Scientific Achievements of the IMEDS-Methods Program in 2014

Executive Summary

Pre-approval studies of regulated medical products are designed to provide clear evidence of product efficacy. However, these studies are typically underpowered for detecting adverse events (AE), especially rare ones, and they lack sufficient follow-up for detecting AEs with long induction or latency periods. Post-market safety surveillance offers an opportunity to continue accumulating evidence, without depriving people of access to products that can improve their health. Insurance claims data and electronic health records collected as part of health care delivery provide information that can be used for safety monitoring. We need to understand how to best exploit observational datasets in order to generate reliable evidence in support of post-market safety monitoring, bearing in mind that observational studies that utilize these data are subject to bias. To address this need, the IMEDS-Methods program serves public health needs by initiating and facilitating research into the methods of safety evaluation in large observational databases.

IMEDS-Methods projects carried out in 2014 addressed three challenges to obtaining unbiased estimates of risks associated with the use of regulated medical products. The first focused on improved development of case identifying algorithms to identify subjects who experienced adverse events. The second shed light on how to avoid potential pitfalls of a novel approach to calibrating p-values recently proposed in the literature. The third examined potential sources of bias and heterogeneity in study results stemming from the use of different data sources, study designs, and common data models. These projects and additional investigations in collaboration with investigators from the FDA and Mini-Sentinel are helping to inform the best uses and limitations of the FDA's Sentinel System. They also provide value to manufacturers, researchers, patients and clinicians by supplying a valid framework for assessing and interpreting risks associated with the use of regulated products. The IMEDS Research Lab continues to be a vital resource for advancing the development of cutting-edge technologies in support of public health needs. Software, tools, and datasets formatted in the OMOP and Mini-Sentinel Common Data Models support FDA, academic and industry activities throughout the United States and internationally.

1. Introduction

Pre-approval studies of regulated medical products are designed to provide clear evidence of product efficacy. However, these studies are typically underpowered for detecting rare adverse events (AE), and they lack sufficient follow-up for detecting AEs

with long induction or latency periods. Post-market safety surveillance offers an opportunity to continue accumulating evidence, without depriving people of access to products that can improve their health. Insurance claims data and electronic health records collected as part of health care delivery provide information that can be used for safety monitoring. However, observational studies that utilize these data are subject to sources of bias that can affect the reliability and interpretability of the results. We need to understand how to best exploit observational datasets in order to generate reliable evidence in support of post-market safety monitoring.

To address this need, the IMEDS-Methods program serves public health needs by initiating and facilitating research into the methods of safety evaluation in large databases. A suite of projects carried out in 2014 focused on better understanding the uses and challenges of large-scale observational studies. Section 2 below discusses the impact of IMEDS-Methods on improving methods to address bias in observational studies. Section 3 describes ongoing work of a joint IMEDS-FDA-Mini-Sentinel collaboration that is assessing the implications of activities carried out by the Observational Medical Outcomes Partnership (OMOP) from 2008 through 2013 for informing Mini-Sentinel activities. Finally, Section 4 describes how the IMEDS Research Lab is advancing the development of cutting-edge technologies in support of public health needs.

2. Addressing bias in large-scale observational studies

Large-scale electronic claims and medical records databases are an invaluable source of information on real-world treatments and medical events. Yet, in many ways these data are less than ideally suited for studying product safety. Common sources of bias include misclassification of outcomes and exposures, treatment by indication, unmeasured confounding, time-dependent confounding, and missing or censored data. The 2014 IMEDS-Methods projects addressed three facets of this problem.

> *Value to patients, doctors, manufacturers and regulators*. Large-scale observational studies are an established source of information for evaluating the safety of medical products. Scientific controversy over possible limitations of such studies has sometimes pushed the decision-making about their use to the realm of opinion and sometimes the force of personality. IMEDS has a central role in setting the evaluation of these resources in an empirically derived and consistent framework, so that public health expertise can focus on the public health questions.

