OUTCOMES RESEARCH USING CLAIMS DATABASES: A CRITICAL REVIEW AND CASE STUDY

By

Eric T. Wittbrodt

A Master's Paper submitted to the faculty of
the University of North Carolina at Chapel Hill
in partial fulfillment of the requirements for
the degree of Master of Public Health in
the Public Health Leadership Program.

Chapel Hill

2013

_____

Advisor: William A. Sollecito, DrPH

_____

Second Reader: Robin S. Turpin, PhD

_____

Date

ABSTRACT

Eric T. Wittbrodt: Outcomes Research Using Claims Databases: A Critical Review and Case Study
(Under the Direction of William A. Sollecito, DrPH)

Outcomes research (OR) is an evolving and critical part of health sciences research due to its focus on the improvement of patient care in the practice setting. Evidence included in OR spans all available clinical information, from data generated by randomized controlled trials to real-world clinical, economic, humanistic, patient-reported, and patient satisfaction data. Comparative effectiveness research (CER) is a branch of OR that compares the overall benefit or cost of two active interventions. The use of claims databases has emerged as a valid, rigorous, and efficient source of data for CER. Due to the inability to control for all confounders, techniques such as propensity score matching and quality checks are necessary to maintain data accuracy and validity. A case study comparing two oral medications used in the treatment of gout will be used to illustrate the techniques, advantages, and disadvantages of using claims databases in OR.

TABLE OF CONTENTS

LIST OF ABBREVIATIONS AND TERMS

ACO	accountable care organization
adherence	the act of a patient following an agreed-upon medication regimen
AHRQ	Agency for Healthcare Quality and Research
AMCP	Academy of Managed Care Pharmacy
CDC	Centers for Disease Control and Prevention
CER	comparative effectiveness research, the process of comparing in the aggregate the overall benefit or cost of two active interventions (typically pharmaceuticals or medical devices); evidence would include randomized controlled trials (RCTs), economic analyses, observational studies, registries, patient-reported outcomes (PROs), and others
CKD	chronic kidney disease
claims databases	existing information that has been captured and stored in electronic form for the primary purpose of processing payment claims for healthcare services rendered
CMS	Center for Medicare and Medicaid Services
confounder	a variable inherent in a population being studied that contributes to random error and obscures the effect of the intervention being tested on the outcome of interest
CONSORT	CONsolidated Standards Of Reporting Trials
CPT	Current Procedural Terminology
DACON	daily consumption (of medication)
DHHS	Department of Health and Human Services
EBM	evidence-based medicine
effectiveness	impact of an intervention on a variable of interest that is demonstrated in routine clinical practice or a “real world” setting

efficacy	impact of an intervention on a variable of interest in a highly restricted and limited fashion for a discrete span of time; most closely associated with data generated by randomized controlled trials (RCTs)
EHR	electronic health record
ESRD	end stage renal disease
expenditures	financial resources allocated for the payment of a good or service
FBX	febuxostat, an oral drug approved for treating hyperuricemia in patients with gout
FDA	Food and Drug Administration
FY	fiscal year
GE	General Electric Healthcare database
gout	a form of painful joint inflammation caused by the deposition of uric acid crystals in the lubricating fluid of the joint space
HCPCS	Healthcare Common Procedure Coding System
health outcome	the quality of patient care based on the result rather than the input
HIPAA	Health Insurance Portability and Accountability Act (of 1996)
hyperuricemia	elevation in serum acid concentration beyond the normal range
ICD-9-CM	International Classification of Diseases, 9th Revision, Clinical Modification
IDN	Integrated (Healthcare) Delivery Network
IOM	Institute of Medicine
ISPOR	International Society for Pharmacoeconomics and Outcomes Research
MDD	major depressive disorder
MPR	medication possession ratio

Meaningful Use	the clause of the Patient Protection and Affordable Care Act (PPACA) that provides CMS-based incentives and guidance for health care providers and facilities to promote the use of electronic health records (EHR)
NDC	national drug code
NIH	National Institutes of Health
OR	outcomes research, the study of the end result of health services that takes patients' experiences, preferences, and values into account and is intended to provide scientific evidence for the purpose of shaping decisions made by all who participate in health care
PCORI	Patient-Centered Outcomes Research Institute
pharmacoeconomics	the subset of health economics that focuses on the cost of pharmacotheraepeutic interventions and the services that support them
PHI	protected health information
PPACA	Patient Protection and Affordable Care Act (of 2010)
PRO	patient-reported outcome
propensity score	a nonparametric method of patient matching for the purpose of mitigation of random error in claims database studies in which a probability of receiving a given intervention is assigned to each patient; the number ranges from 0 to 1 and is derived from all measured covariates
PV	pharmacovigilance, post-marketing research in which the ongoing surveillance of real-world information has a specific emphasis on drug safety
QIO	quality improvement organization
quality	the benefit of an intervention or system that is directly measurable with discrete and validated instruments and can be repeatedly assessed to gauge improvement
RCT	randomized controlled trial
real-world	routine clinical practice in an uncontrolled environment in a diverse patient group

STROBE	STrengthening the Reporting of OBservational studies in Epidemiology
triple aim	improved quality, reduced cost, and improved patient experience of health care encounters, a primary driver of value-based health care
ULT	urate-lowering (drug) therapy
US	United States
value	high quality care that produces the best outcomes in a wholly cost-conscious manner; calculated as the clinical benefit of an intervention divided by its total cost
value-based care	a network of hospitals and providers that are united by a shared responsibility for improved patient outcomes and enhanced fiscal accountability, examples include accountable care organizations (ACOs) and integrated healthcare delivery networks (IDNs)
XOI	xanthine oxidase inhibitor
WHO	World Health Organization

Chapter 1: OUTCOMES RESEARCH

Introduction

Health sciences research, especially that which is conducted by the pharmaceutical, diagnostic, and medical device industries, may take many forms, but all share a common purpose: to determine the intervention (typically a drug, device, procedure, diagnostic tool, or delivery system) that yields the best possible health outcome for individual patients or for a specific population of affected individuals. Traditional research methods include randomized controlled trials (RCT) in which an active intervention is directly compared with a placebo, sham, or active control in two or more groups of subjects, usually in parallel fashion. The RCT is usually prospective in nature, and is designed to occur in controlled environments such that the impact of potential interfering factors, or confounding variables, on the outcome of interest are minimized. Due to regulatory requirements and their generally accepted scientific rigor, RCTs have continued to occupy a pivotal role in health sciences research. In addition, the adoption of evidence-based medicine (EBM) is rooted in the development of care protocols and guidelines that rely on information gleaned from the findings of RCTs often coupled with expert opinion. But, there remains an unmet need for data generated from the use of interventions in less well-controlled environments. Such "real-world" studies focus on the evaluation of health outcomes in the patient care system. For example, the lack of robust and long-term safety data in many RCTs has prompted the US Food and Drug Administration (FDA) to mandate pre- and post-marketing safety studies with specific types of interventions, such as oral antidiabetic drugs, or in selected groups of patients, such as individuals with a high risk of coronary vascular events. The science of pharmacovigilance (PV), as will be discussed, is another example of post-marketing research in which the ongoing surveillance of real-world information has a specific emphasis on drug safety. This is also an area where public health and individual health care intersect, focusing on the prevention of adverse events and other patient harms, occupying a

broader scope than traditional RCTs, and often including large populations or communities followed over long periods of time.

