Innovation in Medical Evidence Development and Surveillance (IMEDS)

IMEDS-Methods Research Agenda

Reagan-Udall Foundation for the FDA (RUF)

January 9, 2014

TABLE OF CONTENTS

BACKGROUND

Following is a summary of the events and considerations that led to the formation of the IMEDS-Methods Research Agenda. Readers who are most interested in the Agenda itself (beginning on page 12) can to refer to this section for context.

1. Purpose

This IMEDS-Methods Research Agenda is intended to give direction and provide a course to conduct methods research within the IMEDS program under the Reagan-Udall Foundation for the FDA. It is anticipated that $5M – $7M will be needed in Year 1 (2014) to evaluate and extend the results of the Observational Medical Outcomes Partnership (OMOP) experiment, and in parallel to this analysis, begin to prepare the infrastructure for further projects to be conducted. These parallel activities include IMEDS continuation of the IMEDS Lab with the current data sources and the IMEDS (formerly OMOP) Common Data Model (CDM), undertaking data conversions from the IMEDS CDM to the Mini-Sentinel CDM and vice versa as necessary for specific projects, developing and issuing requests for proposals for project work noted in the research agenda, and on-boarding an IMEDS Scientific Director to lead and manage the IMEDS-Methods Research Agenda.

2. The FDA's Sentinel Initiative and IMEDS-Methods

The FDA introduced the Sentinel Initiative as a national electronic system intended to transform FDA's ability to track the safety of marketed drugs, biologics, and medical devices. Launched in May 2008 by the FDA, the Sentinel Initiative aims to develop and implement a proactive system that will complement existing systems that the Agency has in place to track reports of adverse events linked to the use of its regulated products.[1]

IMEDS-Methods is a program within the Reagan-Udall Foundation that supports the FDA's mission by initiating and facilitating research into the methods of safety evaluation in large databases. IMEDS-Methods aims to improve the tools for conducting post-marketing safety surveillance using automated healthcare data and to foster the adoption of its findings as appropriate. In meeting this mission, IMEDS-Methods will also be able to add to the general body of knowledge for using automated health data for broader post-market evidence generation on regulated products.[2]

A core goal of the IMEDS-Methods Research Agenda is to support the FDA's full Sentinel system. IMEDS will work to create and foster an inclusive environment that enhances continued engagement and participation and roles of stakeholders as Sentinel becomes operational and is committed to ensuring transparency of both process and findings. IMEDS will provide ready public access to accurate archives of documentation and datasets.

---

[1] FDA's Sentinel Initiative. http://www.fda.gov/Safety/FDAsSentinelinitiative/ucm2007250.htm accessed September 27, 2013

[2] Innovation in Medical Evidence Development and Surveillance (IMEDS) Charter. http://www.reaganudall.org/our-work/safety-and%20better-evidence/imeds-program/imeds-charter/ accessed September 30, 2013. First Draft Completed: April 2013 Proposed Revised Draft: August 2013; Section 3.1

IMEDS understands the FDA to need the following support with respect to the methods that might be employed in a national, electronic system that is a tool for assessing the safety of marketed drugs, biologics, and medical devices.

1. Knowledge of the operating characteristics of Sentinel and Mini-Sentinel for deriving information that can inform the regulatory process. Key elements are sensitivity, positive predictive value, speed, and cost.
2. Guidance as to when these systems are most and least informative. This guidance needs to be formulated in terms that can be applied even before the nature and dimensions of a safety problem are fully understood. The guidance should apply as well for routine monitoring in the absence of any specific safety concern ("proactive surveillance").
3. Improvements that will:
    a. Make Sentinel more sensitive in signaling consequential safety problems.
    b. Raise the probability that those signals that do emerge from the system correspond to consequential safety problems.
    c. Enhance the speed with which new problems are correctly identified.
    d. Enhance the efficiency of the systems operations, thereby promoting goals 3a-c (above) at unchanged or reduced cost.
4. Improved components of the system, including innovations in:
    a. Sources from which data will be obtained
    b. Data capture from the sources
    c. Data format for surveillance and investigation
    d. Data structure and preprocessing for rapid access
    e. Statistical methods for analysis
    f. Presentation and display of data
5. Expansion of a community of experts who can respond to new regulatory questions, data and technical developments to support the FDA[3].

3. Document History

This document is the result of many iterations of discussion among stakeholders. The topics presented here emerge most immediately from the September 6, 2013, meeting of the IMEDS Scientific Advisory Committee (SAC). The SAC's deliberation took as its starting point a draft Research Agenda that IMEDS staff and advisors had assembled on the basis of conversations with the FDA, legacy OMOP investigators, and both commercial and academic stakeholders. The considerations included comments on early draft IMEDS agendas and comments submitted concerning the last full OMOP Research Agenda (drafted for OMOP

[3] C.f. About the FDA – What We Do. http://www.fda.gov/AboutFDA/WhatWeDo/default.htm, accessed September 28, 2013

Executive Board review in March 2013). SAC discussions revolved around the best ways to synthesize, distill, and prioritize the areas. A draft document created after the SAC meeting was circulated among stakeholders, including the SAC itself, members of the FDA involved in Mini-Sentinel program, Mini-Sentinel investigators, legacy OMOP investigators, as well as representatives of organizations that had supported OMOP. The circulated document was not confidential, and recipients were encouraged to share it and submit comments.

Finally the modified document was presented to the IMEDS Steering Committee for discussion at its meeting on October 11, 2013. In addition to review and active discussion during the meeting, members of the Steering Committee submitted written comments. The Steering Committee voted to accept the IMEDS-Methods Research Agenda as presented and with additional points made in discussion at its October 11 meeting. The resulting document has been again circulated widely and opened for public comments, which have been incorporated as possible..

IMEDS anticipates that the Research Agenda will undergo periodic modification, and anticipate that revisions may occur as needed. The program must adapt to growing information generated through IMEDS projects and across the larger research community. The FDA's needs, too, will evolve with time.

4. Legacy of the OMOP Experiment

An important input into the research areas and costs associated with the IMEDS-Methods Research Agenda is the OMOP Project. OMOP's approach to the question of what could be done with electronic data systems is experimental. The investigators asked what results would emerge from applying a variety of standard epidemiologic designs to a large number of drug-effect relations that were known "ground truths," as evidenced by drug labeling and standard medical references.

Many results that emerged from applying epidemiologic techniques to medical data derived from insurance claims and from electronic medical records were as expected, but a disturbingly large number did not replicate the anticipated ground truths.

Appendix I shows a summary of findings from the OMOP experiment that seemed to highlight the fragility of standard approaches to observational research with large databases for drug safety evaluation and regulation. The findings that posed a challenge included:

- Inconsistent results across data sources and analytic techniques to a greater extent than a purely biologic process could explain.
- Results that were consistent across data sources and analytic approaches but that provided answers that were themselves at variance with what is believed to be true. These include known drug-outcome relations that were not detected and drug-outcome pairs strongly believed to have no direct causal relation, but for which associations were repeatedly found.

The second of these findings has been interpreted as an inherent bias, for which the best corrective would be the statistical approach of redefining the definition of no effect, or equivalently adjusting statistical significance levels ("p-value calibration"). Alternative approaches involve attempts to identify and correct the sources of bias. Subsequent experiments have in fact demonstrated that there are a variety of strategies for reducing either bias or heterogeneity between data sources including partitioning by outcome,

tailoring the analysis to the outcome, and using only results from databases with statistical power to answer the question at hand. Understanding how the results of the OMOP Experiment apply to the approach the FDA takes to the Sentinel Initiative is a pressing matter.

