IBM Watson Health™

WHITE PAPER

IBM MarketScan Research Databases for Health Services Researchers

Contents

Introduction

The IBM® MarketScan® Research Databases are a family of research data sets that integrate de-identified patient-level health data (medical, drug and dental), productivity (workplace absence, short- and long-term disability and workers' compensation), laboratory results, health risk assessments (HRAs), hospital discharges and electronic medical records (EMRs) into data sets available for healthcare research. Data are contributed by large employers, managed care organizations, hospitals, EMR providers, Medicare and Medicaid.

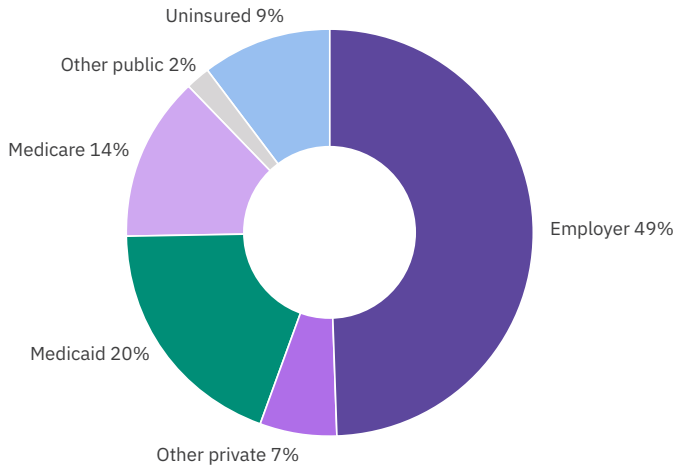
This white paper describes the features and uses of MarketScan Research Databases for health services research. Specifically, the paper illustrates the different attributes of individual MarketScan Research Database, explains how these databases are constructed, describes their uses and highlights examples of published studies based on IBM® MarketScan® data.

Filling a data need: Origins of the MarketScan Research Databases

In response to rising costs, fundamental changes occurred in the US healthcare system in the late 1980s as healthcare delivery shifted toward managed care arrangements. At the same time, there was growing interest in greater accountability for care through quality improvement. Stakeholders sought data on how these changes impacted costs, quality of care, health outcomes and cost-effectiveness. As the purchasers and payers for the privately insured segment of the US population, employers and health plan administrators were interested in accurate and timely information on the drivers of cost growth and the returns on investment for initiatives designed to improve employee health and well-being. Healthcare policymakers and practitioners were interested in the prevalence, incidence and costs of specific diseases, as well as the effectiveness and cost implications of interventions, clinical guidelines and quality improvement initiatives. Providers, healthcare facilities and health services companies were interested in the cost-effectiveness of different therapies in real-world clinical care.

At the time, data sources to support these analyses were typically inadequate. Importantly, there was a lack of reliable healthcare data on privately insured patients and their families. This group continues to comprise the largest segment of US healthcare users—nearly half of the total US population (see Figure 1).

Figure 1: US population distribution by insurance status – 2016



Source: Kaiser Family Foundation estimates. http://kff.org/other/state-indicator/total-population/. Accessed March 2017.

Truven Health Analytics®, acquired by IBM in 2016 and now part of the IBM Watson Health™ business, created MarketScan to address the need for better healthcare data on privately insured Americans. Since its creation, MarketScan has evolved into a suite of proprietary databases that contain one of the longest-running and largest collection of privately and publicly insured, de-identified patient data in the United States. Claims data reflects real-world treatment patterns and costs by tracking millions of patients as they travel through the healthcare system; therefore, they offer researchers detailed information about multiple aspects of care. Data from individual patients are integrated from providers of care, and healthcare utilization and cost record connections are maintained at the patient level.

Over the years, the original claims-centric databases have been enriched and integrated with the addition of absence, disability, workers' compensation, health risk, lab, dental, EMR and hospital data.

Features of MarketScan data

MarketScan databases offer several distinct advantages over other types of data sources.

Very large sample size allows for research on unique patient populations

MarketScan databases offer some of the largest convenience samples available in proprietary US databases—with over 245 million unique patients since 1995. In the most recent full data year, MarketScan databases contain healthcare data for more than 43.6 million covered individuals—large enough to allow creation of a nationally representative data sample of Americans with employer-provided health insurance. The size of the databases helps to maintain analytically sufficient cohort sizes when deep segmentation of patients is required, particularly for orphan diseases.

