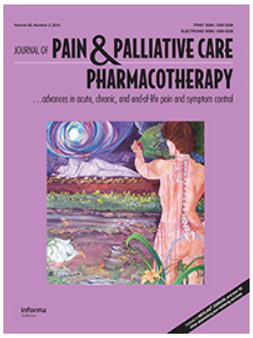

Taylor & Francis
Taylor & Francis Group

Journal of Pain & Palliative Care Pharmacotherapy

ISSN: 1536-0288 (Print) 1536-0539 (Online) Journal homepage: http://www.tandfonline.com/loi/ippc20

Real-World Evidence in Pain Research: A Review of Data Sources

Brandon K. Bellows, Kuan-Ling Kuo, Eman Biltaji, Mukul Singhal, Tianze Jiao, Yan Cheng & Carrie McAdam-Marx

To cite this article: Brandon K. Bellows, Kuan-Ling Kuo, Eman Biltaji, Mukul Singhal, Tianze Jiao, Yan Cheng & Carrie McAdam-Marx (2014) Real-World Evidence in Pain Research: A Review of Data Sources, Journal of Pain & Palliative Care Pharmacotherapy, 28:3, 294-304, DOI: 10.3109/15360288.2014.941131

To link to this article: https://doi.org/10.3109/15360288.2014.941131

Journal of Pain & Palliative Care Pharmacotherapy. 2014;28:294–304.

ISSN: 1536-0288 print / 1536-0539 online
DOI: 10.3109/15360288.2014.941131

informa
healthcare

PHARMACOECONOMICS AND OUTCOMES IN PAIN AND PALLIATIVE CARE

Real-World Evidence in Pain Research: A Review of Data Sources

Brandon K. Bellows, Kuan-Ling Kuo, Eman Biltaji, Mukul Singhal, Tianze Jiao, Yan Cheng, and Carrie McAdam-Marx

ABSTRACT

Outcomes research studies use clinical and administrative data generated in the course of patient care or from patient surveys to examine the effectiveness of treatments. Health care providers need to understand the limitations and strengths of the real-world data sources used in outcomes studies to meaningfully use the results. This paper describes five types of databases commonly used in the United States for outcomes research studies, discusses their strengths and limitations, and provides examples of each within the context of pain treatment. The databases specifically discussed are generated from (1) electronic medical records, which are created from patient-provider interactions; (2) administrative claims, which are generated from providers' and patients' transactions with payers; (3) integrated health systems, which are generated by systems that provide both clinical care and insurance benefits and typically represent a combination of electronic medical record and claims data; (4) national surveys, which provide patient-reported responses about their health and behaviors; and (5) patient registries, which are developed to track patients with a given disease or exposure over time for specified purposes, such as population management, safety monitoring, or research.

KEYWORDS databases, observational studies, outcomes research, pain, treatment

INTRODUCTION

Approximately 116 million adults in the United States (US) suffer from some form of common chronic pain, including low back pain, arthritis, and headache.[1] This is significantly higher than the prevalence of cardiovascular disease (83.6 million),[2] diabetes (25.8 million),[3] and cancer (13 million).[4] Also in 2008, the annual total cost of pain in the US was estimated to be between $560 and $635 billion.[5]

Despite the high prevalence and costs of pain, significant gaps in pain research remain. According to the Food and Drug Administration (FDA), "Many experts in analgesic drug development believe that it is the design of the clinical trials that is at fault in this [evidence gap] . . ."[6] Randomized controlled trials (RCTs) are considered the gold standard for establishing causation in health research. With rigorous methodology, RCTs demonstrate efficacy, or that a treatment *can* work. However, the subjects included in RCTs and the controlled nature of RCTs mean that the trials often do not represent patients treated in actual clinical settings. For example, pain-related RCTs often exclude elderly patients, and those with comorbid psychological disorders, multiple pain disorders, or those taking other medications.[1,7] Furthermore, RCTs are often evaluated over a short time horizon, e.g., 4 to 14 weeks, which does not represent the typical duration of chronic pain treatment.[1]

Therefore, RCTs do not demonstrate effectiveness, or that a treatment *does* work in real-world settings. Outcomes research studies attempt to fill this gap by determining effectiveness, often using clinical

Brandon K. Bellows, PharmD, is a Research Associate in the Pharmacotherapy Outcomes Research Center; Kuan-Ling Kuo, BS, is a PhD Candidate and Eman Biltaji, RPh, MS, Mukul Singhal, BS, Tianze Jiao, BS, and Yan Cheng, MS, are Graduate Students in Pharmacotherapy Outcomes and Health Policy; and Carrie McAdam-Marx, RPh, PhD, is an Associate Professor, Department of Pharmacotherapy, University of Utah, Salt Lake City, Utah, USA.

Address correspondence to: Brandon Bellows, PharmD, Pharmacotherapy Outcomes Research Center, University of Utah, 30 S 2000 E, Room 4963, Salt Lake City, UT 84112, USA (E-mail: brandon.bellows@pharm.utah.edu).

and administrative data generated in the course of patient care, or patient survey data.[8] Although the use of real-world data can help to overcome RCT generalizability limitations, it also creates challenges due to biases and confounding that are avoided in clinical trials by randomization. However, given the high prevalence and costs of pain and pain management, it is important to determine if treatments are clinically effective in real-world settings. To do this, health care providers need access to data from outcomes research, as well as the ability to evaluate and interpret the results of outcomes research studies. An important component of evaluating outcomes research studies is being able to consider the limitations and strengths of the real-world data sources used.