The question that needs to be resolved is the extent to which the OMOP Experiment's findings represent general cautions about the use of observational data derived from insurance claims and electronic medical records for assessing drug safety. Points of difference between the OMOP experiments and the current implementation of the Sentinel Initiative (as "Mini-Sentinel") are the following:

- **Data sources.** The participants in the OMOP Lab and the current IMEDS Lab differ from those in Mini-Sentinel.
- **Data model.** Data from different participants (data partners) were standardized to different common structures in OMOP and Mini-Sentinel.
- **Epidemiologic design and statistical methods and implementation.** OMOP and Mini-Sentinel drew from the same statistical and epidemiologic theory, but the two groups carried out implementation separately.
- **Parameterization.** The implementation of any study design or statistical method requires a host of definitions. These include matters such as timing of presumed drug exposure in relation to dates of drug dispensing, time in database prior to drug exposure to assure that preceding health characteristics are captured, the assumed duration of drug effects in relation to exposure, the combinations of service codes used to define outcome conditions or health status, and the dates of occurrence that should be inferred from codes. The OMOP implementation dealt with such questions algorithmically and made the values of elements of the algorithms into variables whose values were the subject of its experiments. Mini-Sentinel has typically used expert opinion, adapted individually to different research questions.
- **Comparator Selection.** For analysis that required that a comparator drug be selected, the OMOP experiments employed a set of pre-specified algorithms to select comparator drugs. Mini-Sentinel chose comparator drugs based on expert opinion.

5. Who will carry out the IMEDS-Methods Research Agenda?

Research projects will be released publicly in the form of Requests for Proposals (RFPs), and IMEDS anticipates that responses will come from many qualified sources. Any researcher or group may respond. For example, OMOP investigators represent an important pool of individuals and institutions with relevant expertise. In addition, current and past Mini-Sentinel investigators are another group with excellent qualifications for the proposed research. Beyond these groups with a strong historical involvement in the topics of the IMEDS-Methods Research Agenda, there will very likely be other appropriate applicants from the research community. Request for proposals, describing the scope of work and proposal evaluation criteria, will be developed and made publicly available.

IMEDS will entertain investigator-initiated proposals for projects not outlined in the Research Agenda, so long as they fit with the IMEDS and RUF missions. Investigators who wish to propose activities not listed in the Agenda may want to review this section in order to understand the genesis of the priority areas for research.

6. Funding of IMEDS-Methods Projects

A number of possible IMEDS-Methods projects are presented in detail below. The marginal costs of funding these are driven largely by average costs of full-time equivalents (FTEs), including benefits and institutional overhead. At a rough average across the mix of senior and junior investigators and technical experts who have historically been involved in OMOP projects, the cost for an FTE is estimated to be $200,000 annually. Apart from OMOP-style experiments, which involve replicating variations of complex study protocols thousands of times over thousands or tens of thousands of individuals, the marginal costs of computing associated with projects are typically at least an order of magnitude smaller than personnel costs, given that the infrastructure exists for the protocol to be executed.

Beyond the work specified in the IMEDS-Methods Research Agenda, the IMEDS Charter envisions projects for which the genesis and funding come from outside IMEDS. Researchers who have already secured funding for their own time and for Lab costs (for example, from their employer or home institution) can apply to use the IMEDS Lab as a research venue. Researchers who have not yet secured funding commitments might seek the cooperation of the Lab for preliminary work and for statements of willingness to serve as a research venue as they apply for funding from outside sources, such as NIH or PCORI. Non-IMEDS funded projects are subject to the approval process documented in the IMEDS policies and procedures.[4] Investigators with outside funding may avail themselves of support available from the Lab (see the Open Lab Project, below). The Lab costs that they will assume are a fair and proportionate share of the cost of the analytic environment and any support given from IMEDS personnel. The standards of timely transparency that hold for internally funded research will apply equally to externally funded research.

7. The IMEDS Lab

The IMEDS Lab is a cloud configuration of five commercially available data sets, four from Truven Health Analytics and one from Quintiles, see Appendix II for descriptions.

There are two areas in which the IMEDS Lab will need to be expanded. The first is to facilitate investigation of Mini-Sentinel operating characteristics by including data from current Mini-Sentinel partners. The second is to include both US and non-US data sources that currently do not participate in either IMEDS or Mini-Sentinel.

Incorporating as much Mini-Sentinel data as possible into the IMEDS Lab environment is of the highest priority in IMEDS plans for the immediate future. Integration of the Mini-Sentinel data could take different forms.

1. Replicate the distributed analysis structure of the Mini-Sentinel Operations Center.

In this case, the data partners maintain their own data and an Operations Center provides them with template programs to extract study-specific aggregated results. The returned results may be, for example, detailed tabulations of the numbers of individuals with specified exposure and outcome patterns. These center-specific results can be combined as

[4] See http://imeds.reaganudall.org/On-Boarding

in a meta-analysis.[5] The Operations Center checks and assembles the responses and then further processes the information to produce the necessary analyses. Creating a Lab-based coordinating center for requesting partner-based analyses and processing of partner-derived information would have Lab costs similar to those for incorporating partner data directly into the Lab's cloud computing environment. The cost for a participating database partner, including data licensing, CDM formatting, running methods, and staffing at the partner institution to execute data requests has in the past been in the range of $500,000 – $900,000 per partner per year.

2. Transfer some or all data from the Mini-Sentinel partners to the IMEDS Lab under license. (One of the Mini-Sentinel partners, Optum, sells research data commercially that are essentially the same as those data used in Mini-Sentinel.)

This is a hybrid model, with some partners participating as distributed data resources, others simply providing their data under license, to be managed by the IMEDS Lab. As necessary for some activities, this model requires the conversion of existing IMEDS Lab data (Truven and Quintiles) currently in the IMEDS (formerly OMOP) CDM to the Mini-Sentinel CDM for about $100,000. Addition of current Mini-Sentinel data to the IMEDS Lab will require decisions about governance and fees, contracts and payments, and a technical period of integration into the lab. The technical phase, given data access, should take approximately six months for existing Mini-Sentinel partners.

IMEDS is actively investigating both fully distributed and hybrid models for including Mini-Sentinel partner data and has not yet decided on a preferred course. Under either model for some of the work in the Research Agenda, it may be necessary to translate and make available Mini-Sentinel structured data in the IMEDS CDM.

Some of the work in the IMEDS Lab has already involved translation of data between different structures. The availability of common data sets in two CDMs will permit both Agenda-based and investigator-initiated research in characterizing and comparing the results obtained using each, given identical data and similar epidemiologic designs. The availability of common data sets in two CDMS permits full use of tools developed for use in each. To this end, the Lab will need to incorporate analysis routines and modules developed by Mini-Sentinel. Comparisons of the CDMs and comparison of results under the CDMs will be a step in ultimately converting all of IMEDS research to a single, widely accepted CDM.