Complete episodes of care can support more inclusive cost and treatment studies

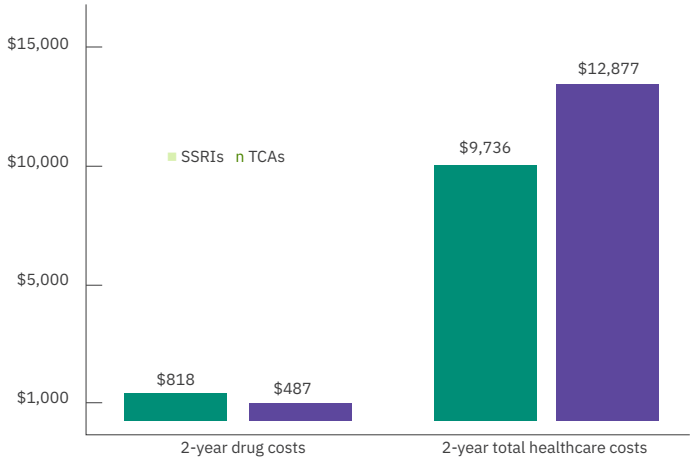
MarketScan databases capture the continuum of care: physician office visits; hospital stays; retail, mail order and specialty pharmacies; and carve-out care, such as mental health services. The importance of complete episodes of care is illustrated by the following cost offset case study.

Case study: Cost offset in treating depression

Establishing the value of a medication based on total healthcare costs as opposed to the direct cost of medication alone can be critical to assessing cost-effectiveness. Why? Because a more expensive drug therapy may produce better overall health outcomes and reduce long-term medical costs.

The study illustrated in Figure 2 compared the cost of two different prescription therapies for treating depression: selective serotonin reuptake inhibitors (SSRIs) and tricyclic antidepressants (TCAs). Using MarketScan data, researchers revealed that the two-year average prescription cost of TCAs was lower, but the overall cost of treatment using TCAs was higher. This demonstrated that although drug costs may be lower, it is important to look at overall healthcare costs when evaluating the cost-effectiveness of a drug treatment.

Figure 2. Managing populations: Cost offset for treating depression with SSRIs versus TCAs



Source: Crown W, Treglia M. Anti-depressant selection, drug use patterns, and two-year health care expenditures. International Health Economics Association Meeting, June 6-9, 1999, Rotterdam, Netherlands

Linked data can enhance research in many disease areas

MarketScan has the ability to link healthcare data from a variety of sources.

- Linking hospital discharge records with claims data at the patient level can fill the gap in a patient's drug therapy between inpatient and outpatient settings.
- EMRs, when linked with claims data, can add significant clinical richness to the healthcare history of patients.
- Lab results can assist in studies of diseases such as diabetes, where HbA1c is a critical indicator of treatment outcome.
- Indirect cost research can be conducted because MarketScan claims data come predominantly from employers that also provide data on absenteeism, disability and workers' compensation, all linked to the corresponding employee's claims.
- Medical data linked to dental claims, allowing research in the area of cardiovascular disease and its relationship to oral health, for example.
- Health services customers have also asked our team to undertake unique linking projects between MarketScan data and registry and in-house customer data.

Detailed prescription drug information

MarketScan databases contain robust information on outpatient prescriptions. Through the Early View subscription option, data are available within 45 days of the end of the service month, with monthly or quarterly updates. The MarketScan databases may afford distinct advantages over others that track only prescription fills because MarketScan data allows identification of the type of disease from medical claims, and they can be used to determine whether clinical, demographic and/or provider characteristics influence prescribing patterns.

Prescription fills for individual patients are recorded so that therapies prescribed concurrently—and presumably used in combination—can also be identified. This provides helpful information about actual drug use patterns, as opposed to individual drug prescription trends.

IBM® MarketScan® Hospital Drug Database provides researchers with inpatient drug utilization data from hospital discharge records. These data allow researchers to evaluate drug use in the inpatient and outpatient settings, including hospital use patterns, switching behavior, combination therapy and patient characteristics. This information can be used to determine if introduction or early use of a product could improve clinical and overall cost outcomes. These data can also be used to analyze diagnosis volumes. When researchers need to evaluate the impact of a hospitalization on prescription drug use, the IBM® MarketScan® Inpatient Drug Linked Data Set links outpatient data from claims with inpatient drug data from the MarketScan Hospital Drug Database, enabling researchers to do an analysis review.

High-quality coding

We believe an advantage of MarketScan data involves robust and high-quality coding. Key examples include:

- Fully paid and adjudicated claims
- Payment and financial information, including the amount that is the patient's responsibility
- Outpatient prescription drug information, including patient copayments, mail order prescriptions, information about injectable treatments, data from specialty pharmacies and carve-out services, manual and electronically submitted claims and plan or formulary summaries

Numerous and widely published research applications

MarketScan-based research has made a contribution to the body of literature used to formulate policy decisions and help improve healthcare for Americans. The first publication appeared in 1990 in the New England Journal of Medicine (J.B. Hillman, et al.). [1] Since then, more than 1,400 articles have appeared in major peer-reviewed journals. MarketScan data have supported a range of health services research conducted by government, academic and private researchers. Studies have been in the areas of:

- Economic burden of illness
- Clinical research
- Economic costs of health risks
- Health and workforce productivity
- Dental research
- Benefit plan design and adherence
- Adverse event rates
- Treatment outcomes
- Population studies
- Comparative effectiveness research

Limitations of the data

As with any data source, MarketScan data have limitations. Some limitations have to do with the nature of claims data and others with the nature of the MarketScan sample population. Limitations include:

- MarketScan databases are based on a large convenience sample. Because the sample is not random, it may contain biases or fail to generalize well to other populations. However, these data can complement other data sets or be used as benchmarks against them.
- Data come mostly from large employers; medium and small firms may be underrepresented, although the MarketScan Research Databases include a large amount of data contributed from health plans.
- Accessing the data requires data management software or programmer support.

Overview of MarketScan data

How the data sets are built

MarketScan databases are constructed by collecting data from employers, health plans and state Medicaid agencies who are our customers and have agreed to be data contributors. Data comprise service-level claims for inpatient and outpatient services and outpatient prescription drugs. All claims have been fully paid and adjudicated. We standardize financial, clinical and demographic fields and add contributor-specific fields. Drug detail (for example, therapeutic class, therapeutic group, manufacturer's average wholesale price and a generic product identifier) and clinical detail (for example, disease episode grouper) are also added.

A unique enrollee identifier is assigned to each individual in a MarketScan database. This identifier is created by encrypting information provided by data contributors. We then combine the standardized fields of the individual databases and create links between years of data and across data types. Data are collected for the MarketScan annual database releases when nearly 100 percent of claims have been paid; this removes the need for completion factors and helps improve the reliability and accuracy of the data.

Protecting the privacy of patient data, as well as the privacy of our customers, is one of our core principles. Therefore, the MarketScan Research Databases are designed to address the requirements of the Health Insurance Portability and Accountability Act of 1996 (HIPAA). The MarketScan Research Databases meet the criteria for a limited-use data set[2] and contain none of the data elements prohibited by HIPAA for such data sets.[3] In addition, we have taken steps to go beyond these HIPAA requirements. The MarketScan databases have undergone statistical analysis by a third party to verify that they meet HIPAA requirements for fully de-identified data sets.[4]

Although meeting these requirements is optional given the current MarketScan licensing process, this additional step demonstrates our commitment to HIPAA compliance and to helping safeguard the confidentiality of patient- level and provider-level data. Finally, all patient- level and provider-level data within the MarketScan Research Databases contain synthetic identifiers to help safeguard the privacy of individuals and data contributors.

We perform additional enhancements to the data during database creation. These include:

- Comparing diagnosis and procedure codes to codes that were in effect when the raw data were collected; editing the diagnosis and procedure codes, if necessary
- Adding major diagnostic categories (MDCs) and diagnosis-related groups (DRGs) to claims, along with the application of other classification systems, such as outpatient treatment groups and disease staging
- Identifying the type of plan, such as health maintenance organization (HMO), preferred provider organization (PPO) and point-of-service (POS) or comprehensive plans
- Verifying that both the experience (claims) and the denominator populations (eligible enrollees) exist for data contributed to the database

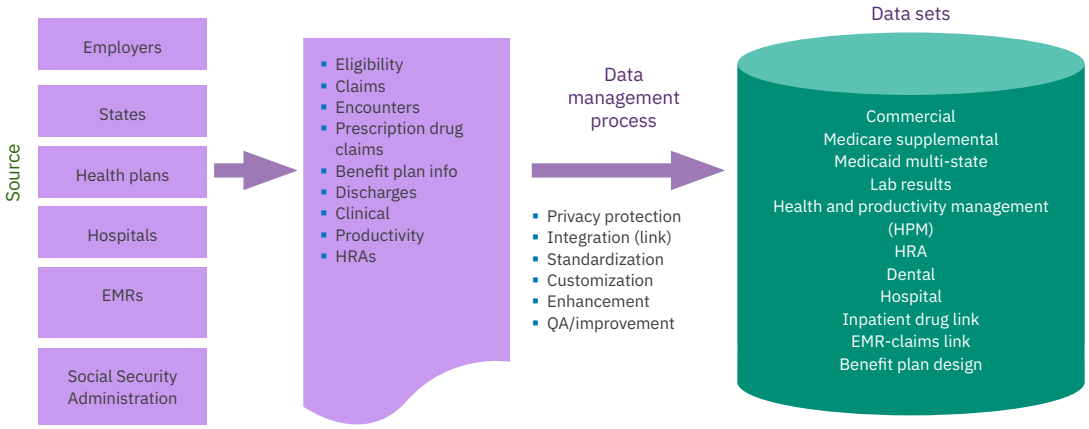
The MarketScan Research Databases: Fully integrated databases

The end product is one of the largest collections of de-identified US patient data available for healthcare research, featuring:

- An opportunity sample from multiple sources (for example, employers, states and health plans)
- More than 32 billion service records
- Over 245 million covered individuals
- More than 260 contributing employers and 40 contributing health plans
- Representation from more than 350 unique carriers

The MarketScan family consists of three core claims databases, a hospital discharge database and an EMR database, as well as several linked databases, data sets and files that combine claims data with other patient and employee data at the patient level, as illustrated in Figure 3.

