[24] Szklo M, Nieto FJ. Combined selection/information biases. Epidemiology: Beyond the Basics. 2nd. ed. Boston: Jones and Bartlett, 2007; 133-146.

[25] McMahon AD, Evans JMM, McGilchrist MM, et al. Drug exposure risk windows and unexposed comparator groups for cohort studies in pharmacoepidemiology. Pharmacoepidemiol Drug Saf. 1998;7:275-80.

[26] Food and Drug Administration. Guidance for Industry and Food and Drug Administration Staff - Factors to Consider When Making Benefit-Risk Determinations in Medical Device Premarket Approvals and De Novo Classifications. March 28,2012. http://www.fda.gov/MedicalDevices/ DeviceRegulationandGuidance/ GuidanceDocuments/ ucm267829.htm Accessed: July 26,2012.

[27] Shrank WH, Choudhry NK, Fischer MA, et al. The Epidemiology of Prescriptions Abandoned at the Pharmacy. Ann Intern Med. 2010;153:633-640.

[28] Schneeweiss S. A basic study design for expedited safety signal evaluation based on electronic healthcare data. Pharmacoepidemiol Drug Saf. 2010 Aug;19(8):858-68.

[29] Glymour MM, Greenland S. Causal diagrams. In: Rothman KJ, Greenland S, Lash TL, eds.,Modern Epidemiology. 3rd ed. Philadelphia: Lippincott Williams and Wilkins, 2008; 183-209.

[30] Greenland S, Lash TL. Bias Analysis. In: Rothman KJ, Greenland S, Lash TL, eds.,Modern Epidemiology. 3rd ed. Philadelphia: Lippincott Williams and Wilkins, 2008; 345-380.

[31] Lash TL, Fox MP, Fink AK. Applying Quantitative Bias Analysis to Epidemiologic Data. New York: Springer. 2009.

[32] Rothman KJ, Greenland S, Lash TL. Confounding. In: Rothman KJ, Greenland S, Lash TL, eds.,Modern Epidemiology. 3rd ed. Philadelphia: Lippincott Williams and Wilkins, 2008; 129-134.

[33] Szklo M, Nieto FJ. Incomplete adjustment: residual confounding. Epidemiology: Beyond the Basics. 2nd. ed. Boston: Jones and Bartlett, 2007; 282-284.

[34] Schneeweiss S, Wang PS. Association between SSRI use and hip fractures and the effect of residual confounding bias in claims database studies. J Clin Psychopharmacol. 2004 Dec;24(6):632-8.

[35] Lash TL, Fox MP, Fink AK. Unmeasured and unknown confounders. In: Applying Quantitative Bias Analysis to Epidemiologic Data. New York: Springer. 2009; 59-62.

[36] Poikolainen K. The magic to make the 'preventive effect' of alcohol disappear and reappear. Addiction. 2008 Nov;103(11):1905; author reply 1905-7.

[37] Rothman KJ, Greenland S, Lash TL. Restriction. In: Rothman KJ, Greenland S, Lash TL, eds.,Modern Epidemiology. 3rd ed. Philadelphia: Lippincott Williams and Wilkins, 2008; 169.

[38] Greenland S, Rothman KJ. Introduction to stratified analysis. In: Rothman KJ, Greenland S, Lash TL, eds.,Modern Epidemiology. 3rd ed. Philadelphia: Lippincott Williams and Wilkins, 2008; 258-282.

[39] Schneeweiss S, Rassen JA, Glynn RJ, et al. High-dimensional propensity score adjustment in studies of treatment effects using health care claims data. Epidemiology 2009; 20: 512–522.

[40] Poole C. Low P-values or narrow confidence intervals: which are more durable? Epidemiol. 2001 May;12(3):291-4.

[41] Duffy SW, Warwick J, Williams AR, et al. A simple model for potential use with a misclassified binary outcome in epidemiology. J Epidemiol Community Health. 2004 Aug;58(8):712-7.

[42] Lash T, Fink AK. Semi-automated sensitivity analysis to assess systematic errors in observational data. Epidemiology 2003; 14: 451–458.

[43] Fox MP, Lash TL, Greenland S. A method to automate probabilistic sensitivity analyses of misclassified binary variables. Int J Epidemiol 2005; 34: 1370–1376.

[44] Lash TL, Schmidt M, Jensen AØ, Engebjerg MC. Methods to apply probabilistic bias analysis to summary estimates of association. Pharmacoepidemiol Drug Saf. 2010 Jun;19(6):638-44.

[45] Greenland S. Multiple bias modeling for analysis of observational data. J R Stat Soc A 2005;168:1–25.