The second major option for expanding the Lab is to include data that currently are neither in the IMEDS Lab nor in Mini-Sentinel. The most important among these is data from the US Centers for Medicare/Medicaid Services (CMS), whose inclusion would greatly expand the numbers of individuals with data in the Lab and would bolster the correspondence between those persons with data in the Lab and Americans generally. Prominent among the non-US sources of individual health information are the Clinical Practice Research Datalink (CPRD) in the UK whose data have been cast into the OMOP CDM, The Health

[5] For an example of current research on this topic in Mini-Sentinel, see: Toh S, Reichman ME, Houstoun M, Ding X, Fireman BH, Gravel E, Levenson M, Li, L, Moyneur E, Shoaibi A, Zornberg G, Hennessy S. Multivariable confounding adjustment in distributed data networks without sharing of patient-level data. Pharmacoepidemiol Drug Saf. 2013 Jul 23 [Epub ahead of print]

Improvement Network, also in the UK, for which there has likewise been the development of an OMOP CDM.[6]

8. Costs of the IMEDS Lab

Current and ongoing fixed expenses for the IMED Lab are approximately $200,000 per month. Costs include staff and contractors, licenses, computing resources, and other related costs. Projects that include running new experiments, adding new data sources, and converting "raw" data into a CDM adds computing time and consequent cost. These costs do not include non-Lab related personnel or other costs.

9. Priorities

The IMEDS SAC and the stakeholders consulted were largely in accord that the four broad areas presented below should be given priority in funding and timing.

In its review on September 6, 2013, the IMEDS SAC proposed the following priority areas.

1. Evaluate and extend the results of the OMOP experiment, whose findings of many discordant drug-outcome results raise questions about the nature of observational research as conducted historically for safety-assessment of drugs and biologics.
    a. Examine findings of interest from the OMOP experiment to identify drug-outcome relations that may be suitable for evaluation in large databases using current methodologies.
    b. Assess these findings of the OMOP experiment using Mini-Sentinel data formatted in the Mini-Sentinel CDM.
    c. Propose new drug-outcome pairs to assess learning from a/b.
    d. Systematically evaluate the protocol-based Mini-Sentinel findings using IMEDS data, tools, CDMs, and analytic methods.

2. Evaluate PROMPT (Prospective Routine Observational Monitoring Program Tools) in the IMEDS Lab.
    a. Assess performance characteristics based on known drug-outcome pairs, including ones in which there is anticipated bias from confounding and other sources such as surveillance bias or misclassification error.
    b. Assess performance of PROMPT and other tools for sequential surveillance.
    c. Assess settings under which PROMPT can be applied effectively for proactive surveillance, which would involve ongoing repeated assessments of new data over time.

[6] Conversion of UK databases into the OMOP/IMEDS CDM requires more effort than conversion of a US claims database such as one derived from CMS. See for example Zhou X, Murugesan S, Bhullar H, Liu Q, Cai B, Wentworth C, Bate A. An evaluation of the THIN database in the OMOP Common Data Model for active drug safety surveillance. Drug Saf. 2013 Feb;36(2):119-34; see also Matcho A, Ryan PB. Transformation of the Clinical Practice Research Datalink (CPRD) to the OMOP CDM Version 4; available at http://omop.org/2013Symposium. See http://omop.org/cdm .

3. Examine methods for detecting non pre-specified adverse events. (For any drug, the nature and the potential number of outcomes are not specified in advance.)

4. Document the relative merits and uses of data types other than insurance claims to supplement or complement Mini-Sentinel queries.

10. Timing

IMEDS intends to initiate all of the projects listed below in the calendar years 2014 and 2015, with the first RFPs to be issued in the first quarter of 2014. It is anticipated, that the majority of the projects listed can be completed within a year of initiation. The pace with which RFPs are issued will correspond to the availability of IMEDS funds to pay for contracted research as well as infrastructure costs.

An external constraint on many of the projects is the data content of the IMEDS Lab. Although one of the IMEDS Lab databases (Truven Clinical Claims and Encounters) is available formatted under the Mini-Sentinel CDM, there are no current Mini-Sentinel data sources in the Lab. Major parts of the IMEDS-Methods Research Agenda require work in the data of Mini-Sentinel's data partners, using the Mini-Sentinel CDM. This will not be feasible until those data become part of the IMEDS Lab. Consequently the ordering of items in the research agenda must reflect first feasibility and then the anticipated utility of the research product.

IMEDS-METHODS RESEARCH AGENDA

This section lists proposed projects that would be funded by IMEDS as part of the IMEDS-Methods Research Agenda. Where activities are listed in detail, the projects should be considered as examples of work that RUF proposes to fund. While the ordering of the projects reflects an intended sequence, the pace at which projects give rise to RFPs and are implemented will depend on the timing and extent of IMEDS's funding and data available.

IMEDS will provide a collaborative environment for academia, government, and industry stakeholders to work together to define and develop methods most suitable for monitoring drug safety. Applicants may wish to propose other work that ties closely to the four areas proposed by the IMEDS SAC, Applicants may also propose alternative paths to the same end. As noted earlier, in addition to the activities listed below, investigator-initiated research projects will be considered for funding with a view to fulfilling the RUF and IMEDS missions.

Summary of Listed Projects

#	Title	Activity
1	The IMEDS Open Lab	Make IMEDS data accessible for research, and foster a research community.
2	Mini-Sentinel and IMEDS discussion and review of Overlapping Work	Compare and build on the congruent findings on angioedema and ACE inhibitors in OMOP and Mini-Sentinel.
3	Comparative presentation of the Mini-Sentinel and IMEDS CDMs	White paper and publication that systematically compares the Mini-Sentinel and IMEDS CDMs.
4	Methods for creating outcome definitions	Catalogue, systematize and investigate methods for defining health outcomes in administrative databases and EMRs.
5	Recreate Mini-Sentinel protocol-based findings	Systematically evaluate selected protocol-based Mini-Sentinel findings using IMEDS data, tools, CDMs, analytic methods.
6	Mini-Sentinel cross-check of OMOP findings	Assess unresolved findings of the OMOP experiment using Mini-Sentinel data formatted in the Mini-Sentinel CDM.
7	Mini-Sentinel new outcomes	Evaluate previously unstudied drug-outcome pairs with presumed known associations to assess learning from the previous projects.
8	New data sources	Document the relative merits and uses of data types other than insurance claims to supplement or complement Sentinel queries.
9	PROMPT assessment	Evaluate PROMPT (Prospective Routine Observational Monitoring Program Tools) in the IMEDS Lab.
10	Non-pre-specified adverse events	Propose and evaluate methods for detecting non-pre-specified adverse events in large databases.
11	Methods to conduct "experiments" in distributed data	Develop and demonstrate efficient methods to conduct many sophisticated variants on analyses in a distributed data environment.
12	Performance standards	Develop standards for evaluating research systems for drug safety.

1. The IMEDS Open Lab

Make IMEDS date accessible for research, and foster a research community.

The IMEDS Open Lab provides for continuity of the ongoing (legacy OMOP) Research Agenda and the associated encouragement of a research community skilled in the use of IMEDS data. The Lab consists of data, vocabulary, methods, and computing resources.

The cost of the Open Lab project, beyond the Lab infrastructure costs, is anticipated to be 1.5 FTEs, plus annual meeting expenses (including travel) of $150,000, and computing costs of $100,000. The policies and procedures under IMEDS have not been finalized, but a reasonable operating model is to assume that qualified investigators will continue to approach the IMEDS Lab with proposals for work. The IMEDS SAC and SC need to agree that the work is aligned with the IMEDS mission. The investigators provide funding for their own time and for any computing activities outside the Lab. They would have an IMEDS Lab account, and IMEDS would provide a negotiated computing budget, which includes access to technical support. The investigators will be required to agree to a policy of "radical transparency," under which all work done within the Lab, including work in progress and results, will be publicly documented.