2.1 Improved development of case-identifying algorithms

A first step in understanding whether a drug increases the risk for an AE is to identify patients who experienced the event (cases). Some AEs are easily diagnosed, while others are not. AEs that always require medical attention are more likely to be captured

in the data than those that do not. Case-identifying (CI) algorithms are procedures for using the recorded data to distinguish cases from non-cases. Their accuracy depends on clinical characteristics and the way in which data are collected and stored. An algorithm that works well in one context may not be accurate in another. Correctly assessing whether a particular drug increases risk for an AE is much harder when occurrences are under (or over) counted. The use of a sub-optimal CI algorithm increases the bias in risk estimates, thereby reducing the reliability of the evidence. Work sponsored by IMEDS-Methods instructs investigators on how to develop and evaluate CI algorithms. This work requires a detailed understanding of the care-seeking behavior in the study population, clinical diagnosis and treatment, and the data source. A checklist comprised of each relevant factor guides researchers through a considered design process. This work is the topic of an upcoming symposium to be held at the 2015 Annual Meeting of the International Society for Pharmacoepidemiology (ISPE). A forthcoming journal paper describing these best practices will be posted to the IMEDS website upon publication.[1]

> *Value to patients, doctors, manufacturers and regulators.* Accurate designation of health outcomes permits safety studies that are unclouded by concerns that the outcome measure itself (rather than the disease it is supposed to represent) might be associated with the use of medical products. This plausible concern arises because the amount of insurance claims or EHR data is itself correlated with use of medical products.

2.2 Limitations of empirical calibration of p-values using empirical data

When residual bias in a study result exists, it is important to understand whether it is large enough to change the implied substantive conclusion. A recent proposed approach for quantifying the likely bias, known as *p-value calibration,* was demonstrated to reduce the number of statistically significant findings in the analysis of drugs already known to be unassociated with the outcome under study (false positives). False positives are of regulatory concern because further investigation wastes scarce resources, and publicity surrounding an FDA investigation can unnecessarily reduce use or adherence. In a comparison with standard p-values, calibrated p-values exhibited superior control of the type-I error rate for studies of regulated drugs known not to increase the risk of the AE under study. How well this gain in performance generalizes to studies of newly marketed products was not discussed.

A project sponsored by IMEDS-Methods and the NIH examined empirical p-value calibration, and investigated the method's performance under violations of the underlying assumptions that may occur in practice. P-value calibration rests on the ability to identify a sufficiently large number of *negative controls*, drugs that are known

[1] S Lanes, AM Walker, J Brown, K Haynes, M Pollack. Identifying Health Outcomes in Electronic Healthcare Databases (under review at *Pharmacoepidemiology and Drug Safety*).

to be unassociated with the outcome under study, and are subject to identical unaddressed sources of bias as the exposure drug of interest. The IMEDS-Methods project underscored the importance of establishing an appropriate set of negative controls. When the controls are not subject to the same sources of bias as the exposure of interest, calibrated p-values can increase the number of false positives and provide worse type-I error control than standard p-values. In addition, even when assumptions are met, p-value calibration often reduces the power for detecting true safety signals. The implications of increasing the chance of a false negative finding, and the availability of data on a sufficient number of valid negative controls are important considerations for deciding whether to use calibrated p-values in practice.[2]

> *Value to manufacturers and regulators.* This work raises the larger issue of the role of p-values in interpreting findings from analyses of large datasets. A focus on random variation as the primary source of error produces a profoundly distorted impression of the reliability of inferences. Chance (i.e. coin-flip) kinds of errors are negligibly small in comparison to substantial uncertainties that arise from variations in data collection and representation in different care delivery systems. Incorporating bias assessment into p-values may complicate the interpretation of the results and leads to a false conclusion that a problem has been taken care of.

3. Assessing the implications of OMOP activities for the Mini-Sentinel system and IMEDS-Evaluation

In 2014 the IMEDS-Methods program launched a suite of projects to investigate how the activities of the Observational Medical Outcomes Partnership (OMOP) carried out from 2008 – 2013 can inform the FDA's Mini-Sentinel active surveillance system. These investigations highlight the importance of coupling medical and epidemiologic knowledge with an understanding of the information available in the data. Expert input is vital for formulating questions and study design. Expertise is also needed to provide meaningful interpretations to the results that are obtained.