[46] Food and Drug Administration. FDA Adverse Event Reporting System (FAERS) (formerly AERS) http://www.fda.gov/Drugs/GuidanceComplianceRegulatoryInformation/Surveillance / AdverseDrugEffects/ default.htm. Accessed Oct 22, 2012.

[47] Schneeweiss S. Chapter 11. Confounding. In: Pharmacoepidemiology and Therapeutic Risk Management. Eds: Hartzema AG, Tilson HH, Chan KA. Harvey Whitney Books Company. Cincinnati, OH. 2008;273-299.

[48] Petri H, Urquhart J. Channeling bias in the interpretation of drug effects. Stat Med. 1991;10:577-581.

[49] Schneeweiss S, Avorn J. A review of uses of health care utilization databases for epidemiologic research on therapeutics. J Clin Epi. 2005;58:323-337.

[50] Steiner JF. Can we identify clinical predictors of medication adherence . . . and should we? Med Care. 2010;48(3):193-5.

[51] McMahon AD, Evans JMM, McGilchrist MM, et al. Drug exposure risk windows and unexposed comparator groups for cohort studies in pharmacoepidemiology. Pharmacoepidemiol Drug Saf. 1998;7:275-80.

[52] Van Staa TP, Abenhaim L, Leufkens H. A study of the effects of exposure misclassification due to the time-window design in pharmacoepidemiologic studies. J Clin Epidemiol. 1994;47:183-9.

[53] Faich GA, Stadel BV. The future of automated record linkage for postmarketing surveillance: A response to Shapiro. Clin Pharmacol Ther. 1989;46:387-9.

[54] Herbst AL, Ulfelder H, Poskanzer DC. Adenocarcinoma of the vagina. Association of maternal stilbestrol therapy with tumor appearance in young women. J Engl J Med 1971;284:878-881

[55] Smalley WE, Ray WA, Daugherty JR, Griffin MR. Nonsteroidal anti-inflammatory drugs and the incidence of hospitalizations for peptic ulcer disease in elderly persons. Am J Epidemiol. 1995 Mar 15;141(6):539-45.

[56] Slater EE, Merrill DD, Guess HA, Roylance PJ, et al. Clinical profile of angioedema associated with angiotensin converting-enzyme inhibition. JAMA. 1988;260(7):967-70.

[57] 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.

[58] Mullooly J, Drew L, DeStefano F, et al. Quality of HMO Vaccination Databases Used to Monitor Childhood Vaccine Safety. Am J Epidemiol. 1999; 149(2):186-194.

[59] Xu S, Zhang L, Nelson JC, et al. Identifying optimal risk windows for self-controlled case series studies of vaccine safety. Stat Med. 2011 Mar 30;30(7):742-52.

[60] Virtanen M, Peltola H, Paunio M, Heinonen OP. Day-to-day reactogenicity and the healthy vaccinee effect of measles-mumps-rubella vaccination. Pediatrics. 2000 Nov;106(5):E62.

[61] Wilson K, Hawken S, Potter BK, et al. Patterns of emergency room visits, admissions and death following recommended pediatric vaccinations - a population based study of 969,519 vaccination events. Vaccine. 2011 May 12;29(21):3746-52.

[62] Fine PE, Chen RT. Confounding in studies of adverse reactions to vaccines. Am J Epidemiol. 1992 Jul 15;136(2):121-35.

[63] Fine PE, Williams TN, Aaby P, et al. Epidemiological studies of the 'non-specific effects' of vaccines: I--data collection in observational studies. Trop Med Int Health. 2009 Sep;14(9):969-76.

[64] Chen RT, Glanz JM, Vellozzi C. Pharmacoepidemiologic studies of vaccine safety. In: Strom BL, Kimmel SE, Hennessy S, eds. Pharmacoepidemiology. 5th ed. Sussex, England: Wiley; 2011.p 423-468.

[65] Farrington CP, Firth MJ, Moulton LH, et al. Epidemiological studies of the non-specific effects of vaccines: II--methodological issues in the design and analysis of cohort studies. Trop Med Int Health. 2009 Sep;14(9):977-85.

[66] Centers for Disease Control and Prevention. Recommended immunization schedules for persons aged 0–18 years—United States, 2012. MMWR 2012;61(5).

[67] Food and Drug Administration. What does it mean for FDA to "classify" a medical device? http://www.fda.gov/AboutFDA/Transparency/Basics/ucm194438.htm. Accessed August 16, 2012.

[68] Food and Drug Administration. Manufacturer and User Facility Device Experience Database - (MAUDE). http://www.fda.gov/MedicalDevices/DeviceRegulationandGuidance /PostmarketRequirements /ReportingAdverseEvents/ucm127891.htm. Accessed August 16, 2012.