Many productive uses of the Lab also require the integration of data from Mini-Sentinel partners. Inclusion of data from Mini-Sentinel partners is a high priority for IMEDS. Individuals or groups interested in using the IMEDS Lab should consult the Lab policies documents available at**:** http://imeds.reaganudall.org/On-Boarding

Current projects being conducted (or completed) in the IMEDS Lab and presented at the 2013 OMOP-IMEDS Symposium are listed in Appendix III. The institutions that have partnered with OMOP or have used the Lab (including OMOP results) include:

Advocate Medical Group	National Institute of Statistical Sciences
American Academy of Family Physicians	Olmsted Medical Center
AstraZeneca	Oracle Health Sciences Global Business Unit
Bristol-Myers Squibb Co.	Pfizer
Children's Hospital Colorado	RAND Corporation
Clinical Data Interchange Standards Consortium	Regenstrief Institute
Columbia University, Columbia University Medical Center	RTI International
Department of Veterans Affairs, PBM Services Center for Medication Safety	Sanofi
Duke Clinical Research Institute	Siemens Healthcare
Ephir, Inc.	Truven Health Analytics
FDA, Center for Drug Evaluation and Research	Department of Biomathematics, David Geffen School of Medicine at UCLA
GlaxoSmithKline	United BioSource Corporation
Harvard School of Public Health	University of Colorado School of Medicine
HealthCore, Inc.	University of Massachusetts Boston
Janssen Research and Development	University of Pittsburgh School of Medicine
Laboratory for Informatics Development, NIH Clinical Center	University of Utah
MedImmune	University of Washington
Merck Sharp & Dohme	USC Schaeffer Center for Health Economics and Policy
Midwest Heart Specialists	

Open Lab Project Activities

1. External Investigator-initiated projects
2. Annual symposium
3. Publication

Products

1. Publications
2. Creation and maintenance of a user community within stakeholders

2. Mini-Sentinel and IMEDS Discussion and Review of Overlapping Work

Compare and build on the congruent findings on angioedema and ACE inhibitors in OMOP and Mini-Sentinel.

Among the findings in which Mini-Sentinel and OMOP found closely similar results was in the elevated risk for angioedema following initiation of angiotensin converting enzyme (ACE) inhibitors in OMOP and the larger class consisting of angiotensin converting enzyme inhibitors and aliskiren in Mini-Sentinel.[7] IMEDS and Mini-Sentinel investigators will jointly interpret and report the results.

One aspect of the OMOP results that was surprising was a large range in the magnitude of the observed risks of angioedema in different settings. Investigators will explore whether such effects are also observed using methods and approaches applied within the Mini-Sentinel assessment. For both Mini-Sentinel and OMOP findings, participants will attempt a detailed description (e.g., by comparator, specification, and data source) of the angioedema result.

This project does not involve the conduct of new assessments by Mini-Sentinel or IMEDS, but will use the existing results from the two settings to better understand OMOP findings and their relevance to Mini-Sentinel activities.

Products

1. Joint assessment for publication of the consonance of different data bases and approaches in identifying the angioedema signal
2. Final public report in the form of a manuscript publishable in the peer-reviewed medical literature, with appendices as needed
3. Manuscript submission

[7] Toh S, Reichman ME, Houstoun M, Ross Southworth M, Ding X, Hernandez AF, Levenson M, Li L, McCloskey C, Shoaibi A, Wu E, Zornberg G, Hennessy S. Comparative risk for angioedema associated with the use of drugs that target the renin-angiotensin-aldosterone system. Arch Intern Med. 2012 Nov 12;172(20):1582-9. Erratum in: Arch Intern Med. 2013 Jan 14;173(1):14.

3. Comparative presentation of the Mini-Sentinel and IMEDS CDMs

White paper and publication that systematically compares the Mini-Sentinel and IMEDS CDMs.

In Mini-Sentinel and in OMOP there has been extensive work and thoughtful development of the respective CDMs. There have been reports comparing the models in specific applications,[8,9] but there remains a need for comprehensive presentation in a single technical white paper, which would be the product of this activity. The research community's concern about the impact of differing data models extends beyond the IMEDS and Sentinel activities, and the white paper should be sufficiently detailed such that other groups could use it as a sole reference document. This paper would detail the different structures, standard vocabularies and algorithms that each CDM is using as its foundation. It would include direct comparisons of the same data coded under both CDMs. The paper would discuss the assumptions that are necessary to interpret data encoded under each model and the research implications of having either model as the starting point for research or surveillance activities. In light of other available research on the most appropriate data structures for the research use of routinely collected medical information, the white paper would make recommendations about the structures of Sentinel and IMEDS data, proposing ultimately a well-suited, common data structure.

Products

1. Final public report in the form of a manuscript publishable in the peer-reviewed medical literature, with appendices as needed
2. Manuscript submission

4. Methods for Creating Outcome Definitions

Catalogue, systematize and investigate methods for defining health outcomes in administrative databases and EMRs.

Both OMOP and Mini-Sentinel have relied on published literature and *ad hoc* expert committees to form definitions of health states that could serve as the "outcomes" in an analysis of drug effects. In preparation for research on methods for outcome definition, there is a need to synthesize what has been learned about the process of creating outcome definitions. Some specific points include:

- Nosologies of disease implicit in ICD-9, ICD-10, SNOMED and other nomenclatures
- Considerations in comparing results using different nomenclatures and in translating definitions between nomenclatures
- Role of claims and results from diagnostic procedures and laboratory tests
- Role of treatments as a reflection of provider's opinions about the presence of disease

---

[8] Ogunyemi OI, Meeker D, Kim HE, Ashish N, Farzaneh S, Boxwala A. Identifying appropriate reference data models for comparative effectiveness research (CER) studies based on data from clinical information systems. Med Care. 2013 Aug;51(8 Suppl 3):S45-52

[9] Kahn MG, Batson D, Schilling LM. Data model considerations for clinical effectiveness researchers. Med Care. 2012 Jul;50 Suppl:S60-7

- Role of relative timing of different pieces of information, as well as repetition of diagnoses or tests
- Medical urgency of treatment
- Health states that can be characterized as incident events
    - Medical scope
    - Defining the time of onset
    - The role of late-appearing information
- Health states with insidious onset
    - Medical scope
    - Accumulation of sufficient criteria
    - Interpretation of the time periods in personal histories before sufficient criteria have accrued
- Contributions to and degradations of sensitivity and predictive power
- Techniques for identifying optimal combinations of available data elements
- Necessity for review of underlying medical text (medical charts)
- Distinctions (if any) between algorithms for health states used as outcomes, as covariates for confounder control and as definitions for study populations

Activities

1. Review existing literature on the methodology of creating outcome definitions.
2. Prepare a report
    a. Giving state-of-the-art discussion of the elements listed above
    b. Propose guidelines for creation of outcome definitions
    c. Propose research that will fill in gaps in knowledge.
3. Select a group of drug-outcome scenarios with a spectrum of complexity of outcome definitions and levels of confounding to serve as motivating examples.

Products

1. Final public report in the form of a manuscript publishable in the peer-reviewed medical literature, with appendices as needed
2. Publication

5. Recreate Mini-Sentinel Protocol-Based Findings

In order to assess whether or not there are any differences in results between the Mini-Sentinel system and the IMEDS system (based in the IMEDS CDM and non-Mini-Sentinel data sources), *this project will systematically evaluate selected protocol-based Mini-Sentinel findings using IMEDS data, tools, CDMs, analytic methods.* This project should also serve as a model for the assessment of between-data source and between-study discrepancy in research results. It also provides an opportunity to investigate the extent to which the data models themselves introduce apparent relations or obscure ones that may be observable with analyses that begin with the raw data elements.