> *Value to researchers, manufacturers and regulators.* The OMOP program performed groundbreaking assessments of the utility of commonly used data sources and research techniques. Because they moved quickly, the research community has been slow to evaluate the quality of OMOP work and the general applicability of its conclusions. This set of assessments directly informs the use of Mini-Sentinel in evaluating the safety of regulated products.

[2] S Gruber and E Tchetgen Tchetgen. Limitations of empirical calibration of p-values using observational data. (under review at *Statistics in Medicine*)

3.1 Potential sources of heterogeneity in study results

OMOP 2010 and 2011 – 2012 Experiments generated results for over 6 million studies of the relationships between selected drugs and adverse health outcomes. OMOP demonstrated heterogeneity in study results across databases, and across study design parameters. IMEDS-Methods is investigating suspected sources of heterogeneity to better understand their impact on OMOP study findings. The data sources used in OMOP studies differs from Mini-Sentinel's distributed network of data partners. Each group uses data formatted in its own common data model (OMOP CDM, MS CDM). OMOP investigated a wide variety of study designs, some of which are not typically carried out by Mini-Sentinel. In light of these differences, IMEDS-Methods was interested in understanding which sources of heterogeneity potentially affect Mini-Sentinel studies.

3.1.1 Common Data Model (CDM)

A CDM standardizes data from disparate sources. This allows a single analysis to be implemented the same way, across multiple sites, and multiple sources of electronic healthcare data. The goal in designing a CDM is to faithfully store relevant information from the original data in a consistent format. Software written to exploit this format can be used repeatedly to analyze any formatted dataset. This saves time, reduces programming errors, and facilitates comparison of study results. The OMOP and MS CDMs have much in common, but programs written for one cannot be applied to the other.

In order to understand whether the choice of CDM can impact study results, IMEDS-Methods sponsored a project to compare and contrast the OMOP and MS CDMs, which discusses their impact on the development and interpretation of analyses conducted within each data model. The resulting paper describes the underpinnings of each CDM, their commonalities, and their differences. [3] OMOP and MS developed their respective CDMs for somewhat different purposes. Each specifies a set of database tables that are linked by internal identifiers (e.g., patient demographic data is stored in one table, and can be linked to another table storing multiple records per patient, one for each visit to a health care provider). Both CDMs capture most of the same information from the source data, but they differ in how the tables are organized internally. Other differences involve the granularity and interpretation of codes indicating drug prescriptions, dispensing, and dosage, interpretation of lab values (changes in units, normal range), and the recording of death records. These and other distinctions change the way a data analyst will access the underlying data, but are unlikely to contribute to observed differences in study results. An investigation to empirically confirm this conclusion is currently underway. Comparing study findings from analyses of the same data

[3] M Weiner, P Embi, M Khan. OMOP and Mini-Sentinel Data Models : Form Fits Functions. (in preparation)

formatted in both CDMs is expected to rule out the choice of CDM as a source of heterogeneity.

> *Value to researchers, manufacturers and regulators.* The OMOP CDM supports essentially the same analyses as the Mini-Sentinel CDM. Groups that have already invested in one do not need to contemplate conversion to the other, given a goal of accurate research. The choice between these rests more on compatibility with collaborating groups and the ongoing development of software that has been tailored to one or the other CDM.

3.1.2 Data sources

OMOP demonstrated that applying a fixed study design to different databases could produce a range of results. Sources of these disparities include differences in the populations represented in each database (e.g., geriatric vs. pediatric), and patterns of care delivery among health care organizations and providers. Claims data and electronic medical records (EMR) differ in the types of patient-level information available. For example, lab values contained in medical records are absent from claims data. Mini-Sentinel distributed data partners offer a mix of claims and EMR data, allowing studies to make use of data that are best fit for purpose. OMOP findings illustrate that study results do not always generalize to other populations. They also remind us to that combining populations from multiple data sources may produce a study population with salient differences from the target population, undermining the goals of the study.