Figure 3. MarketScan Research Databases: Integrated at the patient level



The IBM® MarketScan® Commercial Database consists of medical and drug data from employers and health plans. It contains data for several million individuals annually, encompassing employees, their spouses and dependents who are covered by employer-sponsored private health insurance in the US. Healthcare for these individuals is provided under a variety of fee-for-service (FFS), fully capitated and partially capitated health plans. These include PPOs and exclusive provider organizations (EPOs), POS plans, indemnity plans, HMOs and consumer-directed health plans (CDHPs). Medical claims are linked to outpatient prescription drug claims and person-level enrollment information.

The MarketScan Commercial Database may offer a distinct advantage over other databases for research on medication use. As these data are primarily sourced from employers, claims for mail order prescriptions and specialty pharmacies are also included. Capturing drug data from different types of sources can be particularly important for adherence studies and analyses of injectable drugs.

Case study: Effectiveness of varicella vaccine

The US Centers for Disease Control and Prevention (CDC) have conducted a number of longitudinal studies to determine the impact of vaccines on overall healthcare utilization. Due to their size and robust longitudinal integrity, the MarketScan Research Databases may be ideal for studying the economic and utilization outcomes of treatment over time. CDC researchers investigated the impact of the varicella (chickenpox) vaccine. Although the incidence of the disease dropped substantially since varicella was recommended for routine immunization in 1995, incomplete data made it difficult to track medical visits and expenses related to the disease. CDC researchers used the MarketScan Research Databases to conduct a retrospective, population-based study examining trends in rates and costs for varicella- related hospitalizations and ambulatory visits from 1994 to 2002. During this time period, hospitalizations related to varicella declined by 88 percent, and ambulatory visits declined by 59 percent. Total estimated direct medical expenditures declined by 74 percent, representing a savings of $62.8 million.

Reference: Zhou F, Harpaz R, Jumaan AO, Winston CA, Shefer A. Impact of varicella vaccination on health care utilization. Journal of the American Medical Association 2005; 294(7): 797–802, http://jamanetwork.com/journals/jama/fullarticle/201405

The IBM® MarketScan® Medicare Supplemental Database was one of the first in the US to profile the healthcare experience of retirees with Medicare supplemental insurance paid by employers. The database includes the Medicare-covered portion of payment (represented as Coordination of Benefits Amount or COB), the employer-paid portion and out-of-pocket patient expenses. The MarketScan Medicare Supplemental Database provides detailed cost, use and outcomes data for healthcare services performed in both inpatient and outpatient settings. For most of the population, the medical claims are linked to outpatient prescription drug claims and person-level enrollment data through the use of unique patient or enrollee identifiers.

Beneficiaries in the MarketScan Medicare Supplemental Database have drug coverage; therefore, drug data are available and provide additional, often valuable, information. This feature makes the database a robust tool for pharmacoeconomic and outcomes research, and helps provide insight into the drug use and spending patterns of older Americans.

The IBM® MarketScan® Multi-State Medicaid Database contains the medical, surgical and prescription drug experience of more than 44 million Medicaid enrollees from multiple states. It includes records of inpatient services, inpatient admissions, outpatient services and prescription drug claims, as well as information about long-term care and other medical care. Data on eligibility (by month), service and provider type are also included.

In addition to standard demographic variables, such as patient age and gender, this database includes variables that may be of particular value to researchers investigating Medicaid populations, such as aid category (for example, blind or disabled, Medicare eligible) and race.

Using this database alone or in conjunction with other MarketScan Research Databases, researchers can:

- Analyze disease conditions that may be prevalent among Medicaid populations, such as HIV/AIDS, schizophrenia and diseases of the elderly
- Assess trends in healthcare costs, utilization and outcomes for diseases that strike broadly across populations, such as asthma, cancer and cardiovascular conditions
- Incorporate variables that may not be available in other claims databases, such as race and aid category
- Determine the cost burden of particular diseases or conditions in Medicaid populations

The IBM® MarketScan® Health and Productivity Management (HPM) Database, an example of a linked database, offers the opportunity to combine data on workplace absence, short and long- term disability and workers' compensation with medical/ surgical claims and outpatient drug data. The database allows researchers to assess both the direct and indirect costs associated with a particular condition or treatment.