There have been a number of Mini-Sentinel projects that use a purpose-built protocol to examine a drug-outcome relationship. The motivating belief of using human experts to design a research program a presumption that skilled investigators with knowledge of the data prepare a protocol that represents a best-case scenario for accuracy, since every aspect of data definition and study design in the available databases to produce a correct answer.

This project does not assess the validity of that assumption. It asks the more technical question of whether expert-designed studies of nearly identical structure produce the same results in the two data systems. The first task will be to replicate the Mini-Sentinel protocol as closely as possible in the IMEDS Lab.

A second task would be to identify whether OMOP Experiment findings of the best empirical design are essentially the same as the Mini-Sentinel protocol designs. If they are not, do they lead to substantively different conclusions?

The third, evaluative task that follows from these exercises is to investigate and explain in as much detail as possible any discrepancies between the original Mini-Sentinel protocol results, the Mini-Sentinel protocol replication, and the OMOP-derived design if it differs.

Note that chart review was a part of the Mini-Sentinel protocol assessments and is not generally available in the other databases contributing to the IMEDS Lab. This project is not intended to be an assessment of health outcomes (see Project 3). The selected comparisons will therefore need to be ones for which there is a strong a priori expectation that the outcome is reliably identifiable in insurance claims data, or for which the Mini-Sentinel investigation demonstrated a high predictive value for an algorithm that could be implemented using the elements of the CDM.

Activities

1. Select three Mini-Sentinel protocol-based studies from among those recently posted as complete. According to the Mini-Sentinel website (http://www.mini-sentinel.org/assessments/medical_events/default.aspx), the following Mini-Sentinel projects are either complete or are in progress, and may be complete when this activity is undertaken.
    a. Thromboembolic events after immunoglobulin administration
    b. Influenza vaccines and pregnancy and birth outcomes
    c. Metabolic effects of second generation antipsychotics in youth
    d. Parenteral iron and anaphylactoid reactions
    e. Intussusception risk after rotavirus vaccination in U.S. infants
    f. Influenza vaccines and febrile seizures
    g. Influenza vaccines and birth outcomes
    h. Anti-diabetes drugs & acute myocardial infarction
    i. ACEI/ARBS/Aliskiren/angioedema
    j. Gardasil vaccination and venous thromboembolism

As indicated above, not all are equally suitable for replication. The investigators proposing to do conduct this work would need to justify their choices, to illustrate different aspects of protocol-based database research.

2. Adapt each protocol as nearly as possible to the IMEDS Lab environment using existing OMOP/IMEDS tools
3. Implement the three protocols
4. Examine and evaluate heterogeneity of results across IMEDS-Lab databases
5. Compare the results of the Mini-Sentinel study and the IMEDS Lab study, with interpretation of salient differences and important correspondences

Products

1. Adapted protocols with commentary on changes necessary to implement in the IMEDS Lab environment
2. Statistical analysis plan for each protocol
3. Final public report in the form of a manuscript publishable in the peer-reviewed medical literature, with appendices as needed
4. Manuscript submission

6. Mini-Sentinel Cross-Check of OMOP Findings

Assess unresolved findings of the OMOP experiment using Mini-Sentinel data formatted in the Mini-Sentinel CDM.

This task would focus on findings for which the OMOP Experiment resulted in drug-outcome associations that are not readily explicable in terms of accepted biology and medicine or as clearly understood artifacts of analysis or data. These are presumably relations in which the unexpected findings documented in the OMOP experiment point to the possibility of similarly surprising findings in Mini-Sentinel. The task is to conduct parallel analysis within the Mini-Sentinel data sources that chose to partner with IMEDS.

Activities

1. Identify and document the choice of up to ten findings with presumptively known "ground truths" that were, however, generally not reflected in the OMOP experiment (apparently false positives and negatives). Limit the choice to those in which there is no ready explanation for the discrepancy after review of OMOP data, data format and procedures. Review the basis for the original assignment of ground truth and possible subsequent data, to assure that belief in the effect or its absence is still justifiable. The analysis choices have to be made without knowledge of what results those choices will produce in the Mini-Sentinel, which means that topics already studied within Mini-Sentinel would not be included.
2. Conduct analyses of the ten pairs, including assessment of heterogeneity of results across data partners.
3. Interpret findings.

Products

1. Functioning and documented implementation of Mini-Sentinel modules in the IMEDS Lab environment
2. Interim report of the analyses of the ten selected pairs
3. Final public report in the form of a manuscript publishable in the peer-reviewed medical literature. This should include a proposed scheme for assessing the suitability of drug-disease questions for assessment in databases available to Mini-Sentinel, and a suggestion for further test pairs to evaluate the validity of the proposed scheme.
4. Publication

7. Mini-Sentinel New Outcomes

Evaluate previously unstudied drug-outcome pairs with presumed known associations to assess learning from the previous projects.

One product of the work described above is a set of descriptors of the circumstances under which Mini-Sentinel tools are predicted to yield accurate and reproducible information about a drug-outcome relation. The circumstances include characteristics of drugs, outcomes, timing, concomitant illness and concomitant drugs, and database resources. This project consists of (1) taking proposed new positive and negative test cases meeting the criteria for research questions that are presumed to be amenable to current standard analyses, and (2) conducting analyses of the drug-outcome pairs under a number of plausible scenarios to test whether the presumed consistency between observation and "ground truth" is correct.

Activities

1. Identify up to ten new drug-outcome pairs for evaluation. Document the basis for belief in the proposed ground-truths.
2. Conduct analyses of the ten pairs
3. Compare the results to the predictions of good performance to interpret the reliability of the guidance suggested in the Mini-Sentinel Cross-Check Project
4. Publication (following IMEDS policies and procedures)

Products

1. Interim report on the choice of the ten pairs
2. Interim report on the analytic results
3. Final report in the form of a manuscript publishable in the peer-reviewed medical literature, summarizing pair choice, results, and interpretation

8. New Data Sources

Document the relative merits and uses of data types other than insurance claims to supplement or complement Sentinel queries. A number of initiatives are underway to explore standards and use cases for integrating patient-generated and reported data into healthcare workflows. These new data sources lack accepted research methods and practices for use, especially for use in monitoring medical product safety.

IMEDS research can build upon the disparate public and private activities already underway to explore methods and best practices to support new learning opportunities from these data sources as part of a robust drug surveillance program.

Other work in this area might involve comparisons between investigations of specific questions that have been done both in claims and in EMR data. Do the EMR analyses or analyses based on natural language processing in EMR texts tend to have better performance? Does chart review really add enough accuracy to make it worth the effort and cost? How well do patient-reported outcomes or outcomes identified social media prefigure the results of controlled studies?

In addition to the review and background activities below, IMEDS-Methods welcomes project proposals to carry out integration of data generated by systems other than insurance claims into a safety-surveillance workflow. These might include for example

routine hypothesis generation for PROMPT assessment (see above), projects of immediate follow-back to patients and providers originating observations to generate standardized case reports, and assessment of feasibility of data linkage to capture complete medical care.

This is an active area of research in Mini-Sentinel, and proposals will need to indicate either complementarity or coordination with Mini-Sentinel efforts.

9. PROMPT Assessment

Evaluate PROMPT (Prospective Routine Observational Monitoring Program Tools) in the IMEDS Lab.

