MINI-SENTINEL METHODS

ALTERNATIVE METHODS FOR HEALTH OUTCOMES OF INTEREST VALIDATION

Prepared by: Glen T. Schumock, PharmD, MBA, PhD[1-3]; Todd A. Lee, PharmD, PhD[1-3]; A. Simon Pickard, PhD[1-3]; Wan-Ju Lee, BPharm, MS[1-3]; Hari Patel, PharmD[1-3]; Piyameth Dilokthornsakul, PharmD[1]; Azadeh Shoaibi, MS, MHS[4]; Patrick Archdeacon, MD[4]

Author Affiliations: 1. Center for Pharmacoepidemiology and Pharmacoeconomic Research, University of Illinois at Chicago, Chicago, IL**.** 2. Department of Pharmacy Practice, University of Illinois at Chicago, Chicago, IL. 3. Department of Pharmacy Systems, Outcomes and Policy, University of Illinois at Chicago, Chicago, IL. 4. Center for Drug Evaluation and Research, Office of Medical Policy, US Food and Drug Administration, Silver Spring, MD

August 31, 2013

Mini-Sentinel Methods

Alternative Methods For Health Outcomes
Of Interest Validation

Table of Contents

I.   EXECUTIVE SUMMARY

A.   OVERVIEW OF PROJECT

The FDA Mini-Sentinel project is a prototype for the Sentinel initiative that is conceptualized as a nation-wide medical product safety surveillance system. By applying validated algorithms for adverse events/health outcomes of interest (HOI) to individual patient data, new or existing cases of such HOIs can be detected. Central to this approach is the need to be able to apply algorithms that can accurately identify HOIs based on the available data elements, e.g., International Classification of Disease (ICD) codes. One approach to validating such algorithms is to go to the source medical records of patients to confirm the diagnosis/HOI. However, because of the costly and time-consuming resource requirements for validation of algorithms with medical records, a Mini-Sentinel activity that investigated alternative methods for validating HOIs was proposed. The purpose of this Mini-Sentinel workgroup (WG) activity was to identify HOIs for which there is an alternative electronic data source for confirming cases (such as a patient registry) that could be linked to the Mini-Sentinel Distributed Database (MSDD), and to determine the feasibility of using that alternative source to validate an algorithm for the HOI in the MSDD.

The WG developed a multi-step process that included HOI definition clarification, HOI categorization, database searches, prioritization of HOIs in which an alternative data source exists and linkage may be feasible, further investigation/verification of identified databases, and final recommendations for consideration in phase II. HOI definition clarification was the initial step taken to clarify the clinical context of the HOI. Related HOIs were grouped into disease-based categories or themes, such as blood-related disorders, to better facilitate development expertise on related databases and registries. The literature on existing algorithms and their accuracy (particularly positive predictive value (PPV)) was investigated and summarized, including evidence to support whether the published algorithm(s) were considered sufficiently validated/accurate. Registries and alternative data sources were examined, and feasibility of linkage to MSDD was evaluated. The potential to develop an algorithm for each HOI was rated as: feasible; potentially feasible, unlikely or not feasible. In an iterative process with the FDA, a priority list of HOIs was identified by further investigating the alternative data sources, and weighing criteria such as linkability, cost, accessibility of the data, and overlap of patients with the MSDD. Phase II is a planned separate WG project to conduct a validation study using an alternative database to validate an HOI(s) in the MSDD.

B.   SUMMARY OF FINDINGS

A total of 99 HOIs were investigated in the project. Among the 99, 16 HOIs were deemed to have been well-validated in previous studies (e.g., algorithms with a positive predictive values (PPVs) >0.70) such that additional validation work (alternative or otherwise) was considered unnecessary. An additional 11 HOIs were not considered feasible for alternative validation, and an additional 27 HOIs were considered unlikely for alternative validation. In most cases HOIs were determined to be not feasible or unlikely because of lack of availability of alternative databases or because of major limitations with such databases. Nevertheless, based on the WG's initial review, the WG identified 45 HOIs as potentially feasible or feasible for alternative database validation.

The 45 feasible or potentially feasible HOIs included 10 "lab-based HOIs". These were HOIs that could be validated using laboratory results alone (i.e., the laboratory result represents the gold standard for confirming cases). Among the remaining 35 potentially feasible and feasible HOIs were 17 cancer-related HOIs. Again, the WG considered these to be highly feasible because of the clear availability of linkable alternative databases. However, from a surveillance perspective the WG considered the cancer HOIs as a low priority for phase II of this project. This was primarily because of the time lag between exposure to a medication or other agent that may cause cancer, e.g., latency period, and the related difficulty following patients for such long periods of time in administrative data such as MSDD.

The other 28 potentially feasible and feasible HOIs were ranked by FDA staff as low, medium, or high priority with respect to surveillance importance with the intent that the WG concentrated its further efforts on the highly ranked HOIs. Six HOIs were considered high priority - these included suicide, type 1 diabetes, hypertension crisis, pulmonary fibrosis, pulmonary hypertension and spontaneous abortion.

C.   RECOMMENDATION FOR NEXT PHASE

The WG concluded that the best candidates for alternative validation in Phase II were: 1) suicide, using data from National Death Index as the alternative data source, and 2) type 1 diabetes, using the T1D Exchange Registry as the alternative data source. Note that type 1 diabetes had a second data source – the internal registries in Kaiser Permanente Northern California and Health Partners - which were considered a very good option (but not better than the T1D exchange). Next, the WG considered hypertensive crisis, using data from the Health Maintenance Organization Research Network Hypertension Registry, to be a very good candidate for alternative validation. Last, pulmonary hypertension, using the Registry to Evaluate Early And Long-Term Pulmonary Artery Hypertension Disease Management, was considered a good candidate for alternative validation.

The WG also determined that neither of the data sources identified for spontaneous abortion were viable. These included the Fetal Death Dataset from the CDC National Center for Health Statistics and National Children's Study. The former was determined to be not linkable and the later to have insufficient participant enrollment. Finally, there was one HOI among the six for which the WG could not retrieve sufficient information. That was the pulmonary fibrosis and the data source was the Pennsylvania Idiopathic Pulmonary Fibrosis Registry.

In conclusion, the WG recommends that the FDA and Mini-Sentinel program consider suicide or type 1 diabetes for phase II of this project. Hypertensive crisis and pulmonary hypertension could also be considered. Finally, the 10 lab-based HOIs might also be reasonable for such validation.

II. ABBREVIATION LIST

ADIC: Acute disseminated intravascular coagulation
AHTR: Acute hemolytic transfusion reaction
ALS: Amyotrophic lateral sclerosis
BMI: Body mass index
CBER: Center for Biologics Evaluation and Research
CDC: Centers for Disease Control Prevention
CRN VDW: The Cancer Research Network Virtual Data Warehouse
DP: Data Partner
EHR: Electronic health records
FDA: Food and Drug Administration
HDL: High-density lipoprotein
HOI: Health outcome of interest
IBD: Inflammatory bowel disease
ICD-9: The International Classification of Diseases, 9th Revision, Clinical Modification
ITP: Idiopathic thrombocytopenic purpura
KP: Kaiser Permanente
LDL: Low-density lipoprotein
MSCDM: Mini-Sentinel Common Data Model
MSDD: Mini-Sentinel Distributed Database
NDI: National Death Index
NMS: Neuroleptic malignant syndrome
NPCR: National Program of Cancer Registries
OMOP: Observational Medical Outcomes Partnership
OPTN: Organ Procurement and Transplantation Network
PML: Progressive multifocal leukoencephalopathy
PPV: positive predictive value
SEER: Surveillance, Epidemiology and End Results
SJS: Stevens-Johnson Syndrome
SLE: Systemic lupus erythematosus
T3: Triiodothyronine
T4: Thyroxine
TG: Triglycerides
TRALI: Transfusion-related acute lung injury
TSH: Thyroid stimulating hormone
TTP: Thrombotic thrombocytopenic purpura
US: United States
WG: Workgroup

III. INTRODUCTION

A. BACKGROUND

The FDA Mini-Sentinel project is a prototype for the Sentinel initiative that is conceptualized as a nation-wide medical product safety surveillance system. By applying validated algorithms for adverse events/health outcomes of interest (HOI) to individual patient data, new or existing cases or such HOIs can be detected. Central to this approach is the need to be able to apply algorithms that can accurately identify HOIs based on the available data elements, e.g., International Classification of Disease (ICD) codes.

Accurate identification of HOIs can pose problems in studies of electronic databases, including the Mini-Sentinel Distributed Database (MSDD), that rely on ICD codes, or combinations of ICD codes and other variables (i.e., algorithms) in the data. If these algorithms have low sensitivity or specificity, then significant misclassification of cases can occur. To avoid this, investigators use algorithms that have been previously validated against full-text medical record review as the "gold standard."[1-22] While many algorithms have been validated in this manner, not all have. Furthermore, some validated algorithms are based on data, populations, or exposures that are too different from the MSDD to be useful for the Mini-Sentinel program.

Although Mini-Sentinel has previously conducted HOI algorithm validation against medical records, this process is very resource intensive, both in terms of time and money. Because of the resource requirements for validation via medical records, alternative methods for validating outcomes of interest need to be explored. An alternative to costly and time-consuming medical records validation is the use of an electronic database that contains true cases of an HOI and that can be linked to MSDD (referred to as "alternative validation" hereafter) for validation of an algorithm. Some examples of possible alternate databases that could be used instead of medical records for certain HOIs include patient registries or clinical databases.

Figure 1 depicts the process of alternative validation. The alternative database would be linked to the MSDD. Information contained in the alternative database would serve as the gold standard for presence of the HOI. A set of variables in the MSDD would then be used to construct an algorithm that identifies the occurrence of the HOI. For the sample of patients linked in both databases, the ability of the algorithm to accurately identify patients with the HOI could be assessed. Once the algorithm is validated in this manner, it could then be used for HOI identification in the full MSDD.

There are a number of examples of such alternative methods in the literature. In 2007, Setoguchi and others[23] described use of the Pennsylvania State Cancer Registry data to calculate the sensitivity, specificity, and positive predictive values (PPV) of an algorithm for identifying lymphoma in Medicare claims by linking the two datasets. Registry data such as this may exist for other HOIs and seems to be the most obvious example of linked HOI validation. In fact the Surveillance, Epidemiology and End Results (SEER) cancer registry has been used in a number of published validation studies.[24-27]
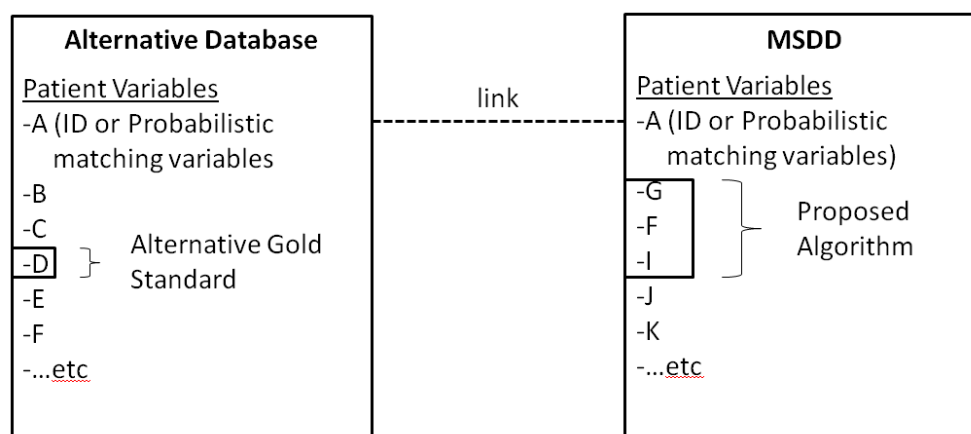
Figure 1. Diagram of linked HOI alternative validation process

In another example, Yuan and colleagues[28] used hospital discharge records as a reference standard to calculate the sensitivity and specificity of diagnoses of atrial fibrillation based on codes in the Health Care Financing Administration Medicare Part A Hospital Discharge Database. Similarly, Coyte et al.[29] used a claims database – the Ontario Health Insurance physician fee service claims database (which distinguishes between primary and revision knee replacements) – to calculate the sensitivity and specificity of an algorithm in the Canadian Institute for Health Information Abstract Master File, held by the Ontario Ministry of Health. A paper by Jollis and others[30] described use of existing clinical database data collected from standardized data forms at the Duke Databank for Cardiovascular Disease to calculate the sensitivity and specificity of ICD-9 code claims-based diagnoses.

It is clear that Mini-Sentinel surveillance activities could benefit from establishing alternative approaches to validation of HOI algorithms in the MSDD to identify true cases in the absence of source record validation. However, it is unlikely that every HOI will have an alternative data source that can serve as a linkable reference standard to MSDD. For example, patient registries may serve as a useful reference standard for certain HOIs but may not exist for others. At present, it is unclear which HOIs are amenable to this alternative validation approach and what electronic databases might exist that can be used for which HOIs.

B. OBJECTIVE

The purpose of this Mini-Sentinel activity was to: 1) identify HOIs for which there is an alternative reference standard (such as registry data) that could be linked to the MSDD; and, 2) determine the feasibility of using that alternative reference standard to validate an algorithm for the HOI in the MSDD.

C. SCOPE

The scope of the current proposal was limited to the objectives above (Phase I). However, conditional on the results and recommendations of Phase 1, a second set of objectives might be pursued (Phase II) with the intent of performing an alternative validation study for an HOI(s) determined to have an appropriate alternative reference standard.

IV. METHODS

In order to accomplish the objectives described above the workgroup (WG) conducted a multi-step process that included HOI definition clarification, HOI categorization, database searches, prioritization of HOIs in which an alternative data source exists and linkage may be feasible, further investigation/verification of identified databases, and final recommendations for consideration in phase II. These are further described as steps 1-6 below. Prior to initiation, the WG, with the help of the Mini-Sentinel Operation Center, fully informed itself on the specifications of the Mini-Sentinel Common Data Model and Distributed Database.

A. STEP 1: CLARIFICATION OF HOI DEFINITION

A list containing 84 HOIs was provided by FDA to the WG (**Table 1**). The first step was to clarify the similarities between different HOIs (e.g., "colitis ischemic" and "ischemic colitis needing surgery") as well as subtypes of HOIs (e.g., cancer or transfusion/graft infections) in the list provided by the FDA. The WG discussed with the FDA the surveillance perspectives and clinical issues associated with each HOI, and then made adjustments of the HOI list by combining or separating HOIs and/or HOI subtypes as appropriate.

B. STEP 2: CATEGORIZATION OF HOI

The WG originally proposed to group HOIs into categories that related to specific organ system(s), syndromes or diseases, and/or drug or drug classes. The WG believed that the categorization approach would be helpful in identifying registries and other alternative linkable electronic data sources for validating HOI algorithms in the MSDD. However, based on WG discussions with FDA, it was determined that only the categorization by organ system was necessary. The WG therefore categorized the HOIs into organ systems using standard medical references and then reviewed and approved these with FDA.

C. STEP 3: SEARCH FOR ALTERNATIVE DATA SOURCE

Registries and other alternative linkable electronic data sources that might have potential to be used for validating HOI algorithms in MSDD were identified in Step 3. There were four sub-parts to this process, described below.

1. Literature Review for Previous Validation Studies

Prior to searching for alternative data sources, the WG first sought evidence of previous validation studies and eliminated those HOIs for which algorithms had already been well validated. Specifically, members of the WG read the systematic reviews of validation studies published in Pharmacoepidemiology and Drug Safety by the Mini-Sentinel Protocol Core, as well as those by the Observational Medical Outcomes Partnership (OMOP). Unpublished work from FDA Center for Biologics Evaluation and Research (CBER) and internal reports from FDA Protocol Core were also reviewed. The WG also conducted literature searches in PubMed using appropriate terms such as "validation," "algorithm," "sensitivity," "positive predict value," plus HOI-specific terms (i.e., various terms for the

Table 1 List of 84 HOIs provided by FDA

Acute disseminated encephalomyelitis	Kawasaki disease
Acute hemolytic transfusion reaction	Lactic acidosis
Acute respiratory failure	Lymphoma
Agranulocytosis	Mania/Bipolar
Amyotrophic lateral sclerosis	Menarche
Aplastic anemia	Menopause
Arterial thrombosis	Myocarditis
Atrial fibrillation	Narcolepsy
Autoimmunity – Consider subtypes	Neuroleptic malignant syndrome
Bell's palsy	Obesity
Birth defects	Optic neurititis
Blindness	Pancytopenia
Brachial neuritis	Pericarditis
Bronchospasm	Progressive multifocal leukoencephalopathy
Cancer (including subtypes)	Pneumonia
Chronic renal failure	Post-transfusion allergic reaction
Cirrhosis	Premature delivery
Colitis ischemic	Pulmonary fibrosis
Congenital anomalies	Pulmonary hypertension
Deafness	Rhabdomyolysis
Depression	Schizophrenia
Disseminated intravascular coagulation	Serotonin syndrome
Dyslipidemias	Spontaneous abortion
Endotoxic shock	Stevens-Johnson Syndrome
Erythema multiforme	Stillbirth
Febrile seizures	Sudden death
Guillain-Barre syndrome	Suicide
Hemmorhagic stroke	Systemic lupus erythematosus
Hemolysis	Tendonopathies
Hemolytic anemia	Thrombotic thrombocytopenic purpura
Henoch-Schonlein purpura	Tics
Hip fracture	Torsades de pointes
Histoplasmosis	Transfusion ABO incompatibility reactions
Hyper/hypothyroidism	Transfusion sepsis
Hyperglycemia	Transfusion/Graft infections
Hypertensive crisis	Transfusion-related acute lung injury
Hypoglycemia	Transverse myelitis
Idiopathic thrombocytopenic purpura	Tuberculosis
Inflammatory bowel disease	Type 1 diabetes
Intussusception	Uveitis
Ischemic colitis needing surgery	Valvulopathy
Juvenile rheumatoid arthritis	Ventricular fibrillation

HOI or related clinical conditions) to identify validation studies for individual HOIs. The WG considered an HOI to be already well-validated if a previous validation study was identified in which: 1) the PPV exceeded 70%[2] and 2) the population was considered generalizable to MSDD. If an HOI was considered well-validated, then it was removed from the HOI list for further investigation. Thus, the search for alternative data sources focused on those HOIs without sufficient evidence of previously validated algorithms.

2. **Search for Registries, Electronic Medical Records or Databases**

The WG conducted a comprehensive search of health care databases. This included registries, data repositories, and other patient-level health data sources. The WG began by targeting websites that provide a listing of health care databases or registries. For example, the Agency for Healthcare Research and Quality Registry of Patient Registries (https://patientregistry.ahrq.gov/), and the "B.R.I.D.G.E to Data" listing of health care databases (http://www.bridgetodata.org/Database-ProfileListing). Next, like the literature search described above, the WG used a combination of search terms including terms referring to both the HOI itself (i.e., various terms for the HOI or related clinical conditions) and to databases (e.g., "registry," "database," "claims data," "electronic medical record") to conduct individual searches for literature describing or using a potentially applicable alternative data sources. This was done primarily using PubMed. Any articles that referred to a database were then reviewed and the database identified was investigated further.

The WG also searched the internet for alternative databases using major search engines (Google, GoogleScholar). Several federal websites (e.g., Center for Disease Control Prevention [CDC], FDA or ClinicalTrials.gov) were also routinely used. The WG also searched websites of professional/medical associations/societies and advocacy groups related to each HOI. These often contained linked to other research-related resources, including registries. Generally, the WG focused only on US-based studies, databases, or registries for inclusion as the potential alternative data sources.

In addition to internet searches, the WG contacted individuals and organizations to identify potential linkable databases. This primarily involved individuals connected to the Mini-Sentinel program, including Mini-Sentinel Data Partners, Mini-Sentinel investigators, and members of the Mini-Sentinel Data Core, Protocol Core, and Planning Board. The WG also sought information on if health provider organizations that contribute to MSDD had other electronic databases or internal registries that might be linkable to MSDD for HOI validation. The WG conducted a survey of the Mini-Sentinel Planning Board for this purpose.

Part of the effort included contacting data vendors that owned or controlled multiple databases, such as IMS Health, Thomson Reuters, and others. While most databases were likely identified in the internet search, the WG also contacted these companies to be sure that none were excluded that might be useful for the purposes of validating HOIs. The WG also attempted to contact and get information from data brokers. Data brokers do not own databases but sell services to help investigators gain access to these. As such, these individuals/organizations are very knowledgeable about both what databases are available and the contents of the databases. An example of a data broker is Health Data Services Corporation (http://www.hdscorp.biz/).

For each HOI, the WG documented the characteristics of each alternative data source identified in a detailed spreadsheet. The information included the name, data source type (registry/electronic medical records/claims database), brief description of the data source, patient population included, start date of the database/registries, estimated number of participants, variable(s) or method(s) used to confirm cases (gold standard) in the data source, region covered (national/regional/local), other comments, contact information and the website links for the data source.

3. Evaluation/Summary

Based on the search results and other considerations of the feasibility of the particular HOI for this alternative method of validation, the WG summarized the alternative data sources found and made initial recommendations for the HOIs. Specifically, for each HOI the WG wrote a summary document (**Appendix A**) that included information about previous validation studies, a summary of the available alternative data sources, possible issues when using alternative methods for validation for this HOI, and an initial recommendation. Recommendations included the categories of "feasible", "potentially feasible", "unlikely", or "not feasible". For those in the not feasible group no alternative data source was found. The unlikely category indicated that a high degree of difficulty exists for validation purposes, even though some databases/registries were identified. For example, previous validation studies failed to identify appropriate algorithms for the HOIs, or the prevalence of the HOI is very low. For those in the potentially feasible group, some alternative data sources were identified but further investigation was considered necessary – such as a better understanding of if the database could be linked to MSDD, number of cases in the database, or if there is sufficient overlap of patients between the database and MSDD. The HOIs that were classified in the feasible group were believed to have the highest probability to be validated via an alternative data source because of the existence of promising databases/registries.

During this process the WG determined that some HOIs could be validated using lab-based data/results and those were separately identified as such in the results. The lab-based HOIs had laboratory values as their gold standard in clinical diagnosis - such as the lipid profile for dyslipidemias, or absolute neutrophil count for agranulocytosis.

4. Workgroup Discussion and Consensus

The WG presented the summary documents and the detailed spreadsheets for each HOI during biweekly conference calls with the FDA and the Mini-Sentinel Operation Center. The findings, process of evaluation, and the rationale of the initial recommendations were discussed. During these calls FDA input was sought and changes made to the initial recommendation for each HOI as appropriate.

D. STEP 4: PRIORITIZATION OF HOI

Step 4 involved prioritization of the HOIs that were initially rated as feasible or potentially feasible, including those that were lab-based, for alternative validation. This prioritization was considered necessary in order to narrow the list of HOIs that were to be investigated further. Several criteria were considered: 1) the importance to FDA from the perspective of surveillance; 2) the degree to which the alternative database was likely to overlap with MSDD in terms of patients/population - i.e., is it likely or unlikely that there will be patients common to both (for example databases/registries with a national

catchment area are more likely to have patients in common with MSDD, regional catchment may depend on area of the country); 3) the overall prevalence of the event (HOI) in the general population; 4) the number of cases in the alternative data source; 5) the degree to which previous validation studies are not generalizable (implies a gap in the evidence); 6) the confidence in the accuracy of the case definition used in the alternative database; and 7) the existence of a registry for the HOI maintained by one of the Mini-Sentinel Data Partners (considered to improve accessibility). After discussion with the FDA, the WG determined the top six HOIs for further investigation (i.e., Step 5).

E.   STEP 5: VERIFICATION OF ALTERNATIVE DATA FOR HIGHLY RANKED HOI

For highly ranked HOIs identified in Step 4, the WG sought additional information by attempting to directly contact the registries/data sources to further investigate the feasibility of using them for alternative validation. Four additional criteria were used in evaluating the feasibility. These are 1) the degree to which the alternative data source(s) contained necessary variables for linking to the MSDD; 2) the degree of accessibility and availability of the alternative data source; 3) the degree of complexity of the process for data acquisition and 4) the cost of the alternative data sources. Other considerations described in Step 4 were also confirmed with the data owners.

F.   STEP 6: FINAL RECOMMENDATION FOR PHASE II

The findings of Step 5 were discussed with the FDA and final recommendations were made by the WG for phase II.

V. RESULTS

A. HOI DEFINITION CLARIFICATION

Based on discussion with the FDA, adjustments were made to the original HOI list based on the focus on surveillance within the Mini-Sentinel program and different clinical issues associated with some of the HOIs (**Table 2Table 2**).

Some HOIs which were considered too broad were separated into multiple HOIs because of different pathologies of the adverse events or other reasons. For example, "hyper/hypothyroidism" was split into hyperthyroidism and hypothyroidism. The HOI "transfusion/graft infection" was split into transfusion infection, tissue graft infection, and solid organ transplant infection. "Cancer" was listed in the original list of HOIs, but the WG agreed to investigate both a general cancer HOI as well as specific site cancers, including anal cancer, bladder cancer, brain/other central nervous system cancer, breast cancer, cervix and uteri cancer, colon and rectal cancer, esophageal cancer, kidney and renal pelvic cancer, leukemia, liver cancer, lung cancer, lymphoma, melanoma, myeloma, pancreatic cancer, prostate cancer, and thyroid cancer.

On the other hand, some HOIs were combined because they were considered similar or otherwise unable to be differentiated with respect to this project. This included "hemolysis" and "hemolytic anemia", which were combined into a single HOI. Similarly, "transfusion ABO incompatibility reactions" and "acute hemolytic transfusion reaction" were combined, as were "colitis ischemic" and "ischemic colitis needing surgery", "erythema multiforme" and "Steven-Johnson Syndrome", and "birth defects" and "congenital anomalies."

For some HOIs, the WG narrowed the HOI to a specific type or sub-type based on the surveillance needs of the FDA. For example, the WG focused on "acute" disseminated intravascular coagulation rather than "chronic" because the clinical symptoms and elevation of laboratory values are more apparent in acute disseminated intravascular coagulation, and because this is more likely to be the HOI of interest in a surveillance system. Similarly, sudden death includes several subtypes (e.g., sudden infant death, sudden death in sports, and sudden cardiac death) but the WG focused on sudden cardiac death since that was of most interest to the FDA for surveillance purposes. Another example was arterial thrombosis, which is generally considered to include ischemic stroke, myocardial infarction, and peripheral arterial embolism. Since ischemic stroke and myocardial infarction algorithms have been previously validated by chart review, the WG focused only on peripheral arterial embolism. The WG also replaced "transfusion sepsis" with the more general "sepsis" HOI because the WG sought to remain as "exposure neutral" as possible. That is, the WG attempted to focus on HOIs independent of an exposure that may cause them. After clarifying all the HOIs, the WG was left with a final list of 99.

B. HOI CATEGORIZATION

These 99 HOIs were categorized into different organ systems including blood, cancer, central nervous system, endocrine/metabolic/renal, gastrointestinal, infectious disease, immune-mediated, musculoskeletal, obstetrics, ocular/otic, and pulmonary (**Table 2**). These categories were then used to focus the WG's searches for alternative databases (Step 3).

C.   PREVIOUSLY VALIDATED HOI

Initial searches for alternative databases (Step 3) involved determining whether the HOI had been previously well-validated. If so, no further investigation was considered necessary (after presentation of the evidence to the FDA). A total of 16 HOIs (**Table 3**) were considered by the WG to be well-validated in previous studies, and thus were not considered candidates for validation through alternative validation methods. Most of the previous validation studies used the medical charts to validate the algorithms tested. For these 16 HOIs the populations studied were considered generalizable to the MSDD, and the PPV for the algorithms tested were greater than 70%.

D.   ALTERNATIVE DATA SOURCES

For each of the remaining 83 HOIs the WG conducted an internet search for alternative data sources as described in the methods above. The detailed findings from these searches are provided as **Appendix B** which is available as a separate Excel document from this report. Also included are the search strategies and results.

In addition to the internet search, the WG conducted a survey of the Mini-Sentinel Planning Board to identify internal registries among Mini-Sentinel Data Partners that might be used for alternative validation. A limited number of Data Partners reported internal registries. Health Partners reported having registries for cancer and diabetes. Cincinnati Children's Hospital reported registries for inflammatory bowel disease, juvenile rheumatoid arthritis, and premature delivery. Kaiser Permanente Colorado reported registries for atrial fibrillation, tumors, hypertensive crisis, and a Perinatal Database which included information on premature delivery. Kaiser Permanente Northwest reported registries for cancer, renal transplant and dialysis (for chronic renal failure) and diabetes (for type 1 diabetes).

E.   SUMMARY/WORKGROUP RECOMMENDATIONS

Information on each HOI, including the nature of previous validation studies, results of alternative data sources identified, any issues regarding validation through alternative methods, and the recommendation of the WG were summarized. These summaries are shown in **Appendix A**.

1.   HOIs That Are Not Feasible

There are 11 HOIs that the WG considered not feasible for alternative validation because no viable data sources were identified (**Table 4**).