The concept of "outcomes research" (OR) in the health sciences is fairly recent, and was initially commonly discussed in the 1990s. Prior to that, research methods documented early evidence that aseptic technique reduced postsurgical mortality and the impact of battlefield casualty treatment on outcomes in soldiers during the Crimean War and World Wars I and II. In a formative paper by Donabedian (1966), outcomes in a medical context were defined as the "measures of the quality of care." The limitations of outcomes assessment in this same paper included the caveat that outcomes may not always be relevant versus the processes employed to achieve them, and that many factors that influence outcomes must be taken into account. Ellwood, in the 1988 Shattuck lecture to the Massachusetts Medical Society, was likely the first individual to publicly mention the phrase *health outcomes* in the context of measuring the quality of patient care based on results rather than inputs. He also envisioned a nationwide repository of information where patients and providers could compare the quality and cost of interventions and delivery systems. While these observations were prescient, an entire decade elapsed before the systematic investigation of the impact of interventions on health outcomes took shape. The phrase "outcomes research" is believed by many to have been coined in print by Clancy and Eisenberg (1998) and was defined by these authors as "the study of the end results of health services that takes patients' experiences, preferences, and values into account…intended to provide scientific evidence relating to decisions made by all who participate in health care" (p. 245). In other words, the drivers of OR are not derived solely from an intervention or provider, but rather are also dependent upon the attributes and needs of the end user, the patient. In 2013, the definition has expanded to include any research activities that seek to better explain, define, or confirm the clinical application of an intervention, especially in comparison with other interventions (including no intervention) in the presence of the

patient-related aspects described above. Also, the set of factors that affect patient outcomes has grown to encompass those that relate to the patient's provider, payer, health care delivery system, geographic location, and social support. This ever-growing list is consistent with the vast expansion of the types of data that are available for a patient's health care encounters that will be discussed later in this paper. In addition, all of these items comport with the well-described social determinants of health, including individual behaviors and characteristics (demographics, genetics, socioeconomic status, educational level, and occupation), interpersonal or social interactions, community conditions, and political or access factors that impact outcomes (CDC, 2011). The discipline of OR, then, helps cast a wider view of the impact of interventions beyond *efficacy*, that effect on a variable of interest in a highly restricted and limited fashion for a discrete span of time that is most closely associated with data generated by RCTs. This is contrasted with *effectiveness*, in which this same effect is demonstrated in a larger and more diverse cross-section of patients in routine clinical practice. Interest in identifying the true clinical merits of two or more interventions in a population has created innovative subtypes of OR including comparative effectiveness research (CER). This paper will address the growing infrastructure to support CER in the US and also some of the misperceptions around its utility and that of OR in general. Some of the primary applications of OR in the realm of public health also will be discussed. Due to the dire need for better evidence to build upon RCTs, the increased availability of vast amounts of health-related data often in electronic form, and enhanced techniques that facilitate rapid and efficient analysis and interpretation of such data, the field of OR is more timely and relevant than ever. Novel data sources, such as secondary databases, offer great versatility to OR investigators. As a result, decision-makers are better equipped, and patients and providers are more empowered. Striving for the "triple aim" of increased quality (with a favorable risk-benefit profile), reduced cost, and improving the patient

experience is now the dominant trend in US health care (Berwick, 2008). Outcomes research is an important tool for the fulfillment of this aim.

1. Goals

   The goals of this paper are as follows:

   a. Explain the concept and branch of science that is OR through a description of its attributes, purpose, and relevance to both individual health care practice and public health.

   b. Describe comparative effectiveness research (CER) and distinguish it from other types of OR, especially in the context of recently implemented health care legislation, including the Patient Protection and Affordable Care Act (PPACA).

   c. Explain the methods and data sources used in OR with a focus on secondary data sources, and administrative claims databases in particular.

   d. Illustrate the applications, advantages, and drawbacks of claims database research when investigating an OR question through an example case study design and available data sources comparing two different drug therapies for gout.

2. The Role of Outcomes Research in Health Care

2.1 Relevance to patient care and public health

RCTs are the gold standard for the evaluation of efficacy, but stakeholders who require insight into the effectiveness of treatments in a larger and more diverse group of patients can no longer rely solely on RCTs for guidance. The availability of everyday clinical experience with an intervention is lacking in a

system fraught with variable prescribing patterns and medically complex patients (often from demographic groups vastly underrepresented in RCTs) (Schneeweis & Avorn, 2005) with inconsistent health care coverage and intermittent access to treatment. Such information is essential for well-informed decisions about care. The decisions may affect individual patients, e.g., treatment protocols and prescription coverage, or populations, e.g., access to care and reimbursement. Conducting carefully designed and thoughtfully executed retrospective OR studies using secondary claims databases is one common approach to augment this body of knowledge. Since OR is a relatively new field some experts have expressed disagreements amongst themselves about its scope, relevance, and purpose. For example, Krumholz (2008) attempted to debunk common "myths" about OR in a recent editorial. He suggested that this type of research is not monolithic in its approach to scientific inquiry and therefore is not an individual field of study, rather it is multidisciplinary by design. The applied health sciences are traditionally at its core, but also techniques from the fields of statistics, epidemiology, and the social and behavioral sciences (among others) comprise essential components of outcomes investigations. Krumholz also posited that OR must be aligned with existing health care problems in the presence of a feedback loop that contributes to the impetus and blueprint for subsequent investigations. The interconnectedness, dynamism, and progressively building characteristics inherent in OR explain its growing relevance and applicability to vexing and contemporary health care issues.

In terms of public health, OR has been instrumental in defining areas of need for intervention programs in multiple disease states such as heart failure and myocardial infarction (Roger, 2011). Such conditions contribute tremendously to the clinical and economic burden of the patients affected and the health care system in which their diseases are managed. Identification of needs in at-risk or vulnerable populations based on measured health disparities is often rooted in geography, socioeconomic status, and race, is an important function of OR. The population health aspects of this type of approach illustrate the synergistic relationship between epidemiology, health education, and OR principles.

Another example of the intersection of OR and public health is in the aforementioned science of pharmacovigilance (PV). The numerous withdrawals of drugs from the US market due to excessive safety signals prompted a growing need for the continuous monitoring of both anticipated and unforeseen adverse events in the population exposed to an intervention of interest. The World Health Organization (WHO) included these monitoring activities in its definition of PV, and extended it to include the interpretation of drug safety data in an effort to develop programs to prevent further harm (WHO, 2013). The FDA also uses PV information to determine if a prospective trial is needed to assess the long-term safety of a product in clinical use. The use of OR techniques in PV activities is essential, because both depend on the evaluation of routinely collected data in the everyday setting.

2.2 Evolution of pharmacoeconomic research to outcomes research

The branch of study called health economics is dedicated to the evaluation of the financial burden or cost of health care services. Pharmacoeconomics is the subset of health economics that focuses on the cost of pharmacotheraepeutic interventions and the services that support them (Bootman et al., 1996). The historical function of pharmacoeconomic data is to support informed decisions about the allocation of resources for various drug therapies. Such data are not intended to be considered in isolation, but rather in conjunction with clinical evidence for efficacy, safety, and other noneconomic factors. The budgets for the various components of a health care system or payer, e.g., hospitalization, ambulatory care, pharmacy, radiology, mental health services, and clinical laboratory services, have traditionally occupied distinct and separate compartments or "silos." The department that is held accountable for its own budget separate from the whole has not usually been concerned with the overall economic impact of an intervention to the system. This mindset has undergone a significant shift as the quality of care delivered by the entire system begins to drive reimbursement for health care services. Thus, decision-makers who focus solely on cost of an intervention, especially in the current era of high budgetary

restraint, demonstrate short-sightedness, because the cost of care extends far beyond this one item. One must consider all of the downstream cost implications of an intervention in a holistic fashion. Consider an example where the cost of drug A may be twice that of drug B, but the collective evidence demonstrates three fewer hospitalizations for disease recurrence with drug B. The cost of each hospitalization is determined to far exceed the difference in cost between the therapies. In this example, neglecting to consider the impact of an intervention on the total cost of care may positively affect the drug budget (due to selection of the less expensive agent), but have the reverse effect on the health care system overall. To be wholly cost-conscious, but also deliver high quality care that produces the best outcomes, is known as the pursuit of *value*. More specifically, value is defined as the clinical benefit of an intervention divided by its total cost (Chassin & Galvin, 1998). Clinical benefit is evidenced by statistically or meaningfully significant improvement in the health status of a patient after a specific intervention. The positive change in health status is denoted by a measured variable or variables of interest that ultimately is expected to yield better outcomes for the individual. The variable can be an outcome itself, such as disease-free survival, or can be a surrogate outcome, such as a decrease in blood cholesterol concentration that is correlated with a significantly decreased risk of heart disease.