The objective of this paper is to assist clinicians and other stakeholders in evaluating pain-related outcomes research by describing databases commonly used in the US for outcomes research studies, discussing data strengths and limitations, and providing examples of each within the context of pain treatment.

TYPES OF DATABASES

Electronic Medical Record Databases

The primary function of electronic medical record (EMR) systems is to capture clinical information generated from patient-provider interactions. Traditionally, EMRs were developed and used by localized health systems. More recently, it is common for providers and health systems to license EMR systems. This has allowed EMR vendors to collect deidentified clinical data from users and consolidate the data into large research databases.

Researchers now recognize the wealth of the real-world clinical information in EMR data sets and are using them to study clinical outcomes in a variety of diseases. The distinguishing value of EMRs as a data source is that they are clinically rich and include patient-level information on features such as height, weight, chief complaint/diagnosis, vital signs, prescription orders, medication history, and laboratory test results. These well-defined data elements are generally captured as structured data, or information that is codified in a discrete field. This allows records to be queried and sorted on these data fields. Additional clinical detail, such as patient history, disease status, and treatment rationale, is captured in unstructured text fields, which are considerably more difficult to query.

In addition, EMR systems capture data at the point of care and EMR research data sets support the examination of treatment patterns and outcomes across time. Thus, EMR data can be used to assess longitudinal patient outcomes, which is particularly useful when studying chronic diseases and conditions. Additionally, EMRs capture data on all patients seen by providers or clinics regardless of the method of payment, which may increase the generalizability of study findings. Another strength is that some EMR data sets include data on millions of patients across the US provided by thousands of health care providers. This allows for the examination of subpopulations, rare diseases, and uncommon treatments.

There are also several limitations of EMR data sets in regards to supporting outcomes research. First, EMR data sets may provide an incomplete and fractured medical history for any given patient. This is due to the fact that EMR systems and, therefore, EMR data sets may be limited to a specific provider or health system and other providers may use different EMRs. Furthermore, EMR systems may be developed to support care in a specific setting, such as ambulatory care clinics or hospitals. Although providers may document other significant encounters in a patient note if reported by either the patient or other health care providers, clinical detail will likely be lacking.

Another limitation of EMR databases is that they contain information on physician orders, but may not have data on whether or not the order was carried out. For instance, an EMR captures prescription orders, but generally does not have data on medication fills. Additionally, patients may be "lost to follow-up" in an EMR system for any number of reasons, including death, moved to a new location, or switched to a new health care provider. These may also lead to an incomplete medical history and make it difficult to study longer-term outcomes, such as mortality.

Finally, as mentioned above, important clinical detail may be captured as unstructured data and is not easily identified for research. Capturing information from unstructured data is complex and requires efforts such as manual chart reviews or advanced computerized techniques, such as natural language processing (NLP).

From a pain treatment perspective, EMR research data sets provide information on diagnoses and pain-related comorbidities, prescribed medications, and orders for procedures or related care such as physical therapy. With large national data sets, there is often good generalizability of results, as the EMRs will not be limited to a specific payer type, such as Medicaid or managed care.

A key limitation in using EMR data for pain research is that EMRs often lack a discrete field for documenting pain scores or severity. Thus, pain

information may not be structured for easy identification, but may be available to researchers with resources to conduct chart reviews or NLP.[9] However, when pain scores can be accessed from an EMR, many factors can then be examined to determine if they are associated with changes in patients' pain severity. For example, Miller et al. were able to assess the correlation between pain scores and physician global assessments in patients with juvenile idiopathic arthritis using data obtained from an EMR.[10] It should be noted that unstructured data, such as clinical text notes, have the potential to contain patient identifying information. Thus, although manual chart reviews and NLP may be used to access information, not all EMR databases provide unstructured data for analysis.

Another limitation is that patients with chronic pain often see multiple providers, including emergency departments. Thus, if the patient is not seeing providers within the same health system, their EMR data will likely be incomplete. Finally, without prescription fill data, it is not generally possible with EMR data sets to assess utilization of prescribed medications. However, many pain medications are controlled substances, which have limits on refills and require frequent provider prescriptions. This may lead to more complete documentation of medication use than medications that can be prescribed with multiple refills.

As would be expected based on how EMRs were traditionally developed, there are many EMR systems used across the US. Only two EMR research databases will be discussed briefly here (Table 1). The first is the General Electric (GE) Centricity EMR, which is used by over 35,000 clinicians across the United States and contains over 30 million patient records.[11] GE Centricity was primarily designed for ambulatory care practices, and approximately 60% of users are primary care providers. Specialty care providers are underrepresented in the GE EMR research data set. Whereas GE Centricity is used across the nation, an example of an EMR data set from a single, regional health system is the University of Utah. The University of Utah has developed an Electronic Data Warehouse (EDW) that has consolidated clinical data captured across the University Health System from numerous EMR and other complementary clinical systems such as laboratory and pharmacy. The EDW contains clinical records for over 1.4 million patients.[12]

A recent example of outcomes study in pain used a rule-based NLP algorithm to derive pain scores using EMRs in patients with metastatic prostate cancer.[9] All medical records for each patient, including paper records that were converted into an electronic format, were requested and compiled into a database. The NLP algorithm identified terms associated with pain in the clinical notes and then used the context in which it was used to determine if the patient was experiencing no pain, some pain, controlled pain, or severe pain. As may be expected for patients with metastatic cancer, the study found that being in the last year of life was significantly associated with having severe pain (odds ratio [OR]: 2.52, 95% confidence interval [CI]: 1.46–4.36; $P = .001$). Although the sample size was small ($N = 33$), the authors were able to automatically access critical components of the clinical notes using NLP. This study highlights the valuable pain information contained in EMRs that can be leveraged when appropriate resources, such as NLP, are available.