Using the MarketScan HPM Database, researchers can:

- Assess the direct and indirect costs associated with a clinical condition
- Measure the impact of diseases on absenteeism, short and long-term disability, and workers' compensation
- Track total healthcare costs across both medical and workers' compensation systems
- Estimate the potential return on investment in wellness or disease management programs
- Assess the impact a child or spouse's illness might have on employee absence
- Determine the relative costs of alternative pharmaceutical and medical device interventions, considering both group medicalcosts and absenteeism costs
- Develop predictive models that help define relationships between demographic factors and HPM outcomes

Case study: Indirect costs for non-adherence in bipolar patients

In one study, researchers examined the association between non-adherence to bipolar medications and lost productivity costs by employers. Adult patients with a bipolar diagnosis and at least one prescription claim for a mood stabilizer or atypical antipsychotic were selected for study.

In the selected cohort, only 35.3 percent of patients were adherent to their medication, as determined by a medication possession ratio of greater than or equal to 80. Non-adherent patients had higher adjusted indirect costs of $771.41 due to absence, $285.72 in short-term disability and $360.62 in workers' compensation. Extrapolating these findings to a fictional employer with 70,000 employees and an incidence rate of 3.3 percent for bipolar disorder, this employer could potentially save $578,378 in combined absence, short- term disability and workers' compensation indirect costs if all employee patients adhered to their bipolar treatment.

Bunn WB 3rd. Indirect costs associated with nonadherence to treatment for bipolar disorder. Journal of Occupational and Environmental Medicine 2010; 52(5): 478–85, https://www.readbyqxmd.com/read/20431414/indirect-costs-associated-withnonadherence-totreatment- for-bipolardisorder

The IBM® MarketScan® Benefit Plan Design (BPD) Database contains detailed information about benefit plan characteristics for a subset of the health plans represented in the MarketScan Commercial and MarketScan Medicare Supplemental Databases. This data asset has undergone a major transformation. Previously the dataset was generated by obtaining and manually abstracting Summary Plan Descriptions (SPDs) from our employer client contributors and using free text internal plan designations in the data to manually map the individual abstractions to the MarketScan databases. We now are using a new methodology for creating the MarketScan BPD Database: the variables in the database are no longer being created by manual SPD abstraction and mapping, but are being generated by plan-by-plan statistical analysis directly from the claims data. This greatly automates the development of the dataset, allowing us to increase by over 400 percent the number of lives mapped to BPD data and also create an inherent assurance that the linkage to claims is complete and correct.

The MarketScan BPD Database allows researchers to do the following:

- Evaluate the impact of health plan features on healthcare utilization
- Assess the relative performance of plan types with varying managed care features
- Include detailed plan provisions, such as copayments, deductibles and coverage options in analysis of healthcare cost and use

Case study: Cost sharing and adherence

Using patient-quarter data from the MarketScan Commercial Database, researchers used generalized estimating equations to determine the effects of patient cost-sharing on adherence to second generation antipsychotic medications. Results demonstrated that higher cost-sharing was inversely associated with high adherence, particularly when cost-sharing levels were above $30. Higher cost-sharing was associated with shorter time to discontinuation.

Reference: Gibson TB, Jing Y, Kim E, Bagalman E, Wang S, Whitehead R, Tran QV, Doshi JA. Cost-sharing effects on adherence and persistence for second- generation antipsychotics in commercially insured patients. Managed Care 2010; 19(8): 40–7, https://www.ncbi.nlm.nih.gov/pubmed/20822071

The IBM® MarketScan® Lab Results Database, a linked claims-lab results database, includes inpatient and outpatient drug data, as well as enrollment and laboratory test results. In some diseases, results from clinical research are surrogate markers for risk reduction and disease management.

The MarketScan Lab Results Database helps researchers evaluate:

- How well a drug is performing in the real-world clinical setting
- Diagnostic test results administered prior to initiation of drug therapy
- Laboratory test results as indicators of drug therapy effectiveness
- Frequency of safety monitoring laboratory tests while a patient is on drug therapy
- Differences in treatment patterns between patients whose disease is under control versus not under control

Case study: A1c monitoring in diabetes

For diabetes patients to achieve and maintain control, the American Diabetes Association guidelines recommend that A1c tests should be repeated within three months following an out-of-range test result or a change in therapy.[5] Researchers used the MarketScan Lab Results Database, which includes patients covered by both commercial and Medicaid insurance, to compare the frequency of retesting between groups of patients at varying A1c levels, based on an initial test. They also asked whether changes in therapy or insurance coverage were associated with more frequent retesting.

The results showed that retesting within six months is no more frequent after out-of-range tests or changes in therapy. In fact, retesting rates were lower for patients covered by Medicaid than for those with commercial insurance. Thirty-five percent of Medicaid patients with an initial A1c result of greater than 9 percent were retested, versus 51.7 percent of patients with commercial insurance. These results supported a previous finding of clinical inertia in response to poor glycemic control.