Value is not identical to *quality*, because the latter is the benefit that is directly measurable with discrete and validated instruments and can be repeatedly assessed to gauge improvement. Another definition of quality involves the patient experience. That is, the twin facets of efficient delivery of care and effective management of disease are inextricably linked, and the issue of access to quality care becomes paramount in this definition (Campbell, 2000). From the increasing use of these terms (value and quality), one may conclude that over time, the evolution away from a strictly economic focus has occurred in response to often valid criticism that overarching financial concerns needlessly led to worse patient outcomes in many cases. In response to this, the health care environment is rapidly taking a more comprehensive approach to the economic burden of illness to include not only the cost of medical

care (known as direct costs), but also the costs of lost work productivity and time spent in the health care system, and the cost of shortened lifespan and reduced quality of life (known as indirect costs).

Moreover, the demonstration of better outcomes for the health care dollars spent is now being demanded by multiple sectors, including payers, employers, and health care systems. In fact, much of the impetus for the PPACA, aside from the vast numbers of uninsured Americans, is the fact that the US health care expenditure per capita on health care was in excess of $7,500 US in 2008 (Kaiser Family Foundation, 2008). This amount is more than for any other nation, yet overall health outcomes are far from desired (the US is ranked 36th for life expectancy among the world's nations) (WHOSIS, 2009). Although multiple factors may be implicated in suppressing life expectancy at birth, it is a disturbing trend that continues to worsen despite the economic largesse dedicated to US health care. Growing acknowledgement of this gap in cost and quality necessitated the transformation of pharmacoeconomics from a prominent decision-making tool to become but one facet of a multipronged approach that also includes consideration of clinical, humanistic, and, now, patient satisfaction data. The sum total of these approaches can be colloquially thought of as OR.

2.3 Comparative effectiveness research (CER) and the PCORI initiative

Strictly isolating the assessment of patient benefit to clinical efficacy in a controlled setting such as an RCT perhaps coupled with cost minimization is an approach now being questioned by many sectors of the health care system. This is due to the realization that the quality of care in ordinary care settings and the patient experience represent glaring omissions from this equation. Adoption of effectiveness as a critical facet of evaluating new interventions reflected a sea change in thinking about this task. A new phrase, *comparative effectiveness research (CER),* was invoked by the Institute of Medicine (IOM, 2007) and others to describe the process comparing in the aggregate the overall benefit or cost of two active interventions (typically pharmaceuticals or medical devices). The body of available evidence used in CER

evaluations includes RCTs, economic analyses, observational studies, registries, patient-reported outcomes (PROs), and others. This list continues to expand as the understanding of scientific methods in CER improves and different types of data become more readily available in electronic form. For example, the impact of disease management programs within a health system can be captured if a connection exists for electronic health records (EHR) between the inpatient and outpatient facilities, as in many systems. Documentation of specific or overall health aspects of a patient's care outcomes and experience with health care delivery can be recorded using PRO tools usually in web-based and other electronic formats. This information can be linked to the EHR for a particular patient and incorporated into a CER analysis. The PROMIS tool developed by the National Institutes of Health (NIH) is one example of a PRO (NIH, 2013). The highest utility of PROs occurs when patients are engaged in their development and testing before implementation; this helps achieve one of the hallmarks if CER: increased patient-centeredness.

The enactment of the PPACA in 2010 ushered in many stepwise changes to the US health care system, particularly in the areas of health care financing, reimbursement, and delivery. However, one feature of the PPACA not widely appreciated is the creation of the Patient-Centered Outcomes Research Institute (PCORI), a nonprofit corporation whose defining purpose is to identify knowledge inadequacies in health care in an effort to systematically foster the design and execution of evidentiary analyses (Clancy & Collins, 2010). Although PCORI is a non-governmental entity, its financial support is provided by the general fund appropriated by Congress as a function of the PPACA and also by an annual fee of $2 per person for all individuals covered by Medicare and commercial (including self-insured) insurance plans. This support will continue through the end of FY 2019 (PCORI, 2013). The ultimate goal of PCORI is to better equip health care decision-makers with more robust and well-considered information based on the totality of clinical, economic, humanistic, and patient-centered data that are available. If significant gaps exist in these data, then PCORI is charged with the facilitation of additional research to bridge such

knowledge deficits. PCORI is closely aligned with the Agency for Health Care Quality and Research (AHRQ) and the NIH both of which will assist in the rapid dissemination of research results from investigators who are supported with grants awarded by PCORI. The development of the research plan for PCORI (called the National Priorities and Research Agenda) occurs through assessment of US public health needs, disease state prevalence and burden, and solicitation of expert opinion and public comment, including those from patients and consumers. The public source of funding for PCORI contributes to its mission to address the most pressing CER-related issues that affect a large number of US citizens. The current agenda consists of five broad categories:

1. Prevention, diagnosis, and treatment options;
2. Healthcare system improvement;
3. Communication of research;
4. Health disparities; and
5. Fostering patient-centered OR and methods innovation (PCORI, 2013).

To date, PCORI has awarded 71 research grants totaling $114 million to investigators in 35 states and the District of Columbia. In sum, the establishment of PCORI and the financial investment in patient-centered OR represents an important juncture in US health care, because it is redefining the types of projects that are relevant to advances in the field. More importantly, redirection of the focus to the recipient of care, the patient, is moving to the forefront of PCORI-sponsored research which will help set the tone for much of health-related research in the future. This is a major departure from the status quo of health care decision-making, because the expanded breadth of research data emanating from PCORI and other purveyors of CER adds a much-needed dimension to the evidence-based process. In this manner, PCORI has clearly defined and resourced the development, validation, and promotion of PROs that extends far beyond what had been done before. For example, the creation of PROs in major depressive disorder (MDD) reflects a patient-centered approach to assessment of the effectiveness of

antidepressant medications, and is a useful supplement to traditional clinician-based rating scales for MDD. Other conditions which lack objective clinical endpoints, such as irritable bowel syndrome and chronic pain syndrome, also are amenable to research that includes tools used in CER.

2.4. Design, Data Sources, and Methods Considerations in Outcomes Research

2.4.1. Advantages of Claims Databases

One increasingly emergent trend in OR is the evaluation of existing information that has been captured and stored for the primary purpose of processing payment claims for health care services rendered. The employment of these secondary or administrative databases is a distinctly useful method. Many of the benefits of these databases, such as greater efficiency, lower expense, and reduction of bias have been described previously (Hall et al., 2011) and often outweigh their limitations, including incomplete or missing data and a lack of integration of different types of claims data (medical versus mental health versus prescription drug claims). This positive balance of attributes affords investigators a tool with great versatility and convenience. The lack of head-to-head active treatment comparison trials, especially with interventions past their patent expiration and many newer agents, can be addressed by evaluating multiple clinical and economic effects of interventions in a real-world environment using database studies. Evolving research design and analysis techniques for claims databases and the minimization of bias coupled with the rapid availability of results collected from patient encounters in a realistic setting have all combined to increase its popularity and acceptance in the scientific community. The “Meaningful Use” clause of the PPACA provides CMS-based incentives and guidance for health care providers and facilities to promote the use of electronic health records (EHR) (DHHS, 2013). This initiative builds upon the existing infrastructure of continuous quality improvement in Medicare, and the initiation of the Quality Improvement Organization (QIO) program, in 2002. The QIO program

coordinates the collection and housing of health care data for Medicare patients across the country in an effort to improve care. The data are collated and analyzed by the QIO and reported back to the provider with a focus on improving the delivery and outcomes of health care (Schenck, 2013). As such, the QIO program provides an example of how claims databases have been successfully used to answer relevant research questions and drive the improvement of care. Multiple QIOs endeavor in a collaborative manner to standardize procedures for data sharing and analysis. This approach has been demonstrated to be a valid means of generating evidence targeted at improving quality in patient care systems. As health care data become more complete in electronic form and disparate sources of information can be connected for an individual patient (also known as integration), the cost- and time-intensive nature of this type of research will become more attenuated. In turn, the accessibility of more comprehensive datasets to a greater number of investigators will drive a demand for knowledge of the best practices for the execution and interpretation of OR using these claims databases.