2.   HOIs That Are Unlikely

A total of 27 HOIs were considered unlikely to be validated through alternative methods (**Table 5**). Although some registries or databases were found for these HOIs, there were limitations identified that would preclude effective validation. Insufficient overlap of patients between the data source and the MSDD was one of the more common reasons for HOIs being considered unlikely to be able to be validated using an alternative database. For example, the registries that the WG found for acute respiratory failure may not contain enough patients who are also in the MSDD. Another reason for classifying an HOI as unlikely was the failure of databases to contain variable(s) necessary for identifying

Table 2 Categorization of clarified HOIs

Health outcomes of interest	Organ system
Acute disseminated intravascular coagulation	Blood
Agranulocytosis	Blood
Aplastic anemia	Blood
Hemolysis/Hemolytic anemia	Blood
Idiopathic thrombocytopenic purpura	Blood
Pancytopenia	Blood
Post-transfusion allergic reaction	Blood
Sepsis	Blood
Solid Organ Transplant Infections	Blood
Thrombotic thrombocytopenic purpura	Blood
Tissue graft infections	Blood
Transfusion ABO incompatibility reactions/Acute hemolytic transfusion reaction	Blood
Transfusion infections	Blood
Transfusion-related acute lung injury	Blood
Cancer	Cancer
Cancer – Anus	Cancer
Cancer – Bladder	Cancer
Cancer – Brain and other central nervous system	Cancer
Cancer – Breast	Cancer
Cancer – Cervix and uteri	Cancer
Cancer – Colon and rectum	Cancer
Cancer – Esophagus	Cancer
Cancer – Kidney and renal pelvic	Cancer
Cancer – Leukemia	Cancer
Cancer – Liver	Cancer
Cancer – Lung	Cancer
Cancer – Lymphoma	Cancer
Cancer – Melanoma	Cancer
Cancer – Myeloma	Cancer
Cancer – Pancreas	Cancer
Cancer – Prostate	Cancer
Cancer – Thyroid	Cancer
Atrial fibrillation	Cardiovascular
Dyslipidemias	Cardiovascular
Hemorrhagic stroke	Cardiovascular
Hypertensive crisis	Cardiovascular
Myocarditis	Cardiovascular
Pericarditis	Cardiovascular
Peripheral arterial embolism	Cardiovascular
Sudden cardiac death	Cardiovascular
Torsades de pointes	Cardiovascular
Valvulopathy	Cardiovascular

Health outcomes of interest	Organ system
Ventricular fibrillation	Cardiovascular
Bell's palsy	Central nervous system
Brachial neuritis	Central nervous system
Depression	Central nervous system
Febrile seizures	Central nervous system
Mania/Bipolar	Central nervous system
Narcolepsy	Central nervous system
Neuroleptic malignant syndrome	Central nervous system
Schizophrenia	Central nervous system
Serotonin syndrome	Central nervous system
Suicide	Central nervous system
Tics	Central nervous system
Transverse myelitis	Central nervous system
Chronic renal failure	Endocrine/metabolic/renal
Hyperglycemia	Endocrine/metabolic/renal
Hyperthyroidism	Endocrine/metabolic/renal
Hypoglycemia	Endocrine/metabolic/renal
Hypothyroidism	Endocrine/metabolic/renal
Obesity	Endocrine/metabolic/renal
Type 1 diabetes	Endocrine/metabolic/renal
Cirrhosis	Gastrointestinal
Colitis ischemic/Ischemic colitis needing surgery	Gastrointestinal
Inflammatory bowel disease	Gastrointestinal
Intussusception	Gastrointestinal
Acute disseminated encephalomyelitis	Immune-mediated
Amyotrophic lateral sclerosis	Immune-mediated
Autoimmunity – Consider subtypes as appropriate	Immune-mediated
Erythema multiforme/Steven-Johnson Syndrome	Immune-mediated
Guillain-Barre syndrome	Immune-mediated
Henoch-Schonlein purpura	Immune-mediated
Kawasaki disease	immune-mediated
Lactic acidosis	Immune-mediated
Systemic lupus erythematosus	Immune-mediated
Endotoxic shock	Infection disease
Histoplasmosis	Infection disease
Progressive multifocal leukoencephalopathy	Infection disease
Hip fracture	Musculoskeletal
Juvenile rheumatoid arthritis	Musculoskeletal
Rhabdomyolysis	Musculoskeletal
Tendinopathies	Musculoskeletal
Birth defects/Congenital anomalies	Obstetrics
Menarche	Obstetrics
Menopause	Obstetrics

Health outcomes of interest	Organ system
Premature delivery	Obstetrics
Spontaneous abortion	Obstetrics
Stillbirth	Obstetrics
Blindness	Ocular/Otic
Deafness	Ocular/Otic
Optic neuritits	Ocular/Otic
Uveitis	Ocular/Otic
Acute respiratory failure	Pulmonary
Bronchospasm	Pulmonary
Pneumonia - Community acquired	Pulmonary
Pneumonia - Hospital acquired	Pulmonary
Pulmonary fibrosis	Pulmonary
Pulmonary hypertension	Pulmonary
Tuberculosis	Pulmonary

Table 3 Well-validated HOIs

Health outcomes of interest	Performance characteristics/comments	References
Amyotrophic lateral sclerosis	84 - 100%	Benatar et al., 2011[31]
Atrial fibrillation	70 - 96%	Jensen et al., 2012[9]
Bell's palsy	81%	Brandenburg et al., 1993[32]
Cancer – Breast	88%	Nattinger et al., 2004[26]
Chronic renal failure	97.5%	Winkelmayer et al., 2005[33]
Hemorrhagic stroke	>80%	Andrade et al., 2012[1]
Hip fracture	98%	Hudson et al., 2013[34]; Ray et al., 1992[35]
Inflammatory bowel disease	81% - 95%	Liu et al., 2009[36]
Intussusception	To be validated by PRISM	Yih et al., 2011[37]
Kawasaki disease	86%	Kao et al., 2008[38]
Pneumonia - Community acquired	>80%	Barber et al., 2013[39]
Premature delivery	87%	Andrade et al., 2013[40]
Stillbirth	99 - 100%	Yasmeen et al., 2006[41]
Sudden cardiac death	85.3%; 86.8%	Hennessey et al., 2010[42]; Chung et al., 2010[43]
Transverse myelitis	75.7%	Klein et al., 2010[44]
Ventricular fibrillation	92%	Tamariz et al., 2012[19]

cases. For example, the gold standard for diagnosis for rhabdomyolysis includes both laboratory results (creatine kinase) and clinical symptoms. In the absence of these, validation studies have yielded very low PPVs. The databases/registries the WG identified for rhabdomyolysis did not contain the clinical information required for confirmation of the diagnosis. Therefore, rhabdomyolysis is unlikely to be validated by linking to external registries.

Table 4 HOIs that are not feasible to be validated by alternative method

Health outcomes of interest
Acute disseminated encephalomyelitis
Autoimmunity
Brachial neuritis
Colitis ischemic/Ischemic colitis needing surgery
Endotoxic shock
Guillain-Barre syndrome
Henoch-Schonlein purpura
Peripheral arterial embolism
Progressive multifocal leukoencephalopathy
Tendinopathies
Tics

As another example, the WG identified a registry that is maintained by the Toxicology Investigators Consortium (ToxIC) that is referred to in the summary of several HOIs, but which had limitations that made most of those HOIs unlikely to be validated using this dataset. ToxIC is maintained by the American College of Medical Toxicology. Cases pertinent to the following HOIs may be included in this registry: hypertensive crisis, Torsades de pointes, hypoglycemia, chronic renal failure, febrile seizures, neuroleptic malignant syndrome, serotonin syndrome, lactic acidosis, rhabdomyolysis, idiopathic thrombocytopenic purpura, acute respiratory failure, Stevens-Johnson Syndrome, hemolysis, and pancytopenia. While the registry seemed promising for alternative validation for these HOIs, several limitations were identified. The first is linkability. The registry does not include patient identification information and the lack of patient demographic information would make probabilistic matching difficult/impossible. Second, related to above, is the likelihood of patient overlap with MSDD. Because the registry is national it is reasonable to assume some overlap of patients, but for any given HOI a low prevalence of cases may limit the sample of patients that overlap. Currently in the registry frequencies of cases for selected HOIs are as follows: 341 with serotonin syndrome; 842 with prolonged QTc (possibly for Torsades de pointes); 28 with neuroleptic malignant syndrome; 213 with rhabdomyolysis; 32 with hemolysis; 1433 with acidosis; 131 with blisters/bullae (possibly for Stevens-Johnson Syndrome); 212 with thrombocytopenia, and 1443 with hypertensive crisis. The last issue is a question of generalizability. The cases in the registry are those seen by a toxicologist in a hospital or emergency department setting and presumably have resulted from some intentional or unintentional exposure leading to the event. For any given HOI if patients typically present to an ER/hospital then cases in this registry may be representative. However, for some HOIs patients may present in an outpatient clinic and so this registry may only represent a subset of more severe cases that eventually end-up in the hospital or ER. The WG determined that the limitations noted make it unlikely to be the preferred alternative data source for many HOIs listed in **Table 5**.

3. HOIs That Are Potentially Feasible

For 13 HOIs the WG identified potential data sources that might allow for alternative validation, but for which there was some limitation or unknown information that led the WG to consider these only as "potentially feasible" (**Table 6**). For each, the best alternative data source(s) are identified in **Table 6**. The common reasons for determining an HOI to be potentially feasible (as opposed to "feasible") were

uncertainty about linkability and the number of cases in the database. For example, the database maintained by the Social Security Administration contains information on blind and deaf people that receive federal benefits due to one of these disabilities; however the ability to link the database to the MSDD was unknown based on information the WG could ascertain through the internet-based searches. Also, use of the Organ Procurement and Transplantation Network (OPTN) registry to validate an algorithm of cirrhosis was considered possible but only patients with severe cirrhosis requiring liver transplant are included in the OPTN registry. Failure to capture individuals with mild cirrhosis in the OPTN registry might affect both the generalizability of the results of a validation study and reduce the number of cases available for such.

Table 5 HOIs considered unlikely to be validated by alternative method

Health outcomes of interest
Acute respiratory failure
Bronchospasm
Depression
Erythema multiforme/Steven-Johnson Syndrome
Febrile seizures
Hemolysis/Hemolytic anemia
Histoplasmosis
Mania/Bipolar
Menarche
Menopause
Myocarditis
Neuroleptic malignant syndrome
Optic neuritits
Pericarditis
Post-transfusion allergic reaction
Rhabdomyolysis
Schizophrenia
Sepsis
Serotonin syndrome
Systemic lupus erythematosus
Tissue graft infections
Torsades de pointes
Transfusion ABO incompatibility reactions/Acute hemolytic transfusion reaction
Transfusion infections
Transfusion-related acute lung injury
Uveitis
Valvulopathy

4. HOIs That Are Feasible

There were 32 HOIs categorized as "feasible" for alternative database validation. Ten of these were considered "lab-based" HOIs and are described below. For the other 22 the WG identified promising

data sources for use in alternative validation (**Table 7**). Most of the best data sources the WG identified contained a large number of patients and had high probability of successful linkage to MSDD.

The cancer subtype HOIs were considered feasible and for the most part all could be validated using the same alternative database(s). Those databases are described here because they are frequently mentioned in the cancer HOI summaries (**Appendix A**). These include national cancer registries (Surveillance Epidemiology and End Results registry [SEER] or National Program of Cancer Registries [NPCR]) and a large cancer registry that included data from several Mini-Sentinel Data Partners (The Cancer Research Network [CRN] Virtual Data Warehouse [VDW]) were considered the best source for validating either cancer in general or any specific type of cancer.

Table 6 HOIs considered potentially feasible for alternative validation

Health outcomes of interest	Best alternative data source(s)
Aplastic anemia	Center for International Blood and Marrow Transplant Research observational database
Blindness	Database in Social Security Administration
Cirrhosis	Organ Procurement and Transplantation Network registry
Deafness	Database in Social Security Administration
Hypertensive crisis	Health Maintenance Organization Research Network Hypertension Registry
Idiopathic thrombocytopenic purpura	Intercontinental Pediatric and Adult Intercontinental Registry on Chronic ITP
Juvenile rheumatoid arthritis	Childhood Arthritis and Rheumatology Research Alliance Registry
Pneumonia - hospital acquired	CDC's National Healthcare Safety Network, National Trauma Data Bank, The Pediatric Ventilator-Associated Pneumonia Registry
Pulmonary fibrosis	Pennsylvania Idiopathic Pulmonary Fibrosis Registry
Pulmonary hypertension	Registry to Evaluate Early And Long-Term PAH Disease Management
Solid organ transplant infections	Organ Procurement and Transplantation Network
Spontaneous abortion	National Children's Study, National Vital Statistics System - Fetal Death Data
Thrombotic thrombocytopenic purpura	Transfusion Medicine/Hemostasis clinical trial network-TTP registry

The SEER registry is a program sponsored by National Cancer Institute that collects data on patients with cancer from 18 separate geographic areas across the United States. The registry includes individuals that were diagnosed with cancer while residing in these geographic areas. To be included in the registry, the case has to be confirmed by the cancer registrar either through pathology reports or by clinical diagnosis in the absence of a pathology report. The SEER registry has been maintained since 1973.

The NPCR is a CDC sponsored initiative that expands the SEER program to ensure coverage of all 50 states in the United Sates. This program was established in 1992 as a complement to the SEER program and collects similar information to the SEER registry. Currently the NPCR program supports central cancer registries in 45 states, the District of Columbia and Puerto Rico.

The CRN VDW is a distributed data network that maintains a database of patients diagnosed with cancer from institutions participating in the cancer research network. The VDW includes detailed information

on tumor characteristics that are captured by local cancer registrars, similar to information captured in other cancer registries. This information is linked with the healthcare claims data of the individual. The CRN VDW is a subset of institutions that participate in the Health Maintenance Organization Research Network and therefore are contributors to the MSDD. The participating institutions in the CRN VDW include Group Health, Henry Ford, Kaiser Permanente, Marshfield Clinic, Fallon, and Health Partners. The individuals in this dataset would clearly overlap with those in the MSDD. The CRN VDW is fed by institution specific cancer registries. For example, Kaiser Permanente Colorado, Kaiser Permanente Northwest and Health Partners all indicated that they have a local cancer registry. It appears that this is the information that is also available in the VDW as part of the CRN VDW. The CRN VDW captures approximately 38,000 new cases of cancer annually and is considered the most promising data source for cancer and all cancer subtypes.

Table 7 HOIs considered feasible for alternative validation

Health outcomes of interest	Best alternative data source(s)
Birth defects/Congenital anomalies	National Vital Statistics System - Birth data
Cancer and cancer subtype (anus, bladder, brain and central nervous system, cervix and uteri, colon and rectum, esophagus, kidney and renal pelvic, leukemia, liver, lung, lymphoma, melanoma, myeloma, pancreas, prostate, thyroid)	The Cancer Research Network Virtual Data Warehouse, The Surveillance Epidemiology and End Results registry, National Program of Cancer Registries, SEER-Medicare linked database
Narcolepsy	Stanford-based Center for Narcolepsy database
Suicide	National Death Index
Tuberculosis	National Notifiable Diseases Surveillance System
Type 1 diabetes	T1D Exchange, Kaiser Permanente Northern California and Health Partners Registries

The WG identified 10 HOIs for which a case could be verified solely by a laboratory value(s) (
Table 8). In addition, the WG also received input from Mini-Sentinel Data Core (Marsha A. Reable) about the availability of specific laboratory results data from the Mini-Sentinel Data Partners. Most HOIs listed below had a high to moderate possibility to be validated with data available from Data Partners. Some of these laboratory results may even be available within the Mini-Sentinel Common Data Model (currently or in the future). For this reason the WG categorized these HOIs separately.

F.   HOI PRIORITIZATION

A total of 45 HOIs were considered potentially feasible, feasible, or lab-based and feasible. All of these were included in the prioritization step. Notably, although cancer and the cancer subtypes were evaluated as feasible HOIs by the WG, the FDA considered these as a low priority in the alternative method for validation at this point. Therefore cancer and cancer subtypes were excluded from the task of prioritization.

Table 8 HOIs considered as feasible for alternative validation and for which such validation could be based on laboratory results

Health outcomes of interest	Laboratory values	Potential to obtain from all/most DPs[#]
Acute disseminated intravascular coagulation	Platelet count, prothrombin time, activated partial thromboplastin time, thrombin time, plasma fibrinogen, plasma factor V and VIII, fibrin degradation products, D-dimer	Low
Agranulocytosis	Absolute neutrophil count	High
Dyslipidemias	Lipid profile (LDL, HDL, TG)	High (note each test is separate)
Hyperglycemia	Plasma glucose level	MSCDM (all DPs)*
Hyperthyroidism	T_3, T_4, TSH	High
Hypoglycemia	Plasma glucose level	MSCDM (all DPs)*
Hypothyroidism	T_3, T_4, TSH	High
Lactic acidosis	Plasma lactate level	Moderate (will be incomplete as most DPs do not have inpatient labs)
Obesity	BMI (weight and height)	MSCDM (DPs with EHR)*
Pancytopenia	White blood count, neutrophil count, platelet count	High, MSCDM (all DPs)*

* Outpatient labs only with exception of a few Kaiser Permanente sites
[#] Information provided by Marsha A. Raebel from Mini-Sentinel Data Core.

After excluding cancer and the cancer subtypes, the remaining 28 HOIs were reviewed internally by the FDA to determine those of highest priority that should be investigated further. The criteria shown in **Table 9** were provided to assist in this process. Six HOIs were ranked as high priority by the FDA; 12 were considered medium priority, and the other 10, which were all lab value based HOIs, were ranked as low priority. The top six HOIs were suicide, type 1 diabetes, hypertensive crisis, pulmonary fibrosis, pulmonary hypertension and spontaneous abortion, which were selected to be investigated for further information.

Table 9 Summary and criteria for HOIs considered in prioritization process

HOIs	Best data source	FDA priority	WG Initial assessment	Overlap with MSDD (likely/ unlikely)	Prevalence (or incidence) in general population	Size of cohort (N)	Generalizability of existing algorithm	Objectivity of gold standard within DPs (high, moderate, unclear)	DP internal registries (Y/N)
Suicide	National Death Index	High	Feasible	Likely (National)	0.01%	All deaths in US	Not applicable*	Moderate (cause of death = suicide)	N
Type 1 diabetes	T1D Exchange	High	Feasible	Likely (National)	0.33% by age 18	26,000	Previous studies only in children or young adults	High (clinician dx)	Y (Health Partners)
	KP Northern California and Health Partners Registries			Likely (DPs)		232,000 (both T1D&T2D)		Low (difficult to differentiate T1D from T2D)	
Hypertensive crisis	Health Maintenance Organization Research Network Hypertension Registry	High	Potentially Feasible	Likely (DPs)	1.00% in hypertension population	Unclear	Not applicable*	High (based on blood pressure)	Y (KP-Colorado, Northern California, Health Partners)
Pulmonary fibrosis	Pennsylvania Idiopathic Pulmonary Fibrosis Registry	High	Potentially Feasible	Likely (Include Geisinger)	0.14 to 0.47%	Unclear	None	Unclear (patients self-register; unclear about case confirmation)	N
Pulmonary hypertension	Registry to Evaluate Early And Long-Term PAH Disease Management	High	Potentially Feasible	Likely (National)	<65yr: 0.01% 65+yr: 0.05%	3,515 (03/2006-12/2012)	None	High (clinician dx)	N

HOIs	Best data source	FDA priority	WG Initial assessment	Overlap with MSDD (likely/ unlikely)	Prevalence (or incidence) in general population	Size of cohort (N)	Generalizability of existing algorithm	Objectivity of gold standard within DPs (high, moderate, unclear)	DP internal registries (Y/N)
Spontaneous abortion	National Vital Statistics System - Fetal Death Data	High	Potentially Feasible	Likely (National)	8-20% in the first 20 weeks	All in US that meet reporting criteria	Only one study from Denmark, not generalizable to the US	Moderate (some issues with reporting by States)	N
	National Children's Study			Likely (National)		Unclear		Unclear (likely high)	
Birth defects/ Congenital anomalies	National Vital Statistics System - Birth data	Medium	Feasible	Likely (National)	3.00%	All births in US	None	Unclear	N
Narcolepsy	Stanford-based Center for Narcolepsy database	Medium	Feasible	Likely (Local-California)	0.03%	Several thousand	None	High (clinician dx)	N
Tuberculosis	National Notifiable Diseases Surveillance System	Medium	Feasible	Likely (National)	0.04%	Unclear	None	High (Reportable cases with specific criteria)	N
Aplastic anemia	International Blood and Marrow Transplant Research observational database	Medium	Potentially Feasible	Likely (Worldwide)	0.6-6.1 cases per million (incidence)	Unclear	Not applicable*	High (clinician dx)	N

HOIs	Best data source	FDA priority	WG Initial assessment	Overlap with MSDD (likely/ unlikely)	Prevalence (or incidence) in general population	Size of cohort (N)	Generalizability of existing algorithm	Objectivity of gold standard within DPs (high, moderate, unclear)	DP internal registries (Y/N)
Blindness	Database in Social Security Administration	Medium	Potentially Feasible	Likely (National)	0.80%	Unclear	None	High	N
Cirrhosis	Organ Procurement and Transplantation Network registry	Medium	Potentially Feasible	Likely (National)	0.15%	16,516	End-stage liver disease and VA population	High (clinician dx; transplant professionals confirmed diagnosis)	N
Deafness	Database in Social Security Administration	Medium	Potentially Feasible	Likely (National)	0.30% are "functionally deaf"	Unclear	None	High	N
Idiopathic thrombocytopenic purpura	Intercontinental Pediatric and Adult Intercontinental Registry on Chronic ITP	Medium	Potentially Feasible	Likely (Worldwide)	0.01%	2,410 (Worldwide)	None	High (clinician dx)	N
Juvenile rheumatoid arthritis	Childhood Arthritis and Rheumatology Research Alliance Registry	Medium	Potentially Feasible	Likely (National)	0.40%	2,571 (05/2010-06/2011)	Adults	High (International League Against Rheumatism criteria)	Y (Cincinnati Children's Hospital)

HOIs	Best data source	FDA priority	WG Initial assessment	Overlap with MSDD (likely/ unlikely)	Prevalence (or incidence) in general population	Size of cohort (N)	Generalizability of existing algorithm	Objectivity of gold standard within DPs (high, moderate, unclear)	DP internal registries (Y/N)
Pneumonia (Hospital-acquired)	CDC's National Healthcare Safety Network	Medium	Potentially Feasible	Likely (National)	0.55% of hospitalizations	Likely large	None	High (healthcare report, diagnosis may vary by location of hospital reporting)	N
	National Trauma Data Bank			Likely (National)		> 5 million records		High (clinical assessment during hospitalization)	
	The Pediatric Ventilator-Associated Pneumonia Registry			Unlikely		552 (01/2009-07/2009)		High (clinical diagnosis using PNU1, specific criteria for pneumonia)	
Thrombotic thrombocytopenic purpura	Transfusion Medicine/ Hemostasis clinical trial network-TTP registry	Medium	Potentially Feasible	Likely (National)	0.000015%	Unclear	Not applicable*	Unclear	N
Solid organ transplant infections	Organ Procurement and Transplantation Network	Medium	Potentially Feasible	Likely (National)	>50% in the 1st year of transplantation	Unclear	Aspergillosis infection	Unclear	N

HOIs	Best data source	FDA priority	WG Initial assessment	Overlap with MSDD (likely/ unlikely)	Prevalence (or incidence) in general population	Size of cohort (N)	Generalizability of existing algorithm	Objectivity of gold standard within DPs (high, moderate, unclear)	DP internal registries (Y/N)
Acute disseminated intravascular coagulation	Mini-Sentinel DPs	Low	Lab-based	Likely (DPs)	30-50% in sepsis; 1% of hospital admissions	Unclear	None	High (Lab data)	N
Agranulocytosis	Mini-Sentinel DPs	Low	Lab-based	Likely (DPs)	0.00025% (incidence)	Unclear	None	High (Lab data)	N
Dyslipidemias	Mini-Sentinel DPs	Low	Lab-based	Likely (DPs)	15.00%	Unclear	None	High (Lab data)	N
Hyperglycemia	Mini-Sentinel DPs	Low	Lab-based	Likely (DPs)	5.40%	Unclear	None	High (Lab data)	N
Hyperthyroidism	Mini-Sentinel DPs	Low	Lab-based	Likely (DPs)	1.30%	Unclear	None	High (Lab data)	N
Hypoglycemia	Mini-Sentinel DPs	Low	Lab-based	Likely (DPs)	35-100% of diabetics	Unclear	Three different hospital ERs	High (Lab data)	N
Hypothyroidism	Mini-Sentinel DPs	Low	Lab-based	Likely (DPs)	4.60%	Unclear	None	High (Lab data)	N
Lactic acidosis	Mini-Sentinel DPs	Low	Lab-based	Likely (DPs)	0.01%	Unclear	None	High (Lab data)	N
Obesity	Mini-Sentinel DPs	Low	Lab-based	Likely (DPs)	35.70%	Unclear	Women	High (Lab data)	N
Pancytopenia	Mini-Sentinel DPs	Low	Lab-based	Likely (DPs)	17.70% in pediatrics	Unclear	None	High (Lab data)	N

* There were validation studies but performance was low.

G. VERIFICATION OF ALTERNATIVE DATA FOR HIGH-RANKING HOI

For the six HOIs that were ranked as high priority by the FDA, information about the linkability, accessibility, complexity, and cost was further collected as described in the methods section. Detailed findings of this process are listed below. The summary of the findings are also shown in **Table 10**.

<u>**HOI #1:** Suicide</u>

<u>**Alternative Data Source:** National Death Index (NDI plus service)</u>

1. Overlap with MSDD. Definite. NDI is based on all deaths in the United States.

2. Prevalence of the HOI in general population. Suicide occurs at a rate of 11/100,000 (0.01%) in the United States.[45]

3. Size of cohort in the data source. NDI includes all Americans who died during 1979-2011. Data for 2012 will be available in spring 2014.

4. Generalizability of existing algorithm. See Summary document in **Appendix A**. There were several validation studies in the literature but PPV was considered insufficient by Mini-Sentinel Protocol Core.

5. Objectivity of gold standard used in the data source. Moderate. NDI is based on death certificates; there may be some under-reporting of suicide on death certificates.

6. Linkability between the data source and MSDD. The NDI can definitely be linked to MSDD using at least one of following combinations of data items: 1) First name and Last name and Social security number; 2) First name and Last name and Month of birth and Year of birth; 3) Social security number and Full date of birth and Sex; 4) Probabilistic matching approach with score of probabilistic matching will be provided.

7. Accessibility of the data source. NDI is easily accessible. NDI users need to submit an application form with a current IRB approval document to NDI. The application will be distributed to 12-member panel to review and comment. It takes approximately 2-3 months for the application to be reviewed and approved. After approval, user subject's record on CD must be sent to NDI staffs by mail and the search result will be sent back to users within 2 weeks. Search results will be provided in 10 files (5 PDF and 5 standard text format files). A text file (named causes) will provide death status (State of death, date of death, death certificate number, age at death), demographic data on NDI records (name, date of birth, sex, race, state of residence, state of birth, etc.), and underlying causes of death. Additional detailed information is available on the NDI website: http://www.cdc.gov/nchs/ndi.htm.

8. Complexity of the data source. The data provided from NDI is in standard text format which is not hard to transform to any statistical program. Moreover, NDI provides some summary statistics in PDF format which may help users to more understand the data. There is standard format for NDI data with a user guide available. However, the NDI data have important limitation especially in causes of death. NDI is authorized to release only the coded causes of death for either of the following two types of NDI record matches: 1) matches that are ranked first in the NDI list of possible NDI record matches, 2) any

match (regardless of its rank) that has a high-enough probabilistic score to be assigned a Status Code of "1" (meaning true match; assumed dead).

9. Costs. The charges for NDI data are based on the number of subjects requested (not the number of records). There are two types of NDI services - NDI Routine (not included causes of death) and NDI plus (included causes of death). For this HOI the NDI plus service would be required. In the NDI plus service, charges are as following: 1) Service charges for initial submission: $350; 2) Service charges for each subsequent submission: $100; 3) Charges for subject with UNKNOWN vital status: $0.21 per subject per year searched: ex. 200 subjects for 2001-2010 (10 years) = 200*10*0.21 = $420; 4) Charges for subject who are KNOWN decedents: $5.00 per subject (regardless of year searched and assuming researcher has no death certificates); 5) Charges for subject who are KNOWN decedents: $2.50 per subject (regardless of year searched and assuming researcher has obtained death certificates)

10. Other issues/considerations. There are several issues and considerations. First, as stated in the summary of suicide HOI (**Appendix A**), there are different measures of suicidal behavior (suicidal ideation, attempted suicide, completed suicide, etc.). The NDI data are only useful for completed suicide. Second, there has been discussion/plan among Mini-Sentinel to obtain NDI data for all patients in MSDD (as part of CDM). If this were to occur it would preclude the need to obtain NDI data for this project separately. Third, the cost of the NDI data will vary with the approach used. The likely approach for Phase 2 of this project (if suicide were selected HOI) would be to obtain data from NDI by sending the identifier of patients identified in the MSDD as being cases (based on algorithm tested). Another possible approach is that sending data of all decedents in MSDD to NDI and collect all possible data including cause of death and use them as gold standard for suicide related death. The problem of this later approach is that it may be more expensive.

<u>HOI #2:</u> Hypertensive Crisis

<u>**Alternative Data Source:**</u> HMORN HTN Registry

1. Overlap with MSDD. Definite. Registry is populated by Kaiser Permanente Northern California, Kaiser Permanente Colorado, and Health Partners, all of which also provide data to MSDD.

2. Prevalence of the HOI in general population. Although an estimated 50 million or more adult Americans suffer from hypertension. About 1-2% of those with hypertension will experience hypertensive crisis (those it can occur in non-hypertensives). Nevertheless, this condition does affect upward of 500,000 Americans each year, and is therefore a not insignificant cause of serious morbidity in the US.[46]

3. Size of cohort in the data source. Unclear. The registry includes all patients from the 3 sites with hypertension spanning the years 2000 through 2009 (based on criteria for entry). This totals 1,745,841 patients. It is not clear how many of these may meet definition of hypertensive crisis. The registry is primarily for essential hypertension and patients must meet one of the following have 1) a hypertension diagnosis or dispensed anti-hypertensive medication, or 2) have elevated BP on 2 or more consecutive outpatient visits.

4. Generalizability of existing algorithm. See Summary document in **Appendix A**. The Mini-Sentinel Protocol Core has investigated validation studies of "hypertensive emergency". Evidence was not sufficient to make a conclusion about a validated algorithm.

5. Objectivity of gold standard used in the data source. High. Blood pressure results are in the database. Hypertensive crisis includes hypertensive urgency and hypertensive emergency. Hypertensive urgency is a systolic blood pressure of 180mmHg or more OR a diastolic blood pressure of 110 or more, with no signs of organ damage. Patients may have symptoms of headache, shortness of breath, nose bleed, and anxiety. Hypertensive emergency is a blood pressure exceeding 180mmHg systolic or 120mmHg diastolic plus signs of impending or progressive target organ dysfunction (kidneys, eyes, brain, heart). While the HMORN HTN Registry has blood pressure results, it does not have information about symptoms or end-organ damage. Therefore, while blood pressure results could be used to define hypertensive crisis generally, it would be difficult to differentiate hypertensive urgency from hypertensive emergency.

6. Linkability between the data source and MSDD. High. The patients can be definitely linked.

7. Accessibility of the data source. High. The data are available to Mini-Sentinel investigators.

8. Complexity of the data source. Low. The database is well described in user manual with tables and fields similar to MSCDM.

9. Costs. Depends on scope of project.

10. Other issues/considerations. There are several issues and considerations. First, as stated in above, this is a hypertension registry - it is not specifically for hypertensive crisis. Without running a query of the database it is not possible to say the number of patients with hypertensive crisis. It is possible that the number is low since the entry criteria for the registry focuses on essential hypertension.

A second issue is the dates of the registry (2000-2009). While this provides plenty of overlap with MSDD (which has data back to 2000), it precludes the ability to examine more recent data. However, this may not be necessary for a validation study.