Evaluating the performance characteristics of the Mini-Sentinel database, analytic methods, and PROMPT tools has the highest priority for FDA because it uses the Mini-Sentinel system as a stream of evidence for regulatory decision making, but we have not empirically evaluated the system's performance in terms of accuracy and reliability.

Some of the metrics for discrimination between positive and negative "ground truths" that might be used for this evaluation include the area under the receiver-operating characteristic curve, net reclassification index or the integrated discrimination index. Other measures might also include quantifications bias or error, or calculation of coverage probability with and without p-value calibration. This evaluation should be based on empirical evidence, which may be in the form of experiments that systematically vary design parameters.

1. Performance characteristics based on known drug-outcome pairs, including ones in which there is anticipated bias from confounding and other sources such as surveillance bias or misclassification error. Determine which types of drug-event combinations are best suited for monitoring via PROMPT. This includes considering drug types, outcomes and their frequency (common vs. rare), sample size, timing and selection of comparator groups. It will be of interest to FDA to perform this experiment in the Mini-Sentinel database using the Mini-Sentinel CDM (MSCDM) as well as in the IMEDS databases (Truven and Quintiles) using MSCDM to evaluate whether the performance will be different in different databases.
2. Assess heterogeneity of PROMPT results across databases.
3. Assess performance of PROMPT and other tools for proactive surveillance.
4. Assess settings under which PROMPT can be applied effectively for proactive surveillance, involving ongoing repeated assessments with new data

According to the Mini-Sentinel members of the SAC the suite of PROMPT analyses has not yet been established. Accordingly the timing and scope of this work is still uncertain.

However, there are a number of questions around surveillance that could be addressed through data in the IMEDS Lab immediately. These include, for example, but are not limited to:

1. Decision rules for initiating and terminating surveillance of specific adverse events in specific products
2. Adaptive designs that redefine, for example, exposure windows and outcome criteria on the basis of observed results
3. Designs for surveillance for safety outcomes with long intervals from first drug exposure to onset

4. Designs for surveillance for safety outcomes whose onset is not captured in a easily-defined event
5. Sensitivity of safety surveillance results to variations in specific elements in outcome definition

Each of these elements might be present in guidelines for the conduct of active surveillance, and IMEDS would welcome a comprehensive proposal for the creation of these.

10. Non-pre-specified adverse events

Propose and evaluate methods for detecting non pre-specified adverse events in large databases.

An important part of drug safety knowledge has emerged from the clinic, in which alert providers recognize what they believe to be a previously unrecognized adverse drug effect. Attempts to carry the process of recognition into data environments that are rich in numbers of individuals, but less dense on level of individual information, have yielded high false-positive rates, so they are not yet appropriate, even as a component of regulatory decision-making.

Regulatory agencies and pharmaceutical companies have no choice but to rely on spontaneous reporting systems and case reports in the medical literature. These methods have important limitations and many safety issues that arise through these mechanisms are later unconfirmed in epidemiologic studies. Recent examples of efforts in this direction include:

- The Tree-Based Scan statistic data mining method, which has been experimentally applied to occupational disease surveillance and drug safety surveillance. Using a Poisson probability model, this method simultaneously evaluates a nested system of outcomes definitions of both a very specific (e.g. acute liver failure) and general (e.g. any type of liver problem) nature. It adjusts the statistical analysis for the multiple testing inherent in the thousands of overlapping outcomes definitions used. PRISM is currently adapting and exploring its use for vaccine safety surveillance.
- An analog to the genome-wide-association-studies set in the context of large databases and drug-outcome pairs. The technique has good predictive value, but does not eliminate false-positive findings. Visualizations highlight class effects and reveal challenges in confounding.[10]

In addition to the review and background activities below, IMEDS-Methods welcomes project proposals to carry out specific data-mining and screening methods (such as those noted above, but not restricted to them), for which prior work supports the possibility of levels of sensitivity and predictive value that would support an effective role in regulatory decision making. Areas of particular interest include validity and completeness of data for patients ages 65 and above; performance of active surveillance algorithms to identify outcomes that are not expected to occur acutely; development of methods that use medical

[10] Ryan PB, Madigan D, Stang PE, Schuemie MJ, Hripcsak G. Medication-wide association studies: Pharmacometrics & Systems Pharmacology (2013) 2, e76; [Epub ahead of print]

chart validation for continuous improvement of algorithms used in claims data to identify outcomes.

Activities

1. A literature survey should be conducted and existing methods should be evaluated in the IMEDS Lab.
2. A taxonomy of data mining alerts must be developed which can be integrated into risk management planning.
3. Recommendations as to how to integrate data mining and follow-up analysis into regulatory decision-making should be provided.

11. Methods to Conduct "Experiments" in Distributed Data

Develop and demonstrate efficient methods to conduct many sophisticated variants on analyses in a distributed data environment.

An innovation of the OMOP Experiment was to approach the assessment of drug safety research techniques as an empirical problem. The OMOP investigators undertook direct examination of performance of many candidate research designs by means of test cases (drug-outcome pairs) with presumed known relations, and examined these under many variations in design parameters in many different databases. Ultimately there were several million instances examined.

The conduct of such experiments requires a CDM across the different data resources, and is greatly facilitated if the data sources effectively exist within a single computing environment. For reasons of institutional governance, the proprietary nature of some medical data (such as detailed costs), and for guarantees of patient confidentiality, it is often necessary to segregate data from difference sources, such as different managed care organizations or different insurers, with each behinds its own firewall. In these circumstances distributed data processing is required, in which individual-level analysis takes place behind each data partner's firewall, and only aggregate results are shared centrally. These methods require *ad hoc* writing of the extraction programs, and often in practice require site-specific tailoring. (See reference 4 for an example of several techniques for distributed analysis with covariate control.)

Continued mathematical development of techniques for distributed data analysis is necessary in order to efficiently conduct large-scale experiments.

IMEDS will work with investigators to develop solutions or research programs leading to computationally efficient methods for conducting large-scale replications and variations on observational analyses in distributed data systems. These should be demonstrated in the Mini-Sentinel partners' data in the IMEDS Lab.

12. Performance Standards

Develop standards for evaluating research systems for drug safety.

Comparison of the merits of systems for assessing drug safety in large populations have largely focused on the elements of the systems rather than the performance of the system overall. Projects 3 (assessment of CDMs) and 4 (outcome definitions) above are examples of this often-necessary particularistic approach. Evaluations of statistical techniques, as seen in Project 5 (distributed data experiments), Project 11 (detection of non-prespecified adverse events) also addresses components rather than systems.

In distinction, this project asks investigators to propose and evaluate standards for assessing an entire research system, which consists of the data collection, the data, the analytic infrastructure, and the suite of methods used for drug safety research. The OMOP investigators proposed one possible approach to system evaluation. They examined the ability of data sources and analytic approaches to sort out "ground truths" of known and known-to-be-absent drug-outcome associations. Other standards are readily imaginable. For example, one might choose to judge a system's potential to fill in specific missing elements crucial to a regulatory decision for which there already exists a groundwork in pharmacology, animal testing, clinical trials, and anecdotal clinical experience. Systems might be graded in terms of their abilities to sort out questions of comparative effectiveness per unit cost or risk. It is likely that systems may be fit for some purposes, such as identifying cumulative toxicity, but inadequate for others such as acute risks.