Administrative Claims Databases

Administrative claims research data sets are created from data on billing and payment interactions between patients, health care providers, and payers. These data sets contain information required for this payment process, including reason for medical visit documented by International Classification of Disease 9th or 10th edition (ICD-9 and ICD-10, respectively) diagnosis codes and service provided documented by Current Procedural Terminology 4th edition (CPT-4) and other service codes. Similarly, when the patient goes to the pharmacy to fill a prescription, the pharmacy submits information on the medication dispensed to the payer for reimbursement.

A key benefit of administrative claims data sets is that they include financial data, including amounts paid by the insurance company and out-of-pocket costs to the patient namely in the form of copayments/coinsurance and deductibles. Claims databases therefore have transaction data for encounters of care from all medical providers when such care is submitted for reimbursement. However, if a patient pays cash or claims are paid by another means, such as workers compensation, then such data will not be captured.

Although there are a large number of health insurance providers in the US, billing requirements are generally uniform. This allows for the creation of very large, administrative claims databases, with data from multiple payers. Researchers have begun to see the potential value of examining health care outcomes using these large, national databases for the comprehensive nature of medical information captured, and for the inclusion of detailed data on service utilization and cost data. Administrative claims databases are therefore used to study health outcomes, including health care costs and utilization.

TABLE 1. Summary of Database Examples

Database/Sponsor/Web site	Population	Coverage			Availability		Data collected				Patient follow-up	Updated	Comments
		L	R	N	Pub	Priv	Clin	PRO	Pharm	Costs			
Electronic medical records (EMRs)													
Centricity EMR • General Electric • http://www3.gehealthcare.com/en/Products/Categories/Healthcare_IT/Electronic_Medical_Records/Centricity_EMR	Used by over 35,000 clinicians across the US and contains over 30 million patient records			√		√	√		√			Quarterly	Nationally representative population with data from 1996
University of Utah Electronic Data Warehouse • University of Utah • http://uuhsc.utah.edu/drc/summary.html	Over 1.4 million patients across the state of Utah		√		√		√		√	√*		Monthly	Regional database with from 1990
Administrative claims													
MarketScan • Truven Health Analytics • http://marketscan.truvenhealth.com/marketscanportal/	Over 100 insurance providers, and about 154 million patients			√		√	√**		√	√		Monthly	National claims database with large population, fully integrated drug and medical claims, and robust data from diverse sources
Clinformatics Data Mart • Optum • http://www.optum.com/life-sciences/differentiate-products/marketing-analytics/clinformatics-data-mart.html	50 million patients across the US			√		√	√**		√	√	36+ months	Semiannually/ Quarterly	National claims database with large population, fully integrated drug, medical claims, and lab test results
Integrated health systems													
Veterans Health Administration • Department of Veterans Affairs (VA) • http://www.va.gov/health/	20 million patients			√	√		√		√	√			Population has a much higher proportion of males and is older.
Geisinger Health System (GHS) • MedMining • http://www.medmining.com	3 million patients		√			√	√	√†	√	√			Regional integrated health system located in central Pennsylvania
National surveys													
Behavioral Risk Factor Surveillance System (BRFSS) • Centers for Disease Control and Prevention (CDC) • http://www.cdc.gov/brfss/	Random sample of each state's noninstitutionalized civilian population 18 years of age and older			√	√		√	√	√		Cross-sectional	Yearly	Information about health-related knowledge, attitudes, and practices with limited patient-reported clinical and pharmacy information

(Continued on next page)

TABLE 1. Summary of Database Examples *(Continued)*

Database/Sponsor/Web site	Population	Coverage			Availability		Data collected				Patient follow-up	Updated	Comments
		L	R	N	Pub	Priv	Clin	PRO	Pharm	Costs			
Medical Expenditure Panel Survey (MEPS) • Agency for Research and Health Quality (AHRQ) • http://meps.ahrq.gov/mepsweb/	Random sample of noninstitutionalized civilian population and families, their medical providers, and employers across the US			√	√			√	√	√	Panel data (2 years for each panel)	Every 2 years	Provides national and regional estimates for health utilization, insurance coverage, health care expenditures, and health status
National Health and Nutrition Examination Survey (NHANES) • Centers for Disease Control and Prevention (CDC) • http://www.cdc.gov/nchs/nhanes.htm	Random sample of noninstitutionalized civilian population of all ages with over sampling people of 60 and older, African Americans, and Hispanics			√	√		√	√	√		Cross-sectional	Yearly	Demographic information, socioeconomic status, dietary, and health status from interviews; and measures medical, dental, physiologic outcomes, and labs from physical examinations
Registries													
PAIN OUT • PAIN OUT consortium and the International Association for the Study of Pain (IASP) • http://www.pain-out.eu/	Patients, ≥18 years old, with postoperative pain			√‡	√		√	√			Minimum of 1 day	N/A	Data are collected regarding the postoperative pain, pain management, and quality of life of patients.