This study won an ISPOR Best Podium Award in 2007 at the 12th Annual International Conference.6

Reference: Gibson TB, Jing Y, Kim E, Bagalman E, Wang S, Whitehead R, Tran QV, Doshi JA. Cost-sharing effects on adherence and persistence for second- generation antipsychotics in commercially insured patients. Managed Care 2010; 19(8): 40–7, https://www.ncbi.nlm.nih.gov/pubmed/20822071

The IBM® MarketScan® Health Risk Assessment (HRA) Database, a claims-HRA linked data set, provides specialized data that can help researchers to evaluate the contribution of patient behaviors to health outcomes. HRAs can also be invaluable for researchers, as they provide self-reported data on clinical variables that may otherwise be unavailable. The MarketScan HRA Database standardizes and links HRA data with the claims experience of patients; this feature presents an opportunity for innovative research. In addition to medical and drug claims, absence, short-term disability and workers' compensation data, HRAs can provide key data inputs for analyzing the health and productivity

of patient cohorts. There is significant overlap between the MarketScan HRA Database and the MarketScan HPM Database; this feature can enrich health and productivity management analyses. Researchers examining diabetes, cardiovascular disease, insomnia and smoking cessation may find these data valuable.

Case study: The economic cost of obesity

Self-reported data found in the MarketScan HRA Database can help to identify the societal burden of obesity in the US. Researchers used MarketScan HRA and MarketScan HPM data to quantify the direct and indirect costs of obesity to US self-insured employers. Body mass index (BMI) derived from HRA survey results allowed researchers to divide patients into risk groups, as defined by the CDC and the World Health Organization. Analysis of direct costs from claims and indirect costs from the same patients in the MarketScan HPM Database yielded results that are consistent with previous research: Patients classified as obese or severely obese had higher overall healthcare costs. Overweight, obese and severely obese patients lost more work time than individuals with normal weight. This study provided payer cost estimates attributable to BMI categories.

Reference: Durden ED, Huse D, Ben-Joseph R, Chu BC. Economic costs of obesity to self-insured employers. Journal of Occupational and Environmental Medicine 2008; 50(9): 991–7, http://journals.lww.com/joem/Abstract/2008/09000/Economic_Costs_of_Obesity_to_Self_Insured.2.aspx

The IBM® MarketScan® Dental Database is one of the only integrated medical, drug and dental databases of its kind. This database links dental claims with medical claims, including the continuum of medical and dental care. It helps researchers to investigate the relationship between dental care, use of pharmaceuticals for oral health and patients' medical conditions, such as:

- Respiratory tract infections
- Chronic sinus infections
- Diabetes
- Chronic acid reflux
- Liver or kidney problems
- Infective endocarditis
- Cardiovascular disease
- Preterm birth

Case study: Dental care and cardiovascular disease

Observational studies have suggested a relationship between cardiovascular disease and periodontitis caused by systemic inflammation that may impair the vascular system. In October 2008, the American Academy of Periodontology reported that patients with periodontal disease were twice as likely to have coronary artery disease as those without the disease.[7]

Researchers examined the relationship between periodontal disease, statin (HMG-CoA) use and cardiovascular disease using the MarketScan Dental Database. Using a matched control group, patients identified as having periodontal disease by ICD-9 code or related dental procedures were divided into two cohorts: those treated with statins and those not treated with statins. The patients were followed for 12 months to observe evidence of cardiovascular events. There were no cardiovascular events in 99 percent of patients with periodontal disease. Twenty-five percent of patients with the disease were on statin medications. The rate of cardiovascular events for statin users and non-statin users were 0.42 percent and 1.16 percent, respectively. Controlling for differences in demographics and preexisting clinical conditions, researchers observed that patients without statin treatment were at a greater risk for cardiovascular events than patients treated with statins in the 12 months following the periodontal diagnosis (OR = 2.77, 95 percent CI, $p<0.0001$).

Reference: Misra A, Hansen LG, Chang S. Periodontal disease, statin use, and cardiovascular events. ISPOR 12th Annual European Congress; October 24–27, 2009; Paris, France, https://www.ispor.org/congresses/

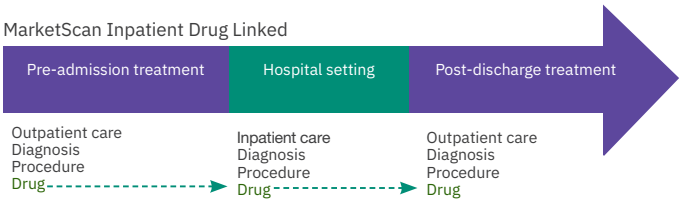
The IBM® MarketScan® Hospital Drug Database is derived primarily from hospital billing systems from US hospitals. This database provides some of the most detailed and comprehensive data available for understanding hospital care, including drug utilization in the inpatient setting.