Third, it is not clear that the HMORN HTN Registry provides any advantages over the MSDD itself assuming that the MSDD CDM includes (or will include) vital signs and specifically both diastolic and systolic blood pressure, and therefore could be used to identify people above the threshold blood pressure for hypertensive crisis. In essence this could be considered a "laboratory-based HOI." However, it is unclear the frequency with which elevated blood pressures are recorded in people when the crisis occurs versus afterward when treatment might have already started.

HOI #3: Type I Diabetes

Alternative Data Source: Kaiser Permanente Northwest and Health Partners Diabetes Registries

1. Overlap with MSDD. Likely. Both Kaiser Permanente Northwest (KPNW) and Health Partners have registries for Diabetes. Both of these organizations also provide data to MSDD so there is definite overlap.

2. Prevalence of the HOI in general population. While the prevalence of diabetes in general is on the raise, Type 1 diabetes is the much less common form. The prevalence of T1D is approximately 1 in 300 in the United States by 18 years of age.[47]

3. Size of cohort in the data source. KPNW: 232,000 (includes both types), Health Partners: 34,147

4. Generalizability of existing algorithm. See Summary document in **Appendix A**. The Mini-Sentinel Protocol Core has not conducted a review on validation studies for type 1 diabetes. The WG found a validation study with high PPV (97%) but it was from a single hospital clinic. The WG also found a Canadian study with a high PPV (>95%). These were both conducted in children or young adults (which is usually when Type 1 diabetes is diagnosed), but it was not clear if these could be generalized to Mini-Sentinel or to adults. The WG also identified issues with the ICD-9 used in these studies. While codes are not age specific the code for Type 1 diabetes is listed as "type 1 (juvenile type)" which may influence who receives the code. It is also important to note that while previous studies used ICD-9 code 250.x, the code that is likely to be of most interest to FDA is 249.x which is "secondary diabetes", defined as "diabetes mellitus (due to) (in) (secondary) (with): drug-induced or chemical induced, or infection."

5. Objectivity of gold standard used in the data source. Low. Both registries include all diabetes patients (Type 1 and Type 2) and are based on claims and EHR data, but in neither registry are the two types differentiated. The KPNW registry includes confirmation of cases by an endocrinologist, but still not by type. In practice that is no clearly definitive way to do differentiate between the two types of diabetes and it is also not necessary for treatment. Type 1 diabetes usually occurs in those <35 years of age but with rise in obesity children are not getting type 2 diabetes. Type 1 patients are usually not obese but many Type 2 diabetics are not either. Urine ketones are often present in Type 1 diabetes but may also be positive in Type 2 if there is severe volume depletion. Anti-glutamic acid decarboxylase (GAD) antibodies, islet cell antibodies, and insulin autoantibodies are present in 85% of patients with Type 1 at the time of diagnosis, but may disappear within a few years, and usually not required for diagnosis. Nevertheless, this may be the only way to confirm Type 1 diabetes but will be negative in some Type 1 patients. In a previous study using the KPNW diabetes registry investigators used the antibody tests to differentiate Type 1 patients along with criteria like age. They identified 129 incident Type 1 diabetics during 1999 to 2005.

6. Linkability between the data source and MSDD. High. The patients can be definitely linked.

7. Accessibility of the data source. High. The registries are available to Mini-Sentinel investigators.

8. Complexity of the data source. Low. The registries are similar data to MSCDM.

9. Costs. Depends on scope of project.

10. Other issues/considerations. None.

Alternative Data Source: T1D Exchange (Type 1 Diabetes Registry)

1. Overlap with MSDD. Likely. The T1D Exchange is populated by physicians 70 pediatric and adult endocrinology clinics with wide geographic distribution across the US (in 32 states). There is likely to be some overlap with MSDD.

2. Prevalence of the HOI in general population. While the prevalence of diabetes in general is on the raise, Type 1 diabetes is the much less common form. The prevalence of T1D is approximately 1 in 300 in the United States by 18 years of age.[47]

3. Size of cohort in the data source. The registry started in 2010. There are 26,000 patients enrolled (as of June 2012).

4. Generalizability of existing algorithm. See Summary document in **Appendix A**. The Mini-Sentinel Protocol Core has not conducted a review on validation studies for type 1 diabetes. The WG found a validation study with high PPV (97%) but it was from a single hospital clinic. The WG also found a Canadian study with a high PPV (>95%). These were both conducted in children or young adults (which is usually when Type 1 diabetes is diagnosed), but it was not clear if these could be generalized to Mini-Sentinel or to adults. The WG also identified issues with the ICD-9 used in these studies. While codes are not age specific the code for Type 1 diabetes is listed as "type 1 (juvenile type)" which may influence who receives the code. It is also important to note that while previous studies used ICD-9 code 250.x, the code that is likely to be of most interest to FDA is 249.x which is "secondary diabetes", defined as "diabetes mellitus (due to) (in) (secondary) (with): drug-induced or chemical induced, or infection."

5. Objectivity of gold standard used in the data source. High. The registry collects core clinical and laboratory data on enrolled patients. Patients are classified as definite or probably based on these data.

6. Linkability between the data source and MSDD. Moderate. The registry does not contain social security number of other identification numbers. However, two options exist for linkage. First, patients in the registry can be contacted by email and asked to consent to participate, in which case they could provide information needed to link. Alternatively, probabilistic matching could be used. The database contains initials, date of birth, zip code, gender, and other variables that could be used for matching (data dictionary is available).

7. Accessibility of the data source. High. The T1D registry desires to make data available to other researchers. Requests for such are reviewed by a steering committee (described as a "semi-formal" process).

8. Complexity of the data source. Low. Described as uncomplicated, and clearly outlined in data dictionary.

9. Costs. Low. The T1D would seek to recover it costs associated with any study. Those costs would vary depending on the nature of the study. For example if patient consent was required then more time on the part of T1D staff would be needed.

10. Other issues/considerations. Probably better than Kaiser and Health Partners Diabetes Registries because of focus on Type 1 Diabetes (and better confirmation of cases). However, clearly would be more effort to link to MSDD.

HOI #4: Pulmonary Fibrosis

Alternative Data Source: Pennsylvania Idiopathic Pulmonary Fibrosis (PA-IPF) Registry

1. Overlap with MSDD. Likely. The PA-IPF registry includes five sites in the state of Pennsylvania that are recruiting patients. One of these sites is the Center for Health Research at Geisinger Health System which is also a participant in the MSDD.

2. Prevalence of the HOI in general population. The estimated prevalence of pulmonary fibrosis is 0.14% to 0.47%.[48]

3. Size of cohort in the data source. Unclear. Unable to contact.

4. Generalizability of existing algorithm. No existing algorithm.

5. Objectivity of gold standard used in the data source. Unclear. Unable to contact.

6. Linkability between the data source and MSDD. Unclear. Unable to contact.

7. Accessibility of the data source. Unclear. Unable to contact.

8. Complexity of the data source. Unclear. Unable to contact.

9. Costs. Unclear. Unable to contact.

10. Other issues/considerations. None

HOI #5: Pulmonary Hypertension

Alternative Data Source: Registry to Evaluate Early and Long-Term PAH Disease Management (REVEAL)

1. Overlap with MSDD. Likely. REVEAL enrolled patients into their multi-center observational registry from 54 participating sites in the United States.

2. Prevalence of the HOI in general population. The estimated prevalence of pulmonary hypertension is 0.0109% among those younger than 65 years of age and 0.0451% in those 65 years and older.[49]

3. Size of cohort in the data source. REVEAL includes information on 3,515 individuals with pulmonary artery hypertension.

4. Generalizability of existing algorithm. No existing algorithm.

5. Objectivity of gold standard used in the data source. High. To be eligible for inclusion, patients needed to be newly diagnosed or previously diagnosed with WHO group I PAH. The clinical criteria that needed to be documented by right heart catheterization any time before study enrollment to identify eligible individuals were: 1) mean pulmonary artery pressure (mPAP) >25mmHg at rest OR mPAP >30mmHg with exercise; AND 2) pulmonary capillary wedge pressure (PCWP) ≤18mmHg; AND 3) pulmonary vascular resistance (PVR) ≥240 dynes × s × cm^{-5b}

6. Linkability between the data source and MSDD. Definite. There are no individual identifiers like social security number or medical record number; however there may be ways to get probabilistic match based on unique information contained in the registry. For example, the registry contains the birth date of the individual as well as the date and location that the right heart catheterization was performed. These elements together may result in a high probabilistic match rate.

7. Accessibility of the data source. Definite. Accessibility of the data is probable. The sponsor of the registry sponsored it as part of the IND for a product they developed. The biggest concern of the sponsor is complying with their existing confidentiality agreements with the patients that are included in the registry.

8. Complexity of the data source. Low. The data can be made available in straightforward files that include only the data elements that are necessary to complete the project. The analysts on the project have experience in creating analytic files from the existing data that has been captured as part of the registry.

9. Costs. Would depend on the scope of the project; however only cost would be funding programmer or analyst to create analytic dataset from existing files. Ballpark estimate of around $10,000.

10. Other issues/considerations. The REVEAL registry enrolled both prevalent and incident cases of disease. Individuals enrolled in the registry within three months of their diagnosis were considered incident cases of pulmonary hypertension. The objectives of the registry were to understand more about the current treatment patterns of patients in the US that were diagnosed with WHO group I pulmonary hypertension and follow these patients to document both short-term and long-term patient outcomes. Patient enrollment began in March 2006 and continued through December 2009. The database was closed on December 31, 2012. Importantly, Group 1 within the WHO pulmonary hypertension classification system includes those with pulmonary artery hypertension and does not include those with pulmonary venous hypertension with left heart disease or pulmonary hypertension associated with respiratory diseases. The registry is sponsored by Actelion.

HOI #6: Spontaneous Abortion

Alternative Data Source: Center for Disease Control (CDC) National Center for Health Statistics (NCHS) Division of National Vital Statistics System (NVSS) Fetal Death Data

1. Overlap with MSDD. Definite. The Fetal Death Data is collected on all fetal deaths over 350g and/or which had pregnancies 20 weeks or over in the United States.

2. Prevalence of the HOI in general population. Incidence is 8 to 20% in pregnancies under 20 weeks with 80% of those being in the first 12 weeks of gestation.[50-52]

3. Size of cohort in the data source. The NCHS data includes all fetal deaths as described above. Data sets from 1982 to 2006 are available for download on the website at: http://www.cdc.gov/nchs/data_access/Vitalstatsonline.htm

4. Generalizability of existing algorithm. See Summary document in **Appendix A**. Only one Denmark study was found, but it is unlikely to be generalizable to the US.

5. Objectivity of gold standard used in the data source. High.

6. Linkability between the data source and MSDD. Definite. Linking to external databases is prohibited per NCHS policy due to concerns of confidentiality loss. The WG wonders if FDA has any means to get around this policy.

7. Accessibility of the data source. High. There is published data readily available from 1982 to 2006 as stated above. The data in these data sets are de-identified but researchers may request additional information (e.g., county of fetal death) on the data following a review of proposed project. A user guide is available with each year on the website as well. The data is available at the website listed above. The process for requesting data is straight-forward and involves following steps: 1) Submission of Project Review Form by investigator, 2) National Association for Public Health Statistics and Information System/National Center for Health Statistics (NAPHSIS/NCHS) will review the data request, 3) If approved, a Data Use Agreement form will need to be completed, 4) If exact dates of events are desired, additional paper work will need to be completed. Full directions of obtaining additional data can be found here: http://www.cdc.gov/nchs/nvss/dvs_data_release.htm

8. Complexity of the data source. Low. Each published year of the Fetal Death data is in SAS format, for easy upload and analysis of data.

9. Costs. None. There is no cost to access the published years' data nor does it appear to cost the investigator for access of the additional information on the published data (e.g., county of fetal death). If the request for access to additional information is denied, the Research Data Center at the CDC can analyze the data for the principal investigator for a fee.

10. Other issues/considerations. There are several issues and considerations. First, only fetal deaths with fetuses weighing over 350g or carried to at least 20 weeks of pregnancy are required to be reported by the states. Thus, a large portion of the spontaneously aborted fetal deaths may not be captured. However, the biggest issue is the prohibition of linking with external databases; this restriction forces this data source to be unusable in this project unless the FDA has some way around this.

Alternative Data Source: National Children's Study

1. Overlap with MSDD. Likely. The National Children's Study recruits (planned) women from all around the country so there should be overlap with MSDD.

2. Prevalence of the HOI in general population. Prevalence was not found; incidence is 8 to 20% in pregnancies under 20 weeks with 80% of those being in the first 12 weeks of gestation.[50-52]

3. Size of cohort in the data source. In the pilot study 5,000 participants were/are being recruited (with 20% attrition) from 40 locations across US. The main study (which won't start recruitment until mid-2015) will be much larger but depends on IOM and Congressional review that is to occur in 2014.

4. Generalizability of existing algorithm. See Summary document in **Appendix A**. Only one Denmark study was found, but it is unlikely to be generalizable to the US.

5. Objectivity of gold standard used in the data source. High. Clinician confirmation.

6. Linkability between the data source and MSDD. Likely. Study is designed to meet data standards and at a minimum will have information necessary for probabilistic matching.

7. Accessibility of the data source. High. A plan for data access is in place.

8. Complexity of the data source. Low. Standard data sets and uniform data standards.

9. Costs. Unclear.

10. Other issues/considerations. The timeline makes the use of this data impractical in the near term.

VI. CONCLUSIONS AND RECOMMENDATIONS

As described in detail in this report, the WG followed a six step process to investigate the feasibility of using an alternative electronic data source for validation of each of 99 HOIs. Among these 99 HOIs, the WG found 16 to have been well-validated in previous studies (**Table 3**), and so no future validation (alternative or otherwise) was considered necessary. In addition, 11 HOIs were deemed "not feasible" for alternative validation (**Table 4**), and an additional 27 HOIs were considered unlikely for alternative validation (**Table 5**). In most cases, HOIs were determined to be not feasible or unlikely because of lack of availability of alternative databases or because of major limitations with such databases. Nevertheless, based on the initial review, the WG did identify 45 HOIs to be potentially feasible or feasible for alternative database validation (**Table 6**, **Table 7**, **Table 8**).

The 45 feasible or potentially feasible HOIs included 10 "lab-based HOIs" (
Table 8). These were HOIs that were considered by the WG to be able to be validated using laboratory results alone (i.e., as the gold standard for confirming cases identified by the algorithm tested). The WG also determined that many or all of these laboratory results could be obtained from Mini-Sentinel Data Partner internal databases. Because of this, the WG viewed these HOIs as easy targets for alternative validation, with good accessibility to the data, likely low cost, and clear linkability. The WG felt that it might even be possible to incorporate many or all of these HOIs into a single, efficient validation study.

Among the remaining 35 potentially feasible and feasible HOIs were 17 cancer-related HOIs. Again the WG considered these to be highly feasible because of the clear availability of linkable alternative databases – specifically the cancer registries described above. Further, cancer is one of the few examples where this method of validation has been previously attempted and published. However, from a surveillance perspective, the cancer HOIs were deemed low priority for phase II of this project. This was primarily because of the time lag between exposure and cancer diagnosis, and the related difficulty following patients for long periods of time in administrative data such as MSDD.

The other 28 potentially feasible and feasible HOIs were ranked by FDA staff as low, medium, or high priority with respect to surveillance importance with the intent that the WG concentrated its further efforts on the highly ranked HOIs. Six HOIs were considered high priority - these included suicide, type 1 diabetes, hypertension crisis, pulmonary fibrosis, pulmonary hypertension and spontaneous abortion.

Further investigation, including direct contact and discussion with the database vendors, was conducted for each of these six HOIs to confirm the feasibility determined in the initial search, and to allow for a ranking among the six HOIs (**Table 10**) for phase II recommendations.

The WG concluded that the best candidates for alternative validation in Phase II are 1) suicide, using data from National Death Index as the alternative data source, and 2) type 1 diabetes, using the T1D Exchange Registry as the alternative data source. Note that type 1 diabetes had a second data source –

Table 10 Summary of information collected for the final six HOIs and WG final assessment

HOIs	Best data source	Linkability	Accessibility	Complexity	Cost	Other	WG final assessment
Suicide	National Death Index	Definite	Definite, fully accessible	Low	Well defined	Completed suicide only	1
Type 1 diabetes	T1D Exchange	Definite	Define, process well-defined	Low	Unclear	Better than KP to identify Type 1 but more effort to link	1
	Kaiser Permanente Northern California and Health Partners Registries	Definite	Definite, fully accessible	Low	Low	Difficult to differentiate Type 1 v. Type 2	2
Hypertensive crisis	Health Maintenance Organization Research Network Hypertension Registry	Definite	Definite, fully accessible	Low	Unclear	Not better than MSCDM	2
Pulmonary fibrosis	Pennsylvania Idiopathic Pulmonary Fibrosis Registry	Unclear	Unclear	Unclear	Unclear	Not able to get information	Unknown
Pulmonary hypertension	Registry to Evaluate Early And Long-Term PAH Disease Management	Definite	Definite, but with confidential concern	Low	Around $10,000 (depends on scope)	Include both incident and prevalent case. Note for incident cases, only PAH was included.	3
Spontaneous abortion	National Vital Statistics System - Fetal Death Data	Prohibited	Define, process well-defined	Low	Free	Most recent data are 2006 (may be updated to 2008 soon)	Not feasible
	National Children's Study	Definite	Definite	Low	Unclear	Patient recruitment for main study will not start until mid-2015	Not feasible

Rating 1= best candidate, 2 = very good candidate, 3 = good candidate

internal registries in Kaiser Permanente Northern California and Health Partners - which was considered a very good option (but not better than the T1D exchange). Next, the WG considered <u>hypertensive crisis</u>, using data from the Health Maintenance Organization Research Network- Hypertension Registry, to be a <u>very good candidate for alternative validation</u>. Last, <u>pulmonary hypertension</u>, using the Registry to Evaluate Early And Long-Term PAH Disease Management, was considered a <u>good candidate for alternative validation</u>.

The WG also determined that neither of the data sources identified for spontaneous abortion were viable. These included the Fetal Death Dataset from the CDC National Center for Health Statistics and National Children's Study. The former was determined to be not linkable and the later to have insufficient participant enrollment. Finally, there was one HOI among the six for which the WG could not retrieve sufficient information. That was the pulmonary fibrosis and the data source was the Pennsylvania Idiopathic Pulmonary Fibrosis Registry.

In conclusion, the WG recommends that the FDA and Mini-Sentinel program consider suicide or type 1 diabetes for phase II of this project. Hypertensive crisis and pulmonary hypertension could also be considered. Finally, the 10 lab-based HOIs lend themselves to identification/validation using electronic databases.

VII.   ACKNOWLEDGEMENTS

VIII.   APPENDIX A

HOI: <u>Acute disseminating encephalomyelitis</u>

Previous Validation Studies: The FDA CBER has conducted a systematic review for identifying acute disseminating encephalomyelitis (ADEM) using administrative or claims data (unpublished). Two previous validation studies of ADEM were identified. Neither was sufficient for this HOI to be considered previously validated. In a study conducted by Leake et al.,[53] incident cases of ADEM in subjects less than 20 years of age were ascertained via three mechanisms: 1) database search of ICD-9 codes 052.0, 055.0, 136.9, 323.5, 323.6, 323.8 and 323.9; 2) systematic review of radiology reports; and 3) prospective identification by study participants. Cases of ADEM were defined as subjects experiencing acute or subacute abnormal neurological symptoms with central nervous system demyelination not explained by another illness. Sixty-four ADEM cases were identified through this strategy, however, the number of cases specifically identified by use of the ICD-9 codes rather than the review of radiology reports or prospective identification was not reported. Therefore, although 42 cases of ADEM were verified after medical record review, the positive predictive value (PPV) of the ICD-9 codes alone was not available. However, the PPV of the entire case finding algorithm (using ICD-9 codes, radiology report review, and clinical reporting) was 66%.

The second study meeting the workgroup (WG)'s inclusion criteria also described incident cases of ADEM. This study by Langer-Gould et al.[54] identified cases of ADEM occurring in children less than 18 years of age in the Kaiser Permanente Southern California health maintenance organization. This organization includes approximately 3.2 million members with greater than 900,000 in the study's targeted age range. The medical record database was searched using the ICD-9 code 323.61 (ADEM) to identify cases, although the authors simultaneously collected cases of other acquired demyelinating syndromes (ADS), including optic neuritis and variations of multiple sclerosis. ADEM was defined as the presence of encephalopathy in addition to multifocal neurological deficits through the use of pediatric consensus definitions. Fifteen cases were confirmed through medical record review, but the initial number identified using the ICD-9 code for ADEM was not reported. Therefore, the accuracy of this ICD-9 code to identify ADEM cases was not available.

Summary of Search of Linkable Databases: The WG was unable to identify any registries or databases that could be used for alternative validation.

Other Issues/Comments: The ICD-9 codes used in the study conducted by Leake et al.[53] are not all specific to ADEM and can include encephalitis and/or myelitis. Diagnosis of ADEM requires brain MRI and presence of specific clinical symptoms.
Definition: ADEM is a neurologic disorder with abrupt onset of multifocal neurologic deficits from likely autoimmune destruction of myelinated cells in the central nervous system, with clinical features potentially including encephalopathy, weakness, sensory loss, and seizures. This disorder is more common in children than adults, and there is often a history of preceding infection or immunization.

WG Recommendation: Not feasible. Linkage to alternative databases is not feasible because there are no registries or external databases to identify cases with ADEM.

HOI: Acute disseminated intravascular coagulation

Previous Validation Studies: The WG conducted a systematic search and found no previous validation studies for acute disseminated intravascular coagulation (ADIC).

Summary of Search of Linkable Databases: The WG was able to identify only one registry that included patients in the US with ADIC. This was a global-registry of patients receiving recombinant-activated factor VII (rFVIIa). The registry included patients with multiple diagnoses that could lead to treatment with rFVIIa, one of which was ADIC. However, the registry is quite small and contains few patients with ADIC, all of whom were treated with a specific intervention. For this reason it is not a viable source for a alternative validation study.

Other Issues: The diagnosis of ADIC is objectively based on laboratory data including platelets, prothrombin time, fibrinogen, and D-dimer. Databases which have laboratory result data available may be potential sources for a validation study. The WG's investigation indicates that all of the Mini-Sentinel Data Partners have data on results of platelet count tests. However, the availability of the other laboratory tests is not clear. If a Mini-Sentinel Data Partner(s) has laboratory result information on the remaining lab tests than this is an HOI that could be validated based on lab values.

WG Recommendation: Feasible – lab based. This is an HOI that would seem best validated using laboratory values within the Mini-Sentinel Data Partners if the lab values exist in their databases.

HOI: Acute respiratory failure

Previous Validation Studies: The Mini-Sentinel Protocol Core has published a systematic review for identifying this HOI using administrative or claims data.[11] Only two studies were identified in the search that used ICD-9 codes for the identification of acute respiratory failure and neither algorithm was considered validated. Therefore, there appear to be no previously validated algorithms for identification of acute respiratory failure.

Summary of Search of Linkable Databases: The WG identified three registries in the WG's search for alternative data sources for acute respiratory failure. The Extracorpeal Life Support Organization (ELSO) registry includes patients that receive ECMO as part of their treatment. The registry includes more than just respiratory failure patients. However, a group of patients with acute respiratory failure will most likely be in the registry. Note, this is a group that would likely have very severe disease. Moreover, the registry does not appear to be linkable as it contains no patient health information. The second registry that was identified was a registry of patients from three medical centers. One of the centers is Geisenger Medical Center, which is a Mini-Sentinel Data Partner, but no information on the size of the registry or the timeframe of the registry could be found. The final registry identified was the ToxIC registry. The ToxIC registry collects patient information about toxicological exposure (including medications) and clinical symptoms which are confirmed by laboratory values and toxicologists. Acute respiratory failure is one of the potential clinical conditions which could be measured in ToxIC registry. However, it seems unlikely due to the concern of linkability, overlap with MSDD and the generalizability to those in the MSDD (see Section E.2 in Chapter V).

Other Issues: None.

WG Recommendation: Unlikely. Because it is unclear how many patients are included in the registry in which Geisenger Medical Center participates, the WG would conclude that validation of this HOI is unlikely using an alternative dataset.

HOI: Agranulocytosis

Previous Validation Studies: The WG conducted a systematic search and found no previous validation studies for this HOI.

Summary of Search of Linkable Databases: The WG found nine registries associated with severe neutropenia or agranulocytosis. Among those, seven registries were product-related. Six of the registries were clozapine-related registries that require users of clozapine to be included in a registry to monitor for occurrence of agranulocytosis. The remaining product-related registry was a registry for daptomycin patients. This registry is also used to monitor for the occurrence of agranulocytosis among daptomycin using patients. These product specific registries could be a potential source for validation of the occurrence of agranulocytosis. However, the main limitation is that these are exposure-based registries and it is not clear how generalizable an algorithm based on these registries would be to other cases of agranulocytosis.

Two additional registries were identified in the search. The first one was the Severe Chronic Neutropenia International Registry (SCNIR). It is a registry for severe chronic or congenital neutropenia patients. It included four kinds of neutropenia including Kostmann Syndrome, Cyclic Neutropenia, Idiopathic Neutropenia and Autoimmune Neutropenia. The patients in this database are not the types that would be of interest for an active surveillance system and therefore this registry would not be useful as an alternative data source. The second registry is the Prospective Nationwide Registry for severe neutropenia. It included patients who started a new chemotherapy regimen for several kinds of cancers including breast, lung, colorectal, and ovarian cancers and lymphoma from 137 community oncology practice sites across USA. Like the other registries noted above, this is a registry that is based on exposure. This is not exposure to a specific medication but rather exposure to chemotherapy that can result in neutropenia. Approximately 24% of included patients (>4,000 patients until 2005) developed severe neutropenia. This may be a potential alternative data source.

Other Issues: None.

WG Recommendation: Feasible – lab based. This is an HOI that would seem best validated using laboratory values within the Mini-Sentinel Data Partners if the lab values exist in their databases. However, there are other registries that contain patients that have experienced agranulocytosis. The primary limitation is these are exposure based registries and algorithms validated with these events may not be generalizable to other cases of agranulocytosis.

HOI: Amyotrophic lateral sclerosis

Previous Validation Studies: The WG conducted a systematic search and found a study conducted by Wittie et al.[55] on methods for combining data from multiple sources to facilitate the successful establishment of a US National Amyotrophic Lateral Sclerosis (ALS) registry; but no PPV was provided. A classification and regression tree (CART)-based methodological approach[56] to developing an algorithm for classifying ALS cases using electronic records has been developed for the HMO research network group (with participating sites: Kaiser Permanente Northern California (KPNC), Southern California (KPSC) and Geisinger Health System (GHS) (N=769; 454; 141)) that correctly classified 93% of ALS subjects.

Summary of Search of Linkable Databases: Several national registries were found, including the National US Veterans ALS registry (n=2400 as of 2008) and The National ALS Registry, which was established in 2010. The later was described as the single largest ALS research project and is designed to identify ALS cases from throughout the entire US. The registry is overseen by the Agency for Toxic Substances and Disease Registry/Centers for Disease Control and Prevention (ATSDR/CDC). However, ATSDR/CDC is prevented from sharing general registry information with the public, including the number of people enrolled in the registry, the number of people enrolled from specific states, average ages, race, and other general information. Nevertheless, it is possible that Mini-Sentinel researchers could work with the ATSDR/CDC.

Other Issues: ALS is a clinically based diagnosis that relies on clinical criteria such as progressive paralysis, amyotrophy, hyperreflexia, and spasticity; during the course of disease, dysphagia, dyspnea, depression, pain and sleep disorders can occur. To confirm diagnoses, all cases are reviewed by an experienced neurologist according to the El Escorial criteria.[57]

WG Recommendation: Well-validated. Validated algorithms have been published with acceptable to excellent PPV.

HOI: Aplastic anemia

Previous Validation Studies: The Observational Medical Outcomes Partnership (OMOP) has reported a systematic review for identifying this HOI using administrative or claims data.[58] There were three out of four articles included in the review used ICD-9 codes as 284.x (284, 284.9, and 284.8) for aplastic anemia (AA). However, the PPV ranged widely from 5% to 59%. The conclusion was that there is no clear consensus on the optimal coding strategy for identifying patients with AA.

Summary of Search of Linkable Databases: The Center for International Blood and Marrow Transplant Research (CIBMTR) observational database and the Autologous Blood and Marrow Transplant Registry (ABMTR) may be useful for validation of aplastic anemia. The CIBMTR joined together the research programs of the National Marrow Donor Program (NMDP) and the International Bone Marrow Transplant Registry (IBMTR) at the Medical College of Wisconsin. The database is widely used for research purposes. Approximately, 330,000 transplant patients were in the database but the actual number of patients with AA is unclear. The ABMTR is a voluntary organization of more than 250 transplantation centers primarily in North and South America that report data on consecutive auto transplantations.

Other Issues: The both registries are held Froedtert Hospital and the Medical College of Wisconsin Clinical Cancer Center. Because the both registries are blood and marrow transplant registry, only AA patients requiring transplantations (severe AA patients) have been included in the registries.

WG Recommendation: Potentially feasible. This HOI could be validated using an alternative database. Recommend focus on severe aplastic anemia and use CIBMTR.

HOI: Autoimmunity

Previous Validation Studies: Validation studies that have developed algorithms focus on specific autoimmune conditions, such as systematic lupus erythematosus (SLE) or inflammatory bowel disease. As such, if relevant they are separately identified on the HOI list.

Summary of Search of Linkable Databases: Two registries were identified as autoimmune condition registries as a broadly encompassing term, both of which lacked details. One registry was based at Kaiser Permanente, which is a Mini-Sentinel Data Partner, but appears to have been funded between 2009-2011 and does not describe a gold standard for ascertaining cases. The second registry is the Hospital for Special Surgery's Autoimmune Disease Registry and Repository, which began in 2009. It includes confirmed cases of SLE and/or anti-phospholipid syndrome. It is unclear how many/if any patients have been enrolled to date. A second issue is that SLE is already considered a HOI independent of the autoimmunity as an HOI.

Other Issues: It would be difficult to develop an algorithm that includes all autoimmune conditions. Validating such would be nearly impossible.

WG Recommendation: Not feasible. Autoimmune conditions as a general HOI is too non-specific to reasonably expect to develop an algorithm with acceptable PPV using an alternative registry or database.

HOI: <u>Birth defects</u>

Previous Validation Studies: The WG conducted a systematic search and found no previous validation studies for this HOI. Some literature exists on algorithms for identification of specific defects (e.g., neural tube defects) but not for birth defects generally.