L = local; R = regional; N = national; Pub = publicly available; Priv = privately available; Clin = clinical data; PRO = patient-reported outcomes; Pharm = prescription variables; Costs = medical or pharmacy cost data.

*Charges.

**Selected laboratory values on subset of population.

†Ability to survey patients.

‡Multinational.

One of the strengths of administrative claims databases is that they contain a very large number of patients. It is not uncommon for patient populations to reach close to 100 million when the claims data are pooled across many insurance providers. This allows for the examination of a wide variety of diseases and medications, including relatively rare diseases.

Administrative claims databases also reflect a nearly complete history of a patient's health care resource utilization and costs as covered by a given payer. This is in contrast to EMR data sets, which may lack information on care delivered by other providers. Along these lines, administrative claims databases include claims for each time a prescription was filled and paid for by the health insurance company. This allows for the assessment of medication adherence calculated from pharmacy claims. However, even with a nearly complete capture of medical encounters and pharmacy dispensing data, there are potential gaps in claims data that must be considered. In patients with chronic pain, but especially in opioid abusers, patients may exhibit shopping behavior or pay cash for prescriptions and office visits. Although a small number of patients may exhibit opioid shopping behavior, those that do are more likely to pay cash for prescriptions and office visits.[13] An encounter or prescription that was paid for with cash may not have a claim submitted and may not be captured in claims data. Additionally, patients may qualify for coverage with multiple payers, and claims from one system would not generally include information from claims in another system.

One of the limitations of administrative claims databases is the absence of clinical data. Although claims data sets contain diagnoses, procedures, and prescriptions, the data do not include clinical variables such as height, weight, body mass index (BMI), vital signs, pain scores, or generally laboratory measures. Some administrative claims databases have integrated limited laboratory test result data from large laboratory providers, but a large majority of patients have no or only sporadic data. Given the general lack of clinical data, researchers using administrative claims data must rely on changes in related resource utilization and costs (e.g., change in use of rescue medications of emergency department visits) as surrogate outcomes. Another limitation of administrative claims databases is in the accuracy of diagnosis codes. For instance, providers may submit claims with rule-out diagnoses when trying to diagnose nonspecific symptoms, which may lead to classification bias. This limitation may be overcome by requiring a second "confirmatory" diagnosis of the disease of interest.

Generalizability is also an issue with claims databases, as they are developed based on patients having insurance or a particular type of payer. Patients with Medicaid or Medicare may or may not be included in a largely commercial administrative claims database, whereas patients with commercial insurance will not be in a Medicaid data set. The uninsured or cash-paying patients will not be captured. Additionally, when patients or their employer switch health insurance companies, that patient will no longer be followed in the database. As a result, administrative claims data sets may be very complete at any point in time, but they generally lack long-term follow-up data.

Given the strengths of administrative claims data, these data sets have been used in pain outcomes research studies. The complete capture of medical encounter and pharmacy dispensing data is a benefit especially when multiple providers treat patients. This is a common occurrence for patients with chronic conditions who may be treated by primary care and specialty providers. This situation is likely exacerbated in chronic pain, as patients may see multiple providers and have high emergency department utilization, making claims databases well suited for pain outcomes research.

Administrative claims data limitations must be recognized, but may be mitigated in pain outcomes research. A key limitation is the lack of pain scores. However, researchers may consider utilization of pain-related care as a measure of outcomes. For example, in a study by Ivanova et al., the effectiveness of duloxetine for the treatment of chronic low back pain was measured by the utilization of other treatment modalities after initiation of duloxetine.[14]

One example of a large administrative claims database used in research is the Truven Health Analytics' MarketScan Claims database (Table 1). MarketScan is a nationally representative database that includes medical and pharmacy claims data from over 100 insurance providers and 170 million unique patients.[15] MarketScan is largely composed of patients with employer-sponsored insurance coverage, but MarketScan also has Medicare and Medicaid components. MarketScan also contains selected laboratory values for a subset of patients and productivity data for employees, which is valuable in pain research. Another example is the Optum Clinformatics Data Mart, which is a large commercial insurer database that includes approximately 50 million patients with both medical and pharmacy claims data from across the US.[16] Optum also has some data from Medicare and selected laboratory test results. Additionally, Optum has recently purchased a large EMR data set and

is working to integrate the clinical data for some of the patients with claims data.

Administrative claims data sets have been frequently used in pain and pain treatment outcomes research. A recent example includes a study by Pergolizzi et al.[17] This study used the MarketScan Commercial Claims and Encounters database to examine the prevalence of major drug-drug interactions in patients prescribed long-acting opioids and the health care costs associated with these interactions. Major drug-drug interactions with opioids were identified based on cytochrome P450 interactions and the potential to cause major adverse drug reactions. The study found that 5.7% of patients prescribed long-acting opioids were exposed to major drug-drug interactions. The most frequently identified drug-drug interaction was between oxycodone and fluconazole. Additionally, it was estimated that major drug-drug interactions increased monthly health care costs by $609. As there is a paucity of data available about opioid-related drug-drug interactions, this study highlights the potential to use claims databases to examine important questions regarding drug utilization and costs in pain.