Scientific and professional associations such as the International Society for Pharmacoeconomics and Outcomes Research (ISPOR) (Motheral et al, 2003) and the Academy of Managed Care Pharmacy (AMCP) have developed and promulgated guidance documents for researchers interested in conducting analyses of claims databases. Commonly used and validated study designs from the field of epidemiology are used in claims database research and include retrospective cohort, prospective cohort, cross-sectional, case-control, and case-crossover (Aschengrau & Seage, 2008). The retrospective cohort design is especially attractive for this type of analysis, because it relies upon existing data and can be carried out relatively quickly and inexpensively. Also, databases can be effectively "de-identified," that is, devoid of any personal information that can be traced back to an individual, thereby satisfying institutional policies and good research practices for safeguarding the privacy of protected health information (PHI) as stipulated in the Health Insurance Portability and Accountability Act of 1996 (HIPAA). Specifically, in 2013 the US Department of Health and Human Services released additional

guidance regarding the protection of PHI in the so-called Final Rule which enlarged the scope of HIPAA protections beyond the "covered entities" in the original law to include any firm that "creates, receives, maintains, or transmits PHI." (GPO, 2013). Entities that conduct research with or provide access to PHI within databases are clearly covered under this rule, and it is incumbent upon such groups to develop, implement, and test systems to protect identifiable PHI. The creation of mobile applications where access to such information is readily available to patients, providers, and third parties has outpaced regulations to address potential security breaches (Wang, 2013). This is a pivotal issue in health services research, particularly using claims databases, that requires additional guidance.

The robustness of a data sample can be enhanced by including information from multiple sources pooled into a single database, and most large national health insurance payers, for example, offer the ability to sufficiently power a comparison to detect outcome differences between two or more interventions even in a patient population narrowly defined by specific characteristics, e.g., demographics and co-morbid conditions. Other proprietary databases, such as those owned by GE Healthcare and Truven, contain data from multiple payers, including the Center for Medicare and Medicaid Services (CMS), in order to capture both private (commercial) and public (government) payer data. This mix of healthcare financing mechanisms is quite desirable because it reflects a representative cross-section of US patients, and is likely to include data from across the spectrum of age and income. This fulfills one major advantage over RCTs which is the inclusion of a diverse sample of individuals with varied access to health care and myriad disease histories. Large databases also offer researchers sufficient sample sizes to evaluate rare diseases for which prospective RCTs are unfeasible due to the lengthy study period necessary for such investigations.

2.4.2. Disadvantages of Claims Databases

Several limitations of claims databases must be acknowledged (van Walraven & Austin, 2012). The overriding concern arises from the immutable fact that these repositories exist for the primary function of the collection, adjudication, and storage of insurance claims. Use of this information for research purposes is secondary in nature, and much of the controversy around database research is linked to this fundamental problem. This realization has prompted a number of enhancements to claims databases in an attempt to provide a better platform for investigational work. For example, the integration of medical and pharmacy claims data, and, in turn, further coupling with patient-specific EHR can greatly streamline the research process. The risk of bias by unmeasured confounders is inherent in database research; this can be minimized by careful identification of the comparison groups so that they are well-matched by potential confounding characteristics. If only commercially insured patients are included, for instance, then selection bias based on access to health care and income may be present. This type of bias may not be avoidable if the data are limited as such; however, this fact must be acknowledged as a flaw by the investigators. The dataset can be enriched with patients insured by Medicaid, the Veterans Administration, or Medicare if a more economically diverse sample is desired. Continuous enrollment for a specified period of time, such as one year, to qualify as an eligible patient for inclusion into a database study, overcomes some of the concern about intermittent coverage. Such a requirement is recommended, because the annual turnover or "churn" of managed care enrollees in a single plan has been reported to be as high as 40% (Short, 2003). However, the assertion of a "healthy patient bias," a specific type of selection bias, has been implicated in some commercial insurance claims databases, e.g., United Health Care, because long-term privately insured patients have better access to care and tend to seek medical attention earlier, have fewer uncontrolled chronic conditions, and engage in healthier practices overall. As evidenced by the health care reform debate, uninsured lower-income Americans

usually delay addressing chronic conditions and seek emergent treatment only after disease severity has become untenable.

Access to a rich source of data often creates a temptation for investigators to "mine" the information and search for significant associations *ad libitum*. This is a common technique in market research but is unacceptable when deriving conclusions which are applied to patient care and public health decisions that can profoundly impact the lives of real people. However, data mining is often useful in exploratory studies in which the principal thrust is to generate hypotheses or questions to drive future research (Berger, 2009). In any event, adherence to scientific standards of hypothesis-testing database research requires the development of an analysis plan *a priori* that is followed through to completion after the data are collected. In this manner, convenient and *post hoc* analyses are strongly discouraged in that "fishing" for significant associations can be justly alleged. This plan is transparently communicated to reviewers of the protocol (for funding and patient protection) and to the peer community in the study report and any publications that result.

Other analytic considerations are worth mentioning. Since retrospective database analyses are not randomized and unmeasured confounding variables are impossible to completely eliminate (residual confounding), especially in real-world settings, investigators have utilized some techniques to mitigate random error. Propensity score matching is a nonparametric and commonly used method for this purpose in which a probability of receiving a given intervention is assigned to each patient; the number ranges from 0 to 1 and is derived from all measured covariates (Cox, 2009). Patients with identical or nearly identical propensity scores who receive different treatments are then matched for comparison. In other words, propensity score matching controls for undetected confounding baseline characteristics inherent in the data (Rubin, 1997). This is a more sophisticated and thorough method of cohort matching than traditional procedures that compare similar patients based on individual features such as age, sex, co-morbidity, or race/ethnicity.

2.4.3. Other Considerations

The specific classification and identification of health care encounters by diagnosis and intervention (procedure, drug, or device) is a universal method of capturing a paid claim for health services. Diagnostic codes, e.g., the International Classification of Disease – Ninth Revision, Clinical Modification (ICD-9-CM) (CDC, 2013) and procedure codes, e.g., Current Procedural Terminology (CPT) (AMA, 2013) are well-accepted and commonly used data points in retrospective claims analyses. These coding systems along with the National Drug Code (NDC) (FDA, 2013) have been harmonized across the US health care system to facilitate consistency between care providers, payers, and systems. The availability of this information is highly amenable to multiple types of observational research. However, the congruence between a patient's diagnostic code in the claims record and the clinical information found in the medical record (EHR) at the point of care (the patient's chart) may be legitimately brought into question (Jollis et al., 1993). Therefore, the internal validity of the data collection process can be tested by comparing the two. For example, the set of ICD-9-CM codes assigned to the five stages of chronic kidney disease may be validated by taking a sample of patients with linked records (between claims and EHR) and verifying that the clinical measures of kidney function, such as serum creatinine or estimated glomerular filtration rate (eGFR) corroborate the former. In this manner, a vital quality control mechanism is fulfilled. Other routinely performed "quality checks" are necessary to determine the usefulness of the data (Johnson et al, 2009). For example, missing values are commonplace, especially among prescription claims records (Lauffenberger et al., 2013), and methods such as imputation or censoring of missing data must be transparently reported. Also, duplicate values and changes in disease coding and insurance coverage may compromise data validity.