Summary of Search of Linkable Databases: There are numerous registries that may be useful for linking to the MSDD. Both pregnancy registries and birth defect registries exist. Pregnancy registries follow women who are pregnant forward and identify complications with the mother and child. Pregnancy registries are typically either medical condition or exposure specific. The FDA website includes a listing of nearly 100 such registries. Birth defect registries include cases of various types of defects and typically use self-report or interview methods to ascertain exposures that may have occurred prior to birth. Most states maintain birth defect registries. The CDC tracks and funds many of these but they are not combined into a national registry. There is a "National Birth Defects Registry" that is operated by a private organization but it is not well organized or very large. Last, the National Vital Statistics System collects data on live births, including the presence of a birth defect in the newborn. The data collection form includes identification of parents and "medical record number" of newborn thus might be linkable to MSDD.

Other Issues: Birth defects are not a single HOI. There are many types of birth defects and this needs to be considered in algorithm validation.

WG Recommendation: Feasible. This HOI could be validated using an alternative database but the registry/database selected for linkage to MSDD would depend on the specific exposure (drug) and/or type of birth defect of interest to the FDA. Birth data in National Vital Statistics System is the recommended database.

HOI: Blindness

Previous Validation Studies: The WG conducted a systematic search and found no previous validation studies for this HOI. Javitt et al.[59] tried to validate their algorithm of ICD-9 codes in Medicare database by comparing the prevalence of blindness in the study cohort to that in the Eye Disease Prevalence Research Group study. A similar prevalence of blindness was found and therefore the authors claimed their algorithm validated. However, the WG did not found studies that use the medical chart as the gold standard to validate this or other algorithms for blindness.

Summary of Search of Linkable Databases: One potential alternative data source for validation of a blindness algorithm is the Social Security Administration Information on disabilities, including blindness, is maintained by the Social Security Administration for purpose of determining benefit eligibility. The definition for blindness used is considered legal blindness and the confirmation is strict. However, it is unclear if this database could be linked to the Mini-Sentinel Distributed Database. In addition to the Social Security Administration, blindness registries maintained by state governments were also found in approximately 19 states. One pharmaceutical company is also developing a registry for blind patients with sleep-related problems. Other options include the National registry of drug-induced ocular side effects, and the United States Eye Injury Registry database, but these also have limitations, such as the uncertainly about linkability, overlap with MSDD or generalizability to MSDD.

Other Issues: The legal definition of blindness in US is the best corrected visual acuity less than 20/200 in the better-seeing eye, which is the same as the criteria used in determination of benefit eligibility for blindness in Social Security Administration.

WG Recommendation: Potentially feasible. The Social Security Administration might be a potential feasible source for alternative database validation of blindness as an HOI.

HOI: Brachial neuritis

Previous Validation Studies: The WG conducted a systematic search and found no previous validation studies for this HOI.

Summary of Search of Linkable Databases: The WG was unable to identify any registries or databases that could be used for alternative validation.

Other Issues: Brachial neuritis (BN), also known as neuralgic amyotrophy, is a rare syndrome of unknown etiology affecting mainly the lower motor neurons of the brachial plexus and/or individual nerves or nerve branches. BN usually is characterized by the acute onset of excruciating unilateral shoulder pain, followed by flaccid paralysis of shoulder and parascapular muscles several days later. The incidence of brachial neuritis is approximately 1-2 cases per 100,000 person-years.

WG Recommendation: Not feasible.

HOI: Bronchospasm

Previous Validation Studies: The WG conducted a systematic search and found no previous validation studies for asthma exacerbation. Previous studies have focused on acute exacerbations of asthma as an HOI which could be considered a type of bronchospasm. These studies reported a wide range of PPVs (6-96%)[60,61] and were only conducted in children and thus lack generalizability to all patients that could experience bronchospasm. According to information from the Mini-Sentinel Protocol Core there is currently a PRISM systematic review for this HOI that is pending.

Summary of Search of Linkable Databases: The WG identified the ToxIC registry as a potential source for individuals with bronchospasm. The ToxIC registry collects patient information about toxicological exposure (including medications) and clinical symptoms which are confirmed by laboratory values and toxicologists. Bronchospasm is one of the potential clinical conditions which could be measured in ToxIC registry. This could be particularly relevant for those events that are a result of an exposure and could include drug-induced bronchospasm. However, it seems unlikely due to the concern of linkability, overlap with MSDD and the generalizability to those in the MSDD (see Section E.2 in Chapter V).

Other Issues: The Mini-Sentinel Protocol Core stated in their report that "CBER is interested in acute wheezing. Hard to identify all codes that may relate to this outcome." If the focus of this HOI is on drug-induced bronchospasm then there do not appear to be any available alternative data sources.

WG Recommendation: Unlikely. This HOI is not one that would be able to be validated using an alternative data source.

HOI: Cancer - General

Previous Validation Studies: The WG conducted a systematic search and found no previous validation studies conducted on "cancer" generally as an HOI. There are previous validation studies for specific types of cancers (see separate summaries included in the appendix).

Summary of Search of Linkable Databases: Several registries were identified that contain multiple types of cancers and may serve as a viable alternative data source for many of the different cancers. The Surveillance Epidemiology and End Results (SEER) registry is a program sponsored by National Cancer Institute that collects data on patients with cancer from 18 separate geographic areas across the US. The registry includes individuals that were diagnosed with cancer while residing in these geographic areas. To be included in the registry, the case has to be confirmed by the cancer registrar either through pathology reports or by clinical diagnosis in the absence of a pathology report. The SEER registry has been maintained since 1973.

The second registry identified was the SEER-Medicare linked database that includes all Medicare eligible patients identified through SEER and their Medicare claims data. This database also includes a 5% sample of non-cancer cases of Medicare beneficiaries.

The next national registry is the National Program of Cancer Registries (NPCR) which is a CDC sponsored initiative that expands the SEER program to ensure coverage of all 50 states in the United Sates. This program was established in 1992 as a complement to the SEER program and collects similar information to the SEER registry. Currently the NPCR program supports central cancer registries in 45 states, the District of Columbia and Puerto Rico.

In addition to the national registries noted above, there are state-based central cancer registries. These registries are maintained at the state level and serve as another potential alternative data source for identification of patients with cancer. The North American Association of Central Cancer Registries (NAACCR) is an organization that oversees many of the state-based central cancer registries.

The National Cancer Database is a database that was established through a partnership with the American College of Surgeons and the American Cancer Society to track cancer-related outcomes in the US. However, this registry only contains information at the institution level and would not be linkable to the Mini-Sentinel Data Partners.

The Cancer Experience Registry is a self-reported cancer registry. The Cancer Genetics Network Core database is a national registry that includes individuals with a personal or family history of breast, prostate, colorectal cancer or melanoma. This is a registry used predominantly to study genetic issues surrounding cancer and would not be a viable alternative data source for the purposes of validation of an algorithm.

The VA Central Cancer Registry and the Automated Central Tumor Registry (ACTUR) are general cancer registries that include persons diagnosed with cancer in the VA health care system and the armed services, respectively. Finally, the Transplant Cancer Match is a database that focused on patients with organ transplants that have developed cancer.

The Cancer Research Network (CRN) Virtual Data Warehouse (VDW) is a distributed data network that maintains a database of patients diagnosed with cancer from institutions participating in the cancer research network. The VDW includes detailed information on tumor characteristics that are captured by local cancer registrars, similar to information captured in other cancer registries. This information is linked with the healthcare claims data of the individual. The CRN VDW is a subset of institutions that participate in the Health Maintenance Organization Research Network and therefore are contributors to the MSDD. The participating institutions in the CRN VDW include Group Health, Henry Ford, Kaiser Permanente, Marshfield Clinic, Fallon, and HealthPartners. The individuals in this dataset would clearly overlap with those in the MSDD. The CRN VDW is fed by institution specific cancer registries. For example, Kaiser Permanente Colorado, Kaiser Permanente Northwest and HealthPartners all indicated that they have a local cancer registry. It appears that this is the information that is also available in the VDW as part of the CRN VDW. The CRN VDW captures approximately 38,000 new cases of cancer annually.

Other Issues: None.

WG Recommendation: Feasible. Cancer is an HOI that is feasible to validate with several viable options. The alternative data source that might be most feasible for cancers with a relatively high incidence rate is the CRN VDW since many of these participants are already part of the MSDD. For cancers that are rare that will require a larger registry to ensure sufficient numbers of individuals with the cancer, SEER, NPCR and SEER-Medicare being the most viable options.

HOI: Cancer - Anus

Previous Validation Studies: The WG conducted a systematic search and found no previous validation studies for this HOI.

Summary of Search of Linkable Databases: No anal cancer databases or registries were found. Because anal cancer occurs more frequently in HIV-positive individuals, especially men who have sex with men, a more directed search was done. Two databases were found: the HIV/AIDS Match Study and Collection and Verification of Data for Matched Records from US Cancer and HIV/AIDS Registries, which is a one-time merge of several registries. Both databases, although possibly nationwide, are unfortunately specific to HIV/AIDS patients and anonymous because they were compilations of other registries and data scrubbed, making both unable to be directly linked to other databases and poorly generalizable.

Other Issues: Anal cancer is rare, occurring in about 1 in 100,000 patients in the general population. Additionally, the various types of lower GI cancers may be difficult to distinguish, e.g., colon, rectal, and anal cancers.

WG Recommendation: Feasible. Anal cancer is an HOI that is feasible to validate with several viable options. The alternative data source that might be most feasible is the CRN VDW since many of these participants are already part of the MSDD. If there is an insufficient number of cases in the CRN VDW, then SEER, NPCR and SEER-Medicare are alternatives that would be feasible for use in validation (see general cancer summary and detailed report).

HOI: Cancer - Bladder

Previous Validation Studies: The WG conducted a systematic search and found no previous validation studies for this HOI.

Summary of Search of Linkable Databases: Several databases and registries related to bladder cancer were found. The bladder cancer specific alternative databases included Familial and Atypical Urothelial Cancer Registry, A Prospective Registry to Assess the Effectiveness and Local Tolerability of Intravesical Valrubicin in Subjects with Non-muscle Invasive Bladder Cancer (NMIBC), Study of Outcomes after Surgery/Treatment to Treat Bladder Cancer (PORCH), and The Drake Health Registry Study.

The Familial and Atypical Urothelial Cancer Registry is an institutional registry. The NMIBC is a treatment specific registry. The PORCH and the Drake Health Registry Study are institutional registries. None of these registries are viable alternative data sources for purposes of algorithm development in the MSDD cohort.

Other Issues: None.

WG Recommendation: Feasible. Bladder cancer is an HOI that is feasible to validate with several viable options. The alternative data source that might be most feasible is the CRN VDW since many of these participants are already part of the MSDD. If there is an insufficient number of cases in the CRN VDW, then SEER, NPCR and SEER-Medicare are alternatives that would be feasible for use in validation (see general cancer summary and detailed report).

HOI: Cancer - Brain and central nerve system cancer

Previous Validation Studies The WG conducted a systematic search and two previous studies were identified, one of which was from Scotland and not based on administrative data. The other, published in 2009 by Eichler and others,[62] examined the accuracy of ICD-9 codes in administrative claims data for identifying patients with "secondary neoplasm of the brain and spinal cord." Administrative data came from Partners Healthcare System (MA) and validation was conducted using medical records. The PPV ranged from 91 to 93%. However, since this focused on secondary/metastatic cancer rather than primary brain cancers it is not sufficient to consider this HOI to be previously well validated.

Summary of Search of Linkable Databases: Only one brain cancer specific database could be found. That was the Central Brain Tumor Registry of the United States (CBTRUS). CBTRUS compiles data from state registries and it includes both malignant and non-malignant primary brain tumors. The data are used for describing population-based incidence and survival patterns of brain tumor cases, to evaluate diagnosis and treatment, and to conduct etiologic studies.

Other Issues: Brain cancer is not a single type. It includes all tumors inside the cranium or in the central spinal canal. Brain cancers involve the brain itself, but also the lymphatic tissue, blood vessels, the cranial nerves, the meninges, skull, and pituitary gland, or pineal gland. Within the brain itself, the involved cells may be neurons or glial cells (which include astrocytes, oligodendrocytes, and spendymal cells). Tumors, benign or malignant, can occur in different parts of the brain, and may or may not be primary tumors. A primary tumor is one that has started in the brain, as opposed to a metastatic tumor, which is something that has spread to the brain from another part of the body. Metastatic tumors are more common than primary tumors by 4:1. While the most important risk factors for brain cancer are age and family history/genetics, radiation/radiotherapy and previous chemotherapy have been associated, along with various other exposures. Diagnosis of the type of brain/CNS cancer is complicated and ICD-9 codes are numerous. ICD-9 code 191 is malignant neoplasm of the brain and it includes 191.0 through 191.9 depending on the location in the brain. ICD-9 code 192 is malignant neoplasm of other and unspecified parts of the nervous system, including cranial nerve (192.0), cerebral meninges (192.1), etc. There are also codes for benign neoplasm of the brain and other parts of the nervous system (225). ICD-9 code 198.3 refers to secondary malignant neoplasm of the brain and spinal cord.

WG Recommendation: Feasible. Brain and CNS cancer is an HOI that is feasible to validate with several viable options for alternative databases. The alternative data source that might be most practical is the CRN VDW since many of these participants are already part of the MSDD. Brain cancer is rare so if there is an insufficient number of cases in the CRN VDW, then SEER, NPCR and SEER-Medicare would be feasible for use in validation (see general cancer summary and detailed report). Last the CBTRUS is also an option.

HOI: Cancer - Breast

Previous Validation Studies: The WG conducted a systematic search and found that there were three previous validation studies that all used a version of the SEER-Medicare data to validate algorithms for breast cancer. The algorithm with the best test characteristics was an algorithm developed by Nattinger and colleagues[26] that had sensitivity >80%, specificity >99% and positive predictive values of 88% and higher. Freeman and colleagues[24] evaluated four separate algorithms for the identification of incident breast cancer cases. In addition, there was a subsequent evaluation of the performance of this algorithm and other algorithms for the identification of breast cancer in a newer SEER-Medicare cohort.[25] In this study, the Nattinger algorithm had a PPV of 82.6% while the algorithm developed by Freeman and colleagues had a PPV of 93.2%.

Summary of Search of Linkable Databases: There were several breast cancer specific alternative data sources that were identified. Of these, some were limited in the stage of disease included in the registry, for example the ASCO Breast Cancer Registry pilot program included only patients in Stage 1 to 3. The Breast Cancer Surveillance Consortium has a registry of women that underwent screening mammography and some of which had breast cancer. The women identified with breast cancer have been linked to the SEER data where available. The limitation of this registry is that it only contains cases of breast cancer where the woman underwent a screening mammography and therefore does not include the women that were diagnosed through another pathway. The Breast and Prostate Cancer Data Quality and Patterns of Care Database is an existing cohort of patients with breast and prostate cancer that have their healthcare claims and tumor information linked from seven states. However, it only contains data from patients that were diagnosed in 2004. Several other registries are from single sites or include patients that self-identify as having breast cancer.

Other Issues: None.

WG Recommendation: Well-validated. This is a cancer type has an algorithm that was previously well-validated (using SEER-Medicare data).

HOI: <u>Cancer - Cervix and uteri</u>

Previous Validation Studies: The WG conducted a systematic search and found no previous validation studies for this HOI.

Summary of Search of Linkable Databases: While many other countries have registries specific to cervical cancer none were identified in the US.

Other Issues: The cervix uteri is the lower part of the uterus extending from the isthmus of the uterus into the vagina; neck of uterus; neck of womb. Cervical cancer and uterine cancer are two different types of cancer. ICD-9 code 179 is "malignant neoplasm of the uterus, part unspecified", while 180 is "malignant neoplasm of cervix uteri".

WG Recommendation: Feasible. Cervical cancer is an HOI that is feasible to validate with several viable options. The alternative data source that might be most feasible is the CRN VDW since many of these participants are already part of the MSDD. If there is an insufficient number of cases in the CRN VDW, then SEER, NPCR and SEER-Medicare are alternatives that would be feasible for use in validation (see general cancer summary and detailed report).

HOI: Cancer - Colon and rectum

Previous Validation Studies: The WG conducted a systematic search and was able to identify two validation studies for identification of cancer among patients with colon cancer in the US.[63,64] However, both of these studies focused on the identification and validation of algorithms for metastatic disease among patients with colon cancer. Therefore, these algorithms are not relevant for the identification of new cases of colon or colorectal cancer.

Summary of Search of Linkable Databases: Several registries specific to colon or colorectal cancer were found. Many of these registries focus on the familial or hereditary component of colorectal cancer and will include both those with and without colorectal cancer. There were several registries that were essentially a collaboration of multiple single institution registries. The Colon Cancer Family Registry included patients and families of those with colon cancer from six sites in North America and multiple international sites. The other registry that includes several single institution registries is the Collaborative Group of the Americas on Inherited Colorectal Cancer. This registry began in 1995 and is a collaboration of many cancer sites from across the US. In addition, there are several single institution colorectal cancer registries.

Other Issues: Colon and rectal cancer are often combined as colorectal cancer when a tumor occurs in either site. Some of the registries identified were specific to colon cancer but for the most part the registries included colon and rectal cancers.

WG Recommendation: Feasible. Colon cancer is an HOI that is feasible to validate with several viable options. The alternative data source that might be most feasible is the CRN VDW since many of these participants are already part of the MSDD. If there is an insufficient number of cases in the CRN VDW, then SEER, NPCR and SEER-Medicare are alternatives that would be feasible for use in validation (see general cancer summary and detailed report).

HOI: <u>Cancer - Esophagus</u>

Previous Validation Studies: The WG conducted a systematic search and found no previous validation studies for this HOI.

Summary of Search of Linkable Databases: Two national clinical databases contain information about patients undergoing esophageal cancer resection were identified: The Society of Thoracic Surgeons General Thoracic Surgery Database (STS GTDB) and The American College of Surgeons National Surgical Quality Initiative Program (ACS NSQIP). Clinical data and outcome of esophageal cancer resections were collected and confirmed by physicians. The participants with esophageal cancer resection were 6,740 and 1,030, respectively. A registry specific to esophageal cancer was also found. The Mayo Clinic Esophageal Adenocarcinoma and Barrett's Esophagus (EABE) Registry enrolled patients with long segment Barrett's esophagus, squamous cell cancer of the esophagus, and esophageal adenocarcinoma. However, only around 600 patients were identified in this registry. The overlap with MSDD may not be sufficient for validation purpose.

Other Issues: None.

WG Recommendation: Feasible. Esophageal cancer is an HOI that is feasible to validate with several viable options. The alternative data source that might be most feasible is the CRN VDW since many of these participants are already part of the MSDD. If there is an insufficient number of cases in the CRN VDW, then SEER, NPCR and SEER-Medicare are alternatives that would be feasible for use in validation (see general cancer summary and detailed report).

HOI: Cancer - Kidney and renal pelvic

Previous Validation Studies: The WG conducted a systematic search and found no previous validation studies for this HOI.

Summary of Search of Linkable Databases: Several databases and registries related to kidney and renal cancer were found. The kidney cancer specific data sources included the Mayo Clinic Registry, Tracking Renal Tumors After Cryoablation Evaluation (TRACE), the Kidney Cancer Association Registry, Proleukin Observational (PROCLAIM) Registry, Familial and Atypical Urothelial Cancer Registry.

The Mayo Clinic Registry, and Familial and Atypical Urothelial Cancer Registry are institutional registries. The TRACE and PROCLAIM are treatment specific registries and the Kidney Cancer Association Registry is a registry which contained only self-report cases of kidney cancer. Thus, they are not optimal choices for an alternative database for future validation studies.

Other Issues: None.

WG Recommendation: Feasible. Kidney/renal cancer is an HOI that is feasible to validate with several viable options. The alternative data source that might be most feasible is the CRN VDW since many of these participants are already part of the MSDD. If there is an insufficient number of cases in the CRN VDW, then SEER, NPCR and SEER-Medicare are alternatives that would be feasible for use in validation (see general cancer summary and detailed report).

HOI: Cancer - Leukemia

Previous Validation Studies: The WG conducted a systematic search and two previous studies were identified, and both were in pediatric cancers. One study was completed in acute myeloid leukemia (AML)[65] while the other was conducted in the setting of acute lymphoblastic leukemia (ALL).[66] The AML study was conducted using ICD-9 codes and validated using chart review in a subset of patients; however ICD-9 codes alone resulted in a poor PPV of 31%, and after inclusion of chemotherapy review, the PPV increased to 100%.[65] The ALL study used a 3 step process involving inclusion using ICD-9 codes, exclusion of certain patients who did not have newly diagnosed ALL, and verification using billing of ALL induction therapy. The PPV 95% confidence interval ranged from 90% to 96%.[66] Although both studies show good positive predictive values, since both studies focused on pediatric leukemia, it is difficult to extrapolate the algorithms to leukemia as a whole.

Summary of Search of Linkable Databases: Several potentially linkable registries were found. Three of the registries cover familial leukemia and methods of genetic transmission between generations. These are: The International Familial Childhood Leukemia Registry (IFCLR), Familial Chronic Lymphocytic Leukemia Study (FCLL), and The Genetic Factors in Familial Hematologic Malignancies (GFFHM). The IFCLR enrolls patients with any kind of childhood leukemia or lymphoblastic lymphoma. The FCLL covers only chronic lymphocytic leukemia, whereas the GFFHM covers all hematological malignancies. Several registries were also patient group specific in terms of patient age, cancer type and/or location. The Pediatric Chronic Myeloid Leukemia Registry (PCMLR) is especially limiting due to the age and leukemia type restrictions. Other age specific but not cancer specific registries include Enrollment on the Childhood Cancer Research Network (CCRN) of the Children's Oncology Group (COG) and Carolina Senior: UNC Registry for Older Cancer Patients. Cancer specific registries include Connect™ CLL: The Chronic Lymphocytic Leukemia (CLL) Disease Registry and the Connect® MDS/AML: The Myelodysplastic Syndromes (MDS) and Acute Myeloid Leukemia (AML) Disease Registry. SCRI Tissue Testing Registry is not specific to one type of cancer, but it is localized to Nashville, Tennessee. All aforementioned registries are unlikely to be generalizable to all leukemia patients due to the imposed limitations on the patients enrolled.

Other Issues: There are many types of leukemia; they include cancer such as ALL, AML, chronic myeloid leukemia, chronic lymphocytic leukemia, and others. Additionally, hematologic malignancies as a whole also includes lymphoma. Thus, one database covering all leukemias without including lymphomas was not found. Additionally, leukemia ICD-9 codes are numerous. For example, ICD-9 code 204.xx is for ALL, although it may be relapsed or newly diagnosed. Thus, for the many other types of leukemia, there must be a multitude of ICD-9 codes. Additionally, the medication regimens for leukemia are complicated and vary greatly between types; thus, it may be more difficult to validate the diagnosis of leukemia using chart reviewer.

WG Recommendation: Feasible. Leukemia is an HOI that is feasible to validate with several viable options. The alternative data source that might be most feasible is the CRN VDW since many of these participants are already part of the MSDD. If there is an insufficient number of cases in the CRN VDW, then SEER, NPCR and SEER-Medicare are alternatives that would be feasible for use in validation (see general cancer summary and detailed report).

HOI: <u>Cancer - Liver</u>

Previous Validation Studies: The WG conducted a systematic search and found no previous validation studies specific to liver cancer.

Summary of Search of Linkable Databases: Only two registries specific to liver cancer were found. They are Oregon Liver Tumor Registry and A National Registry of Patients with Hepatocellular Carcinoma. Both two registries are institutional-level registries. Thus, they may not be alternative databases for validation studies.

However, because liver cancer can lead liver transplantation, the national or international registries of patients with transplantation may be useful. These include the International Registry of Hepatic Tumors in Liver Transplantation, and the United Network for Organ Sharing (UNOS) database. Similarly, it may be possible to identify liver cancer cases in databases of patients with hepatic virus infection, such as the VA Hepatitis C Clinical Case Registry, because hepatic virus infection is strongly related to liver cancer.

Other Issues: None.

WG Recommendation: Feasible. Liver cancer is an HOI that is feasible to validate with several viable options. The alternative data source that might be most feasible is the CRN VDW since many of these participants are already part of the MSDD. If there is an insufficient number of cases in the CRN VDW, then SEER, NPCR and SEER-Medicare are alternatives that would be feasible for use in validation (see general cancer summary and detailed report). Last, the databases listed above might also be used.

HOI: Cancer - Lung

Previous Validation Studies: The WG conducted a systematic search and found no previous validation studies for this HOI.

Summary of Search of Linkable Databases: The WG identified the Stacey Scott Lung Cancer Registry, which is specific to patients lung cancer that have undergone autoflorescence bronchoscopy and spiral CT scans. The registry includes biologic samples and patient data that are banked from contributors around the world. It is unclear if the patients would overlap with Mini-Sentinel Data Partners or if linkage would be possible.

Other Issues: None.

WG Recommendation: Feasible. Lung cancer is an HOI that is feasible to validate with several viable options. The alternative data source that might be most feasible is the CRN VDW since many of these participants are already part of the MSDD. If there is an insufficient number of cases in the CRN VDW, then SEER, NPCR and SEER-Medicare are alternatives that would be feasible for use in validation (see general cancer summary and detailed report).

HOI: Cancer - Lymphoma

Previous Validation Studies: The Mini-Sentinel Protocol Core has published a systematic review for identifying this HOI using administrative or claims data.[8] Several algorithms that had been used to identify patients with lymphoma in claims data were reported. However, they were able to identify only one paper that validated algorithms for the identification of patients with lymphoma. Setoguchi and colleagues[23] used data from the Pennsylvania State tumor registry to validate lymphoma identified in claims data. They evaluated four potential algorithms for the identification of patients with lymphoma. The test characteristics of the four algorithms were: Algorithm 1 – sensitivity 55.2%, specificity 99.9%, positive predictive value 61.5%; Algorithm 2 – sensitivity 79.8%, specificity 99.8%, positive predictive value 62.8%; Algorithm 3 – sensitivity 88.3%, specificity 99.7%, positive predictive value 56.6%; and Algorithm 4 – sensitivity 88.7%, specificity 99.3%, positive predictive value 34.7%.[23] Based on the PPVs none of the algorithms would be considered validated for the purposes of this project.

Summary of Search of Linkable Databases: As with most HOIs, several international registries exist that include patients with lymphoma. In addition, there were several registries specific to lymphoma that were found which included sites in the US. Many of the registries that were identified focused on specific types of lymphoma and no one specific registry included a general group of patients with lymphoma. Of the registries that were found for patients with lymphoma, many focus on the familial or hereditary component of lymphoma and will include both those with and without colorectal cancer. There was also a registry that aimed to collect treatment data on those with Peripheral T-cell lymphoma.

Other Issues: The previous Mini-Sentinel Protocol Core review[8] provides a good summary of the complexities associated with the identification of patients with lymphoma in claims data and the difficulties in validating an algorithm. First, there are multiple types of lymphoma and several classification systems that exist for categorizing the types of lymphoma. As noted in the Mini-Sentinel Protocol Core review, these multiple classifications do not align particularly well with existing ICD-9 codes and that is a potential explanation for the inability to develop a strong algorithm. For example, lymphomas can be separated based on etiology into T-cell and B-cell lymphomas or they can be separated based on expected outcomes (e.g., curable and non-curable) or they can be separated by the relative aggressiveness of the cancer. In addition, there is a historical categorization of Hodgkin's and non-Hodgkin's lymphoma. Importantly, the use of these various classification systems appears to have changed over time and therefore it is important to consider the time period of an alternative data source and how that might influence the "types" of lymphoma included in the data source.

WG Recommendation: Feasible. Lymphoma is an HOI that is feasible to validate with several viable options. The alternative data source that might be most feasible is the CRN VDW since many of these participants are already part of the MSDD. If there is an insufficient number of cases in the CRN VDW, then SEER, NPCR and SEER-Medicare are alternatives that would be feasible for use in validation (see general cancer summary and detailed report).

HOI: Cancer - Melanoma

Previous Validation Studies: The WG conducted a systematic search and found no previous validation studies for this HOI.

Summary of Search of Linkable Databases: Two melanoma-specific registries were identified, however, both are single center databases so are of limited use. The Duke Melanoma and Tumor Registry is one of the largest melanoma registries in the US beginning three decades ago. However, it only contains patients diagnosed with melanoma who were treated or diagnosed at Duke. Another database is the Melanoma Research Registry at the Huntsman Cancer Institute at the University of Utah. Enrollment is voluntary and also limited to a single center.

Other Issues: None.

WG Recommendation: Feasible. Melanoma is an HOI that is feasible to validate with several viable options. The alternative data source that might be most feasible is the CRN VDW since many of these participants are already part of the MSDD. If there is an insufficient number of cases in the CRN VDW, then SEER, NPCR and SEER-Medicare are alternatives that would be feasible for use in validation (see general cancer summary and detailed report).

HOI: <u>Cancer - Myeloma</u>

Previous Validation Studies: The WG conducted a systematic search and found no previous validation studies for this HOI.

Summary of Search of Linkable Databases: The WG could find only two active registries that are specific to multiple myeloma or related plasma cell dyscrasia. The Ohio State Multiple Myeloma and Amyloidosis Data Registry is restricted to Ohio and therefore not likely useful. The Celgene Corporation maintains the "Connect MM – Multiple Myeloma Disease Registry". This is a national registry and hopes to enroll 1500 patients by 2017. Limited information is available on-line but it appears to have utility for use for validation studies within MS.

Other Issues: Myeloma, also known as multiple myeloma, or plasma cell myeloma, is a cancer of plasma cells, a type of white blood cell. In multiple myeloma, collections of abnormal plasma cells accumulate in the bone marrow where they interfere with the production of normal blood cells. Myeloma is diagnosed with serum protein electrophoresis, bone marrow examination, urine protein electrophoresis, and X-rays of commonly involved bones. While genetic differences are most commonly associated with myeloma, exposure to certain viruses and toxic substances, including radiation, have also been associated with development of the cancer. Myeloma has a single ICD-9 code 230 - multiple myeloma and immunoproliferative neoplasm.

WG Recommendation: Feasible. Multiple myeloma is an HOI that is feasible to validate with several viable options. The alternative data source that might be most feasible is the CRN VDW since many of these participants are already part of the MSDD. If there is an insufficient number of cases in the CRN VDW, then SEER, NPCR and SEER-Medicare are alternatives that would be feasible for use in validation (see general cancer summary and detailed report). Last, the Connect MM registry may be another alternative.

HOI: <u>Cancer - Pancreas</u>

Previous Validation Studies: The WG conducted a systematic search and only one previous validation study was identified. The study was conducted in Indiana University School of Medicine. The author found that PPV of an ICD-9 code as 157.xx for pancreatic cancer was 38%.[67] Since there is only one validation study conducted in a single medical center with low PPV, this HOI is not considered to be previously well validated.