Integrated Health Systems

EMR and administrative claims databases each provide important data for outcomes research, but neither provides a complete picture of patient care and outcomes. Given the fragmented health care system in the US and the lack of common health identifier, linking EMR and claims data is exceedingly difficult. However, this is not an issue when the payer and provider are a single entity, as in integrated health systems. These are typically local or regional health care systems that have both administrative claims and clinical EMR information on a subset of their total patient population.

An obvious strength of integrated health systems is that they have comprehensive data for patients with health plan coverage, as it combines the completeness and economic data of a claims data set with the rich clinical data found in an EMR. With a more complete medical picture, researchers can examine important outcomes and address potential biases that may arise in observational studies. For example, medication adherence from claims data can be considered while examining clinical outcomes such as pain scores from the EMR. An additional strength is that, in some areas, integrated health systems are the predominant providers of both health care and health insurance. This may lead to patients being in the health system for a relatively long period time and allows the study of the long-term impact of health interventions.

One of the limitations of integrated health systems is that they are often local or regional and may have limited generalizability. Additionally, integrated health systems may have different systems of care in place that may not be representative of the care received by the general population. For example, an integrated system may have incentive-driven quality improvement programs in place that can lead to a greater standardization of care than is seen outside the integrated care setting. Similarly, due to their local or regional nature, integrated health systems databases may have fewer patients with a given treatment or disease than national administrative claims or EMR databases. Thus, sample size may be a limiting factor depending on the disease, treatment, or outcome of interest.

Integrated health systems can be a valuable source of data for pain-related outcomes research. Many of the limitations seen with claims or EMR data can be overcome because of the ability to integrate clinical data with utilization data when assessing outcomes. An example would be the ability to control for hemoglobin A_{1c} levels and medication adherence when assessing outcomes of medication treatment for diabetic painful neuropathy. Some limitations of claims and EMR data are not, however, overcome with integration, such as the lack of pain scores in a structured data in EMRs, or gaps in utilization data for patients who are doctor shopping or paying cash for opioid prescriptions.

One example of a large integrated health system is the Department of Veterans Affairs (VA), which primarily provides health care for qualified, former members of the armed forces (Table 1). The VA is the largest integrated health system in the US and had approximately 83.6 million outpatient visits and over 700,000 inpatient stays in 2012. The VA has data for more than 20 million patients.[18] Another example of an integrated health system is the Geisinger Health System (GHS) in central Pennsylvania. GHS comprises over 800 physicians and provides care for over 3 million patients, of which approximately one third of the patients also have coverage with the Geisinger Health Plan.[19]

Hausmann et al. used integrated health system data in a recent pain-related outcomes study.[20] Their historical study used the electronic health records from the Pittsburgh VA health care system to assess racial disparities in applying recommended guidelines for opioid monitoring and treatment. Patients with noncancer pain who were prescribed opioids for over 90 days were included. Using this database, they were able to look at all aspects of patient care, including inpatient and outpatient visits, pharmacy utilization, laboratory testing, and referrals. The study

found that, when compared with white patients, pain was documented less frequently among black patients (61.1% vs. 73.3%; $P < .001$) and they were referred to pain specialists less often (15.0% vs. 22.1%; $P = .01$). Additionally, compared with white patients, black patients were subjected to more urine drug tests (mean: 5.8 vs. 3.5; $P = .04$) and referrals for substance abuse assessments (7.9% vs. 3.6%; $P < .001$). This study highlights the ability to access and analyze rich data in integrated health systems.

National Surveys

National health surveys represent another source of data for outcomes research. These surveys are typically initiated and funded by federally sponsored organizations such as the Centers for Disease Control and Prevention (CDC) or the Agency for Healthcare Research and Quality (AHRQ). The primary purpose of national surveys varies, but they all provide information on the health of the US population as a whole, most of which is patient reported. The surveys are typically conducted at regular intervals and draw a random sample of thousands of individuals or households, which are then weighted to be representative of the entire US population. The data are usually collected through face-to-face or telephone interviews with individuals and households. However, select surveys gather additional data varying from physical and laboratory examinations to contacting health care providers and employers. Most of the surveys are cross-sectional in nature and interview participants only once, but there are several that interview participants multiple times over a specific time period.

The information gathered from patients differs depending on the survey purpose. Most collect demographic information to correctly weight the population. Other information varies, but may include current and past medical history, prescription medication use, and the costs associated with receiving health care. Additionally, participants may be asked questions regarding family members or household, socioeconomic status (e.g., marital status, education, employment status), health-related risk behaviors (e.g., smoking history, alcohol consumption, sexual activity, drug abuse), quality of life, and others (e.g., health care access, immunizations, disease awareness, dietary information, physical activity, seatbelt use). Many national surveys also collect data on pain, which will be discussed below.

A major strength of national survey data is that, when weighted correctly, they are representative of the general population of the US. This allows examination of disease prevalence over time and provides insight into associations between many risk factors and diseases. Additionally, these databases include individuals who may otherwise not be captured because they are not seeking care. By their very nature, EMRs, administrative claims, and integrated health systems databases require an encounter or interaction with the health care system to be included. Furthermore, these databases may contain information not readily accessible or available in EMRs, administrative claims, or integrated health systems data, such as information on health-related behaviors and self-reported health status. Another benefit of these databases is that they are usually free to the public (or with nominal fees), whereas other databases may cost a significant amount for data access.