Issues relevant to reporting the results of retrospective database research have been described. Due to the large sample sizes of patients included in these analyses, statistically significant but numerically small between-group differences may be over-interpreted. For example, consider a database analysis of

100,000 patients with Type 2 diabetes that is powered to detect differences in blood glucose concentrations of at least 1 mg/dL between two different treatments. The clinical importance of this small change in the measured endpoint is dubious where the threshold for significance may even be within the measurement error of the test (which may be as high as 10%), such as a laboratory value or a physiologic measurement like heart rate or blood pressure. While the risk of finding a significant difference between groups when none truly exists (known as a Type I error) may be very small (<5%, thereby satisfying statistical norms), the definition of a detectable difference needs to be rooted in practice-based reality. This criterion becomes more germane when the results of a retrospective database analysis conflict with previously conducted RCTs. The investigators of the former are then obligated to discuss reasons for this discordance and provide specific recommendations for interpretation of the data. Attempts to codify the proper description of OR results have recently occurred. The European Science Foundation convened an expert panel of researchers who developed the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) guidelines that were published in 2007 (von Elm, 2007). Modeled on the CONSORT guidelines for the reporting of RCT results (Schulz, 2010), STROBE supplied a checklist of 22 items for investigators to consider as "essential for good reporting of observational studies." The domains include the title and abstract, introduction, methods, results, discussion, and funding information. Of note, the key results with objectives fulfilled, limitations, interpretation, and generalizability of the study are denoted as core components of the discussion section, arguably the most revealing information in a research report. The underlying assumption of STROBE is that observational investigators are availed with ever increasing amounts of data from large numbers of patients, and it is incumbent upon this community to proceed with the same rigor and transparency as RCT investigators. Due to these enhancements in the conduct and reporting of OR studies, greater weight is being placed on the evidence that is generated. Notwithstanding the limitations of OR studies enumerated earlier in this paper, more credence than ever is placed in

observational study results, and the gravity of decisions made on their bases increases. Thus, it is imperative that the scientific, health care, and public health arenas must have the ability to attach the same reasonable certitude to the findings of observational studies as they do to RCTs.

What follows is an example of an OR study that employs many of the techniques described above to assess the overall health economic impact of two pharmacotherapeutic interventions for hyperuricemia in gout patients.

Chapter 2. Case Study: A Comparative Evaluation of the Health Economic Outcomes of Febuxostat and Allopurinol in Gout Patients with Chronic Kidney Disease

1. Disease State Background: Gout

Gout is a form of arthritis (joint inflammation) whereupon supersaturation of uric acid leads to crystal deposition in the lubricating fluid of the joint space (synovial fluid). This in turn activates the recruitment of specific types of white cells to the site, and the release of chemicals from these cells produces a localized, but intensely painful, inflammatory response (Schmerling, 2012).

The management of a gouty attack (also known as acute gout) involves mitigation of the inflammatory response with short-term drug therapy. After the flare resolves, with or without drug therapy, then uric acid lowering treatment (ULT) is indicated for an indefinite period to maintain a normal serum uric acid concentration, defined as less than 6 mg/dL (Khanna, 2012). Current options to treat excess uric acid include drugs that inhibit its enzymatic production, known as xanthine oxidase inhibitors (XOI).

The case study in this paper will focus on the XOI class in particular. Differentiating characteristics exist between the two marketed XOI, allopurinol and FBX. Allopurinol, approved by the FDA in 1966, is

an effective uric acid lowering drug, but must be used cautiously and at lower doses in patients with chronic kidney disease (CKD) due to the increased risk of severe allergic skin reactions. Such reactions are rare, but can become life-threatening. The dosing recommendations for gout patients with CKD remain unclear, leading many prescribers to underdose allopurinol in this group with resultant uncontrolled hyperuricemia. FBX was approved in 2009 and lacks the risk of allergic reactions that are associated with allopurinol. It is also safe to use at either marketed dosage strength (40 mg or 80 mg) in patients with CKD. This distinction from allopurinol allows more aggressive dosing of FBX in a critical subpopulation of gout patients so that serum uric acid may be reduced to the target concentration of less than 6 mg/dL .

2. Study Rationale

Both allopurinol and FBX are considered first-line options for the treatment of hyperuricemia in gout (Khanna, 2012). Due to the shorter duration of time that FBX has been on the market, it costs about 10-14 times as much as allopurinol. On the basis of a direct price comparison alone, one could conclude that allopurinol is more cost-effective than FBX. However, the overall costs associated with either therapy regarding the health care utilization of gout patients would need to be determined to assess their value relative to each other. Unfortunately, no published real-world data currently exist that compare the overall clinical and economic outcomes of XOI in gout patients. This is especially true with respect to other concurrent disease states and the accompanying treatment patterns with ULT. Enhanced knowledge of real world expenditures in gout patients may also help to further clarify the case to be made for the dominant value of FBX versus allopurinol. Data generated by analysis of databases that contain EHR and medical and prescription claims could be used to explore this issue. These data would be very useful to value-based health delivery systems, including managed care organizations and accountable care organizations (ACOs) and also could provide valuable insights for further innovations in

gout management. Therefore, the hypothesis of this study is that treatment with FBX will result in greater economic benefit than allopurinol as demonstrated by improved cost effectiveness and lower overall health care utilization. This hypothesis will be explored in CKD patients with gout, because they may experience the most benefit with respect to these outcomes from FBX as compared with allopurinol (Burns & Wortmann, 2011).

3. Objectives

a. The primary study objective is to determine the economic outcomes (using total direct cost) for patients with both gout and Stage 2 – Stage 4 CKD who are taking FBX compared with allopurinol.

b. The secondary objective is to describe the incidence of gout flares (using diagnostic and procedural codes) for patients with both gout and Stage 2 – Stage 4 CKD who are taking FBX compared with allopurinol.

4. Methods

4.1 Study Design

The study will be a retrospective cohort analysis using information extracted from MarketScan, a medical claims database. The two patient cohorts will both have gout and CKD and will be taking ULT as described later in this paper. Data extracted from MarketScan will include dates of service from January 1, 2005 through December 31, 2012.

4.2 Study population

Enrollment in the study will be episode-based. Each patient may contribute multiple episodes to the analytic dataset provided all of the following criteria are met during each episode:

- Have at least one prescription for allopurinol or FBX between January 1, 2005 and December 31, 2011;
- Have a diagnosis of gout (as coded by ICD-9-CM 274.xx) on at least one inpatient claim or on two outpatient non-diagnostic claims prior to initiation of allopurinol or FBX;
- Are continuously enrolled for at least 12 months prior and 12 months post allopurinol or FBX initiation;
- Have both medical and pharmacy benefit plus complete data availability during both baseline and follow-up periods;
- Have evidence of CKD, stages 2-4 as defined later, in the 12 months prior to index;
- Are at least 18 years of age at the time of initiation of therapy with allopurinol or FBX; and
- Have at least one laboratory value for serum uric acid in the 12-month period prior to initiation of therapy.

Patients will meet the CKD criteria if there is either an:

- ICD 9 CM diagnosis code for CKD, stages 2-4, on at least one inpatient claim or at least two outpatient claims on different days between 30 and 365 days apart; or
- CPT 4, HCPCS or ICD 9 procedure code on at least one medical claim, either inpatient or outpatient.

These claims are intended to capture data for patients that have a diagnosis or procedure coded for billing purposes for CKD with a reasonable expectation that this disease exists in such individuals. Since a

risk of misalignment of the coding for and true presence of CKD exists in any patient, the CKD status and stage will be corroborated with clinical laboratory data where available.

Patients will be followed within each treatment episode until disenrollment from MarketScan, end of the study period, discontinuation of either allopurinol or FBX or a prescription for the alternate agent.

4.3 Propensity score matching

Propensity score matching will be used to minimize selection bias. The propensity model will characterize the probability that an individual will receive allopurinol or FBX on the basis of observed variables. Such variables in the propensity model may include age, type of health plan, geographic region, population density, the year the ULT was initiated, the number of gout flares during baseline, serum uric acid concentrations, the presence of specific clinical conditions, CKD stage, exposure to other gout therapies and/or exposure to other medication classes, e.g. cardiovascular medications. Patients treated with FBX subsequently will be matched with those treated with allopurinol with similar predicted probabilities. Matching will be performed for optimal statistical power and goodness-of-fit, ideally at a match ratio of 1 FBX:5 allopurinol patients.