APPENDICES

Appendix I: Summary of Findings from the OMOP Experiment

Key findings on database and method heterogeneity from the OMOP experiment[11]

From the investigators' description of their methods,

"… we selected 2 widely used epidemiologic designs, namely a new-user cohort design with propensity score adjustment and a self-controlled case series design. … For all 53 drug-outcome pairs (representing 9 "positive" and 44 "negative" controls) … we applied the cohort method in a consistent and typical fashion, ... The comparison drugs in each case were drugs with the same indication as for the target drug but were not in the same drug class. For the self- controlled case series method for all 53 pairs, we considered the first occurrence of each outcome, excluded outcomes occurring on the first day of any exposure period…."

1. Results that were expected to show a harmful (positive) drug-outcome association
    a. Expected harmful associations that were born out in the OMOP experiment in at least one design and not contradicted in the other
        i. ACE inhibitors and angioedema: strong positive association across databases and with both designs. Point estimates were highly variable and substantially higher in self-controlled than in propensity-adjusted cohort designs.
        ii. Antibiotics and acute liver injury uniformly associated in self-controlled designs, effects consistently near the null in propensity-adjusted cohort designs
        iii. Amphotericin B and acute renal failure uniformly associated in propensity-adjusted cohort designs, effects consistently near the null in self-controlled designs
        iv. Warfarin and bleeding: uniformly associated in self-controlled cohort designs, generally found in propensity-adjusted cohort designs
    b. Expected harmful associations that were uniformly and surprisingly protective (a negative association between drug and outcome)
        i. Tricyclic antidepressants and acute myocardial infarction: consistently negatively associated with AMI in both self-controlled and propensity-adjusted cohort designs
    c. Expected positives that were variably clustered around the null ("False negatives")
        i. Atypical antipsychotics and AMI: effects consistently near the null in both designs
        ii. Benzodiazepines and hip fracture: effects consistently near the null in self-controlled design, variable but tending to be positive in cohort

[11] Madigan D, Ryan PB, Schuemie M, Stang PE, Overhage JM, Hartzema AG, Suchard MA, Dumouchel W, Berlin JA. Evaluating the impact of database heterogeneity on observational study results. Am J Epidemiol. 2013 May 5. [Epub ahead of print]

iii. Bisphosphonates and upper GI bleeding: consistently negative in self-controlled designs, clustered around the null in propensity-adjusted cohort designs

2. Expected null results: Pairs with an expected lack of association
    a. Expected null results that were surprisingly and consistently positive ("False positive associations")
        i. Typical antipsychotics and upper GI bleeding: consistently positive in both study designs
        ii. Antibiotics and acute renal failure: consistently positive in both study designs
        iii. Typical antipsychotics and acute renal failure: usually positive (some null) in cohort designs, consistently positive in self-controlled designs
    b. Expected null findings with both positive and negative (protective) results according to study design choice ("False variable associations")
        i. Beta-blockers and hip fracture: generally positive in cohort designs, generally negative in self-controlled designs
        ii. Antiepileptics and acute renal failure: generally positive in cohort design, negative in self-controlled design
    c. Expected null findings that were surprisingly and consistently negative (False protective associations)
        i. ACE inhibitors and hip fracture: consistently and strongly negative by both designs.
        ii. Bisphosphonates and acute liver injury: consistently and strongly negative by both designs.
    d. Expected null findings that were surprising positive or negative by one design, but not the other. (False variable associations)
        i. Warfarin and acute renal failure: negative (protective) by self-controlled design, null by cohort design
        ii. Benzodiazepines and angioedema: strongly negative (protective) by cohort design, modestly positive or null by self-controlled design

Appendix II: Data Sets (CDM v4) in the IMEDS Laboratory

Name	Description	Population	Observation time	Drugs	Conditions	Procedures	Observations
Truven MarketScan Commercial Claims and Encounters (CCAE)	Represents privately insured population and captures administrative claims with patient-level de-identified data from inpatient and outpatient visits and pharmacy claims of multiple insurance plans.	Total:107.8m %male:48.8 Mean Age(SD): 37.5 (18.4)	Pt-years: 227.8m 2003-2011	Records : 1,942.1m NDC from pharmacy dispensing claims HCPCS/CPT/ICD9P procedures from inpatient/outpatient medical claims	Records : 2,876.2m ICD9 from inpatient / outpatient medical claims	Records : ,512.7m HCPCS/CPT/ICD9P procedures from inpatient / outpatient medical claims	Not available
Truven MarketScan Multi-State Medicaid (MDCD)	Contains administrative claims data for Medicaid enrollees from multiple states, including inpatient, outpatient, and pharmacy services.	Total:15.5m %male:42.2 Mean Age(SD): 29.0 (22.8)	Pt-years: 40.6m 2002-2011	Records : 622.5m NDC from pharmacy dispensing claims HCPCS/CPT/ICD9P procedures from inpatient/outpatient medical claims	Records 1,084.4m ICD9 from inpatient/outpatient medical claims	Records : 1,467.1m HCPCS/CPT/ICD9P procedures from inpatient / outpatient medical claims	Not available
Truven MarketScan Medicare Supplemental Beneficiaries (MDCR)	Captures administrative claims for retirees with Medicare supplemental insurance paid by employers, including services provided under Medicare-covered payment, employer-paid portion, and any out-of-pocket expenses.	Total:8.3m %male:44.6 Mean Age(SD): 79.9 (8.7)	Pt-years: 22.4m 2003-2011	Records : 606.2m NDC from pharmacy dispensing claims HCPCS/CPT/ICD9P procedures from inpatient/outpatient medical claims	Records 660.6m ICD9 from inpatient/outpatient medical claims	Records : 1,177.2m HCPCS/CPT/ICD9P procedures from inpatient/outpatient medical claims	Not available
Truven MarketScan Lab Supplemental (MSLR)	Represents privately insured population that has at least one recorded laboratory value, with administrative claims from inpatient, outpatient, and pharmacy services supplemented by laboratory results.	Total:4.4m %male:35.9 Mean Age(SD): 44.8 (17.2)	Pt-years: 8.9m 2003-2011	Records : 124.5m NDC from pharmacy dispensing claims HCPCS/CPT/ICD9P procedures from inpatient/outpatient medical claims	Records 224.5m ICD9 from inpatient/outpatient medical claims	Records : 415.8m HCPCS/CPT/ICD9P procedures from inpatient/outpatient medical claims	Records : 234.9m LOINC from outpatient laboratory services
GE Centricity (GE)	The GE MQIC (Medical Quality Improvement Consortium) represents GE Centricity EMR users who contribute data for secondary analytic use. The GE MQIC database reflects events in usual care, including patient problem lists, prescribing patterns and over-the-counter use of medications, and other clinical observations in the ambulatory care setting.	Total:34.5m %male:43 Mean Age(SD): 44 (22.5)	Pt-years: 67.8m 1986-2012** **1986 is first year with more than 10,000 people	Records : 719.2m GPI from medication history and prescriptions written	Records 293.7m ICD9 from problem list	Records : 506.8m CPT from procedure list	Records : 4,137.2m LOINC for laboratory values, SNOMED for chief complaints, signs and symptoms