The IBM® MarketScan® Inpatient Drug Linked Data Set helps answer research questions regarding the potential effect of an inpatient stay on drug utilization. The file matches patients from MarketScan Research Databases (Commercial, Medicare Supplemental and Multi-State Medicaid) and hospital discharge records (Hospital Drug). These data can help researchers evaluate:

- Drug use (spillover), switching and adherence between settings of care
- Pre- and post-hospitalization treatment
- Repeated hospitalizations
- Health outcomes
- Drug-specific and/or total healthcare costs

Claims data frame the picture of the continuum of care before, during and after hospitalization, thus providing rich cross sectional and longitudinal details about patient treatment patterns (see Figure 4). Hospital discharge data provide the inpatient drug component. The result is enriched insights into the transition between inpatient and outpatient treatment.

Figure 4. Tracking prescription drug treatment through a hospitalization



MarketScan linked data: We have deep experience in linking MarketScan administrative claims data to EMRs, patient registries, in-house customer and formulary data in a HIPPA-compliant manner. Our researchers have helped health services customers undertake unique linking projects using these data while maintaining scientific integrity, and IBM is the only company who can link the MarketScan Research Databases to other data.

Case study: Comparison of adverse events reporting between claims and EMR data

In this study, researchers used multiple myeloma as a test case to evaluate the differences in the occurrence of selected adverse events recorded in an oncology EMR database compared with an administrative claims database. Further, they looked at the occurrence of selected adverse events in a linked claims-oncology EMR database.

A total of 278 patients met all study eligibility criteria. Patients were followed through their first line of therapy plus 15 days for the occurrence of selected adverse events including neutropenia, thrombocytopenia, venous thromboembolism (VTE), peripheral neuropathy and/ or diarrhea. Adverse events typically monitored by oncologists either through routine laboratory values (for example, neutropenia or thrombocytopenia) or by direct observations during follow-up visits (for example, peripheral neuropathy) were recorded more often in the EMR. Adverse events such as VTEs often result in inpatient admissions and were more likely to be recorded in claims data. Conditions such as diarrhea, which may be due to a variety of causes including chemotherapy or general gastrointestinal illnesses, were recorded more often in claims data compared to the EMR. Overall, the analysis suggested that the linked claims-EMR database provided the more complete assessment of potential treatment-related adverse events and provided support for the utility of using a linked claims-EMR data source, such as the IBM® MarketScan® Oncology Claims-EMR Linked Data Set, for oncology research.

Reference: Irwin DE, Varker H, Princic N, Farr A. Comparison of treatment-related adverse events recorded in administrative claims data with those recorded in electronic medical records for multiple myeloma patients. Truven Health Analytics. October 2015, The Journal of The International Society for Pharmacoeconomics and Outcomes Research, 11/2015, Vol. 18, Issue 7, pg A683, http://www.valueinhealthjournal.com/article/S1098-3015 (15)04599-4/abstract

IBM® MarketScan® Treatment Pathways

MarketScan Treatment Pathways is a visual interface into MarketScan or other research data that allows users to follow patients forward and backward in time throughout their courses of treatment. Results are obtained rapidly and without programmer support. Underlying MarketScan Treatment Pathways are the medical, surgical, drug and lab data found in the MarketScan Research Databases.

The software graphically represents the patient's journey as a series of events that are sequenced into drawn treatment maps (Figure 5). MarketScan Treatment Pathways may be used by researchers to help them evaluate events leading up to a diagnosis, time treatment, switching patterns and treatment outcomes. Treatment pattern analysis allows users to identify and display multiple continuing patterns of drug and procedural treatments over time.

Pattern transition diagrams show how users transition from one pattern of treatment to another over time. They can be put into motion with time animation to show the nonlinear flow of patients from one treatment modality to another (Figure 6).

Figure 5. Treatment Pathways helps increase the speed of analyzing MarketScan data