4.4 Data extraction and analysis

The primary economic outcome will be direct medical expenditure as measured during each treatment episode. All direct medical expenditures will be captured regardless of fiscal obligation, including health plan insured amounts, coordination of benefits and patient co-payment, deductible, and co-insurance amounts. Expenditures will be stratified as gout-related or not gout-related and may be reported both in expenditures per month on therapy, per episode of therapy or per year as appropriate. Expenditures will be further subdivided as follows:

- Inpatient

- Outpatient
- Office Visits
- Emergency room
- Pharmacy
- Other

Means and standard deviations will be population-based. All costs will be standardized and reported in 2012 US dollars using the medical component of the Consumer Price Index. Differences between all allopurinol and FBX cases will be considered statistically significant ($p < 0.05$ in a two-tailed test) if the 95% CI does not include $0 for expenditures.

The primary clinical outcome will be the incidence and frequency of acute gout episodes or flares. For each clinical outcome, the number and proportion of patients experiencing a gout flare will be reported as will the incidence rate per unit of person time observed and time to event. Kaplan–Meier survival curves will also be generated. In addition, the mean (with standard deviation) number of acute flares per treatment episode will be reported.

The analysis will begin with simple statistics describing accrual into the study followed by the demographic and clinical characteristics of the sample overall, both pre- and post-propensity matching, stratified by exposure cohort. Subsequent tables will detail the frequency (counts and event rates) of acute gout flares in patients as well as first event descriptors, stratified again by exposure cohort. Additional tables will similarly summarize and report both all cause and gout specific direct medical expenditure and utilization statistics, also stratified by exposure cohort.

Basic analyses will include descriptive profiles of all independent and dependent variables. Categorical variables will be summarized in frequency tables. Continuous and other numeric variables will be summarized by presenting the number of observations, mean, standard deviation, and median.

Statistical tests of significance for differences in these distributions will be carried out. Chi-square tests will be used to assess the statistical significance of categorical variables; t-tests and ANOVA will be used for continuous variables.

Chapter 3: DISCUSSION

1. Methods Analysis and Critique

Salient aspects of a claims database analysis include proper selection of the study design, the database, the patient cohorts, analysis plan, and reporting characteristics. The rationale and potential drawbacks of each aspect will be discussed in turn as illustrated by the case study.

The information captured in a claims database has by definition occurred in the past for specific health care encounters. Therefore, the retrospective design is most fitting, and the comparison of two or more cohorts of patients exposed to different interventions (or no intervention) strongly suggest that a retrospective cohort design is most appropriate. In the case study, the objective is to compare the overall direct medical costs of care between two similar groups of patients exposed to one of two treatments for gout. A prospective design would not be able to expeditiously answer this question, because it would require a multi-year duration, and would be influenced by factors that are unforeseen, such as the entry of additional drugs into the gout therapeutic area. Also, the uptake of a new agent such as FBX occurs no sooner than 12 months after FDA approval, so the data available for recipients of FBX in this study are likely limited to the 2010-12 time span.

The data source will be the Truven MarketScan® Commercial Claims and Encounters (Commercial) and the Medicare Databases including the MarketScan laboratory database. The MarketScan® databases are the largest convenience sample available in a proprietary US claims database with 170 million unique

patients since 1995. Nearly half of all US health plans, including many of the largest insurers, contribute to these databases. In the most recent full data year, MarketScan claims databases contain data for 50 million covered lives. Both claims databases include individuals covered under a variety of fee-for-service, point of service, and capitated reimbursement schemes. Laboratory data are included from major national laboratory testing entities, such as LabCorp. The databases are HIPAA-compliant and as such are exempt from Institutional Review Board approval. Other available large commercial databases include the GE Healthcare database, which contains very robust electronic medical records and laboratory data, but are not integrated with prescription drug claims. The latter aspect was an overriding factor that precluded the inclusion of the GE product in the case study, because drug utilization data are essential for hypothesis testing. If prescription drug usage cannot be assessed, in this case study for either allopurinol or FBX, then its impact on outcomes is impossible to isolate. Even though initial fill and refill claims document that a point of care event has occurred, i.e., the dispensing of a prescription drug, this does not provide verification of patient adherence. This is a limitation of all claims database studies, because there is no way to verify that patients have actually taken the medicine they have received from the pharmacy other than contacting each individual in the database. Not only is this a violation of HIPAA policy, but it also is impractical, because there are millions of patients in a large database, and the answer provided would rely on recall of an event that may have occurred months or years in the past. The act of taking one's medication as prescribed is known as adherence, and its assessment is particularly challenging using retrospective data. One method is to estimate the medication possession ratio (MPR) which is calculated as the days supply of drug dispensed for the evaluation period (typically at least 12 months) divided by the total number of days in the evaluation period (Andrade et al, 2006). An alternative to MPR is the daily consumption rate (DACON) which more crudely estimates adherence as the number of oral dosage units taken per day dispensed divided by the time. The DACON is more useful for medications that may be taken more than once daily. For the case

study, both allopurinol and FBX are indicated for once daily dosing; shortening the dosing interval more than that does not offer any therapeutic gain and may elevate the risk of adverse events or intolerance.

Another benefit of the MarketScan databases is the cross-section of claims represented in the sample. For example, the types of different commercial (non-government) insurance plans may vary greatly between patients, based on employer offerings, self-insured status, geography, medical history (in adults, prior to implementation of the PPACA in 2014), and income. Despite the acknowledged limitation that MarketScan does not contain claims from Medicaid, the Veterans Administration, many other commercial plans, and, certainly, uninsured individuals, a reasonably representative mix of insured patients is included from across the US. The last feature assists the researcher in overcoming variations in gout treatment by US geographic sector (Davis & Wreath, 2013). Without this component, there may be a treatment selection bias based on patient location, perhaps skewing the cost data in a positive direction.

Patient selection is quite possibly the part of OR studies requiring the most care and planning. If the cohorts are ill-defined, then incomplete, irrelevant, or unusable data may be collected. The case study requires that adults (age 18 years or greater) with gout be included since the two drugs are only indicated for that age group; allopurinol has FDA-approved indications other than gout, but FBX is approved only for the treatment of hyperuricemia in gout. The use of birthdates and ICD-9 diagnostic codes for gout will capture data for this cohort. The requirement of an additional comorbid condition, chronic kidney disease, further narrows the population of interest to those with a factor that complicates the management of gout in two ways. First, the dosing of allopurinol must be adjusted downward for impaired kidney function and, second, persistent elevations in uric acid are associated with progression of CKD. These patient characteristics form the crux of the hypothesis that FBX is especially well-suited to improve outcomes in this group, because it does not require dosage adjustment in CKD and possesses greater potency at lowering uric acid. Lastly, the National Kidney Foundation CKD

stages are used to define patients with gout who have comorbid CKD (NKF, 2002). The desired cohort would have stages 2, 3, and 4 disease whose severity is mild, moderate, or severe, respectively. Stage 1 patients are excluded, because they have near normal kidney function and would not typically have any clinical findings or treatment associated with this degree of renal impairment. Also, Stage 5 patients are described as end-stage renal disease (ESRD) patients who require dialysis or kidney transplantation to survive. These patients would be excluded from analysis primarily because neither ULT drug has dosing recommendations for their use in this population.

Specification of claims that are episode-based is critical, because the expectation is that patients have an incident gout flare (as signified by a coded diagnosis of gout) followed within 12 months by a prescription claim for ULT. This specific sequence is necessary in order to establish that new users of ULT are included in the analysis. It is most advantageous to have new users of ULT, i.e., previously ULT-naïve, so that any carryover benefit or harm of previous ULT use is removed as a confounding variable. In addition, at least one serum uric acid concentration is required so that hyperuricemia is documented. Both allopurinol and FBX are indicated for the treatment of hyperuricemia in gout patients so both criteria must be present to align with this FDA-approved use.