Summary of Search of Linkable Databases: Several registries specific to pancreatic cancer could be found. Most of them are familial pancreatic cancer registry. They included pancreatic cancer patients and families which have relatives experienced in pancreatic cancer. Among those, there is a network that linked 7 registries together called the Pancreatic Cancer Genetic Epidemiology (PACGENE) consortium. It included Karmanos Cancer Institute and Wayne State University, the Mayo Clinic, University of Texas/M. D. Anderson Cancer Center, Johns Hopkins University, Creighton University, University of Toronto/Mount Sinai Hospital and the Dana-Farber Cancer Institute. Funding for the PACGENE Consortium has been received through a grant from the National Cancer Institute. The consortium supposes to be the largest registry in the US. The rest of registries the WG found are institutional based pancreatic cancer registries that may not be appropriate to be alternative databases for validation study.

Other Issues: None.

WG Recommendation: Feasible. Pancreatic cancer is an HOI that is feasible to validate with several viable options. The alternative data source that might be most feasible is the CRN VDW since many of these participants are already part of the MSDD. If there is an insufficient number of cases in the CRN VDW, then SEER, NPCR and SEER-Medicare are alternatives that would be feasible for use in validation (see general cancer summary and detailed report).

HOI: <u>Cancer - Prostate</u>

Previous Validation Studies: The WG conducted a systematic search and found no previous validation studies for this HOI.

Summary of Search of Linkable Databases: Several databases and registries related to prostate cancer were found. The prostate cancer specific alternative databases included the COMPARE registry which is a registry of patients that have elevated PSA levels following treatment of a primary prostate adenocarcinoma. The Breast and Prostate Cancer Data Quality and Patterns of Care Database is an existing cohort of patients with breast and prostate cancer that have their healthcare claims and tumor information linked from seven states. However, it only contains data from patients that were diagnosed in 2004. The CaPSURE database is an observational database that includes prostate cancer patients from 40 urologic practices from around the US.

In addition, there are a number of treatment specific registries for patients with prostate cancer that track patients following receipt of the intervention. All of these are exposure specific registries which may have limitations in the generalizability of the patients included in the registry.

Other Issues: None.

WG Recommendation: Feasible. Prostate cancer is an HOI that is feasible to validate with several viable options. The alternative data source that might be most feasible is the CRN VDW since many of these participants are already part of the MSDD. If there is an insufficient number of cases in the CRN VDW, then SEER, NPCR and SEER-Medicare are alternatives that would be feasible for use in validation (see general cancer summary and detailed report).

HOI: Cancer - Thyroid

Previous Validation Studies: The WG conducted a systematic search and found no previous validation studies for this HOI.

Summary of Search of Linkable Databases: The WG was able to find one alternative data source specific to thyroid cancer. The Thyroid Tumor and Cancer Registry began in 2008 and includes patients with a diagnosis of thyroid cancer or thyroid nodules. It includes approximately 2000 patients; however it appears to be a single institution registry.

Other Issues: None.

WG Recommendation: Feasible. Thyroid cancer is an HOI that is feasible to validate with several viable options. The alternative data source that might be most feasible is the CRN VDW since many of these participants are already part of the MSDD. If there is an insufficient number of cases in the CRN VDW, then SEER, NPCR and SEER-Medicare are alternatives that would be feasible for use in validation (see general cancer summary and detailed report).

HOI: Chronic renal failure/Chronic kidney disease

Previous Validation Studies: The OMOP has reported a systematic review for identifying renal failure using administrative or claims data and several studies with high PPV for chronic renal failure/disease (CKD) were identified.[58] For example, Winkelmayer et al.[33] conducted a validation study of CKD in 2005 using Medicare data and found high PPV (97.5%) for the best algorithm. A review of kidney disease validation studies was published by Grams et al.[68] in 2011. For CKD sensitivity and specificity were as high as 82 and 100% depending on the study/algorithm. For CKD the gold standard used in most validation studies is eGFR <60 on two SCRs separated by 3-6 months. Note that for ESRD gold standard might also include dialysis.

Summary of Search of Linkable Databases: Several good alternate databases are available for external validation of a CKD algorithm in MSDD. The best is likely the United States Renal Data System (USRDS). The USRDS is a well-established, national database that collects, analyzes, and distributes information on the US end-stage renal disease (ESRD) population, including treatments and outcomes. It is likable and likely has significant overlap with MSDD. Another potential external database is that available from DaVita Healthcare Partners, Inc., one of the largest kidney care companies in the US. This resource is significant as it represents patients in all stages of CKD, not just end-stage renal disease. Lab data from DaVita would clearly be available to identify cases based on glomerular filtration rate (GFR) since GFR < 60 mL/min/1.73m^2 for >3 months is the commonly used gold standard for CKD (eGFR can be calculated from serum creatinine). Another database similar to DaVita is that available from Fresenius Medical Care - another provider of dialysis services for patients with ESRD that makes their electronic data available for research purposes. It may also be possible to validate an algorithm for CKD using data from one or more of the Mini-Sentinel Data Partners. Geisenger, Humana, and Kaiser Permanente all appear to have databases that include patient laboratory data (including SCR).

Other Issues: There are five stages of Chronic Renal Disease – stage 5 is end-stage renal disease (ESRD) and requires dialysis. Stage 3-5 is based on the level of GFR.

WG Recommendation: Well-validated. However, this HOI could be also validated using an alternative database – there are several viable options for databases, including Data partners.

HOI: <u>Cirrhosis</u>

Previous Validation Studies: The WG conducted a systematic search and found two recent validation studies published by Goldberg et al.[69,70] One that focused on end-stage liver disease (ESLD)[69] and the other focused on hepatocellular carcinoma (HCC).[70] The algorithm for ESLD was not focused on all patients with cirrhosis, but rather those with cirrhosis and an event indicative of decompensated liver disease. Thus, the validation study included a subset of all patients with cirrhosis. The study was conducted in two hospitals in the University of Pennsylvania Health System. The algorithms developed had high positive predictive values (PPV) for cirrhosis, with all but the first algorithm having PPVs > 90%.[69] The second validation study Goldberg et al. conducted focused on HCC and not on patients with cirrhosis.[70] In addition, there has been a validation study of cirrhosis and chronic liver disease using Veterans Affairs (VA) data[71], but this study may not be generalizable to the MSDD because of the differences in administrative data between VA and MSDD (encounter data vs. billing data).

Summary of Search of Linkable Databases: A few small registries maintained by academic medical centers were identified, but were too small to be considered a viable source of cases for a validation study. Similarly, there is an alpha-1 antitrypsin deficiency registry that includes a number of patients that developed cirrhosis due to their alpha-1 antitrypsin deficiency but this is cirrhosis due to a very specific cause and is also too small in sample size. The VA maintains a Hepatitis C case registry. Hepatitis C is a leading cause of cirrhosis. However, this registry is not a viable alternative because it likely does not overlap of patients with the Mini-Sentinel Data Partners. The final alternative data source identified was the Organ Procurement and Transplantation Network (OPTN) registry.
This registry contains information on patients that are both pre-transplant and post-transplant. As of Jan 18, 2013 there were 16,516 patients on the liver transplant waiting list. Several of these were on the transplant waiting list as a result of cirrhosis. Thus, it would be possible to identify patients that were cases of cirrhosis that lead to liver failure and necessitated a transplant. There is likely to be overlap with this data and Mini-Sentinel Data Partners.

Other Issues: The biggest limitation with the OPTN data is that it would only be useful in identifying the most severe cases of cirrhosis that had advanced to a stage where liver transplant was necessary. It is not clear that this would be optimal for identifying cirrhosis in an active surveillance system as it may be important to identify earlier or milder cases of cirrhosis.

WG Recommendation: Potentially feasible. It is possible this HOI could be validated with an alternative data source if cases of cirrhosis that lead to the need for a liver transplant are considered to be the types of events of interest for surveillance. The OPTN registry is recommended as the alternative database.

HOI: <u>Colitis ischemic</u>

Previous Validation Studies: The WG conducted a systematic search and found only one validation study for this HOI. The study by Sands et al.[72] used an ICD-9 algorithm for ischemic colitis from administrative claims from a large insurer (United Health Care), and validated using patient charts. The authors reported a PPV of 75% (n=57). However, it is a single small study which may not be sufficient to determine that this HOI is already well-validated.

Summary of Search of Linkable Databases: Most information on ischemic colitis seems to be in reference to its occurrence as a complication of aortic surgery (aneurysm repair), and there are some large studies on this and even a registry in Sweden. Antipsychotics-induced ischemic colitis was studied using a French pharmacovigulance database, but no registries or databases specific to this HOI were identified in the US. Lotronex (alosetron), a drug for irritable bowel syndrome was removed from the market because of cases of ischemic colitis. Later it was allowed to return to the market and all patients were required to be enrolled in a Risk Evaluation and Mitigation Strategy (REMS) program which includes monitoring for cases of ischemic colitis. Limited information is available publically about the program and even if available it would be very exposure specific.

Other Issues: Ischemic colitis (or colonic ischemia, or colitis ischemic) occurs when blood flow to part of the large intestine (colon) is reduced due to narrowed or blocked blood vessels. The ischemia can result in pain, damage to tissues, bowel perforation, and may require surgery. Diagnostic studies must include endoscopic, radiographic, surgical, and/or histopathologic findings. It is sometimes confused with Crohn's disease, ulcerative colitis, diverticulitis, or colorectal cancer. There is one ICD-9 code (557) called "Colonic ischemia" described as "vascular insufficiency of intestine".

WG Recommendation: Not Feasible.

HOI: Deafness

Previous Validation Studies: The WG conducted a systematic search and found no previous validation studies for this HOI.

Summary of Search of Linkable Databases: Although there are many associations related to interpreters, benefits, living aids for deaf, few of them maintain databases or registries. The national level alternative source might be the disability insurance program administered by the Social Security Administration, which is also the most promising source if linkage can be made. State level registries for deafness were found in some states under the State Early Hearing Detection and Intervention program; however, these registries were developed mainly for new-born infants for screening purposes, where deafness is a birth defect.

Other Issues: The diagnosis criteria for deafness is as follows: an average air conduction hearing threshold of 90 decibels or greater in the better ear and an average bone conduction hearing threshold of 60 decibels or greater in the better ear or a word recognition score of 40 percent or less in the better ear determined using a standardized list of phonetically balanced monosyllabic words.

WG Recommendation: Potentially feasible. Not many alternative data sources can be used for this HOI. Information maintained by Social Security Administration might be the promising source for validation if linkage if possible.

HOI: Depression

Previous Validation Studies: The Mini-Sentinel Protocol Core has published a systematic review for identifying this HOI using administrative or claims data.[21] A range of PPVs are reported that in some cases exceed a PPV 90%; however, these are all caveated with various methodological nuances and discussed in detail in that report. Specifically, the authors noted that the quality of clinical depression case recognition is inconsistent and constrains the performance of electronic health information to identify depressive disorders. From the report, the highest agreement with clinically diagnosed depression was achieved by an algorithm that required over a 12 month period at least 2 listings of ICD-9 codes for 296.2 (major depressive episode, single episode), 296.3 (major depressive episode, recurrent episode), 300.4 (dysthymic disorder), or 311 (depression not elsewhere classified) along with a filled prescription for an antidepressant medication. The PPV was 49.1% and the chance corrected agreement of this algorithm was moderate (Kappa: 0.464). The authors concluded that algorithms based on ICD-9 codes for depression are unlikely to achieve acceptable sensitivity in identifying depression, and that the inclusion of prescription claims for antidepressant medications may or may not improve the PPV of algorithms because of the wide range of use of antidepressants.

Summary of Search of Linkable Databases: The National Network of Depression Center (NNDC) is a network comprised of about 16 of the top teaching medical centers whose efforts are combined to create a comprehensive research registry to facilitate studies to develop early intervention or prevention strategies for depression and bipolar disorders. Participants are asked to fill out questionnaires related to mood symptoms, overall sense of well-being, impact of symptoms on work and social activities, and medication side effects. The National Psychosis Registry is an ongoing registry of all veterans diagnosed with psychosis, which includes schizophrenia other than latent, schizoaffective disorder, bipolar disorder, and other non-organic psychoses. In addition, there is a National Registry for Depression (NARDEP) maintained by the VA.

Other Issues: A diagnosis of depression is based on subjective data that may vary based on clinician's interpretation of DSM-IV diagnostic criteria, and there are subtypes of depression that can overlap with other conditions, such as mania. See the summary on bipolar/mania for further discussion. This and other mental health disorders may have greater issues of linkability due to restrictions on patient disclosure/confidentiality.

WG Recommendation: Unlikely. While it may be possible to link to alternative data sources because several do registries exist, issue of patient confidentiality may preclude this. Further, it appears that identification of an algorithm with acceptable performance (PPV>0.70) would be very challenging. It is recommended that the publication by Townsend et al.[21] should be reviewed prior to proceeding further in attempts to validate another algorithm.

HOI: <u>Dyslipidemia</u>

Previous Validation Studies: The WG conducted a systematic search and found no previous validation studies for this HOI.

Summary of Search of Linkable Databases: Diagnosis of dyslipidemia, or high cholesterol levels, is objectively based on the lipid panel as defined by the ATPIII guidelines. In the absence of lipid levels in the MSCDM, validation could occur by linking to these lab values within the Mini-Sentinel Data Partners' databases if such lab values are available. Alternatively there are some commercially available databases that include lab values (ie, Optum, Medstat) and could be used for external validation. Also, The National Cardiovascular Disease Registry includes data on patients with cardiovascular disease (including lipid levels). Patients are enrolled by physicians by sharing electronic health records directly to the registry.

Other Issues: None.

WG Recommendation: Feasible – lab based. This is an HOI that would seem best validated using laboratory values within the Mini-Sentinel Data Partners if the lab values exist in their databases.

HOI: Endotoxic shock

Previous Validation Studies: The WG conducted a systematic search and found no previous validation studies for this HOI.

Summary of Search of Linkable Databases: The WG was unable to identify any registries or databases that could be used for alternative validation.

Other Issues: Endotoxic shock is a subtype of septic shock. Although, there is no registries specific to the endotoxic shock but there are several registries related to sepsis/septic shock, which was reviewed as a separate HOI.

WG Recommendation: Not feasible.

HOI: Erthema multiforme/Stevens-Johnson Syndrome

Previous Validation Studies: The Mini-Sentinel Protocol Core has published a systematic review for identifying erthema multiforme (EM)/Stevens-Johnson Syndrome (SJS) or toxic epidermal necrolysis (TEN) using administrative or claims data.[15] They found a limited literature that focused on ICD-9 code based algorithms that reported PPVs of around 60%. They noted that the studies were 25 years or older and therefore the identified algorithms were outdated because they did not incorporate the October 2008 diagnostic code changes which created an ICD-9 code unique to SJS. A 2012 study by Eisenberg et al.,[73] found difficulties identifying SJS within administrative claims databases from 2000-2007 due to the limitations of the 4 digit ICD-9 code. The PPV for inpatient claims only was 2.00% (95%CI = 0.24%–7.04%), for inpatient claims with 695.1x in first diagnosis field was 4.11% (95%CI = 0.86%–11.54%), and for final decisions of either clinical certainty or probable cases of SJS was 6.00% (95%CI = 3.14%–10.25%).[73]

As noted in Mini-Sentinel Protocol Core review,[15] through September 2008, ICD-9-CM code 695.1 incorporated the EM conditions of erythema iris and herpes iris, SJS,TEN/Lyell's syndrome, and staphylococcus scalded-skin syndrome (SSSS). Because this code is multi-diagnostic, reporting a PPV statistic for each unique disease under its umbrella would be deceptive. Therefore they reported the PPV of the combination of diseases of study interest: EM, SJS, and TEN. SSSS, which was responsible for 15%–16% of cases using this code, was excluded from these PPV calculations because it is no longer incorporated into ICD-9-CM code 695.1. After excluding SSSS, between 53% and 60% of ICD-9-CM code 695.1 reports were validated cases of EM, SJS, or TEN.

Summary of Search of Linkable Databases: The Stevens-Johnson Syndrome Foundation maintains a self-reported registry to gather an accurate count of SJS/TEN patients. In the survey for this registry, patients are asked to report the name of the hospital for any hospitalization due to this condition. The other one identified is ToxIC registry. The ToxIC registry collects patient information about toxicological exposure (including medications) and clinical symptoms which are confirmed by laboratory values and toxicologists. EM/SJS is one of the potential clinical conditions which could be measured in ToxIC registry. However, it seems unlikely due to the concern of linkability, overlap with MSDD and the generalizability to those in the MSDD (see Section E.2 in Chapter V).

Other Issues: As of October 2008 a 5th digit was incorporated into the diagnosis of SJS in the ICD-9 code. ICD-9 code of 695.13 was specific for SJS and not TEN or SJS-TEN. There have not been any validation studies using the new SJS code, and so, further research is necessary.

WG Recommendation: Unlikely. It is unlikely to link to an alternative database with verified cases of EM/SJS.

HOI: Febrile seizures

Previous Validation Studies: The Mini-Sentinel Protocol Core has published a systematic review for identifying seizures, convulsions, or epilepsy using administrative or claims data.[12] Among the reviewed articles, Barlow et al.[74] reported a PPV of 65.4% for both febrile and non-febrile seizures by using an ICD-9 code algorithm for pediatric patients receiving DTP and MMR vaccines. Another study by Shui et al.[75] in 2009 reported various PPVs for emergency department visits due to seizures in those who had received pneumococcal vaccine. Lastly, Klein et al.[76] reported PPVs of 90% (7 to 10 days after vaccination) and 83% (0 -6 days and 10 - 42 days after vaccination) for febrile seizures (identified by ICD 9 algorithm) in children who received MMR and varicella vaccines. In summary, studies focused on determining the incidence of febrile seizures after vaccination in pediatric populations found PPVs between 65 to 90%. No further studies validating this HOI were found after the review by Mini-Sentinel Protocol Core was published.

Summary of Search of Linkable Databases: The International ION channel epilepsy patient registry (IICEPR) was identified as potential registry for patients with febrile seizures. This self-reported registry contains information on patients with febrile seizures starting before age 1 year, first seizure following immunization, and family history which includes febrile and non-febrile seizures. In addition, the WG identified ToxIC registry. The ToxIC registry collects patient information about toxicological exposure (including medications) and clinical symptoms which are confirmed by laboratory values and toxicologists. Febrile seizure is one of the potential clinical conditions which could be measured in ToxIC registry. However, it seems unlikely due to the concern of linkability, overlap with MSDD and the generalizability to those in the MSDD (see Section E.2 in Chapter V).

Other Issues: The diagnosis of febrile seizure includes temperature along with clinical features of seizures. A specific ICD-9 code (780.31) has been allocated for both simple and complex febrile seizures. Previous validation studies have focused only on pediatric patient populations. Furthermore, using the ICD-9 codes has resulted in wide range of PPVs (65 to 90%) in different studies. For future validation studies, using a patient self-reported registry may not be optimal since it does not contain patients with a gold-standard definition of febrile seizure.

WG Recommendation: Unlikely. This HOI is unlikely to be validated using any external data sources given the issues above.

HOI: Guillain-Barré syndrome

Previous Validation Studies: The WG conducted a systematic search and found no previous validation studies for this HOI.

Summary of Search of Linkable Databases: The WG was unable to identify any registries or databases that could be used for alternative validation.

Other Issues: The diagnosis of GBS is dependent on protein levels (> 45 mg/dL) from cerebrospinal fluid and neurologic studies. It is not common for administrative databases to contain specific lab values

WG Recommendation: Not feasible.

HOI: <u>Hemolysis/Hemolytic anemia</u>

Previous Validation Studies: Hemolysis is one of the clinical conditions involved in hemolytic anemia; therefore the WG searched for these two HOIs together. The WG conducted a systematic search and found no previous validation studies for these two HOI.

Summary of Search of Linkable Databases: There are multiple subtypes of hemolytic anemia. The focus of the search for alternative databases for hemolytic anemia was focused on acquired hemolytic anemia with a particular focus on drug-induced hemolytic anemia. The WG was unable to find any suitable data sources for validating hemolytic anemia. Specifically, the WG did not find any registry related to acquired hemolytic anemia. The WG was able to identify registries for sickle cell disease (SCD) and hemolytic uremic syndrome (HUS), which are kinds of hemolytic anemia but they are not acquired hemolytic anemia. These subtypes of hemolytic anemia are inherited hemolytic anemia. Therefore, these were not considered to be useful for validation of this HOI that would be used for active surveillance. The ToxIC registry collects patient information about toxicological exposure (including medications) and clinical symptoms which are confirmed by laboratory values and toxicologists. Hemolysis is one of the potential clinical conditions which could be measured in ToxIC registry. However, it seems unlikely due to the concern of linkability, overlap with MSDD and the generalizability to those in the MSDD (see Section E.2 in Chapter V).

Other Issues: None.

WG Recommendation: Unlikely. There are no promising alternative data sources for validation of hemolytic anemia or hemolysis.

HOI: Henoch-Schonlein purpura

Previous Validation Studies: The FDA CBER has conducted a systematic review for identifying Henoch-Schonlein purpura (HSP) using administrative or claims data (unpublished).In a study by Goodman et al.,[77] the ICD-9 code 287.0 was utilized to identify HSP cases in those between the ages of 16 and 20 who received the meningococcal vaccine. However, the WG could find no validation studies that demonstrate the PPV of and algorithm to identify definite cases of HSP.

Summary of Search of Linkable Databases: The WG was unable to identify any registries or databases that could be used for alternative validation.

Other Issues: There is no specific ICD-9 code for HSP (the closest is 287.0, for allergic purpura).

WG Recommendation: Not feasible.

HOI: Hip fracture

Previous Validation Studies: The OMOP has reported a systematic review for identifying hip fracture using administrative or claims data.[58] They found 3 studies that attempted to validate code-based algorithms from 1999-2009. They concluded that while there is consistency in coding for hip fracture (820.x and 821.x), none of the algorithms had been validated. However, a systematic review performed in 2013 by Hudson et al.[34] identified three algorithms with excellent PPV (two of which were not identified in the OMOP report) and concluded that administrative data can be used to identify hip fractures.

Summary of Search of Linkable Databases: Most of the published literature and searches on registries were non-US based. In the US, four registries were identified, with the Mayo Clinic registry being the most established. Kaiser Permanente has a hip/joint registry, but it is early in development. Registries that focused on hip fractures were not readily identifiable. The Kaiser Permanente hip replacement database could serve as an alternative source of validation but it is unclear how many cases are available, and cause of hip replacement may not be available.

Other Issues: None.

WG Recommendation: Well-validated. Several validated algorithms have been published in high quality studies, so further validation work seems unnecessary.

HOI: <u>Histoplasmosis</u>

Previous Validation Studies: The WG conducted a systematic search and found no previous validation studies for this HOI.

Summary of Search of Linkable Databases: One registry was found which tracked fungal infections, called the Collaborative Exchange of Antifungal Research (CLEAR) registry. It was a multi-center registry which gathered data from 1996-2000 and was maintained by the pharmaceutical company Enzon. Its information was used to track experience and clinical cases with their product Abelcet (amphotericin B lipid complex) in addition to other antifungal agents. This registry is limited by the timeframe in which it gathered data and may not provide overlap with the current MSDD. The exact type of data (lab values, ICD-9 codes, physician diagnoses, etc.) available in this registry was not determined. Additional registries were not identified.

Other Issues: None.

WG Recommendation: Unlikely. It is unlikely this HOI could be validated using an alternative database.

HOI: Hyperglycemia

Previous Validation Studies The WG conducted a systematic search and found no previous validation studies for this HOI.

Summary of Search of Linkable Databases: The WG found no registries specific to hyperglycemia. Hyperglycemia is defined as a blood glucose level of approximately >126 mg/dL, and at higher levels may be accompanied by symptoms such as polyuria, polydipsia, and polyphagia. It is typically diagnosed by measuring serum glucose. Validation of an algorithm for hyperglycemia could be accomplished using an alternative database that contains serum glucose laboratory results. Several commercial insurance claims databases have such lab results. More importantly several of the Mini-Sentinel Data Partners have these data available in their electronic records. Serum glucose is also one of the laboratory values planned for inclusion in the MSCDM.

Other Issues: One issue with this HOI may be lack of a consensus on the exact blood glucose value that defines hyperglycemia. Glucose values that can be considered hyperglycemic may vary on an individual basis.

WG Recommendation: Feasible – lab based. This is an HOI that would seem best validated using laboratory values within the Mini-Sentinel Data Partners if the lab values (serum glucose) exist in their databases.

HOI: <u>Hypertensive crisis</u>

Previous Validation Studies: The Mini-Sentinel Protocol Core has conducted a systematic review for identifying hypertensive emergency using administrative or claims data (unpublished). However, the evidence was not sufficient to make a conclusion about a validated algorithm.

Summary of Search of Linkable Databases: Hypertensive crisis includes hypertensive urgency and hypertensive emergency. Hypertensive urgency is a systolic blood pressure of 180mmHg or more OR a diastolic blood pressure of 110 or more, with no signs of organ damage. Patients may have symptoms of headache, shortness of breath, nose bleed, and anxiety. Hypertensive emergency is a blood pressure exceeding 180mmHg systolic or 120mmHg diastolic plus signs of impending or progressive target organ dysfunction (kidneys, eyes, brain, heart). The Kaiser Permanente Colorado owned an internal registry of hypertension and the identification of hypertensive crisis may be possible. In addition, the MSCDM includes vital signs and specifically both diastolic and systolic blood pressure, and therefore could be used to identify people above the threshold blood pressure. However, it is unclear the frequency with which elevated blood pressures are recorded in people when the crisis occurs versus afterward when treatment might have already started. Further, for hypertensive emergency additional diagnostic evidence (or end organ damage) would be necessary for validation - including kidney damage, stroke, heart failure, aneurysm, pulmonary edema, and others – and not clear how this would be done using an alternative database.

Other Issues: Related terms are "hypertensive urgency" and "hypertensive emergency".

WG Recommendation: Potentially feasible. It is possible to validate hypertensive crisis using the registry in Kaiser Permanente Colorado (which is included in the Health Maintenance Organization Research Network Hypertension Registry).

HOI: <u>Hyperthyroidism</u>

Previous Validation Studies: The WG conducted a systematic search and found no previous validation studies for this HOI.

Summary of Search of Linkable Databases: The WG found no registry or external databases in the US. Most studies using large database or registry are developed in Denmark or Finland.

Other Issues: Serum TSH, T_3 and T_4 are used as a gold standard for hyperthyroidism diagnosis. Note that there are primary hyperthyroidism (low TSH, high T_3, T_4), subclinical hyperthyroidism (low TSH, normal T_3, T_4) and euthyroid hyperthyroxinemia (normal TSH, normal or high T_3, T_4).

WG Recommendation: Feasible – lab based. This is an HOI that would seem best validated using laboratory values within the Mini-Sentinel Data Partners if the lab values exist in their databases.

HOI: Hypoglycemia

Previous Validation Studies The WG conducted a systematic search and found a validation study by Ginde and colleagues.[78] This study utilized ICD-9-CM codes from billing/administrative records to identify hypoglycemic visits within three different hospital Emergency Departments, validated by medical records review. Their algorithm had an 89% positive predictive value for detecting hypoglycemia visits.[78] Though this study had high PPV it may not be sufficiently generalizable to preclude further validation studies by the Mini-Sentinel program.

Summary of Search of Linkable Databases: Hypoglycemia is defined as a plasma glucose level of <70 mg/dL, which may present with or without symptoms of hypoglycemia including sweating, hunger, parasthesias, anxiety, drowsiness, and confusion. Diagnosis of hypoglycemia may be defined by Whipple's triad: (1) hypoglycemic symptoms, as listed above (2) low measured plasma glucose, and (3) relief of symptoms after plasma glucose level is raised. Therefore, like hyperglycemia, validation of an algorithm for hypoglycemia could be accomplished using an alternative database that contains serum glucose laboratory results. Several commercial insurance claims databases have such lab results. More importantly several of the Mini-Sentinel Data Partners have these data available in their electronic records. Serum glucose is also one of the laboratory values planned for inclusion in the MSCDM.

Other Issues: An issue with the identification of hypoglycemia is that a plasma glucose level of <70 mg/dL may not necessarily coexist with hypoglycemic symptoms. Patients may develop symptoms at different levels of plasma glucose. Yet, the gold standard objective measure appears to be a plasma glucose level of <70 mg/dL. Another issue relates to the availability of hospital lab results with Mini-Sentinel Data Partners. It is likely that most laboratory values within Data Partner systems will be ambulatory-based, rather than from an ER visit or hospitalization. If hypoglycemia resulting from exposure to a drug/devise leads to an ER visit or hospitalization (and that is where the serum glucose is tested) then using Mini-Sentinel Data Partner lab results for validation of an algorithm may not allow for identification of these cases.

WG Recommendation: Feasible – lab based. This is an HOI that would seem best validated using laboratory values within the Mini-Sentinel Data Partners if the lab values (serum glucose) exist in their databases.

HOI: Hypothyroidism

Previous Validation Studies: The WG conducted a systematic search and found no previous validation studies for this HOI (few validation studies were conducted in Denmark).

Summary of Search of Linkable Databases: The WG found no registries for this HOI in US. There is a World Thyroid Registry developed by a UK thyroid expert Gordon Skinner, which includes approximately 4,000 enrollees with hypothyroidism or uncontrolled hypothyroidism from all over the world. It is unlikely that there would be much overlap of patients in this registry with MSDD. No other useful external sources were found.

Other Issues: The gold standard of hypothyroidism diagnosis is based on elevated serum TSH and a low serum free T_4. Note that there are primary hypothyroidism (high TSH, low T_4), subclinical hypothyroidism (high TSH, normal T_3, T_4) and euthyroid hypothyroxinemia (normal TSH, normal or low T_3, T_4).

WG Recommendation: Feasible – lab based. This is an HOI that would seem best validated using laboratory values within the Mini-Sentinel Data Partners if the lab values exist in their databases.

HOI: Idiopathic thrombocytopenic purpura

Previous Validation Studies: The WG conducted a systematic search and found no previous validation studies for this HOI.

Summary of Search of Linkable Databases: The WG found eight registries related to idiopathic thrombocytopenic purpura (ITP). Among those, 4 registries were international registries developed by the Intercontinental Cooperative ITP Study Group (ICIS). These registries include patients from the US. The registries this study group has developed, included ICIS-registry I, ICIS-registry II, ICIS splenectomy registry, and ICIS PARC-ITP registry. The first three ICIS registries focused on children with ITP. The ICIS splenectomy registry included children with ITP who were considered candidates for splenectomy. The ICIS PARC-ITP registry is the only one of the ICIS registries to include both adults and children. Currently, the ICIS PARC-ITP registry contains more than 2,410 patients worldwide including in the US. Moreover, it appears to contain patient data that would allow linkage to the MSDD. However, the total number of patients in the US is unclear.