Appendix III: Poster Presentations at the OMOP-IMEDS 2013 Symposium

Poster Title	Authors / Affiliations (**Bold** is presenting author)
Controlling Cost and Execution Time Using MapReduce for OMOP CDM Transformation	Slava Kolisnychenko, Alex Shkop, Don Torok, **Mark Khayter**; Ephir, Inc.
New interfaces to established methods for exploration and analysis of longitudinal healthcare data.	**Mohammad Al-Ansari**, Rave Harpaz, William DuMouchel, Michael Johnston, Leslie Hepburn; Oracle Health Sciences Global Business Unit
OpenFurther: Federating and Generating OMOP Datasets	**Ramkiran Gouripeddi**, N. Dustin Schultz, Ryan Butche, Peter Mo, Richard L. Bradshaw, Randy Madsen, Phillip B. Warner, Bernard A. LaSalle, Julio C. Facelli; University of Utah, Salt Lake City, UT
Applications of the OMOP CDM for Clinical Trial Feasibility Assessment	**Chris Knoll**, Frank DeFalco, Patrick B. Ryan; Janssen Research and Development
A system and user interface for standardized preparation of analytic data sets	**Daniella Meeker,** Christopher Skeels, Laura Pearlman, Karl Czajkowski, Lucila Ohno-Machado; RAND Corporation
Architectural Comparison of Three Healthcare Integrated Data Repositories: Quest for Data Representation Best Practices	**Vojtech Huser,** James J. Cimino; Laboratory for Informatics Development, NIH Clinical Center, Bethesda, MD
Harmonization of the OMOP CDM with the BRIDG Model	**Mitra Rocca**, Wayne Kubick, Rebecca Kush, Patrick Ryan; Food and Drug Administration, Center for Drug Evaluation and Research; Clinical Data Interchange Standards Consortium; Janssen Research and Development
Operationalizing Asthma Analytic Plan using OMOP CDM	**Elias Brandt,** Bethany Kwan, Michael Kahn, Marion Sills, Barbara Yawn, Monica Federico, Lisa Schilling; American Academy of Family Physicians, University of Colorado School of Medicine, Children's Hospital Colorado; Olmsted Medical Center; Children's Hospital Colorado; University of Colorado School of Medicine
An examination of OMOP CDM vocabulary for completeness with respect to drugs, conditions and procedures	**Hugh Kawabata**, Xiao Shao, Christopher Davidson, Perfecto Gayan, Teresa Simon; Bristol-Myers Squibb Co.
OMOP CDM Conversion in AsPEN for SCAN Project	**Yinghong Zhang**, Edward Chia-Cheng Lai, Chantelle Hardy, Li Lin, Soko Setoguchi-Iwata; Duke Clinical Research Institute

Poster Title	Authors / Affiliations (**Bold** is presenting author)
Implementing Premier Alliance hospital data into the OMOP CDM	**R Murray**, T Duan, K Castanos, D O'Hara, J Morrison; United BioSource Corporation, Bristol-Myers Squibb
Building Quality Checks and Measures into the Transformation of Data into the CDM	**Erica Voss**, Qianli Ma, Patrick B. Ryan; Janssen Research and Development
Conversion of the Premier database to the OMOP CDM	**Rupa Makadia**, Patrick B. Ryan; Janssen Research and Development
Mapping Medicaid and Medicare claims to the OMOP CDM: Lessons Learned from Three Data Sources	**Traci Yamashita;** Patricia St. Clair, Lisa Schilling, University of Colorado School of Medicine, USC Schaeffer Center for Health Economics and Policy
Transformation of the Clinical Practice Research Datalink (CPRD) to the OMOP CDM Version 4	**Amy Matcho,** Patrick B. Ryan; Janssen Research and Development
Detecting novel adverse drug reactions (ADRs) by quantitatively mapping relationships between medical conditions and drug exposures during patient's visits	David Blatt, **Nicholas Tatonetti;** Columbia University, Columbia University Medical Center
Effects of Method Parameters and Ground Truth in the OMOP Results Database	**Alan F. Karr;** National Institute of Statistical Sciences
Evaluation of Analytical Method Performance Using Self-Controlled Case Series	**Jasmanda Wu**, Ling Zhang, Patrick Caubel, Juhaeri Juhaeri; Sanofi
Investigating the Impact of OSIM2 Data Generation on Estimator Performance	**Susan Gruber;** Harvard School of Public Health
Predictive modeling	Ben Letham, Cynthia Rudin, **Tyler H. McCormick**, David Madigan; University of Washington
Statistical Method Development for Observational Comparative Effectiveness Research	**Alan Karr,** Bob Obenchain, Stan Young; National Institute of Statistical Sciences
Graphically Examining Potential Heterogeneous Treatment Effects in Binary Outcomes or Endpoints	**Paul Juneau**, Elnara Eynullayeva, Daniel Huse; Truven Health Analytics
Identification and confirmation of IPF cases in an electronic insurance claims database	**Daina Esposito**, Stephan Lanes, Gaurav Deshpande, Crystal Holick, Daniel Mines, Sean O'Quinn, Catrin Wessman, Setareh Williams, Trung Tran; HealthCore, Inc., MedImmune, AstraZeneca

Poster Title	Authors / Affiliations (**Bold** is presenting author)
Use of a multistate Markov modeling approach for investigating time-varying factors associated with lipid lowering medications for primary prevention of cardiovascular disease	**Suzanne L. West**, Barry S. Eggleston, Marianne Kluckman, Amanda Hansen, Ken LaBresh, Vincent Bufalino, Michael F. O'Toole; RTI International, Midwest Heart Specialists, Advocate Medical Group
Sensitivity Analysis of Methods for Active Surveillance of Acute Myocardial Infarction Using Electronic Databases	Xiaochun Li, Cynthia J. Girman, **Susan Ofner**, Changyu Shen, Kimberly G. Brodovicz, Linas Simonaitis, Nancy Santanello; Indiana University School of Medicine; Indiana University School of Public Health, Regenstrief Institute, Merck Sharp & Dohme
Calibrating the strength of association of drug-outcome pairs using the empirical null derived from known non-associations and known positive associations identified in the Observational Medical Outcomes Partnership	**Xiaochun Li,** Susan Ofner, Changyu Shen, Jia Zhan, Zhiwen Liu, Marc Rosenman, Nancy Santanello; Indiana University School of Medicine, Indiana University School of Public Health, Regenstrief Institute, Merck Sharp & Dohme
Extending Bayesian inference in pharmacovigilance beyond point estimates with massive parallelization	**Trevor Shaddox**, Marc Suchard; Department of Biomathematics, David Geffen School of Medicine at UCLA, University of California, Los Angeles
Application of Multivariate Self-Controlled Case Series Method for Active Drug Safety Surveillance in a Signal Screening Framework: Leveraging THIN Data in the OMOP CDM	**Xiaofeng Zhou**, Rongjun Shen, Sundaresan Murugesan, Andrew Bate; Pfizer
The Impact of Censoring Drug Switching in Medication Adherence Measures of Chronic Non-combinational Oral Drugs	**Vivienne J. Zhu**, J. Marc Overhage, Qianli Ma, Patrick B. Ryan; Johnson and Johnson Pharmaceutics, Siemens Healthcare
Mining adverse drug reactions from electronic health records	**Henry Lo;** University of Massachusetts Boston,
Applying OMOP Tools and Methods to Nursing Home Data	**Richard D. Boyce**, Jeremy Jao, Steven M. Handler; University of Pittsburgh School of Medicine
Seasonality in acute liver injury in healthcare claims data	**Rachel Weinstein**, Martijn Schuemie; Janssen Research and Development
Predicting Strokes Using Relational Random Forests	**Zach Shahn;** Columbia University
Shooting a New Target: Application of RIFLE Criteria to Identify Acute Kidney Injury Using Outpatient Laboratory Data	**Aaron Katz**, Patrick B. Ryan; University of North Carolina at Chapel Hill; Janssen Research and Development