The elements of the analysis plan are clearly developed and described a priori in accordance with good research practices for observational studies described previously. The use of propensity score matching, defined previously in this paper, will be employed in the case study, because it is the most reliable and valid method of mitigating potentially confounding patient characteristics. Multiple factors will be entered into the propensity score so as to reduce bias. Only direct medical costs will be included in the case study analysis, because indirect costs such as impact on lost income, reduced quality of life, and years of life lost, are not contained within a claims database. All of the various sectors of health care that potentially generate direct medical claims will be included as described in the methods section; this is so that the most complete accounting of costs can be assessed. The monetary value of total direct

costs will be compared using parametric statistical tests for continuous variables and the incidence of the primary clinical outcome (gout flares) will be compared for each therapy using a parametric test for proportions.

Reporting of the data results will follow the guidance of the STROBE document. This paper contains the elements required for compliance, including the title and abstract, introduction, and methods. The sole funding source is Takeda Pharmaceuticals International, Inc. The results and discussion will follow when available; but preliminary elements of the discussion appear in this paper, such as the limitations and generalizability of the data in general terms.

In spite of the carefully planned design and rigorously executed data analysis in the case study, disadvantages inherent in any retrospective claims database study may be present. Most concerning is the use of prescription claims as a surrogate for patient adherence to therapy. Evidence of initial and repeat filling of prescriptions is not an ideal measure of adherence, but given the current lack of patient-centered data, it is the best available method at this time. A possible follow-on project of the case study would help address this deficiency through direct patient engagement to assess and confirm therapy adherence at the various points of care, such as the clinic, pharmacy, and community health centers. Another potential source of error is the misalignment of the coded diagnosis for gout and the true nature of the patient's health condition. For example, gout may be mentioned in the differential diagnosis for joint pain upon presentation to a health care provider and coded as such. However, the ultimate diagnosis may not be gout. In the case study a documented serum uric acid concentration and prescription for ULT offers additional evidence to support the coded diagnosis. One method that confers the most confidence in the coded data was mentioned previously and involves taking a sample of patients and reviewing the medical record for progress notes to corroborate the presence of gout. This is known as validation, and is very resource-intensive to conduct for the entire cohort studied.

Therefore, in the case study, a sample of 50 patients will be analyzed for alignment of the gout code and medical record notes.

2. Future Steps

The case study constitutes the first phase of a research plan to assess and develop methods to improve the care and outcomes of patients with gout, especially those with comorbid CKD. If the hypothesis is not rejected, then this would be the first analysis in the public domain to offer evidence of holistic economic benefit of one ULT versus another in a robust sample of gout patients. Specific patterns of medication use may be identified in the case study, such as when gout patients tend to stop taking ULT, when gout flares tend to recur and how often, and the types of health care expenditures observed with patients who experience worse gout outcomes. All of these data may serve as a platform for the development of patient engagement initiatives designed to improve adherence to gout therapy regimens. One method would be to implement on a pilot basis a patient outreach program for more active disease management through coordination of gout care between providers such as the hospital (if admitted for treatment of gout flare), primary care provider, consulting rheumatologist, and pharmacy. The data from the case study would assist in the identification of gout patients with CKD who are at the highest risk of nonadherence and resultant worse outcomes such as persistent elevations in uric acid, repeated gout flares, and increased likelihood of joint damage and reduced function and mobility. An ideal setting for this pilot project would be an accountable care organization (ACO) or other integrated health care delivery network (IDN) where a network of hospitals and providers are united by a shared responsibility for both clinical and financial outcomes. In the ACO model, the network is monetarily incentivized to improve the quality of care as evidenced by meeting outcome standards. Even though the gout disease state has not yet risen to the level of prominence where value-based payments are in place, it will be a matter of time before most conditions are paid for this way. The existence of US

guidelines for the management of gout specifies the use of particular ULT drugs to target a serum uric acid concentration of less than 6 mg/dL. Alignment of gout care and outcomes with these guidelines using innovative patient activation and adherence programs is one proposed step toward improvement in not just cost-effectiveness and quality, but also a better patient experience and engagement in their own care. These steps are intended to establish and reinforce the accountability of the health care system, but also that of the patient and his or her caregiver as stakeholders in the process of improving health outcomes.

In summary, OR is driving how currently available health care plans are assessed for quality, efficiency, and patient satisfaction (AHRQ, 2000). The grading of insurance coverage through newly established health care exchanges relies heavily on outcomes data, and this trend is expected to continue into the future as health care evolves into a consumer-centric rather than a provider-centric enterprise.

REFERENCES

Agency for Healthcare Research and Quality. (2000). Outcomes Research. Retrieved November 1, 2013 from http://www.ahrq.gov/research/findings/factsheets/outcomes/outfact/index.html.

American Medical Association. (2013). Current procedural terminology. Retrieved October 12, 2013 from http://www.ama-assn.org/ama/pub/physician-resources/solutions-managing-your-practice/coding-billing-insurance/cpt.page

Andrade, S.E., Kahler, K.H., Frech, F., & Chan, K.A. (2006). Methods for evaluation of medication adherence and persistence using automated databases. *Pharmacoepidemiology and Drug Safety, 15,* 565-574.

Aschengrau, A. & Seage, G.R. (2008). Overview of Epidemiologic Study Designs. In A. Ashengrau, G.R. Seage (Eds.), *Essentials of Epidemiology in Public Health, 2nd ed.* (139-168). Sudbury, MA: Jones and Bartlett.

Berger, M.L., Mamdani, M., Atkins, D., & Johnson, M.L. (2009). Good Research Practices for Comparative Effectiveness Research: Defining, Reporting and Interpreting Nonrandomized Studies of Treatment Effects Using Secondary Data Sources: The ISPOR Good Research Practices for Retrospective Database Analysis Task Force Report—Part I. *Value in Health, 8,* 1044-1052.

Berwick, D. M., Nolan, T.W., and Whittington, J. (2008). The Triple Aim: Health, Care, and Cost. *Health Affairs, 27,* 759-769.

Bootman, J.L., Townsend, R.J., & McGhan, W.F. (1996). Introduction to Pharmacoeconomics. In J.L. Bootman , R.J. Townsend, W.F. McGhan (Eds.), *Principles of Pharmacoeconomics, 2nd ed.* (4-19). Cincinnati, OH: Harvey Whitney Books Company.

Burns, C.M. & Wortmann, R. (2011). Gout therapeutics: new drugs for an old disease. *Lancet, 377,* 165-177.

Campbell, S.M., Roland, M.O., & Buetow, S.A. (2000). Defining Quality of Care. *Social Science and Medicine, 51,* 1611-1625.

Centers for Disease Control and Prevention. (2013). Classification of diseases, functioning, and disability. Retrieved October 12, 2013 from http://www.cdc.gov/nchs/icd.htm.

Chassin, M.R. & Galvin, R.W. (1998). The Urgent Need to Improve Health Care Quality. *JAMA, 280,* 1000-1005.

Clancy, C. & Collins, F.S. (2010). Patient-Centered Outcomes Research Institute: The Intersection of Science and Health Care. *Science Translational Medicine, 2,* 1-3.

Clancy, C.M. & Eisenberg, J.M. (1998). Outcomes Research: Measuring the End Results of Health Care. *Science, 282*, 245-246.

Cox, E., Martin, B.C., van Staa, T., Garbe, E., Siebert, U., & Johnson, M.L. (2009). Good Research Practices for Comparative Effectiveness Research: Approaches to Mitigate Bias and Confounding in the Design of Nonrandomized Studies of Treatment Effects Using Secondary Data Sources: The International Society for Pharmacoeconomics and Outcomes Research Good Research Practices for Retrospective Database Analysis Task Force Report—Part II. *Value in Health, 8,* 1053-1061.