One limitation of national surveys is that many are cross-sectional, which prevents the assessment of longitudinal or temporal relationships between treatment and outcomes (i.e., can assess association but not causation). Another limitation is that, although they are nationally representative, most surveys do not capture institutionalized subsets of the US population such as patients residing in long-term care facilities, incarcerated individuals, and those serving in the military. Similarly, individuals are often contacted to participate via telephone. Individuals without access to a telephone, such as the homeless, will likely not be sampled. Another limitation of national surveys is that the data are often self-reported and many surveys do not include data from physical examinations or laboratory testing.

A few examples of national surveys include the National Health and Nutrition Examination Survey (NHANES), the Behavioral Risk Factor Surveillance System (BRFSS), and the Medical Expenditure Panel Survey (MEPS). These surveys can be used in pain-related research, with noted strengths and limitations (Table 1).

NHANES, initiated by the CDC in 1960, is a nationwide program to assess the health and nutritional status of adults and children in the US. Approximately 5000 individuals are interviewed per year, with a subset of patients undergoing physical examinations and laboratory assessment. NHANES is a widely used source of data to assess the prevalence and trends of chronic disease in the US, with a core component collected each year. However, NHANES also collects data on specific topics that vary over time, with several topics in pain assessed in the past 15 years. For example, data on miscellaneous pain were collected from 1999 to 2004, data on inflammatory arthritis were collected from 2009 to 2010, and data on muscle pain have been collected since 2011.

The BRFSS is a nationwide, state-based health survey to assess health-related risk behaviors, chronic diseases, and use of preventative practices. Initiated

by the CDC in 1984, BRFSS is the largest survey in the world, completing approximately 400,000 interviews every year. BRFSS has questions specifically related to pain and pain control, including cancer pain, joint pain, medication use for pain, and interference with daily activities due to pain. Additionally, states may add their own questions to the BRFSS. For example, a study about chronic pain and its treatment with prescription drugs is based on the 2007 Kansas BRFSS, which contained 10 questions surrounding chronic pain.[21]

MEPS is a set of nationwide surveys administered by AHRQ since 1966, which provides national estimates of health care utilization, insurance coverage, health care expenditures, and health status. Unlike many other surveys, MEPS consists of five interviews with an individual over 2.5 years. This panel approach allows for a longitudinal examination of how changes in health affect medical expenditures. In addition to collecting information on utilization and expenditures, MEPS also includes the 12-question version of the Short Form Health Survey (SF-12), which includes questions about the presence and severity of pain and how pain influences patient activities.

An example of a pain-related study based on national survey data is a case-control study by Lanier et al.[22] The authors identified opioid-related decedents from Utah Office of the Medical Examiner and then utilized BRFSS respondents as a control group to determine risk factors for prescription opioid-related deaths in Utah. The study used interviews of the opioid-related decedents' next-of-kin and the Utah 2008 BRFSS to assess health-related behaviors and risk factors reported by survey respondents. Questions included the use of prescription pain medications during the previous year, patterns of opioid use, source of opioids, and the presence of chronic pain. For the analysis, the BRFSS population was weighted to reflect Utah adult population. The study found that decedents were more likely to have obtained pain medication from a nonprescription source (exposure prevalence ratio [EPR]: 4.8, 95% confidence interval [CI]: 3.6–6.0] and used more medication than was prescribed (EPR: 16.5, 95% CI: 9.3–23.7) than controls. This study highlights the value of survey data in hypothesis generation and in assessing potential associations between pain treatments and outcomes.

Patient Registries

Patient registries are population-specific, longitudinal data sets developed to support specialized care or to track specific information on the population of interest. Registries may be designed to establish the natural history of a disease; to examine the clinical effectiveness or monitor the safety of a health care intervention; to capture the quality of life of patients with a disease; or to examine the cost-effectiveness of treatments.[23] Typically, registries are classified by how the patients are identified and included, such as patients with a specific disease or exposure to a drug, device, or procedure. Registries are also often developed to track data not routinely captured in patient care. However, registries may interface with other electronic health data to avoid duplication of effort and to supplement the registry with additional data to support outcomes assessments.

Registries collect a defined set of information on all registry patients in a predetermined format. Thus, a benefit of registries is that they are, theoretically, complete, specific, and uniform. However, information captured by registries is not necessarily limited to the required fields. Other information may be collected, but may not be standardized and structured in the same way. Because registries are created for many different purposes, and for a wide variety of diseases and exposures, the information contained in them varies significantly from one registry to the next.

Although registries may be designed to meet a specific local need, nationwide patient registries, often for uncommon or rare diseases or adverse events, also exist. Furthermore, registries may be created by state or federal agencies with mandatory reporting requirements to capture data that would not otherwise be captured.

One of the limitations of patient registries is that they are generally designed for a specific purpose and collect a specific, and generally limited, set of data. Thus, exploring outcomes or treatment trends, or adjusting for other patient variables, may be difficult when these were not considered during the registry design. Another limitation of registries is that most registry reporting is voluntary, thus missing data from patients and underreporting can be an issue. Further, patients may choose to no longer participate, and clinics or providers may not adequately follow up with patients to capture data on desired outcomes.