Another ITP registry identified was the North American Chronic ITP Registry (NACIR) which enrolled children and adolescents with chronic ITP. The registry enrolls patients from 16 sites across the US and Canada. In addition to the multi-center registries there was an ITP registry from the state of Oklahoma. This registry would not be a viable choice for linkage with the MSDD. The last ITP specific registry was the Promacta® Pregnancy Registry which enrolled pregnant ITP patients who were exposed to Promacta®.

The final registry that includes patients with ITP is the ToxIC registry. The ToxIC registry collects patient information about toxicological exposure (including medications) and clinical symptoms which are confirmed by laboratory values and toxicologists. ITP is one of the potential clinical conditions which could be measured in ToxIC registry (platelet count < 150x109/L). However, it seems unlikely due to the concern of linkability, overlap with MSDD and the generalizability to those in the MSDD (see Section E.2 in Chapter V).

Other Issues: None

WG Recommendation: Potentially feasible. This HOI is considered potentially feasible with the PARC-ITP registry being the most viable alternative data source.

HOI: Inflammatory bowel disorder: Crohn's disease and Ulcerative colitis

Previous Validation Studies: Six validation studies of algorithms for identifying inflammatory bowel disorder (IBD) were identified.[36,79-83] Among these six, only three were US based. The most notable is Liu et al.[36] in which an IBD registry was constructed from computerized data in Kaiser Permanente from 1996-2002. A total of 2,906 persons were sampled. After confirming with the medical chart, the PPV was reported as 81%-95%. The validity of computerized data for identifying subtypes of IBD included 88% for Crohn's disease (CD) versus 87% for ulcerative colitis (UC). A second validation study was conducted in the Department of Veterans Affairs by Thirumurthi et al.[79] including 3,827 patients. Diagnosis was confirmed by manual chart abstraction. The PPV was 88-100% for CD and 50-93% for UC. The greatest limitation of this study is the lack of patient population diversity (older and mostly males). A third validation study conducted by Herrinton et al.[80] included 400 patients with IBD between 1999-2001 from HMORN in the US. The PPV of the case-finding algorithm was 81%-84%. These studies seem to suggest that this HOI is already well-validated.

Summary of Search of Linkable Databases: In addition to the validation studies discussed above, alternative patient databases were found as well. The TREAT Registry is an observational research program specific to CD with the main objective to document the variety of treatment regimens currently employed in the management of Crohn's diseases. A total of 6,273 patients have been enrolled. Other registries included the CCFA Pediatric PROTECT Study and the Pediatric IBD Collaborative Research Group Registry which focused on pediatric patients. Furthermore, the SECURE Cimzia Post-Marketing Registry and the Humira Ulcerative Colitis Registry were drug-specific patient registries. SECURE is focused on the effectiveness of Cimzia on CD while the Humira Registry concerns the safety and effectiveness of adalimumab in patients with CD and UC.

Other Issues: The main forms of IBD are CD and UC. The gold standard for diagnosis of IBD is a colonoscopy. While CD and UC share many symptoms, they are treated differently medically and surgically. The main distinction between CD and UC is that for CD the location of the inflammation may occur anywhere along the digestive tract from the mouth to the anus. Meanwhile, for UC the large intestine is typically the only site that is affected. In the colonoscopy, CD has a "cobblestone" appearance with patches of healthy tissues in the diseased section while in UC the colon wall is thinner and shows continuous inflammation with no patches. Moreover, granulomas – inflamed cells that become lumped together to form a lesion – are present in CD but not in UC. Thus, histological analysis is also of use in the diagnosis of CD vs. UC.

WG Recommendation: Well-validated. Based on the completed validation studies, there should be no need to further consider IBD by the current project. However, alternative databases were also examined in case more information is needed.

HOI: Intussusception

Previous Validation Studies: The WG conducted a systematic search and found no previous validation studies for this HOI. However, on the Mini-Sentinel website a document was found called "Monitoring for intussusception after two rotavirus vaccines by the PRISM program.[37]" Aim #2 of this proposal is "To determine through medical chart review the positive predictive value of an ICD-9 code-based algorithm for identifying intussusception." If this project is on-going then there would seem to be no need to consider it further under the current project.

Summary of Search of Linkable Databases: No patient registries or databases specific to intussusception were found. There are some exposure-specific registries or databases. Intussusception has been associated with rotavirus vaccination in children. The FDA/CDC maintains the Vaccine Adverse Event Reporting System and patient identification is reported in some cases. There is also the Vaccine Safety Datalink (VSD) - a collaborative effort between CDC's Immunization Safety Office and 9 managed care organizations (MCOs) in the US. The VSD was established in 1990 to monitor immunization safety and address the gaps in scientific knowledge about rare and serious events following immunization. The VSD includes a large linked database that uses administrative data sources at each MCO. Each participating site gathers data on vaccination (vaccine type, date of vaccination, concurrent vaccinations), medical outcomes (outpatient visits, inpatient visits, urgent care visits), birth data, and census data. Because some Mini-Sentinel Data Partners participate in the VSD there would likely be good overlap of patients for a validation study.

Other Issues: Intussusception has been associated with rotavirus vaccination in children. It has also been associated with antibiotic use in children. In adults, it has been cited as a complication of gastric bypass surgery.

WG Recommendation: If a Mini-Sentinel validation project is already in progress then there should be no need to consider intussusception by the current project. If not then the exposure specific databases above may be options for a validation study, potentially feasible.

HOI: Juvenile rheumatoid arthritis

Previous Validation Studies: The FDA CBER has conducted a systematic review for identifying rheumatoid arthritis using administrative or claims data (unpublished). However, no validation studies focused on juvenile rheumatoid arthritis (JRA).

Summary of Search of Linkable Databases: Four registries were identified, including; 1) Childhood Arthritis and Rheumatology Research Alliance (CARRA), 2) Juvenile Rheumatoid Arthritis Research Registry, 3) Juvenile Idiopathic Arthritis Registry, and 4) Rheumatology Clinical Registry.

The CARRA registry includes patients from 59 various medical centers in US who were reported to have JRA by a rheumatologist. The registry is operated by Duke University and funded by National Institute of Arthritis and Musculoskeletal and Skin Diseases (NIAMS). The use of the CARRA registry for validation studies is dependent upon the cross-linkage of patients within the registry to those who are part of MSDD. More information is needed to confirm the feasibility of this registry for future validation studies.

Both Juvenile Rheumatoid Arthritis Research and Juvenile Idiopathic Arthritis registries are enrolling patients for prospective observational studies (see clinicaltrials.gov). Further information about the registries and its use for research purposes was not available.

The Rheumatology Clinical Registry is developed by the ACR to provide an easy-to-use tool for quality improvement initiatives. The registry includes de-identified patient-level data, hence its appropriateness for validation studies may be limited.

Other Issues: Clinical Information -- JRA is diagnosed using the criteria developed by the American College of Rheumatology (ACR). The three major subtypes of JRA are diagnosed as following: 1) Systemic: Persistent intermittent fever (>103 F) with or without rheumatoid rash or any organ involvement; 2) Pauciarticular: Arthritis in four or fewer joints during the first six months of the disease; and 3) Polyarticular: Arthritis in five or few joints during the first six months of the disease. Since JRA is also known as juvenile idiopathic arthritis (JIA), the WG's search strategy included both terms.

WG Recommendation: Potentially feasible. This HOI is potentially feasible to be validated using alternative databases if the CARRA registry can be accessed.

HOI: Lactic acidosis

Previous Validation Studies: The WG conducted a systematic search and found no previous validation studies for this HOI.

Summary of Search of Linkable Databases: There are not any registries or databases for lactic acidosis.

Other Issues: There is not an ICD-9 code specifically for lactic acidosis (pH<7.35). The use of ICD-9 code of 276.2 is used for acidosis, but it includes metabolic, respiratory, and "not other specified acidosis" in addition to lactic acidosis. Serum lactate is a value that can only be found in medical charts/labs. Studies by Selby et al.[84] and Brown et al.[85] identified cases of lactic acidosis in Kaiser Permanente's hospital discharge database using criteria that included an ICD-9 code of 276.2 and a serum lactate level of greater than or equal to 5 mmol/L in the absence of ketones to classify patients as "probable", or if 2-5mmol, they were termed "possible". Only one probable case was identified; no PPVs were reported.

Lactic acidosis is typically observed in an acute care setting as it is not a chronic disease; it is a life-threatening condition if not dealt with. Lactic acidosis is typically observed as an adverse drug event rather than a disease state. The body may eventually compensate itself or the use of sodium bicarbonate can be used to treat lactic acidosis. It is most commonly associated with the use of metformin in patients with T2DM, although this association is now questioned. Stang et al.[86] studied the incidence of lactic acidosis in metformin users and reported that during the study period, 11,797 residents received one or more metformin prescriptions, resulting in 22,296 person-years of exposure. There were 10 subjects who had hospital discharges with a diagnosis of acidosis. However, primary record review revealed only two cases with laboratory findings of elevated blood lactate levels, for an incidence rate of 9 cases per 100,000 person-years of metformin exposure. In both cases, other factors besides metformin could have contributed to the lactic acidosis.

The WG also investigated the possibility of using laboratory results from Mini-Sentinel Data Partners to validate this HOI. According to M. Raebel, there is only a moderate likelihood of obtaining plasma lactate levels from Data Partners because most do not have inpatient lab results.

WG Recommendation: Feasible – lab based. This is an HOI that would seem best validated using laboratory values within the Mini-Sentinel Data Partners for those that have inpatient lab results.

HOI: Mania/Bipolar

Previous Validation Studies: The WG conducted a systematic search and found no previous validation studies for this HOI.

Summary of Search of Linkable Databases: Several potential linkable databases/registries were identified. The Stanley Center Voluntary Bipolar Disorder Case Registry is a self-reported registry in which some patients had a previous physician diagnosis. There was agreement 93% of the time between self-reported diagnostic status and Structured Clinical Interview for DSM-IV (SCID) diagnosis.

The Veterans Healthcare Administration (VHA) National Psychosis Registry, is a registry of veterans diagnosed with psychosis who have received VHA services at any time since 1988. It is maintained by the Serious Mental Illness Treatment Resource and Evaluation Center (SMITREC) which maintain two patient registries, the National Psychosis Registry (NPR) and the National Registry for Depression (NARDEP) which include data from multiple VHA sources for all VHA patients with these diagnoses.

A national survey, the NIMH Epidemiological Catchment Survey, contains confirmed cases of different types of bipolar disorder/mania based on a clinical interview and algorithms to classify the patients.

The National Network of Depression Center (NNDC) is a network comprised of about 16 of the top teaching medical centers whose efforts are combined to create a comprehensive research registry to facilitate studies to develop early intervention or prevention strategies for depression and bipolar disorders. Participants are asked to fill out questionnaires which ask about mood symptoms, overall sense of well-being, impact of symptoms on work and social activities, medication side effects, and other issues.

Other Issues: The SCID is used to obtain a diagnosis of bipolar spectrum disorder, and to obtain a differential diagnosis an algorithm is applied to distinguish between major depressive disorder, major depressive disorder in partial remission, bipolar disorder, dysthymic disorder, and depression not otherwise specified. Separation of these conditions is challenging. For the manic-hypomanic-subsyndromal symptom (SSM) spectrum, given that depressive manifestations are known to be nearly universal in manic and hypomanic individuals, overlap with the bipolar spectrum must be considerable.[87] For instance, in the NIHM ECA database, as reported by Regier et al.[88] both the bipolar I and bipolar II groups have relatively high prevalence of comorbid alcoholism, drug abuse, and major depressive episodes. This and other mental health disorders may have greater issues of linkability due to restrictions on patient disclosure/confidentiality.

WG Recommendation: Unlikely. All registries/potentially linkable databases had limitations: the NIHM ECS was initiated 30 years ago and unlikely to be current; the Veterans registry may have limited generalizability; the Stanley Center registry relies on self-report. The NNDC has limited information available online and linkability is unclear.

Previous Validation Studies: The WG conducted a systematic search and found no previous validation studies for this HOI.

Summary of Search of Linkable Databases: The WG found several large women's studies such as the Women's Health Initiative (women aged 50-79), the Nurses' Health Study (nurses aged 30-55 years) and the Women's Health Study (female health professionals aged > 45 years). Self-reported age at menarche was included in these studies and the study participants are nationally representative. In addition, the WG found a dataset developed by Breast Cancer Surveillance Consortium that combined seven mammography registries of women aged 35-84 years since 1996. The dataset al.so includes a variable for self-reported age at menarche but this is often not collected or not reported. One registry called the Breast Cancer Family Registry also contained information of age at menarche. The Breast Cancer Family Registry enrolled families with multiple or early-onset cases of breast or ovarian cancer in Australia, Canada and US.

Other Issues: No algorithm for menarche was found in previous studies. Most studies used self-reported age at menarche to evaluate the relationship between menarche and potential risks (e.g., breast cancer). The WG searched "menarche"-related ICD-9 codes (http://icd9cm.chrisendres.com/index.php) and found only one code (256.39 Delayed menarche).

One issue with many of the data source the WG identified is the time lag from actual menarche to the date of self-report to the database. Related to this is the time difference between the date of the HOI in these databases (assume >20 years ago) and the data available from MSDD (assume now).

WG Recommendation: Unlikely. Due to the difficulties of developing an algorithm, issues of time lag, and inaccuracy of self-report of menarche, this HOI might not be validated through alternative sources.

HOI: Menopause

Previous Validation Studies: The WG conducted a systematic search and found no previous validation studies for this HOI.

Summary of Search of Linkable Databases: The WG found some registries and databases that include information on menopause status or age at menopause. Large women's studies such as the Women's Health Initiative (women aged 50-79), the Nurses' Health Study (nurses aged 30-55 years) and the Women's Health Study (female health professionals aged > 45 years) contained self-reported age at menopause. The study participants are nationally representative. One dataset developed by Breast Cancer Surveillance Consortium, which combined seven mammography registries of national representative women aged 35-84 years, also included data of self-reported menopause status. Other registries of patients with cancer (UCSF Cancer Registry) or patients post myocardial infarction (Prospective Registry Evaluating Outcomes After Myocardial Infarction: Events and Recovery) also contained self-reported information regarding menstrual history and menopause status. However, the generalizability is relatively limited.

Other Issues: A disorder of menopause (i.e., premature menopause, menopausal and premenopausal disorders, symptomatic menopause) should be defined since "menopause" itself is a natural biomedical status rather than an adverse outcome. Although there is an ICD-9 code indicating the age-related natural postmenopausal status (ICD-9-CM V49.81), seldom do pharmacoepidemiological studies use this algorithm to define patients with menopause or postmenopausal status. Therefore a robust algorithm for menopause might not be easy to identify. Last, the date of menopause is self-reported and subject to recall bias in all databases.

WG Recommendation: Unlikely. Due to the difficulties of algorithm development and potential recall bias, the WG does not recommend this HOI as a good candidate for alternative validation.

HOI: Myocarditis

Previous Validation Studies: The FDA CBER has conducted a systematic review for identifying myopericardits using administrative or claims data (unpublished) and suggested that no good validation studies exist.

Summary of Search of Linkable Databases: There are not good registries for myocarditis. Like Pericarditis there is the ACAM2000® (smallpox vaccine) Myopericarditis Registry which is operated by Sanofi which includes active duty, Reserve, or National Guard service members with either myocarditis or pericarditis associated with receipt of ACAM2000 vaccine. Since this is for military personnel it is unclear what degree of overlap would exist with MSDD. There is also a Vaccine Associated Myopericarditis Registry operated by the US military but it would have the same issue of overlap with MSDD. Myocarditis is also an adverse event associated with clozapine and patients with this HOI might be recruited in the clozapine registries maintained by the pharmaceutical companies, but only if exposed to clozapine.

Other Issues: Myocarditis or inflammatory cardiomyopathy is inflammation of heart muscle (myocardium). It can be caused by infection, autoimmune reactions, and has been associated with some vaccines, chemotherapy, antipsychotics, and other drug exposure. It can lead to chest pain, heart failure, or sudden death. The diagnosis is made by endomyocardial biopsy. Myocardial inflammation can be suspected on the basis of ECG results, elevated CRP and/or ESR. Markers of myocardial damage (troponin and creatine kinase cardiac isoenzymes) are elevated. The gold standard is a biopsy of the myocardium. Because of the biopsy being the gold standard it is possible that myocarditis could be validated using a lab/test results.

WG Recommendation: Unlikely. Validation may be feasible with lab/test result data from one or more Data Partner, but lab result necessary are multiple. Otherwise do not pursue because of limited alternate data sources and limited overlap with MSDD.

HOI: <u>Narcolepsy</u>

Previous Validation Studies: The WG conducted a systematic search and found no previous validation studies for this HOI.

Summary of Search of Linkable Databases: The National Narcolepsy Registry (NNR) was designed for researchers, and the registry contains clinical information, medical data on the family and DNA samples. It is unclear whether this registry can be linked to MSDD due to the lack of a gold standard in diagnosing narcolepsy. The Stanford Center for Narcolepsy treats hundreds of narcoleptic patients each year and many patients volunteer to be in their research protocol. If a database of these patients is available, it would contain patients who meet the gold standard criteria.

Other Issues: A polysomnograph (PSG) and a multiple sleep latency test (MSLT) must be done at a sleep lab and are essential in order to properly diagnose narcolepsy. ICD-9 code of 347.0 is used for narcolepsy.

Narcolepsy with cataplexy (NC) is caused by almost complete loss of hypocretin (orexin) neurons in the hypothalamus. These neurons produce the sleep-wake and REM sleep-regulating neuropeptides hypocretin-1 and hypocretin-2 (orexin-A and orexin-B). Several studies have detected a low level of hypocretin-1 in the cerebrospinal fluid (CSF hcrt-1) in the majority of NC patients as well as in some patients with narcolepsy without cataplexy (NwC). Thus, determination of CSF hcrt-1 may be of diagnostic value for narcolepsy, and it has been included as a diagnostic tool in the current International Classification of Sleep Disorders (ICSD-2). In a study by Knudsen et al.,[89] investigators reported that "…even if the ICSD-2 criterion for low CSF hcrt-1 was a very sensitive and specific diagnostic tool across different populations, it could be argued that its clinical use would be limited: low CSF hcrt-1 mainly detects clear-cut NC patients who are already relatively easy to recognize and diagnose with the conventional diagnostic tools (ICSD-2-defined cataplexy, PSG, and MSLT)."

WG Recommendation: Feasible. Feasible to link to alternative database; depends on ability to access the Stanford-based Center for Narcolepsy database.

HOI: Neuroleptic malignant syndrome

Previous Validation Studies: The WG conducted a systematic search and found no previous validation studies for this HOI.

Summary of Search of Linkable Databases: There were no registries or external databases identified that were specific to NMS. The North American Malignant Hyperthermia (NAMH) Registry may contain patients with NMS (as a related syndrome in conjunction with malignant hyperthermia). NAMH is a self-reported patient registry that allows researchers to obtain de-identified patient data but it is likely not linkable to the MSDD. In addition, the WG also identified the ToxIC registry. The ToxIC registry collects patient information about toxicological exposure (including medications) and clinical symptoms which are confirmed by laboratory values and toxicologists. NMS is one of the potential clinical conditions which could be measured in ToxIC registry. However, it seems unlikely due to the concern of linkability, overlap with MSDD and the generalizability to those in the MSDD (see Section E.2 in Chapter V).

Other Issues: Although NMS and malignant hyperthermia (MH) share many aspects of clinical presentation in common, they are considered different. A few studies consider NMS to be a neurological presentation of MH. The diagnosis of NMS is confirmed with symptoms that include fever, muscle rigidity, altered mental status, and autonomic dysfunction. For future validation studies, using a patient self-reported registry may not appropriate since it does not contain patients with a gold-standard definition of NMS.

WG Recommendation: Unlikely. This HOI is unlikely to be feasible because there is no good alternative datasets available.

HOI: Obesity

Previous Validation Studies: The WG conducted a systematic search and found one study by Andrade et al.[90] that validated ICD-9 codes in medical claims data among women (pre-pregnancy) from HMO Research Network (HMORN) Center for Education and Research and Therapeutics, including Group Health Cooperative, Harvard Pilgrim Health Care, and Kaiser Permanente Colorado. The codes were: morbid obesity ICD-9-CM 278.01; obesity ICD-9-CM 278, 278.0, 278.00, 649.10, 649.11, 649.12, 649.13, 649.14; overweight ICD-9-CM 278.02. These were validated by assessing actual body mass index (BMI) from the patient charts. Overall, 93% of women with a coded diagnosis for obesity/morbid obesity had a documented BMI of ≥30kg/m2; the PPVs were 92% - 95%. This population (women) may not be considered generalizable enough to consider that this HOI has been well validated (missing men). Also, the sensitivity of the algorithm was low (33%).

Summary of Search of Linkable Databases: Obesity if a common condition but while there are some registries/databases related to it, none are optimal for validation purposes. The National Weight Control Registry is the largest, with over 10,000 participants, but it focuses on individuals who that maintained a weight loss of at least 30 pounds for at least a year. Supported by NIH, the registry data are self-reported. Because the focus of this registry is weight control it may not be optimal for FDA's purposes. Much effort has focused in children, and at least one state has started a child obesity registry (Michigan). The National Children's Study (previously discussed) may also be a good source for childhood obesity in the future. There are a number of survey projects, such as the National Survey of Children's Health, or the National Health and Nutrition Examination Survey, that may track obesity in the population but are not linkable for validation purposes. Last, there are some registries specific to obese patients who have received bariatric surgery – mostly focusing on the outcomes of surgery. While these could be used to identify obese patients

Other Issues: Obesity is commonly defined by BMI, which is a person's weight in kilograms divided by the square of height in meters (kg/m^2). WHO and NIH define overweight as BMI 25-29.9 kg/m^2, and obesity as BMI ≥ 30 kg/m^2. Assuming that an algorithm includes only ICD-9 codes then this HOI can be externally validated using height and weight from Data Partners. Note that height and weight are also part of the MSCDM.

BMI may not always be the best indicator of obesity status, particularly in persons such as athletes with a lot of muscle mass. Therefore, some aspect of clinical judgment is necessary in these situations to determine if the patient is actually obese. There is also current debate about using other tests to determine obesity, such as skinfold thickness, but these tests are often too expensive to do within clinics.

Another question is if "obesity" is the HOI of most interest, or if what might be more important is "weight gain". Weight gain itself may not impart additional risk, whereas one might consider the obesity does.

WG Recommendation: Feasible – lab based. This is an HOI that would seem best validated using laboratory values/vital statistics (weight, height) within the Mini-Sentinel Data Partners.

HOI: Optic neuritis

Previous Validation Studies: The FDA CBER has conducted a systematic review for identifying optic neuritis using administrative or claims data (unpublished). A study conducted by Winthrop et al.[91] evaluated the risk of optic neuritis associated with anti-TNF therapy (SABER study) and reported a high PPV of 100%; however, only 135 patients were included in the study all of whom were from Oregon Health & Science University and the Portland Veteran's Administration Medical Center. In another study by Langer-Gould et al.,[54] they identified the optic neuritis by using the ICD-9 codes 377.30 (unspecified optic neuritis), 377.32 (acute retrobulbar neuritis), and 377.39 (other optic neuritis) but only 29 cases were confirmed through reviewing the medical records and no PPV was reported. Therefore, no sufficient evidence exists for validating this HOI.

Summary of Search of Linkable Databases: The Optic Neuritis Treatment Trial (ONTT) database was developed from a randomized trial that evaluated the value of corticosteroids in the treatment of acute optic neuritis. However, the data are old (the trial was conducted from 1988-2003). In a registry of patients with multiple sclerosis (North American Research Committee on Multiple Sclerosis, NARCOMS), 3,798 enrollees reported a history of optic neuritis. However, the WG was not sure whether the identification of optic neuritis was determined solely by self-report data and the generalizability might be a concern, and the data were from 2008. The last potential source is the National Registry of Drug-Induced Ocular Side Effects (NRDIOSE), but the registry is composed of case reports from the US FDA and WHO. The patient overlap with MSDD, confirmation of the optic neuritis, and linkability all might be the major issues.

Other Issues: None.

WG Recommendation: Unlikely. The validation of this HOI is probably not feasible, but if NRDIOSE is linkable then possible. Other alternative databases have limitations that preclude their use.

HOI: Pancytopenia

Previous Validation Studies: The WG conducted a systematic search and found no previous validation studies for this HOI.

Summary of Search of Linkable Databases: The WG was able to identify the ToxIC registry. The ToxIC registry collects patient information about toxicological exposure (including medications) and clinical symptoms which are confirmed by laboratory values and toxicologists. Pancytopenia is one of the potential clinical conditions which could be measured in ToxIC registry. However, it seems unlikely due to the concern of linkability, overlap with MSDD and the generalizability to those in the MSDD (see Section E.2 in Chapter V). No other registries related for patients with pancytopenia were identified.

Other Issues: The diagnosis of pancytopenia is based on laboratory data which includes hemoglobin, platelets, and white blood cell count. Therefore, linkable databases which contain all of these laboratory parameters may serve as an alternative source for validation. Currently, the MSCDM only includes hemoglobin values and is therefore insufficient for identification of pancytopenia. An option for validating this HOI would be to use laboratory results data that are not part of MSCDM from existing Data Partners. Alternatively, commercially available datasets that include lab results (e.g., OptumInsights, Medstat) could be used as a source for validating the results externally.

WG Recommendation: Feasible – lab based. This is an HOI that would seem best validated using laboratory values within the Mini-Sentinel Data Partners if the lab values exist in their databases.

HOI: <u>Pericarditis</u>

Previous Validation Studies: The WG conducted a systematic search and found no previous validation studies for this HOI.

Summary of Search of Linkable Databases: Only one registry was identified, the ACAM2000® (smallpox vaccine) Myopericarditis Registry which is operated by Sanofi which includes active duty, Reserve, or National Guard service members with either myocarditis or pericarditis associated with receipt of ACAM2000 vaccine. Since this is for military personnel it is unclear what degree of overlap would exist with MSDD.

Other Issues: Pericarditis is a condition in which the sac-like covering around the heart (pericardium) becomes inflamed. The Diagnosis includes a variety of tests including auscultation, ECG, echocardiography, Chest X-ray and blood test (CRP, ESR, LDH, leukocyte, Trop-I, CK-MB). Tuberculosis is a common cause of pericarditis and at least one international registry exists of TB pericarditis. Some vaccines have been associated with pericarditis, as have tetracyclines.

WG Recommendation: Unlikely. Do not pursue because of limited alternate data sources and limited overlap with MSDD.

HOI: <u>Peripheral arterial embolism</u>

Previous Validation Studies: The HOI was originally listed as arterial thrombosis. Based on early discussion it was agreed to focus on peripheral arterial embolism. The WG conducted a systematic search and found no previous validation studies for this HOI.

Summary of Search of Linkable Databases: The WG was unable to identify any registries or databases that could be used for alternative validation. The only database found that is related is called DEEP EMBOLI (Distal embolic event protection using excimer laser ablation in peripheral vascular interventions) registry, which was referred to in a 2009 paper.[92] No other information about this registry could be found.

Other Issues: None.

WG Recommendation: Not feasible. No alternative data sources are available.

HOI: Pneumonia

Previous Validation Studies: A recent paper conducted by Barber et al.[39] systematically reviewed the literature for validated algorithms for several serious infections, which included pneumonia. The authors identified several studies that validated algorithms for community-acquired pneumonia using ICD-9 codes in both general populations and those 65 years of age and older, many of which had PPVs greater than 80%. Therefore, community acquired pneumonia appears to be a validated HOI.

Summary of Search of Linkable Databases: The databases identified as potential sources for validation were mostly limited to ventilator-associated pneumonia (VAP) databases. The WG was unable to identify an alternative database that would be useful for validating pneumonia as an outcome. When focused on VAP, the National Trauma Data Bank (NTDB) is an aggregation of trauma registry data from the across the US that includes trauma patients, a subset of whom experience VAP during their hospitalization. Therefore, this database is limited to trauma patients that ultimately acquire VAP. The CDC's National Healthcare Safety Network (NHSN) is a network of healthcare facilities that report healthcare-acquired infections (HAIs) to the CDC for use in surveillance and other activities. The infections that are reported to the CDC will include VAP and other healthcare acquired pneumonias. However, it is not clear if this database would contain data elements that would facilitate linkage to the MSDD. Finally there is an alternative data source specific to children that experience VAP. The Pediatric Ventilator-Associated Pneumonia Registry (VAPoR) is a US based registry that contains cases of VAP in mechanically ventilated patients between ages 0 and 18 years.

Other Issues: None.

WG Recommendation: Well-validated for community acquired pneumonia and potentially feasible for ventilator-associated pneumonia. The data in CDC's NHSN, NTDB or VAPoR are potential alternative sources for validating VAP but further information is needed to identify the best data source.

HOI: Post transfusion allergic reactions

Previous Validation Studies: The WG conducted a systematic search and found no previous validation studies for this HOI.

Summary of Search of Linkable Databases: There is limited data on databases/registries specific for this HOI alone. Most of the blood registries and databases found also include information about other HOIs (e.g., ABO Incompatibility, A**cute hemolytic transfusion reaction (AHTR), Transfusion-related acute lung injury (**TRALI), etc). This makes it difficult to assess the number of patients with a specific HOI in those databases.

The US Biovigilance Network through the American Association of Blood Banks (AABB) may be useful for validation of reported cases of Post-Transfusion allergic reactions. The source is a collaborative effort amongst the Health Human Service (HHS), CDC and different organizations involved in the collection and transfusion of blood. Numerous institutions nationwide feed information up to this national initiative.

Other issues: Overall, post-transfusion allergic reactions are often minor (irritation of the skin and/or mucous membranes) and may not routinely be reported to health care providers. Serious symptoms such as difficulty breathing and death from this HOI are rare.

WG Recommendation: Unlikely. The Biovigilance Network is the best available resource to use for this HOI; however given the problems noted above related to the potential under-reporting and rarity of severe events the WG feels alternative validation of this HOI is unlikely.

HOI: Premature delivery

Previous Validation Studies: The WG conducted a systematic search and found one study conducted by Andrade et al.[40] using the Medication Exposure in Pregnancy Risk Evaluation Program (MEPREP) data to validate pregnant related conditions (including preterm birth) in the health care claims. The collaborators of this program included FDA, the HMO Research Network, Kaiser Permanente Northern and Southern California, and Vanderbilt University. Eleven health plans (within FDA Mini-Sentinel Data Partners) contributed the administrative claims data and the data were linked to the birth certificate files. The algorithm of preterm birth was developed and validated through the review of medical charts. A total of 151 patients were identified and a high PPV was reported as 87% (95%CI 81%-92%).