[69] Food and Drug Administration. MedSun: Medical Product Safety Network. http://www.fda.gov/MedicalDevices/Safety/MedSunMedicalProductSafetyNetwork/default.htm. Accessed August 16, 2012.

[70] Paxton EW, Inacio MC, Khatod M, et al. Kaiser Permanente National Total Joint Replacement Registry: aligning operations with information technology. Clin Orthop Relat Res. 2010 Oct;468(10):2646-63.

[71] Paxton EW, Inacio MC, Kiley ML. The Kaiser Permanente implant registries: effect on patient safety, qualify improvement, cost effectiveness, and research opportunities. Perm J 2012;16(2):36-44.

[72] Miller M, Ulisney K, Baldwin T. INTERMACS (Interagency Registry for Mechanically Assisted Circulatory Support): A New Paradigm for Translating Registry Data Into Clinical Practice. J Am Coll Cardiol. 2010;56;738-40.

[73] Baim DS, Mehran R, Kereiakes DJ, et al. Postmarket surveillance for drug-eluting coronary stents: a comprehensive approach. Circulation. 2006 Feb 14;113(6):891-7.

[74] Williams DO, Holubkov R, Yeh W, et al. Percutaneous Coronary intervention in the current era compared with 1985–1986: the National Heart, Lung, and Blood Institute registries. Circulation. 2000; 102: 2945–2951.

[75] Resnic FS, Normand ST. Postmarketing Surveillance of Medical Devices — Filling in the Gaps. N Engl J Med. 2012;366(10):875-7.

[76] Food and Drug Administration. FDA News Release: FDA proposes unique device identification system for medical devices. http://www.fda.gov/NewsEvents/Newsroom/PressAnnouncements/ucm310505.htm. Accessed August 16, 2012.

[77] Ramsay CR, Wallace SA, Garhwaite PH, et al. Assessing the learning curve effect in health technologies. Lessons from the non-clinical literature. Int J Technol Assess Health Care 2002; 18: 1-10.

[78] Angrist JD, Imbens GW, Rubin DB. Identification of causal effects using instrumental variables. J Am Stat Assoc 1996; 91: 459-62.

[79] de Vries RR, Faber JC, Strengers PF, Haemovigilance: an effective tool for improving transfusion practice. Vox Sang. 2011; 100(1): 60-7.

[80] Centers for Disease Control and Prevention. The National Healthcare Safety Network (NHSN) Manual: Biovigilance Component. 2011. http://www.cdc.gov/nhsn/pdfs/hemovigmoduleprotocol_current.pdf. Accessed Aug 22, 2012.

[81] Menis M, Izurieta HS, Anderson SA, et al. Outpatient transfusions and occurrence of serious noninfectious transfusion-related complications among US elderly, 2007-2008: utility of large administrative databases in blood safety research. Transfusion. 2012 Feb 8. [Epub ahead of print]

[82] Daniel GW, Menis M, Sridhar G, et al. Immune globulins and thrombotic adverse events as recorded in a large administrative database in 2008 through 2010. Transfusion. 2012 Mar 12. [Epub ahead of print]

[83] Strom BL. Overview of automated databases in pharmacoepidemiology. In: Strom BL, Kimmel SE, Hennessy S, eds. Pharmacoepidemiology. 5th ed. Sussex, England: Wiley; 2011.p 158-162.

[84] Naus J, Wallenstein S. Temporal surveillance using scan statistics. Stat Med. 2006; 25(2):311–324.

[85] McClure DL, Xu S, Weintraub E, Glanz JM. An efficient statistical algorithm for a temporal scan statistic applied to vaccine safety analyses. Vaccine. 2012 Jun 8;30(27):3986-91.

[86] http://sites.google.com/site/biasanalysis/. Accessed September 13, 2012.

[87] Orsini, N.; Bellocco, R.; Bottai, M.; Wolk, A.; Greenland, S. A tool for deterministic and probabilistic sensitivity analysis of epidemiologic studies. Stata J 2008, 8, 29-48.

[88] Abramson, J.H. (2011) WINPEPI updated: computer programs for epidemiologists, and their teaching potential. Epidemiologic Perspectives & Innovations 2011, 8:1.

[89] McGettigan P, Henry D. Cardiovascular risk with non-steroidal anti-inflammatory drugs: systematic review of population-based controlled observational studies. PLoS Med. 2011 Sep;8(9):e1001098.

[90] Schneeweiss S, Glynn RJ, Tsai EH, et al. Adjusting for unmeasured confounders in pharmacoepidemiologic claims data using external information: the example of COX2 inhibitors and myocardial infarction. Epidemiology. 2005 Jan;16(1):17-24.