Given the design and purpose, registry data can be use for pain-related outcomes research. For instance, a pharmaceutical company may be required by the FDA to establish a registry to monitor the safety and effectiveness of a new drug as part of its risk management system. In this scenario, physicians or patients may be asked to submit information to the registry regarding pain relief and adverse events. The benefit of a pain-specific registry is the opportunity to collect patient-reported pain outcomes, such as pain scores and quality of life. Such data are often not collected in routine care or not captured in a systematic manner.

A registry can also be designed to better capture data on the adverse or unwanted effects of pain treatment that would likely not be captured or inadequately captured in a claims data set or EMR. However, the ability to assess outcomes with registries is highly dependent on the design of the registry, and as such, registry data may be inflexible in terms of the scope of outcomes that can be assessed.

An example of a patient registry in pain is the PAIN OUT registry (Improvement in postoperative PAIN OUTcome), which is operated by the PAIN OUT consortium and the International Association for the Study of Pain (Table 1).[24] The purpose of the PAIN OUT registry is to improve the clinical care of patients with postoperative pain. PAIN OUT is the largest database on postoperative pain, with more than 200 hospitals participating from around the world. PAIN OUT uses a Web-based questionnaire, available in 20 languages, to collect patient-reported outcomes and clinical data. Participating hospitals are then able to receive feedback about their quality of care and respond accordingly. The PAIN OUT registry has recently been used to compare postsurgical pain management and patient-reported outcomes in hospitals in America with hospitals in Europe.[25] PAIN OUT was also used to examine the association between early postoperative pain and health-related quality of life in patients.[26]

Using PAIN OUT registry, Chapman et al. were able to compare patient-reported outcomes of postoperative treatment and pain control after orthopedic surgery in American institutions with their European counterparts.[25] Patient-reported outcomes included characteristics of pain, emotional comfort, and satisfaction with care provided. The study found that, when compared with the European patients, more American patients received opioids on the first day after surgery (98.3% vs. 70.2%; $P < .001$) and more stated they wanted more treatment (34.7% vs. 15.3%; $P < .001$). Additionally, American patients reported to have higher "worst pain" scores than European patients (mean 7.39 vs. 5.40; $P < .001$). This study shows the potential to use registries to collect and analyze data in pain that is not routinely captured in other data sources.

DISCUSSION

The prevalence and costs of chronic pain and chronic pain management in the US is high, with almost a third of the US population suffering from long-term pain. Thus, comprehensive examination of the economic, clinical, and humanistic outcomes in patients with pain from observational studies, considered alongside data from RCTs, is necessary to provide a more comprehensive picture of the impact of health care interventions and management strategies.[27]

Conducting real-world analyses has notable challenges, including the identification of appropriate data sources to address research questions. This paper aims to facilitate this effort by describing characteristics of data sources commonly used in outcomes research in the US and to characterize general strengths and limitations of various data types for use in pain research. Specifically, five categories of databases were discussed: (1) EMRs, (2) administrative claims, (3) integrated health systems, (4) national surveys, and (5) patient registries. Although each of these databases has their own unique strengths and limitations, with appropriate methodology they all can be used to answer important outcomes research questions in pain.

One of the factors behind the limitations of secondary data sets stems from the fact that many these data sources were not designed for research purposes. Therefore, they may not contain all of the information desired. Thus, choosing data sources and using appropriate study designs is vital to reduce the risk of confounding and other biases. Another potential limitation of using secondary data sets for research is accessibility. Many of the data sources discussed are available commercially (e.g., Truven Health MarketScan, GHS through MedMining), but may have significant costs. Commercially available claims databases may range from $60,000 to $80,000, whereas EMRs may range from $90,000 to $100,000. Other noncommercially available data sources (e.g., VA, University of Utah) may have very limited availability to researchers outside the organization.

In conclusion, pain-related outcomes research attempts to demonstrate the effectiveness and cost-effectiveness of pain management interventions as they are used in everyday practice settings. However, using real-world data for outcomes studies comes with challenges that must be recognized and addressed to generate meaningful evidence. Clinicians and health care administrators should be aware of these challenges. They should also be familiar with the strengths and limitations of data sources to accurately interpret real-world evidence and apply the evidence to patient- and population-level pain-management decisions.

Declaration of interest: The authors report no conflicts of interest. The authors alone are responsible for the content and writing of the paper.

REFERENCES

[1] Institute of Medicine. Relieving Pain in America: A Blueprint for Transforming Prevention, Care, Education, and Research. Washington, DC: The National Academies Press; 2011.

[2] Go AS, Mozaffarian D, Roger VL, et al. Heart disease and stroke statistics—2014 update: a report from the American Heart Association. *Circulation.* 2014;*129*:e28–e292.

[3] Cowie CC RK, Byrd-Holt DD, Gregg EW, et al. Prevalence of diabetes and high risk for diabetes using A1C criteria in the U.S. population in 1988–2006. *Diabetes Care.* 2010;*33*:562–568.

[4] Howlader N, Krapcho M, Garshell J, et al., eds. SEER Cancer Statistics Review, 1975–2010. Available at: http://seer.cancer.gov/csr/1975_2010/. Accessed March 11, 2014.

[5] Gaskin DJ, Richard P. The economic costs of pain in the United States. *J Pain.* 2012;*13*:715–724.