Summary of Search of Linkable Databases: Not investigated because well-validated algorithms already exist.

Other Issues: None.

WG Recommendation: Well-validated. This HOI has been well validated within Mini-Sentinel Data Partners. No need to be validated through alternative data source.

HOI: <u>Progressive multifocal leukoencephalopathy</u>

Previous Validation Studies: Amend and colleagues[93] determined the incidence of progressive multifocal leukoencephalopathy (PML) in patients without HIV in a claims database. As part of this study, the authors conducted a chart review to validate ICD-9 code 046.3 for identification of patients with PML. The PPV for this was 30%.

Summary of Search of Linkable Databases: No databases containing patients with PML were found. A search was also conducted to identify AIDS databases that might contain PML as a secondary diagnosis, but none were found.

Other Issues: PML is a rare viral disease causing demyelination throughout the central nervous system. It is primarily diagnosed by MRI. Virtually all patients have underlying immunosuppression, with 80% of cases occurring in patients with AIDS.

WG Recommendation: Not Feasible. PML is an HOI that cannot be validated with an existing alternative database.

HOI: Pulmonary fibrosis

Previous Validation Studies: The Mini-Sentinel Protocol Core has published a systematic review for identifying pulmonary fibrosis (PF) and interstitial lung disease (ILD) using administrative or claims data.[10] The investigators were unable to identify any validated algorithms for the either PF or ILD. The review identified five previous studies that had used diagnostic claims (e.g., ICD-9 codes) to identify patients with PF or ILD. The conclusion of the review was that additional work was needed to validate algorithms for the identification of patients with PF or ILD.

Summary of Search of Linkable Databases: There are a few potential data sources that may be useful for validation studies of PF. As noted in the Mini-Sentinel report above,[10] PF is a sub-type of ILD. Therefore, PF was the focus of the search of the alternative databases that might be useful for validation.

In general, there were two broad categories of alternative sources that might be useful for validation of PF. The first category is registries and there were two registries identified that seemed most promising as alternative data sources. The Pennsylvania Idiopathic Pulmonary Fibrosis Registry is a state-wide registry of PF cases that includes contributions from several academic medical centers and the Geisinger Health System. Because Geisinger is a partner in the MSDD it would seem likely to have patient overlap and be linkable. The second registry is the Organ Procurement and Transplantation Network (OPTN). This registry contains a list of patients on the waiting list for an organ transplant and the reason/diagnosis leading to the need for a transplant. Pulmonary fibrosis can lead to the need for lung transplantation and therefore this registry provides a potential source of patients with severe PF that would necessitate a transplant.

The second category of data or databases that might be useful for validation is observational cohort studies. There have been several observation cohort studies conducted on patients with PF. In the US there is a network of 26 research centers that conduct work on PF (IPFNet). This network has conducted several observational cohort studies of PF and may serve as a source for identification of patients with PF that could be potentially linkable to MSDD. It is not clear as to the extent of overlap with MSDD partners but there would likely be some given the geographic spread across the US.

Other Issues: None.

WG Recommendation: Potentially feasible. This is an HOI that could potentially be validated using an alternative data source. The most promising source may be the PF registry from the state of Pennsylvania that includes Geisinger as a source of PF patients (Pennsylvania Idiopathic Pulmonary Fibrosis Registry). It will be important to determine the sample size of PF patients available in this network.

HOI: <u>Pulmonary hypertension</u>

Previous Validation Studies: The WG conducted a systematic search and found no previous validation studies for this HOI.

Summary of Linkable Databases: There were several registries that were identified in the search for alternative databases that included patients with pulmonary hypertension. Several of the databases were single healthcare institution databases and are therefore limited by their sample sizes. Others were registries that are no longer active and the timeframe for which patients were included may not overlap sufficiently with the data available in the MSDD. The REVEAL (Registry to Evaluate Early And Long-Term PAH Disease Management) may provide the best opportunity for validation of diagnosed cases of pulmonary hypertension. Data are available for 3,515 patients from 55 participating centers in the US, with enrollment beginning in 2006. For pediatric populations, the Extracoporeal Life Support Organization Registry is an international database with over 45,000 patients, including patients from 40 individual states within the US. Another useful pediatric registry may be Tracking Outcomes and Practice in Pediatric Pulmonary Hypertension (TOPP), which is an international registry which also includes data from the US.

Other Issues: None.

WG Recommendation: Potentially feasible. This HOI could be validated using an alternative database depending on the amount of overlap between REVEAL and the MSDD.

HOI: Rhabdomyolysis

Previous Validation Studies: Several studies have reported the development of ICD-9 code based algorithms. Andrade et al.[94] evaluated ICD-9 code based algorithms to identify cases of myopathy and rhabdomyolysis, reporting a PPV of 74% (n = 26 out of 35). The algorithm included primary or secondary discharge code for myoglobinuria, primary code for "other disorders of muscle," or a secondary code for "other disorders of muscle" along with a claim for a CK test within seven days of hospitalization or a discharge code for acute renal failure. However, another study by Floyd et al.[95] reported a PPV of 7.2% when using the new ICD-9 code of 728.88 for rhabdomyolysis after recent use of statins. These studies are not sufficient to consider this HOI previously validated.

Summary of Search of Linkable Databases: The ToxIC registry collects patient information about toxicological exposure (including medications) and clinical symptoms which are confirmed by laboratory values and toxicologists. Rhabdomyolysis is one of the potential clinical conditions which could be measured in ToxIC registry. However, it seems unlikely due to the concern of linkability, overlap with MSDD and the generalizability to those in the MSDD (see Section E.2 in Chapter V). No other registries were found.

Other Issues: The diagnosis of rhabdomyolysis is confirmed by CK levels along with clinical presentation of symptoms. While CK levels are useful, they are insufficient. Studies using CK-levels along with ICD-9 codes have reported poor PPVs. Clinical features, as obtained from chart reviews, are necessary to make a definitive diagnosis of rhabdomyolysis.

WG Recommendation: Unlikely. This HOI is unlikely to be validated using an alternative data source.

HOI: <u>Schizophrenia</u>

Previous Validation Studies: The WG conducted a systematic search and found no previous validation studies for this HOI.

Summary of Search of Linkable Databases: Several registries were identified. The Electronic Schizophrenia Treatment Adherence Registry (e-STAR) is an international registry established for patients using risperidone (injectable). According to clinicaltrials.gov website, there were 230 patients enrolled in this study. Further information about the registry and its use for research purposes was not available. It is unclear whether any patients in that database are from the US.

The Clinical and Translational Science Institute (CTSI) of Indiana has an active self-reported registry for patients with schizophrenia. The purpose of the registry is to collect information so that subjects can be contacted for any future research studies. This registry relies on a self-reported diagnosis of schizophrenia, which is a limitation.

Additionally, five registries which included patients taking clozapine were identified. Drug manufacturers (including TEVA, Mylan, Novartis, and Azur Pharma) are mandated by the FDA to collect specific laboratory values and patient identification information as part of the Risk Evaluation and Mitigation Strategy (REMS). Pharmacist or physicians enroll patients into the registry when dispensing or prescribing clozapine. Information regarding the diagnosis of schizophrenia or other off-label indication uses is not reported. Since the enrollment of patients in the registry is based on a prescription for clozapine and not a diagnosis of schizophrenia, the use of this registry to identify patients with a gold-standard definition of schizophrenia is questionable. Hence, these registries are limited in terms of their use for any future validation studies.

Other Issues: This and other mental health disorders may have greater issues of linkability due to restrictions on patient disclosure/confidentiality.

WG Recommendation: Unlikely. This HOI is unlikely to be linkable to databases that contain confirmed cases based on a verifiable gold standard.

HOI: <u>Sepsis and septic shock</u>

Previous Validation Studies: The Mini-Sentinel Protocol Core has published a systematic review for identifying transfusion sepsis using administrative or claims data.[3] The authors found that the ICD-9 code 038.x had a high PPV for sepsis (>80%) in two separate validation studies. The authors recommended that further research on sepsis code validation should focus on performance of codes other than 038.x to identify sepsis or septicemia, such that an optimal combination of codes could be determined. Overall, the number of studies on the validity of sepsis algorithms is relatively small with some inconsistent results. Further research on sepsis algorithms could be useful.

Summary of Search of Linkable Databases: The WG found 6 registries for sepsis or septic shock. One of them, the Promoting Global Research Excellence in Severe Sepsis (PROGRESS) Registry is a global registry but the number of patients of US in the registry is relatively low (762 patients in 2008). Other 4 registries are small, single setting registries, including The Ohio State University Sepsis Registry, the STOP sepsis Registry (Loma Linda university), the Sepsis/ARDS Registry (Emory university school of medicine), and the Hospital of the University of Pennsylvania Registry (Hospital of the University of Pennsylvania). Another registry the WG found is called EM Shock Net, which is a research network focused on clinical issues related to various forms of shock, particularly septic shock and undifferentiated shock. All these registries may have potential use to validate algorithms for sepsis and septic shock. However, because most of them are small-single setting registries, sample size and linkability might be the issues.

Other Issues: None.

WG Recommendation: Unlikely. All registries and databases identified have limitations.

HOI: <u>Serotonin syndrome</u>

Previous Validation Studies: The WG conducted a systematic search and found no previous validation studies for this HOI.

Summary of Search of Linkable Databases: No databases or registries specific to patients with serotonin syndrome were identified. The ToxIC registry collects patient information about toxicological exposure (including medications) and clinical symptoms which are confirmed by laboratory values and toxicologists. Serotonin syndrome is one of the potential clinical conditions which could be measured in ToxIC registry. However, it seems unlikely due to the concern of linkability, overlap with MSDD and the generalizability to those in the MSDD (see Section E.2 in Chapter V).

Other Issues: The diagnosis of serotonin syndrome is made based on a wide range of clinical features that may include increased heart rate, fever, high blood pressure, dilated pupils, tremors, and twitching among others symptoms.

WG Recommendation: Unlikely. This HOI is unlikely to be feasible to be validated using alternative databases unless the ToxIC registry is viable and sufficient cases are identified.

HOI: Solid organ transplant infections

Previous Validation Studies: The Mini-Sentinel Protocol Core has published a systematic review for identifying infections related to blood products, tissue grafts or solid organ transplants using administrative or claims data.[6] The review was able to identify only a single study that validated an algorithm for aspergillosis infection in transplant recipients. The investigators found that ICD-9 codes for aspergillosis had a PPV of 71% and a code for pneumonia with aspergillosis had a PPV of 88%; however the sensitivities were relatively low with both of these codes. The conclusion of the review focused on the heterogeneity of this particular HOI and that it may be difficult to create an algorithm that can be broadly applied to the identification of infections that are conditional on a particular exposure having occurred. Finally, the increased risk of infection due to the condition that necessitated the graft/transfusion (e.g., trauma) makes identifying the actual source of infection difficult. Immunosuppressant uses in transplant patients also make it difficult to identify if infection is from immunosuppression or from the transplant itself.

Because of the concerns noted in the previous Mini-Sentinel review,[6] the WG attempted to identify alternative databases for each of the components of this HOI rather than focusing on the HOI as a whole. The primary justification for this is because of the need for an exposure prior to the HOI and the WG felt that identifying alternative databases through the exposure may result in more options than identification of a general infection-related alternative data source.

Summary of Search of Linkable Databases: The registry that is relevant for infections that occur following a solid organ transplant is the Organ Procurement and Transplantation Network (OPTN). This registry contains both patients that are on the waiting list for a transplant as well as individuals that have received a transplant. Information post-transplant is collected at six months and 1 year and then annually thereafter. It is not clear if this information would be linkable to MSDD nor if it contains information on post-transplant infections.

Other Issues: None.

WG Recommendation: Potentially feasible. It is potentially feasible this HOI could be validated using the data in OPTN.

HOI: Spontaneous abortion

Previous Validation Studies: The FDA CBER has conducted a systematic review for identifying stillbirth and spontaneous abortion using administrative or claims data (unpublished). They found ICD-9 codes 632, 634.x, as well as V27.0-V27.7 were used for spontaneous abortion but no validation studies for spontaneous abortion were identified. The WG also conducted a systematic search and only one validation study was identified. It was conducted in Denmark and had a high PPV of 97.4%.[96] However, no US studies were found and Denmark study is not likely generalizable to the US.

Summary of Search of Linkable Databases: The WG could find no registries or databases specific to spontaneous abortion (or miscarriage), which is usually defined as occurring at < 20 weeks (see below). The National Vital Statistics System maintains records on fetal deaths but this is typically >20 weeks gestation. Further, these data have a long lag in availability and the ability to link to these data is unclear. There are many pregnancy registries (see FDA website) but all are based on a specific exposure (drug) or a disease. For this reason these may suffer some limitations in terms of generalizability. However, most of these are national in scope and likely to have overlap with MSDD, and would likely capture spontaneous abortions that occur.

Another possible alternative data source is the National Children's Health Study (NCS). The NCS is a multi-year research study that was authorized by Congress under the Children's Health Act of 2000. That act requires the NIH Eunice Kennedy Shriver National Institute of Child Health and Human Development to conduct the NCS. The NCS will examine the effects of environmental influences on the health and development of more than 100,000 children across the US, following them from before birth until age 21. Women in their first trimester of pregnancy will be invited to participate in the pre-pregnancy portion of the NCS. Women who are not pregnant but have a high probability of becoming pregnant will be asked to participate in the early pregnancy portion of the NCS. All other eligible women will be asked if the study can contact them periodically to assess their pregnancy status. They also will be asked to contact the NCS should they become pregnant. Women identified as pregnant within 4 years after initial screening will be invited to enroll in the NCS. While the main purpose of the study is to follow children, the outcomes of pregnancies will be collected and presumably could be used by MS. The problem is that the status of NCS funding is unclear and the main study may not yet have begun (difficult to determine from website).

Other Issues: A spontaneous abortion (also called "Miscarriage") is the spontaneous end of a pregnancy at a stage where the embryo/fetus is incapable of surviving independently (i.e., early in gestation). Miscarriage is the most common complication of early pregnancy. Spontaneous abortion is a frequently used clarification to distinguish this natural process from an induced abortion. Fetal death refers to the spontaneous intrauterine death of a fetus at any time during pregnancy. Fetal deaths later in pregnancy (at 20 weeks of gestation or more) are also sometimes referred to as stillbirths. In the US, State laws require the reporting of fetal deaths, and Federal law mandates national collection and publication of fetal death data. Most states report fetal deaths of 20 weeks of gestation or more and/or 350 grams birth weight – but these would not be defined as spontaneous abortions. However, a few states report fetal deaths for all periods of gestation. Fetal death data is published annually by the National Center for Health Statistics, in reports and as individual-record data files.

WG Recommendation: Potentially feasible. Pregnancy registries (National Vital Statistics System - Fatal Death Data) would be used for this. NCS may also be available in future.

HOI: Sudden cardiac death

Previous Validation Studies: The Mini-Sentinel Protocol Core has published a systematic review for identifying this HOI using administrative or claims data.[19] The authors concluded that there is/are a good algorithm (validated) for sudden cardiac death. Previous validation studies have been conducted by Hennessey et al.[42] and Chung et al.[43] with high PPV (85.3% and 86.8% respectively). Hennessey was specific to sudden cardiac death and ventricular arrhythmia whereas Chung was sudden cardiac death. Based on these PPVs it would seem that further validation studies are not necessary.

Summary of Search of Linkable Databases: Not investigated because well-validated algorithms already exist.

Other Issues: Sudden cardiac death is well-validated. Sudden death also occurs in infants and is known as Sudden Infant Death Syndrome (SIDS). This is often respiratory/sleep in origin. There is also Sudden Unexpected Infant Death (SUID) with a variety of causes. There is also sudden death in sports which is often associated with young athletes.

WG Recommendation: Well-validated. The WG agreed that the HOI should focus on sudden cardiac death (not other forms of sudden death). Based on work of Mini-Sentinel Protocol Core, there are already well-validated algorithms for sudden cardiac death.

HOI: <u>Suicide</u>

Previous Validation Studies: The Mini-Sentinel Protocol Core has published a systematic review for identifying this HOI using administrative or claims data.[22] The conclusion was that there is insufficient data support specific recommendations regarding a preferred algorithm. Previous validation attempts focused on "completed suicide" or "suicide attempt".

Summary of Search of Linkable Databases: The National Death Index (NDI) may be useful for validation of completed suicide. State-level death record databases also exist. In either case cause of death variable will allow for identification of completed suicides and is linkable to MSDD. Other options include the National Violent Death Reporting System (NVDRS). Data are available from 18 states and focus on violent deaths, including suicide – but may miss suicides not reported. State-level trauma registries also exist that will also include suicide if admitted to a trauma center. This may also for identification of suicide attempts but only if seen at a trauma center. Other countries have registries of suicide attempts but not the US.

Other Issues: Suicide event of interest (completed suicide, suicide attempt, suicidal ideation, etc.) matters in alternative database selection.

WG Recommendation: Feasible. This HOI could be validated using an alternative database. Recommend focus on completed suicide and use NDI.

HOI: Systemic lupus erythematosus

Previous Validation Studies: The FDA CBER has conducted a systematic review for identifying systematic lupus erythematosus (SLE) using administrative or claims data (unpublished). In the report, relatively few studies were identified that provided validated algorithms for the identification of SLE in a broad based population. The PPV of ICD-9 code 710.0 in selected populations is in the range of 70% to 90% and in general populations is in the range of 50% to 60%. Two studies reported the development of algorithms to identify SLE cases validated by medical records. In validation studies ICD-9 710.0 was the initial screening criteria to identify SLE patients, but actual comparison was agreement between clinician diagnosis based on medical record, and use of the ACR criteria.

Summary of Search of Linkable Databases: Two registries were found that were sponsored by the National Institute of Arthritis and Musculoskeletal and Skin Diseases (NIAMS), which is a branch of the National Institute of Health (NIH). These include The Lupus Family Registry and Repository (LFRR), which is made up of participants that were requested permission for their medical records, and the Research Registry for Neonatal Lupus (RRNL), which collects identifying and diagnostic information from enrolled women and their children. Participants from both registries are from all over the US who have been diagnosed with SLE and have been confirmed upon inclusion into the registries. The LFRR and RRNL are sponsored by the NIAMS and may be linkable to the MSDD.

Other Issues: The two registries identified may contain family members in addition to patients with SLE (LFRR includes family members of SLE patients in an effort to find a genetic link for SLE, RRNL includes both mother and child to understand the maternal and fetal contribution to the risk of disease). Previous validation studies using ICD-9 codes had poor PPV. An effective algorithm will likely require additional clinical variables (such as ACR criteria), but these are not part of MSCDM. Thus, for phase 2 of the HOI linkage project, it may be unlikely that an algorithm can be developed with sufficient test characteristics (i.e., PPV > 0.70).

WG Recommendation: Unlikely. It is possible to link to databases with confirmed cases of SLE. However, previous attempts to develop algorithms with acceptable PPV (>0.70) have not been successful, an important consideration in future investment towards a valid algorithm.

HOI: Tendinopathies

Previous Validation Studies: The WG conducted a systematic search of previous validation studies. Using the Ingenix Research Database (source: United Health Care), Seeger et al.[97] reported a 91% PPV for an algorithm for identifying Achilles tendon rupture, using medical records to confirm cases. A study conducted by Wolf et al.[98] used the Defense Medical Epidemiology Database to identify patients with de Quervain's tenosynovitis using an ICD-9 code (727.04) but no PPV was reported. Similarily, a study by Yee et al.[99] used the United Health Care Research Database, a database consisting of 13 health plans, to identify the incidence of tendon or joint disorders secondary to fluoroquinolone use compared with azithromycin but no PPV was reported.

Summary of Search of Linkable Databases: The WG was unable to identify any registries or databases that could be used for alternative validation.

Other Issues: Tendinopathies is a broad and vaguely defined term for various tendon-related disorders, including tendinitis and tendinosis. Achilles tendinopathy and tendon rupture, epicondylitis (tennis and golf elbow), and rotator cuff tendinopathy are just a few examples of tendinopathies. Others include patellar tendinopathies and wrist tendinopathies. In addition, more specific types of tendinopathies were encountered in various publications, such as stenosing flexor synovitis and de Quervain's tenosynovitis. As a result, there is not one ICD-9 code that encompasses all tendinopathies, rather certain types of tendinopathies have their own ICD-9 code.

WG Recommendation: Not feasible. An acceptable algorithm exists for Achilles tendon rupture but not other types of tendinopathy. Linkage to alternative databases or registries is not feasible given none were identified.

HOI: <u>Thrombotic thrombocytopenic purpura</u>

Previous Validation Studies: The WG conducted a systematic search of previous validation studies. Wahl and colleagues[100] published a validation study for using ICD-9 codes to identify thrombotic thrombocytopenic purpura (TTP) using HealthCore data. The PPV for the algorithms developed as were generally low - the highest being 72.3%.

Summary of Search of Linkable Databases: The WG found 4 TTP registries in the US. The first one was a regional TTP registry (Oklahoma TTP-HUS Registry) which was established in 1989. It is a widely used for research purposes. Several publications have included data from this registry. The Oklahoma TTP-HUS Registry includes patients that required plasma exchange from all hospitals in 58 of the 77 Oklahoma counties. The inclusion of only patients from Oklahoma is an important limitation of the registry with respect to potential overlap with MSDD.

The second registry identified is an international TTP registry, which is a collaboration of 7 countries including the US. However, only the University of Oklahoma is included in the collaboration. Thus it is likely similar to the Oklahoma TTP-HUS Registry in terms of patient population.

The third registry is Transfusion Medicine/Hemostasis (TMH) clinical trial network-TTP registry which was established in 2002. This registry is a collaboration of 17 institutions across the US. The goal of the registry is to enroll all TTP patients from these collaborating institutions. This registry seems to be the most promising candidate for an alternative database given its national scope, however it is not clear if the registry can be linked to MSDD.

Last is the International Registry and Biorepository for Thrombotic Microangiopathy (TMA). This registry only includes children with TTP so may not be sufficiently generalizable to the MSDD population.

Other Issues: None.

WG Recommendation: Potentially feasible. The availability of a national registry for patients with TTP (TTP registry in Transfusion Medicine/Hemostasis (TMH) clinical trial network) makes this a potentially feasible HOI to validate with an alternative data source.

HOI: Tics

Previous Validation Studies: The FDA CBER has conducted a systematic review for identifying tics using administrative or claims data (unpublished) and identified four studies focused on identification of patients with tics. This report found three studies that used the data from Center for Disease Control and Prevention Vaccine Safety Datalink database (VSD) and one used data from Thompson Healthcare Marketscan. All four studies focused on identifying tic disorders in neonates and pediatric patient populations. Methods employed in the studies did not include positive predicted values.

Summary of Search of Linkable Databases: The WG was unable to identify any registries or databases that could be used for alternative validation.

Other Issues: Diagnosis of tics is based on symptoms and commonly associated with Tourette's disorder. Exploring Tourette's disorder and tics associated with those patients would be outside of the scope of this HOI.

WG Recommendation: Not feasible. Since there were no registries and external databases identified, this HOI would be difficult to validate.

HOI: Tissue graft Infections

Previous Validation Studies: The Mini-Sentinel Protocol Core has published a systematic review for identifying infections related to blood products, tissue grafts or solid organ transplants using administrative or claims data.[6] The review was able to identify only a single study that focused on validating an algorithm for aspergillosis infection in transplant recipients. The investigators found that ICD-9 codes for aspergillosis had a PPV of 71% and a code for pneumonia with aspergillosis had a PPV of 88%; however the sensitivities were relatively low with both of these codes. The conclusion of the review focused on the heterogeneity of this particular HOI and that it may be difficult to create an algorithm that can be broadly applied to the identification of infections that are conditional on a particular exposure having occurred. Finally, the increased risk of infection due to the condition that necessitated the graft/transfusion (e.g trauma) makes identifying the actual source of infection difficult. Immunosuppressants in transplant patients also make it difficult to identify if the infection is from immunosuppression or from the transplant itself.

Because of the concerns noted in the previous Mini-Sentinel review,[6] the WG attempted to identify alternative databases for each of the components of this HOI rather than focusing on the HOI as a whole. The primary justification for this is because of the need for an exposure prior to the HOI and the WG felt that identifying alternative databases through the exposure may result in more options than identification of alternative data sources for the more general infection HOI.

Summary of Search of Linkable Databases: Similar to the approach for transfusion-related infections, the WG focused on the most common types of tissue grafts performed (e.g., corneal transplants, ACL replacement, etc.) and the infections associated with those grafts. The WG was unable to locate specific registries for tissue grafts that tracked rates of infections associated with those tissue grafts. However, the American College of Surgeons maintains the National Surgical Quality Improvement Program (NSQIP) that tracks surgeries and their outcomes in participating institutions throughout the US. This includes tracking of surgical site infections following procedures. The database would clearly include information on tissue grafts and surgical site infections; however it does not currently contain billing/administrative data nor is it clear if it can be linked with external data sources.

Other Issues: None.

WG Recommendation: Unlikely. It is unlikely this HOI could be validated using an alternative data source.

HOI: <u>Torsades de pointes</u>

Previous Validation Studies: The WG conducted a systematic search and found no previous validation studies for this HOI.

Summary of Search of Linkable Databases: Torsades de pointes means "twisting of the spikes" in French and it refers to a specific, rare variety of ventricular tachycardia that that exhibits distinct characteristics on the electrocardiogram (ECG). It is associated with long QT syndrome where prolonged QT intervals are visible on the ECG. Diagnosis is made based on ECG. A variety of drugs can cause Torsades de pointes and a "registry" of these drugs is available at http://www.azcert.org/. But this is not a patient registry.

The WG could find only one registry/database that might be useful. The ToxIC registry collects patient information about toxicological exposure (including medications) and clinical symptoms which are confirmed by laboratory values and toxicologists. Torsades de pointes is one of the potential clinical conditions which could be measured in ToxIC registry. However, it seems unlikely due to the concern of linkability, overlap with MSDD and the generalizability to those in the MSDD (see Section E.2 in Chapter V). The only other reference the WG could find related to Torsades de points was regarding adverse event reporting systems.

Other Issues: None.

WG Recommendation: Unlikely. Except for the ToxIC registry there are not useful alternative databases identified that could be used for external validation.

HOI: Transfusion ABO incompatibility reaction/Acute hemolytic transfusion reaction

Previous Validation Studies: The Mini-Sentinel Protocol Core has published a systematic review for identifying this HOI using administrative or claims data.[4] However, no validation studies were identified. Since acute hemolytic transfusion reaction (AHTR) is considered a subtype of transfusion ABO incompatibility reactions, the WG combined these two HOIs. The WG also conducted a systematic search and found no previous validation studies for these.

Summary of Search of Linkable Databases: The US Biovigilance Network under the American Association of Blood Banks (AABB) may be useful for validation of reported cases of transfusion ABO incompatibility reactions. The source is a collaborative effort amongst the HHS, CDC and different organizations involved in the collection and transfusion of blood. Numerous institutions provide data information to this national initiative.

Other Issues: Given the testing that occurs prior to a transfusion, this is a very rare event. It may not be a good candidate for validation given the relative infrequency of the occurrence of the HOI.

WG Recommendation: Unlikely. Despite the availability of the US Biovigilance Network, the rarity of this HOI makes it unlikely to be validated using alternative methods.

HOI: Transfusion infections

Previous Validation Studies: The Mini-Sentinel Protocol Core has published a systematic review for identifying infections related to blood products, tissue grafts or solid organ transplants using administrative or claims data.[6] The review was able to identify only a single study that focused on validating an algorithm for aspergillosis infection in transplant recipients. The investigators found that ICD-9 codes for aspergillosis had a PPV of 71% and a code for pneumonia with aspergillosis had a PPV of 88%; however the sensitivities were relatively low with both of these codes. The conclusion of the review focused on the heterogeneity of this particular HOI and that it may be difficult to create an algorithm that can be broadly applied to the identification of infections that are conditional on a particular exposure having occurred. Finally, the increased risk of infection due to the condition that necessitated the graft/transfusion (e.g trauma) makes identifying the actual source of infection difficult. Immunosuppressants in transplant patients also make it difficult to identify if infection is from immunosuppression or from the transplant itself.

Because of the concerns noted in the previous Mini-Sentinel review,[6] the WG attempted to identify alternative databases for each of the components of this HOI rather than focusing on the HOI as a whole. The primary justification for this is because of the need for an exposure prior to the HOI and the WG felt that identifying alternative databases through the exposure may result in more options than identification of alternative data sources for the more general infection HOI.

Summary of Search of Linkable Databases: In searching for alternative databases for transfusion related infections, the WG focused on general bacterial infections related to transfusions and also on the most common types of bacterial and viral infections that have been associated with transfusions. Similar to other transfusion related HOIs, the AABB biovigilance network is a potential source for identification of patients that experienced an infection following their transfusion. It is unclear how many patients with infections are captured in that data system. In addition, the Infectious Diseases Society of America Emerging Infections Network (IDSA/EIN) is a provider-reported network that includes over 1100 infectious disease specialists and is intended to serve as a mechanism for surveillance with emerging infectious diseases. This data resource could potentially include patients with transfusions that experienced an infection.

Other Issues: There is overlap with this HOI and sepsis as it would seem nearly all of the infections that result from a blood transfusion would be a blood-related infection. The exception would be a site-specific infection (e.g., cellulitis).

WG Recommendation: Unlikely. It is unlikely this HOI could be validated using an alternative data source.

HOI: Transfusion-related acute lung injury

Previous Validation Studies: The WG conducted a systematic search and found no previous validation studies for this HOI. There was a single study that compared blood gas results following transfusion to identification of cases by physicians.[101] The study reported a 93% PPV for computer-based detection of TRALI. However this was based on only 14 patients in a single institution.

Summary of Search of Linkable Databases: Like the other blood transfusion-related HOIs, most of the registries and databases identified also included information about other HOIs in addition to TRALI that are transfusion related (eg. ABO incompatibility, acute hemolytic transfusion reaction (AHTR), transfusion allergic reactions, etc). Therefore, it is difficult to assess the number of patients with a specific HOI in these databases. However, because this is an exposure dependent HOI, alternative databases that could be used for validation exist.

The US Biovigilance Network through the American Association of Blood Banks (AABB) may be the most useful resource for validation of reported cases of TRALI. The source is a collaborative effort amongst the FDA, HHS, CDC and different public and private organizations involved in the collection and transfusion of blood. State-level departments of public health and numerous non-government institutions provide data this national initiative.