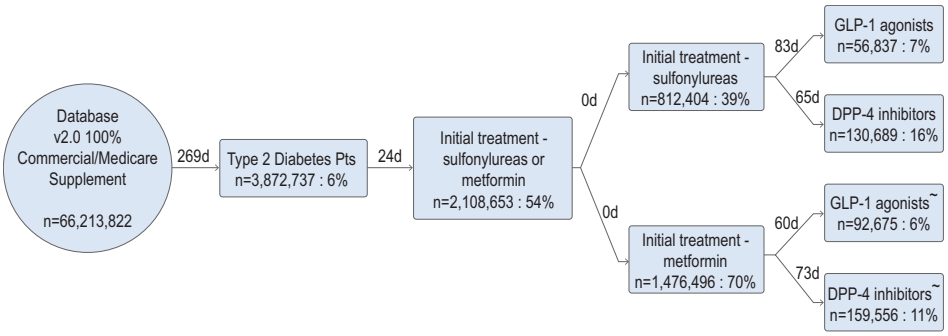
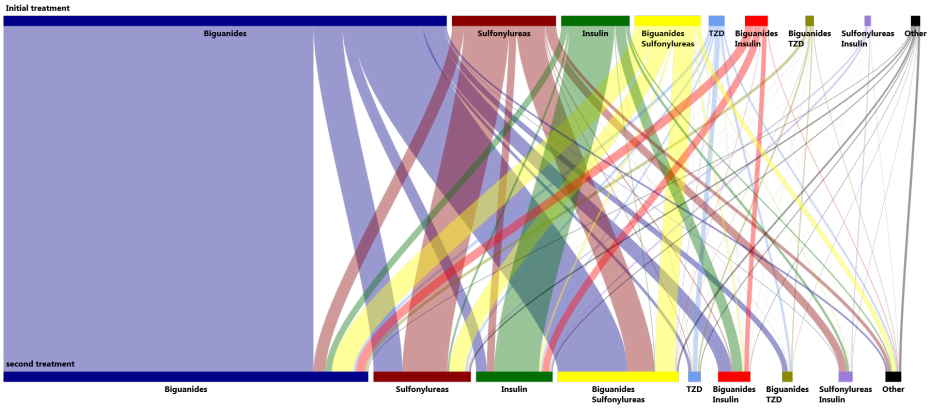


Figure 6. Diagrams show transitions from one treatment pattern to another



Summary

The MarketScan Research Databases and online analytics tools offer robust, flexible resources for health services research. The databases have several distinctive features:

- Integrated, patient-level data that are pooled from diverse points of care, reflecting the realworld continuum and cost of healthcare (including the indirect costs)
- Longitudinal tracking of patient data from all sources of care—one of the strongest in the industry
- Use of MarketScan data in more than 1,400 studies published in peer-reviewed journal articles places the MarketScan Research Databases among the most published in the US

Additional information that may be of interest to you

- MarketScan bibliography: Search peer-reviewed publications for studies using MarketScan data, from 1998 to the present
- Using MarketScan data for health economic modeling studies

About IBM Watson Health

Each day, professionals throughout the health ecosystem make powerful progress toward a healthier future. At IBM Watson Health, we help them remove obstacles, optimize efforts and reveal new insights to support the people they serve. Working across the landscape, from payers and providers to governments and life sciences, we bring together deep health expertise; proven innovation; and the power of artificial intelligence to enable our customers to uncover, connect and act — as they work to solve health challenges for people everywhere.

For more information on IBM Watson Health, visit: ibm.com/watsonhealth

Footnotes

1 Hillman BJ, Joseph CA, Mabry MR, Sunshine JH, Kennedy SD, Noether M. Frequency and costs of diagnostic imaging in office practice—A comparison of self-referring and radiologist -referring physicians, New England Journal of Medicine, 1990; 323: 1604–1608; http://www.nejm.org/doi/full/10.1056/NEJM199012063232306

2 The HIPAA Privacy Rule allows the use of a "limited data set" for research purposes. A limited data set is one in which the direct identifiers have been removed, but certain potential identifiers remain. Use of a limited data set is contingent upon the negotiation of a data use agreement. HIPAA Privacy Rule Limited Data Set: http://www.wichita.kumc.edu/Documents/wichita/researchcompliance/Limited%20Data%20Sets.pdf

3 HIPAA Privacy Rule, U.S. Department of Health and Human Services, National Institutes of Health, https://privacyruleandresearch.nih.gov/pr_08.asp

4 Statistical disclosure review and analyses of Truven's MarketScan CCAE MDCR MSA Geography Data conducted by Daniel C. Barth-Jones, M.P.H., Ph.D., President, dEpid/dt Consulting, Inc., 2/30/2013.

5 Diabetes Management Guidelines, American Diabetes Association (ADA) 2016 Guidelines, ndei.org., http://www.ndei.org/ADA-diabetes-managementguidelines-diagnosis-A1C-testing.aspx.html

6 ISPOR Best Podium Research Presentation Awards, https://www.ispor.org/awards/12meet.asp#podium

7 Position paper of The American Academy of Periodontology: Periodontal disease as a potential risk factor for systemic, August 1998, Researchgate, https://www.researchgate.net/publication/13578175_Position_paper_of_The_American_Academy_of_Periodontology_Periodontal_disease_as_a_potential_risk_factor_for_systemic_diseases

IBM Corporation
Route 100
Somers, NY 10589

Produced in the United States of America
April 2018

HPW03041-USEN-01

IBM Watson Health™