[6] Food and Drug Administration. Analgesic Clinical Trials Innovation, Opportunities, and Networks (ACTION) Initiative. 2010. Available at: http://www.fda.gov/AboutFDA/PartnershipsCollaborations/PublicPrivatePartnershipProgram/ucm231130.htm. Accessed March 19, 2014.

[7] Zulman DM, Sussman JB, Chen X, Cigolle CT, Blaum CS, Hayward RA. Examining the evidence: a systematic review of the inclusion and analysis of older adults in randomized controlled trials. *J Gen Intern Med.* 2011;*26*:783–790.

[8] Park HY, Schumock GT. Overview of Outcome Research. *Book 5: Science and Practice of Pharmacotherapy I and II. Pharmacotherapy Self-Assessment Program*. Fifth ed. Kansas City, MO: American College of Clinical Pharmacy; 2005.

[9] Heintzelman NH, Taylor RJ, Simonsen L, et al. Longitudinal analysis of pain in patients with metastatic prostate cancer using natural language processing of medical record text. *J Am Med Inform Assoc.* 2013;*20*:898–905.

[10] Miller ML, Ruprecht J, Wang D, et al. Physician assessment of disease activity in JIA subtypes. Analysis of data extracted from electronic medical records. *Pediatr Rheumatol Online J.* 2011;*9*:9. doi:10.1186/1546-0096-9-9

[11] Brixner DI, Bron M, Bellows BK, et al. Evaluation of cardiovascular risk factors, events, and costs across four BMI categories. *Obesity.* 2013;*21*:1284–1292.

[12] Jennings BT, Marx CM. Implementation of a pharmacist-managed diabetes program. *Am J Health Syst Pharm.* 2012;*69*:1951–1953.

[13] Cepeda MS, Fife D, Chow W, Mastrogiovanni G, Henderson SC. Opioid shopping behavior: how often, how soon, which drugs, and what payment method. *J Clin Pharmacol.* 2013;*53*:112–117.

[14] Ivanova JI, Birnbaum HG, Kantor E, Schiller M, Swindle RW. Duloxetine use in chronic low back pain: treatment patterns and costs. *PharmacoEconomics.* 2012;*30*:595–609.

[15] Hansen L, Chang S. Health research data for the real world: the MarketScan Databases. A White Paper. 2012. Available at: http://truvenhealth.com/Portals/0/assets/2012_Truven_MarketScan_white_paper.pdf. Accessed May 28, 2014.

[16] McAdam-Marx C, McGarry LJ, Hane CA, Biskupiak J, Deniz B, Brixner DI. All-cause and incremental per patient per year cost associated with chronic hepatitis C virus and associated liver complications in the United States: a managed care perspective. *J Manag Care Pharm.* 2011;*17*:531–546.

[17] Pergolizzi JV, Ma L, Foster DR, et al. The prevalence of opioid-related major potential drug-drug interactions and their impact on health care costs in chronic pain patients. *J Manag Care Pharm.* 2014;*20*:467–476.

[18] Bellows BK, LaFleur J, Kamauu AW, et al. Automated identification of patients with a diagnosis of binge eating disorder from narrative electronic health records. *J Am Med Inform Assoc.* 2014;*21*:e163–e168.

[19] McAdam-Marx C, Bellows BK, Unni S, et al. Impact of adherence and weight loss on glycemic control in patients with type 2 diabetes: cohort analyses of integrated medical record, pharmacy claims, and patient-reported data. *J Manag Care Pharm.* 2014;*20*(7):691–700.

[20] Hausmann LR, Gao S, Lee ES, Kwoh CK. Racial disparities in the monitoring of patients on chronic opioid therapy. *Pain.* 2013;*154*:46–52.

[21] Toblin RL, Mack KA, Perveen G, Paulozzi LJ. A population-based survey of chronic pain and its treatment with prescription drugs. *Pain.* 2011;*152*:1249–1255.

[22] Lanier WA, Johnson EM, Rolfs RT, Friedrichs MD, Grey TC. Risk factors for prescription opioid-related death, Utah, 2008–2009. *Pain Med.* 2012;*13*:1580–1589.

[23] Gliklich RE, Dreyer NA, eds. Registries for Evaluating Patient Outcomes: A User's Guide. 2nd ed. Rockville, MD: Agency for Healthcare Research and Quality; 2010. (Prepared by Outcome DEcIDE Center [Outcome Sciences, Inc. d/b/a Outcome] under Contract No. HHSA29020050035I TO3.)

[24] Rothaug J, Zaslansky R, Schwenkglenks M, et al. Patients' perception of postoperative pain management: validation of the International Pain Outcomes (IPO) questionnaire. *J Pain.* 2013;*14*:1361–1370.

[25] Chapman CR, Stevens DA, Lipman AG. Quality of postoperative pain management in american versus European institutions. *J Pain Palliat Care Pharmacother.* 2013;*27*:350–358.

[26] Taylor RS, Ullrich K, Regan S, et al. The impact of early postoperative pain on health-related quality of life. *Pain Pract.* 2013;*13*:515–523.

[27] Kozma CM, Reeder CE, Schulz RM. Economic, clinical, and humanistic outcomes: a planning model for pharmacoeconomic research. *Clin Ther.* 1993;*15*:1121–1132; discussion 1120.

RECEIVED: 28 May 2013
REVISED: 23 June 2014
ACCEPTED: 30 June 2014