Other Issues: The incidence of TRALI has been sharply declining since the American Red Cross started preferentially distributing plasma from male donors in 2007. Subsequently, there has been an 80% decrease in reported cases of TRALI after plasma transfusion. In 2006, there were six cases of possible TRALI-related fatalities following plasma transfusion; in 2008 and 2009, there were none. There was also a significant reduction in non-fatal events by 2008, and this continued into 2009. Plasma distributions from male donors now exceed 99% for groups A, B, and O, and approximately 60% for group AB.

In the study noted above, TRALI was identified amongst patients that underwent a transfusion where the ratio of PaO_2 to FiO_2 was less than 300 on a blood gas. Therefore, it may be possible to develop an algorithm based on blood gas results as the gold standard if these were available in any of the Mini-Sentinel Data Partners. It is unclear if these lab results are available in electronic data held by Mini-Sentinel Data Partners.

Recommendation: Unlikely. It is unlikely this HOI could be validated with an alternative database because of the rarity of the event. An alternative data source does exist and another option could be the use of blood gas lab results from Mini-Sentinel Data Partners; however, the overall sample may be too small to make this a feasible alternative.

HOI: Tuberculosis

Previous Validation Studies: The Mini-Sentinel Protocol Core has conducted a systematic review for identifying tuberculosis using administrative or claims data (unpublished) and found a wide range of PPV from 0% to 100%. Evidence was not sufficient to make a conclusion about a validated algorithm. The algorithms included diagnostic codes and multi-drug prescription regimen components were recommended by the Protocol Core.

Summary of Search of Linkable Databases: The CDC's National Notifiable Diseases Surveillance System (NNDSS) may be useful for validation of diagnosed cases of tuberculosis. This consists of data provided by individual states-level tuberculosis registries. States have mandatory tuberculosis registries, and that data is voluntarily reported to the CDC. In addition to the data compiled by the CDC for the entire country, each state maintains a tuberculosis registry in their departments of public health. Two of the larger existing registries that have been used in previous research studies and that may also be potential candidates for linkage include the New York City tuberculosis registry from the New York City Department of Health and Mental Hygiene (DOHMH) and the California Reportable Disease Information Exchange (CalREDIE). These databases include cases of tuberculosis reported by healthcare professionals, are potentially linkable, and may have patient overlap with MSDD.

Other Issues: None.

WG Recommendation: Feasible. This HOI could be validated using the alternative database. The CDC's NNDSS is recommended to consider for validating this HOI.

HOI: Type 1 diabetes mellitus

Previous Validation Studies: The WG conducted a systematic search and identified a validation study conducted by Rhodes et al.[102] in 2007. A high PPV (97%) was reported for an algorithm of ICD-9 codes (250.x1 or 250.x3, where x=0-9) from billing data from a Boston hospital clinic that was validated by examining medical charts (all patients <26 years old). A study by Vanderloo and colleagues[103] published in 2012 using a Canadian administrative database reported a high PPV (>95%). A similar study was published by Guttmann and colleagues[104] (also in Canada) in 2010. The aforementioned studies were all in children or young adults. No validation studies were identified for an adult population but type 1 diabetes (T1DM) is usually diagnosed in childhood. Generally ICD-9 codes are not age specific but in this case the code is listed as "type 1 (juvenile type)" which may influence who receives the code. It is also important to note that previous studies used ICD-9 code 250.x whereas the code that would likely be of most interest to FDA is 249.x which is "secondary diabetes", defined as "diabetes mellitus (due to) (in) (secondary) (with): drug-induced or chemical induced, or infection." For this reason it may be best to consider this HOI not yet validated.

Summary of Search of Linkable Databases: The most promising database is the new "T1D Exchange" established in 2010 that focuses solely on patients with T1DM. The T1D Exchange currently includes 66 clinics across the US and continues to expand. As of June 2012, there were 25,000 patients enrolled in the registry. The registry collects clinical and laboratory data from patients with T1DM and would seem to be a reasonable alternative database for linking to Mini-Sentinel and conducting a validation study. Another promising source is the Diabetes Registry at Kaiser Permanente Northwest. Kaiser Permanente is a collaborator of the MSDD. The registry includes only patients within the Kaiser Permanente network and has been in existence since 1993 with a total of 232,000 patients. The registry utilizes a "standardized criteria" to identify diabetes patients, includes both T1DM and T2DM.

Other Issues: A common issue is the lack of a clear means of differentiation between T1DM and type 2 diabetes (T2DM). In practice that is no clearly definitive way to do differentiate between the two types of diabetes and it is also not necessary for treatment. T1DM usually occurs in those <35 years of age, but with the rise in obesity children are now getting T2DM. T1DM patients are usually not obese but many T2DM diabetics are not either. Urine ketones are often present in Type 1 diabetes but may also be positive in Type 2 if there is severe volume depletion. Anti-glutamic acid decarboxylase (GAD) antibodies, islet cell antibodies, and insulin autoantibodies are present in 85% of patients with Type 1 at the time of diagnosis, but may disappear within a few years, and usually not required for diagnosis. Nevertheless, this may be the only way to confirm Type 1 diabetes but will be negative in some Type 1 patients.

WG Recommendation: Feasible. This HOI is feasible for validation using an alternative database. The recommended focus of this HOI is the T1D Exchange and the Kaiser Permanente Registry.

HOI: Uveitis

Previous Validation Studies: The FDA CBER has conducted a systematic review for identifying erthema multiforme (EM) using administrative or claims data (unpublished). Three studies tried to validate an algorithm for uveitis using medical chart review (including one of the Mini-Sentinel Data Partners - Kaiser Permanente Northern California). However, the PPV was not high, ranging from 24.0% to 52.1%. The authors concluded that the appropriate uveitis algorithms need to be further determined.

Summary of Search of Linkable Databases: Since uveitis is prevalent in patients with immune-mediated diseases, there are registries for immune-mediated diseases such as multiple sclerosis or rheumatic disease that might be useful. However, the sample size (around 300/registry) and the generalizability might be issues. One clinical trial called the Systemic Immunosuppressive Therapy for Eye Diseases (SITE) cohort study, funded by National Eye Institute, had developed a database for patients with uveitis, scleritis, or mucous membrane pemphigoid determined by medical chart review. Approximate 6,300 patients were included in the study. This might be an important source for this HOI validation, but it is unclear if the study is still on-going. Another potential data source is National Registry of Drug-Induced Ocular Side Effects. In this national registry uveitis is a common adverse event - though the number of uveitis cases cannot be determined from the website at this stage and it is unclear if the registry is linkable to MSDD.

Other Issues: The identification of appropriate uveitis algorithms rather than the alternative databases might be the main obstacle of this HOI validation.

WG Recommendation: Unlikely. The validation of this HOI may be difficult. Recommended data source includes the database held by SITE trial and the National Registry of Drug-Induced Ocular Side Effects, but these have limitations.

HOI: <u>Valvulopathy</u>

Previous Validation Studies: The WG conducted a systematic search and found no previous validation studies for this HOI.

Summary of Search of Linkable Databases: There are some registries in Europe related to valvulopathy or valvular heart disease but most are exposure specific. None were found in the US. However, in the US there are a couple of databases of cardiac surgery outcomes – including valve repair. One is maintained by the Society of Thoracic Surgery (STS), and another by the State of New York. The Department of Veteran's Affairs apparently also has one though no specific information was found about it. These are typically surgeon or hospital reported. At least for the STS database there is information on the reason for the valve repair so it may be possible to identify cases of exposure-induced valvulopathy. However, it is not clear if there is sufficient patient identification information to allow the database to be linked to MSDD. Furthermore, this data source would restrict any validation study to cases requiring surgical repair.

Other Issues: Valvulopathy is a disease or disorder of the values of the heart. Valvular abnormalities from any cause can lead to hemodynamic overload on ventricles and eventually this leads to myocardial dysfunction, congestive heart failure, and sudden death. Valvular disease may require valve replacement surgery and this is the second most common heart operation performed in the US. Four drug classes are associated with valvulopathy – appetite suppressants (e.g., fenfluramine, dexfenfluramine), dopamine agonists (e.g., pergolide, cabergoline), ergot alkaloids (methysergide, ergotamine), recreational drugs (e.g., Ecstasy). Echocardiography is the key test for the diagnosis of valvulopathy.

WG Recommendation: Unlikely. The STS database is probably best source but unlikely to be linkable and unclear if case of valve disease is well-documented.

IX. APPENDIX B

See Excel spreadsheet titled "Mini-Sentinel_Alternative HOI Validation_Detailed Findings". The document is available at: http://www.mini-sentinel.org/

X. REFERENCES

1.   Andrade SE, Harrold LR, Tjia J, et al. A systematic review of validated methods for identifying cerebrovascular accident or transient ischemic attack using administrative data. Pharmacoepidemiol Drug Saf 2012;21 Suppl 1:100-28.

2.   Carnahan RM. Mini-Sentinel's systematic reviews of validated methods for identifying health outcomes using administrative data: summary of findings and suggestions for future research. Pharmacoepidemiol Drug Saf 2012;21:90-9.

3.   Carnahan RM, Herman RA, Moores KG. A systematic review of validated methods for identifying transfusion-related sepsis using administrative and claims data. Pharmacoepidemiol Drug Saf 2012;21:222-9.

4.   Carnahan RM, Kee VR. A systematic review of validated methods for identifying transfusion-related ABO incompatibility reactions using administrative and claims data. Pharmacoepidemiol Drug Saf 2012;21:230-5.

5.   Carnahan RM, Moores KG. Mini-Sentinel's systematic reviews of validated methods for identifying health outcomes using administrative and claims data: methods and lessons learned. Pharmacoepidemiol Drug Saf 2012;21:82-9.

6.   Carnahan RM, Moores KG, Perencevich EN. A systematic review of validated methods for identifying infection related to blood products, tissue grafts, or organ transplants using administrative data. Pharmacoepidemiol Drug Saf 2012;21:213-21.

7.   Cutrona SL, Toh S, Iyer A, et al. Design for validation of acute myocardial infarction cases in Mini-Sentinel. Pharmacoepidemiol Drug Saf 2012;21:274-81.

8.   Herman RA, Gilchrist B, Link BK, Carnahan R. A systematic review of validated methods for identifying lymphoma using administrative data. Pharmacoepidemiol Drug Saf 2012;21:203-12.

9.   Jensen PN, Johnson K, Floyd J, Heckbert SR, Carnahan R, Dublin S. A systematic review of validated methods for identifying atrial fibrillation using administrative data. Pharmacoepidemiol Drug Saf 2012;21 Suppl 1:141-7.

10.   Jones N, Schneider G, Kachroo S, Rotella P, Avetisyan R, Reynolds MW. A systematic review of validated methods for identifying pulmonary fibrosis and interstitial lung disease using administrative and claims data. Pharmacoepidemiol Drug Saf 2012;21:256-60.

11.   Jones N, Schneider G, Kachroo S, Rotella P, Avetisyan R, Reynolds MW. A systematic review of validated methods for identifying acute respiratory failure using administrative and claims data. Pharmacoepidemiol Drug Saf 2012;21:261-4.

12.   Kee VR, Gilchrist B, Granner MA, Sarrazin NR, Carnahan RM. A systematic review of validated methods for identifying seizures, convulsions, or epilepsy using administrative and claims data. Pharmacoepidemiol Drug Saf 2012;21:183-93.

13.   Moores K, Gilchrist B, Carnahan R, Abrams T. A systematic review of validated methods for identifying pancreatitis using administrative data. Pharmacoepidemiol Drug Saf 2012;21:194-202.

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

15.   Schneider G, Kachroo S, Jones N, et al. A systematic review of validated methods for identifying erythema multiforme major/minor/not otherwise specified, Stevens–Johnson Syndrome, or toxic epidermal necrolysis using administrative and claims data. Pharmacoepidemiol Drug Saf 2012;21:236-9.

16.   Schneider G, Kachroo S, Jones N, et al. A systematic review of validated methods for identifying anaphylaxis, including anaphylactic shock and angioneurotic edema, using administrative and claims data. Pharmacoepidemiol Drug Saf 2012;21:240-7.

17.   Schneider G, Kachroo S, Jones N, et al. A systematic review of validated methods for identifying hypersensitivity reactions other than anaphylaxis (fever, rash, and lymphadenopathy), using administrative and claims data. Pharmacoepidemiol Drug Saf 2012;21:248-55.

18.   Singh JA, Kundukulam JA, Bhandari M. A systematic review of validated methods for identifying orthopedic implant removal and revision using administrative data. Pharmacoepidemiol Drug Saf 2012;21:265-73.

19.   Tamariz L, Harkins T, Nair V. A systematic review of validated methods for identifying ventricular arrhythmias using administrative and claims data. Pharmacoepidemiol Drug Saf 2012;21 Suppl 1:148-53.

20.   Tamariz L, Harkins T, Nair V. A systematic review of validated methods for identifying venous thromboembolism using administrative and claims data. Pharmacoepidemiol Drug Saf 2012;21:154-62.

21.   Townsend L, Walkup JT, Crystal S, Olfson M. A systematic review of validated methods for identifying depression using administrative data. Pharmacoepidemiol Drug Saf 2012;21:163-73.

22.   Walkup JT, Townsend L, Crystal S, Olfson M. A systematic review of validated methods for identifying suicide or suicidal ideation using administrative or claims data. Pharmacoepidemiol Drug Saf 2012;21:174-82.

23.   Setoguchi S, Solomon DH, Glynn RJ, Cook EF, Levin R, Schneeweiss S. Agreement of diagnosis and its date for hematologic malignancies and solid tumors between medicare claims and cancer registry data. Cancer Causes Control 2007;18:561-9.

24.   Freeman JL, Zhang D, Freeman DH, Goodwin JS. An approach to identifying incident breast cancer cases using Medicare claims data. J Clin Epidemiol 2000;53:605-14.

25.   Gold HT, Do HT. Evaluation of three algorithms to identify incident breast cancer in Medicare claims data. Health Serv Res 2007;42:2056-69.

26.   Nattinger AB, Laud PW, Bajorunaite R, Sparapani RA, Freeman JL. An algorithm for the use of Medicare claims data to identify women with incident breast cancer. Health Serv Res 2004;39:1733-49.

27.		Warren JL, Feuer E, Potosky AL, Riley GF, Lynch CF. Use of Medicare hospital and physician data to assess breast cancer incidence. Med Care 1999;37:445-56.

28.		Yuan Z, Bowlin S, Einstadter D, Cebul RD, Conners AR, Jr., Rimm AA. Atrial fibrillation as a risk factor for stroke: a retrospective cohort study of hospitalized Medicare beneficiaries. Am J Public Health 1998;88:395-400.

29.		Coyte PC, Hawker G, Croxford R, Wright JG. Rates of revision knee replacement in Ontario, Canada. J Bone Joint Surg Am 1999;81:773-82.

30.		Jollis JG, Ancukiewicz M, DeLong ER, Pryor DB, Muhlbaier LH, Mark DB. Discordance of databases designed for claims payment versus clinical information systems. Implications for outcomes research. Ann Intern Med 1993;119:844-50.

31.		Benatar M, Wuu J, Usher S, Ward K. Preparing for a U.S. National ALS Registry: Lessons from a pilot project in the State of Georgia. Amyotroph Lateral Scler 2011;12:130-5.

32.		Brandenburg NA, Annegers JF. Incidence and risk factors for Bell's palsy in Laredo, Texas: 1974-1982. Neuroepidemiology 1993;12:313-25.

33.		Winkelmayer WC, Schneeweiss S, Mogun H, Patrick AR, Avorn J, Solomon DH. Identification of individuals with CKD from Medicare claims data: a validation study. Am J Kidney Dis 2005;46:225-32.

34.		Hudson M, Avina-Zubieta A, Lacaille D, Bernatsky S, Lix L, Jean S. The validity of administrative data to identify hip fractures is high--a systematic review. J Clin Epidemiol 2013;66:278-85.

35.		Ray WA, Griffin MR, Fought RL, Adams ML. Identification of fractures from computerized Medicare files. J Clin Epidemiol 1992;45:703-14.

36.		Liu L, Allison JE, Herrinton LJ. Validity of computerized diagnoses, procedures, and drugs for inflammatory bowel disease in a northern California managed care organization. Pharmacoepidemiol Drug Saf 2009;18:1086-93.

37.		Yih K, Lieu T, Kulldorff M, et al. Monitoring for intussusception after two rotavirus vaccines by the PRISM Program. Mini-Sentinel CBER Surveillance Protocol 2011. Available at: http://www.mini-sentinel.org/work_products/PRISM/Mini-Sentinel_PRISM_Rotavirus-and-Intussusception-Protocol.pdf. Accessed [7/15/2013].

38.		Kao AS, Getis A, Brodine S, Burns JC. Spatial and temporal clustering of Kawasaki syndrome cases. Pediatr Infect Dis J 2008;27:981-5.

39.		Barber C, Lacaille D, Fortin PR. A systematic review of validation studies of the use of administrative data to identify serious infections [Epub ahead of print]. Arthritis Care Res (Hoboken) 2013. DOI: 10.1002/acr.21959.

40.		Andrade SE, Scott PE, Davis RL, et al. Validity of health plan and birth certificate data for pregnancy research. Pharmacoepidemiol Drug Saf 2013;22:7-15.

41.   Yasmeen S, Romano PS, Schembri ME, Keyzer JM, Gilbert WM. Accuracy of obstetric diagnoses and procedures in hospital discharge data. Am J Obstet Gynecol 2006;194:992-1001.

42.   Hennessy S, Leonard CE, Freeman CP, et al. Validation of diagnostic codes for outpatient-originating sudden cardiac death and ventricular arrhythmia in Medicaid and Medicare claims data. Pharmacoepidemiol Drug Saf 2010;19:555-62.

43.   Chung CP, Murray KT, Stein CM, Hall K, Ray WA. A computer case definition for sudden cardiac death. Pharmacoepidemiol Drug Saf 2010;19:563-72.

44.   Klein NP, Ray P, Carpenter D, et al. Rates of autoimmune diseases in Kaiser Permanente for use in vaccine adverse event safety studies. Vaccine 2010;28:1062-8.

45.   Centers for Disease Control and Prevention, National Center for Injury Prevention and Control. Web-based Injury Statistics Query and Reporting System (WISQARS). Available at: www.cdc.gov/ncipc/wisqars. Accessed [07/15/2013].

46.   Hebert CJ, Vidt DG. Hypertensive crises. Prim Care 2008;35:475-87.

47.   Maahs DM, West NA, Lawrence JM, Mayer-Davis EJ. Epidemiology of type 1 diabetes. Endocrinol Metab Clin North Am 2010;39:481-97.

48.   Raghu G, Weycker D, Edelsberg J, Bradford WZ, Oster G. Incidence and prevalence of idiopathic pulmonary fibrosis. Am J Respir Crit Care Med 2006;174:810-6.

49.   Kirson NY, Birnbaum HG, Ivanova JI, Waldman T, Joish V, Williamson T. Prevalence of pulmonary arterial hypertension and chronic thromboembolic pulmonary hypertension in the United States. Curr Med Res Opin 2011;27:1763-8.

50.   Harlap S, Shiono PH. Alcohol, smoking, and incidence of spontaneous abortions in the first and second trimester. Lancet 1980;2:173-6.

51.   Wilcox AJ, Weinberg CR, O'Connor JF, et al. Incidence of early loss of pregnancy. N Engl J Med 1988;319:189-94.

52.   Wang X, Chen C, Wang L, Chen D, Guang W, French J. Conception, early pregnancy loss, and time to clinical pregnancy: a population-based prospective study. Fertil Steril 2003;79:577-84.

53.   Leake JA, Albani S, Kao AS, et al. Acute disseminated encephalomyelitis in childhood: epidemiologic, clinical and laboratory features. Pediatr Infect Dis J 2004;23:756-64.

54.   Langer-Gould A, Zhang JL, Chung J, Yeung Y, Waubant E, Yao J. Incidence of acquired CNS demyelinating syndromes in a multiethnic cohort of children. Neurology 2011;77:1143-8.

55.   Wittie M, Nelson LM, Usher S, Ward K, Benatar M. Utility of capture-recapture methodology to assess completeness of amyotrophic lateral sclerosis case ascertainment. Neuroepidemiology 2013;40:133-41.

56.   Albert K, Hoffman S, Quesenberry C, Strickland D, Van Den Eeden S, Hornbrook M. Developing an algorithm for classifying ALS cases using electronic records. 2010. Available at: http://www.hmoresearchnetwork.org/archives/2010/concurrent/B5-04-Albers.pdf. Accessed [7/15/2013].

57.   Nagel G, Unal H, Rosenbohm A, Ludolph AC, Rothenbacher D. Implementation of a population-based epidemiological rare disease registry: study protocol of the amyotrophic lateral sclerosis (ALS)--registry Swabia. BMC Neurol 2013;13:22.

58.   The Observational Medical Outcomes Partnership (OMOP). Health Outcomes of Interest. Available at: http://omop.fnih.org/HOI. Accessed [7/15/2013].

59.   Javitt JC, Zhou Z, Willke RJ. Association between vision loss and higher medical care costs in Medicare beneficiaries costs are greater for those with progressive vision loss. Ophthalmology 2007;114:238-45.

60.   Dawood FS, Kamimoto L, D'Mello TA, et al. Children with asthma hospitalized with seasonal or pandemic influenza, 2003-2009. Pediatrics 2011;128:e27-32.

61.   Mullooly JP, Schuler R, Mesa J, Drew L, DeStefano F. Wheezing lower respiratory disease and vaccination of premature infants. Vaccine 2011;29:7611-7.

62.   Eichler AF, Lamont EB. Utility of administrative claims data for the study of brain metastases: a validation study. J Neurooncol 2009;95:427-31.

63.   Anaya DA, Becker NS, Richardson P, Abraham NS. Use of administrative data to identify colorectal liver metastasis. J Surg Res 2012;176:141-6.

64.   Whyte JL, Engel-Nitz NM, Teitelbaum A, Gomez Rey G, Kallich JD. An Evaluation of Algorithms for Identifying Metastatic Breast, Lung, or Colorectal Cancer in Administrative Claims Data [Epub ahead of print]. Med Care 2013. DOI: 10.1097/MLR.0b013e318289c3fb.

65.   Kavcic M, Fisher BT, Torp K, et al. Assembly of a cohort of children treated for acute myeloid leukemia at free-standing children's hospitals in the United States using an administrative database. Pediatr Blood Cancer 2013;60:508-11.

66.   Fisher BT, Harris T, Torp K, et al. Establishment of an 11-Year Cohort of 8733 Pediatric Patients Hospitalized at United States Free-standing Children's Hospitals With De Novo Acute Lymphoblastic Leukemia From Health Care Administrative Data [Epub ahead of print]. Med Care 2012. DOI: 10.1097/MLR.0b013e31824deff9.

67.   Friedlin J, Overhage M, Al-Haddad MA, et al. Comparing methods for identifying pancreatic cancer patients using electronic data sources. AMIA Annu Symp Proc 2010;2010:237-41.

68.   Grams ME, Plantinga LC, Hedgeman E, et al. Validation of CKD and related conditions in existing data sets: A systematic review. Am J Kidney Dis 2011;57:44-54.

69.   Goldberg D, Lewis J, Halpern S, Weiner M, Lo Re V, 3rd. Validation of three coding algorithms to identify patients with end-stage liver disease in an administrative database. Pharmacoepidemiol Drug Saf 2012;21:765-9.

70.   Goldberg DS, Lewis JD, Halpern SD, Weiner MG, Lo Re V, 3rd. Validation of a coding algorithm to identify patients with hepatocellular carcinoma in an administrative database. Pharmacoepidemiol Drug Saf 2013;22:103-7.

71.   Kramer JR, Davila JA, Miller ED, Richardson P, Giordano TP, El-Serag HB. The validity of viral hepatitis and chronic liver disease diagnoses in Veterans Affairs administrative databases. Aliment Pharmacol Ther 2008;27:274-82.

72.   Sands BE, Duh MS, Cali C, et al. Algorithms to identify colonic ischemia, complications of constipation and irritable bowel syndrome in medical claims data: development and validation. Pharmacoepidemiol Drug Saf 2006;15:47-56.

73.   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;21:760-4.

74.   Barlow WE, Davis RL, Glasser JW, et al. The risk of seizures after receipt of whole-cell pertussis or measles, mumps, and rubella vaccine. N Engl J Med 2001;345:656-61.

75.   Shui IM, Shi P, Dutta-Linn MM, et al. Predictive value of seizure ICD-9 codes for vaccine safety research. Vaccine 2009;27:5307-12.

76.   Klein NP, Fireman B, Yih WK, et al. Measles-mumps-rubella-varicella combination vaccine and the risk of febrile seizures. Pediatrics 2010;126:e1-8.

77.   Goodman MJ, Nordin JD, Belongia EA, Mullooly JP, Baggs J. Henoch-Scholein purpura and polysaccharide meningococcal vaccine. Pediatrics 2010;126:e325-9.

78.   Ginde AA, Blanc PG, Lieberman RM, Camargo CA, Jr. Validation of ICD-9-CM coding algorithm for improved identification of hypoglycemia visits. BMC Endocr Disord 2008;8:4.

79.   Thirumurthi S, Chowdhury R, Richardson P, Abraham NS. Validation of ICD-9-CM diagnostic codes for inflammatory bowel disease among veterans. Dig Dis Sci 2010;55:2592-8.

80.   Herrinton LJ, Liu L, Lafata JE, et al. Estimation of the period prevalence of inflammatory bowel disease among nine health plans using computerized diagnoses and outpatient pharmacy dispensings. Inflamm Bowel Dis 2007;13:451-61.

81.   Bernstein CN, Blanchard JF, Rawsthorne P, Wajda A. Epidemiology of Crohn's disease and ulcerative colitis in a central Canadian province: a population-based study. Am J Epidemiol 1999;149:916-24.

82.   Benchimol EI, Guttmann A, Griffiths AM, et al. Increasing incidence of paediatric inflammatory bowel disease in Ontario, Canada: evidence from health administrative data. Gut 2009;58:1490-7.

83.   Rezaie A. Development of an ICD Coding Definition for Inflammaory Bowel Disease. Calgary, Alberta: University of Calgary; 2007.

84.   Selby JV, Ettinger B, Swain BE, Brown JB. First 20 months' experience with use of metformin for type 2 diabetes in a large health maintenance organization. Diabetes Care 1999;22:38-44.

85.   Brown JB, Pedula K, Barzilay J, Herson MK, Latare P. Lactic acidosis rates in type 2 diabetes. Diabetes Care 1998;21:1659-63.

86.   Stang M, Wysowski DK, Butler-Jones D. Incidence of lactic acidosis in metformin users. Diabetes Care 1999;22:925-7.

87.   Judd LL, Akiskal HS. The prevalence and disability of bipolar spectrum disorders in the US population: re-analysis of the ECA database taking into account subthreshold cases. J Affect Disord 2003;73:123-31.

88.   Regier DA, Farmer ME, Rae DS, et al. Comorbidity of mental disorders with alcohol and other drug abuse. Results from the Epidemiologic Catchment Area (ECA) Study. JAMA 1990;264:2511-8.

89.   Knudsen S, Jennum PJ, Alving J, Sheikh SP, Gammeltoft S. Validation of the ICSD-2 criteria for CSF hypocretin-1 measurements in the diagnosis of narcolepsy in the Danish population. Sleep 2010;33:169-76.

90.   Andrade SE, Moore Simas TA, Boudreau D, et al. Validation of algorithms to ascertain clinical conditions and medical procedures used during pregnancy. Pharmacoepidemiol Drug Saf 2011;20:1168-76.

91.   Winthrop KL, Chen L, Fraunfelder FW, et al. Initiation of anti-TNF therapy and the risk of optic neuritis: from the safety assessment of biologic ThERapy (SABER) Study. Am J Ophthalmol 2013;155:183-9.

92.   Shammas NW, Coiner D, Shammas GA, Christensen L, Dippel EJ, Jerin M. Distal embolic event protection using excimer laser ablation in peripheral vascular interventions: results of the DEEP EMBOLI registry. J Endovasc Ther 2009;16:197-202.

93.   Amend KL, Turnbull B, Foskett N, Napalkov P, Kurth T, Seeger J. Incidence of progressive multifocal leukoencephalopathy in patients without HIV. Neurology 2010;75:1326-32.

94.   Andrade SE, Graham DJ, Staffa JA, et al. Health plan administrative databases can efficiently identify serious myopathy and rhabdomyolysis. J Clin Epidemiol 2005;58:171-4.

95.   Floyd JS, Heckbert SR, Weiss NS, Carrell DS, Psaty BM. Use of administrative data to estimate the incidence of statin-related rhabdomyolysis. JAMA 2012;307:1580-2.

96.   Lohse SR, Farkas DK, Lohse N, et al. Validation of spontaneous abortion diagnoses in the Danish National Registry of Patients. Clin Epidemiol 2010;2:247-50.

97.   Seeger JD, West WA, Fife D, Noel GJ, Johnson LN, Walker AM. Achilles tendon rupture and its association with fluoroquinolone antibiotics and other potential risk factors in a managed care population. Pharmacoepidemiol Drug Saf 2006;15:784-92.

98.   Wolf JM, Sturdivant RX, Owens BD. Incidence of de Quervain's tenosynovitis in a young, active population. J Hand Surg Am 2009;34:112-5.

99.   Yee CL, Duffy C, Gerbino PG, Stryker S, Noel GJ. Tendon or joint disorders in children after treatment with fluoroquinolones or azithromycin. Pediatr Infect Dis J 2002;21:525-9.

100.   Wahl PM, Terrell DR, George JN, et al. Validation of claims-based diagnostic codes for idiopathic thrombotic thrombocytopenic purpura in a commercially-insured population. Thromb Haemost 2010;103:1203-9.

101.   Finlay-Morreale HE, Louie C, Toy P. Computer-generated automatic alerts of respiratory distress after blood transfusion. J Am Med Inform Assoc 2008;15:383-5.

102.   Rhodes ET, Laffel LM, Gonzalez TV, Ludwig DS. Accuracy of administrative coding for type 2 diabetes in children, adolescents, and young adults. Diabetes Care 2007;30:141-3.

103.   Vanderloo SE, Johnson JA, Reimer K, et al. Validation of classification algorithms for childhood diabetes identified from administrative data. Pediatr Diabetes 2012;13:229-34.

104.   Guttmann A, Nakhla M, Henderson M, et al. Validation of a health administrative data algorithm for assessing the epidemiology of diabetes in Canadian children. Pediatr Diabetes 2009;11:122-8.

XI.   BRIEF DESCRIPTION OF FIGURES

Figure 1 shows the linkage between the MSDD and the alternative database containing the gold standard/confirmation of case. Information contained in the alternative database would serve as the gold standard for presence of the HOI. An algorithm constructed by a set of variables in the MSDD would be used to identify the HOI. For the sample of patients linked in both databases, the ability of the algorithm to accurately identify patients with the HOI could be assessed.