3.1.3 Study design

OMOP applied a range of study designs to estimate the association between drug-outcome pairs, with the goal of generating evidence for further study. Different study designs corresponded to different scientific questions of interests (for example, short-term vs. lifetime risk), and little attention was paid to whether a given design could produce a result with a clear, unambiguous interpretation. Heterogeneity in estimation of the association between a drug and an adverse outcome is expected when risk periods, comparator groups, or inclusion/exclusion criteria are varied. In contrast, study results were robust to the choice of tuning parameters of statistical analyses, such as the precision of the prior for a Bayesian analysis, or the number of propensity score strata for an adjusted cohort analysis.

3.2 Characteristics of study design and elements that may facilitate or challenge electronic safety monitoring systems

Obtaining unbiased causal effect estimates from large-scale observational studies is a challenging task. Nevertheless, electronic health databases are a valuable component of the Mini-Sentinel safety surveillance system. In order to make the best use of Mini-Sentinel resources, a joint work group of IMEDS, FDA, and Mini-Sentinel investigators took on the task of characterizing elements of the scientific question, data, and study design that contribute to producing robust study findings. The effect of ACE Inhibitors on angioedema has been discussed extensively in the literature, and studies have been carried out by both OMOP and Mini-Sentinel. The work group used this drug-outcome pair as a platform to discuss characteristics of the drug, the disease, and the drug-disease association that contribute to successful monitoring. These include widespread use of the drug, abrupt onset of disease requiring medical attention, and a strong association between the drug and the outcome. Some factors that are expected to facilitate electronic safety monitoring were not present in this example. These include accurate recording of periods of exposure and non-exposure, and an exposed population that has few co-morbidities or background drug exposures.[4] The existence of many of these factors increases confidence in the reliability of the evidence.

The group is currently working to characterize additional elements of scientific questions that cannot be readily answered using data and study designs available within the Mini-Sentinel system.

> *Value to regulators and manufacturers.* Characterizing examples of assessments that appear to be reliable and ones that appear to be wrong in analyses of large data sources sets the stage for general guidelines as to when to use resources such as Mini-Sentinel routinely, and when they can be used for research-style assessments. They also provide a starting point for experience-based guidelines on when such sources are likely to mislead.

4. Advancing technology in the IMEDS Research Laboratory

The IMEDS Research Lab supports investigators from FDA, NIH, industry, and academia in testing and developing methods to support post-market safety monitoring. The Lab is one of the few venues available to qualified investigators seeking access to large administrative health databases. Offerings include data, software, scalable computing resources, and remote access through any web-browser. Investigators in the IMEDS Lab support one another through a community bulletin board, and present their work to the

[4] C Bell, A Chakravarty, S Gruber, S Heckbert, M Levenson, D Martin, J Nelson, S Pinheiro, B Psaty, C Reich, S Schneeweiss, A Shoaibi, S Toh, A Walker. Commentary - When can electronic drug safety succeed? Pharmacoepidemiology and Drug Safety (2014).

broader community through the monthly IMEDS Community Call webinar series. Several investigators will be presenting their research at a session on Novel Computational Approaches in Safety Surveillance at the next annual Joint Statistical Meetings (JSM, 2015), organized by IMEDS-Methods.

The cutting-edge activities of researchers in the Lab include developing methods for adjusting for time-dependent confounding and unmeasured confounding, using machine learning to improve risk prediction, profiling patient populations to inform clinical trial design, and applications of data-adaptive techniques to adverse event detection. IMEDS-Methods' support of early-stage research has been leveraged by Lab investigators in applications for substantial NIH funding.

The IMEDS Lab will also play an important role in industry-sponsored studies run by the IMEDS-Evaluation program. The Lab will serve as a cost-effective test platform prior to running queries in the IMEDS distributed data network. Furthermore, the Mini-Sentinel operations center is considering using the IMEDS Lab to expand the variety of test platforms before distributing source code to data partners in its distributed network.

> *Value to patients, doctors, manufacturers and regulators.* Large data applied to public health are dangerous in inexperienced hands. The IMEDS Lab is helping the nation create a cadre of investigators with technical skills combined with the directly experience of both successful and overly ambitious approaches to the massive information that will inevitably become a cornerstone of drug evaluation.