Davis, A. & Wreath, J. (2013). Assessing the Increasing Costs to Manage Patients with Gout by State. *Arthritis & Rheumatism, 65*, S826.

Donabedian, A. (1966). Evaluating the Quality of Medical Care. *Milbank Memorial Fund Quarterly, 44*, 166–206.

Ellwood, P. M. (1988). Shattuck lecture--outcomes management. A technology of patient experience. *N Engl J Med, 318,* 1549-1556.Institute of Medicine. (2007). Learning What Works Best. The Nation's Need for Evidence on Comparative Effectiveness in Health Care. Retrieved September 23, 2013 from www.iom.edu/~/media/.../ComparativeEffectivenessWhitePaperF.ashx.

Hall, G.C., Sauer, B., Bourke, A., Brown, J.S., Reynolds, M.W., & Lo Casale, R. (2011). Guidelines for Good Database Selection and Use in Pharmacoepidemiology Research. *Pharmacoepidemiology and Drug Safety.* DOI: 10.1002/pds.2229.

Institute of Medicine. (2007). Learning What Works Best. The Nation's Need for Evidence on Comparative Effectiveness in Health Care. Retrieved September 23, 2013 from www.iom.edu/~/media/.../ComparativeEffectivenessWhitePaperF.ashx.

Johnson, M.L., Crown, W., Martin, B.C., Dormuth, C.R., & Siebert, U. (2009). Good Research Practices for Comparative Effectiveness Research: Analytic Methods to Improve Causal Inference from Nonrandomized Studies of Treatment Effects Using Secondary Data Sources: The ISPOR Good Research Practices for Retrospective Database Analysis Task Force Report—Part III. *Value in Health, 8,* 1062-1073.

Jollis, J.G., Ancukiewicz, M., DeLong, E.R., Pryor, D.B., Muhlbaier, L.H., & Mark, D.B. (1993). Discordance of Databases Designed for Claims Payment versus Clinical Information Systems. Implications for Outcomes Research. *Ann Intern Med, 119*, 844-850.

Khanna, D., Fitzgerald, J.D., Khanna, P.P., Bae, S., Singh, S.K., Neogi, T.,…& Terkeltaub, R. (2012). 2012 American College of Rheumatology Guidelines for Management of Gout. Part 1: Systematic Nonpharmacologic and Pharmacologic Therapeutic Approaches to Hyperuricemia. *Arthritis Care & Research, 64,* 1431-1446.

Krumholz, H. M. (2009). Outcomes Research: Myths and Realities. *Circ Cardiovasc Qual Outcomes, 2,* 1-3.

Lauffenberger, J.C., Balasubramanian, A., Farley, J.F., Critchlow, C.W., O'Malley, C.D., Roth, M.T.,…& Brookhart, M.A. (2013). Completeness of prescription information in US commercial claims databases. *Pharmacoepidemiology and Drug Safety, 22,* 899-906.

Motheral, B., Brooks, J., Clark, M.A., Crown, W.H., Davey, P., Hutchins, D.,…Stang, P. (2003). A Checklist for Retrospective Database Studies—Report of the ISPOR Task Force on Retrospective Databases . *Value in Health, 6,* 90-97.

National Institutes of Health. (2013). PROMIS. Retrieved September 25, 2013 from www.nihpromis.org/.

National Kidney Foundation. (2002). KDOQI Clinical Practice Guidelines for CKD. Retrieved November 1, 2013 from www.kidney.org/professionals/kdoqi/guidelines_ckd/p4_class_g2.htm

Patient-Centered Outcomes Research Insitute. (2013). How We're Funded. Retrived October 12, 2013 from http://www.pcori.org/about-us/how-were-funded/.

Patient-Centered Outcomes Research Insitute. (2013). National Priorities and Research Agenda. Retrieved September 22, 2013 from http://pcori.org/research-we-support/priorities-agenda/.

Roger, V. L. (2011). Outcomes Research and Epidemiology: The Synergy Between Public Health and Clinical Practice. *Circ Cardiovasc Qual Outcomes, 4,* 257-259.Rubin, D.B. (1997). Estimating Causal Effects from Large Data Sets Using Propensity Scores. *Ann Intern Med, 127,* 857-863.

Rubin, D.B. (1997). Estimating Causal Effects from Large Data Sets Using Propensity Scores. *Ann Intern Med, 127,* 857-863.

Schenck, A.P., McArdle, J., & Weiser, R. (2013).Quality Improvement Organizations and Continuous Quality Improvement in Medicare. In Sollecito, W.A. & Johnson, J.K. (Eds.), *McLaughlin and Kaluzny's Continuous Quality Improvement In Health Care, 4th ed*. (421-452). Burlington, MA: Jones & Bartlett Learning.

Schneeweis, S. & Avorn, J. (2005). A Review of Uses of Health Care Utilization Database for Epidemiologic Research on Therapeutics. *J Clin Epi, 58,* 323-337.

Schulz, K.F., Altman, D.G., & Moher, D. (2010). CONSORT 2010 Statement: Updated Guidelines for Reporting Parallel Group Randomized Trials. *Ann Intern Med, 152,* 726-732.

Shmerling, R.H. (2012). Management of Gout. A 57-Year-Old Man With a History of Podagra, Hyperuricemia, and Mild Renal Insufficiency. *JAMA, 308,* 2133-2141.

The Centers for Disease Control and Prevention (2011). Social Determinants of Health. Retrieved September 13, 2013 from http://www.cdc.gov/socialdeterminants/Definitions.html.The Commonwealth Fund. (2003). *Churn, churn, churn: how instability of health insurance shapes America's uninsured problem.* New York, NY: Short, P.F., Graefe, D.R., & Schoen, C.

The Commonwealth Fund. (2003). *Churn, churn, churn: how instability of health insurance shapes America's uninsured problem.* New York, NY: Short, P.F., Graefe, D.R., & Schoen, C.

The Food and Drug Administration. (2013). The National Drug Code Directory. Retrieved October 27, 2013 from http://www.fda.gov/drugs/informationondrugs/ucm142438.htm

The Henry J. Kaiser Family Foundation. (2011). Snapshots: Health Care Spending in the United States & Selected OECD Countries. Retrieved September 13, 2013 from http://kff.org/health-costs/issue-brief/snapshots-health-care-spending-in-the-united-states-selected-oecd-countries/.

US Department of Health and Human Services. (2013). What Is Meaningful Use? Retrieved September 22, 2013 from http://www.healthit.gov/policy-researchers-implementers/meaningful-use.

US Department of Health and Human Services. 45 CFR parts 160 and 164: Modifications to the HIPAA Privacy, Security, Enforcement and Breach Notification Rules Under the Health Information Technology for Economic and Clinical Health Act and the Genetic Information Nondiscrimination Act: Other Modifications to the HIPAA Rules: Final Rule. Retrieved September 27, 2013 from http://gpo.gov/fdsys/pkg/FR-2013-01-25/pdf/2013-01073.

van Walraven, C., & Austin, P. (2012). Administrative database research has unique characteristics that can risk biased results. *J Clin Epi, 65,* 126-131.

von Elm, E., Altman, D.G., Egger, M., Pocock, S.J., Gotzsche, P.C., & Vandenbroucke, J.P. (2007). The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) Statement: Guidelines for Reporting Observational Studies. *PLOS Medicine, 4,* 1623-1627.

Wang, C.J. & Huang, D.J. (2013). A Checklist for Retrospective Database Studies—Report of the ISPOR Task Force on Retrospective Databases. *JAMA, 310,* 1121-1122.

World Health Organization. (2009). Global Health Observatory Data Repository. Retrieved September 27, 2013 from http://apps.who.int/gho/data/node.main.688?lang=en.

World Health Organization (2013). Pharmacovigilance. Retrieved October 12, 2013 from http://www.who.int/medicines/areas/quality_safety/safety_efficacy/pharmvigi/en/.