In the late 1980s, fundamental changes were transforming the U.S. healthcare system. In response to rising costs, healthcare delivery was shifting toward managed care arrangements. At the same time, there was growing interest in greater accountability for care through quality improvement. Many stakeholders sought data on these changes in the system and their impact on costs, quality of care, health outcomes, and the cost-effectiveness of various healthcare arrangements. In particular, employers and health plans, the purchasers and payers for the largest insured segment of the U.S. population (the privately insured), were interested in accurate and timely information on the drivers of cost growth and the return on investments for initiatives to improve employee health and well-being. Healthcare policy makers 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 pharmaceutical companies were interested in the cost-effectiveness of different therapies under normal care conditions.

At the time, data sources to support the requisite analyses were inadequate in various ways. Data about specific conditions that came from randomized clinical trials yielded insight about the efficacy and direct costs of therapies under best-practice conditions, but did not generalize well to broader populations; nor did they provide insight into longer-term outcomes (such as returning to work or avoiding a recurrence of illness). In addition, few clinical trials collected data on costs. Data on U.S. care patterns, health trends, and costs were typically focused on special populations (such as government beneficiaries) or specific care settings (such as hospitals). Most of all, there was a lack of reliable healthcare data on the largest segment of U.S. healthcare users: privately insured patients and their families, who account for an estimated 57 percent of the total population.

MARKETSCAN® DATABASES FROM THOMSON REUTERS

The MarketScan data warehouse, a family of databases, contains individual-level healthcare claims, lab test results, and hospital discharge information from large employers, managed care organizations, hospitals, Medicare, and Medicaid programs. This white paper describes the features and research uses of MarketScan data. Specifically, the paper examines the unique features of individual MarketScan databases, shows how these databases are constructed, describes their uses and data elements, and highlights examples of published studies based on MarketScan data.
To address the need for better data on privately insured Americans, the Healthcare business of Thomson Reuters created the MarketScan data warehouse. Since its creation, this warehouse has evolved into a suite of proprietary databases containing the largest collection of employer-based patient data in the United States. MarketScan claims data reflect the real world of treatment patterns and costs by tracking millions of patients as they travel through the healthcare system, offering detailed information about all aspects of care. Data from individual patients are integrated from all providers of care, maintaining all healthcare utilization and cost record connections at the patient level.

Over the years, the original claims-centric databases have been clinically enriched, fully integrating health and productivity (workplace absence, short-term disability, workers’ compensation), lab test results, and health risk assessment data. In 2006, Thomson Reuters acquired the Solucient® hospital databases. These datasets have become part of the MarketScan family.

**UNIQUE FEATURES OF MARKETSCAN CLAIMS DATA**

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

**Large sample size**
The MarketScan claims databases offer the largest convenience sample available in proprietary databases with 77 million unique patients since 1996. In the most recent full data year, MarketScan claims databases contain data on over 24 million covered lives. Its sample size is large enough to allow creation of a nationally representative data sample of Americans with employer-provided health insurance and Medicaid. In addition, hospital discharge data from Solucient-legacy databases contribute more than 38 million inpatient records available for research.

**Complete episodes of care**
MarketScan claims databases capture the full continuum of care in all settings, including: physician office visits; hospital stays; retail, mail order, and specialty pharmacies; and carve-out care. Linking hospital discharge records with claims data at the patient level has significantly increased the capability of MarketScan data to capture the continuity of a patient’s drug therapy between the inpatient and outpatient setting.

**Strong longitudinal tracking at the patient level**
The stability of MarketScan data sources allows superior continuity of patients over multiple years, generally longer than other claims databases. This is due to the majority of MarketScan data sourced from large employers. Employer-provided data also allow tracking of patients across health plans. This tracking ability is useful because people change health plans more often than they change jobs, and these data are able to capture patients who are “lost” in plan-based data sources—upward of 17 percent of patients in those databases.
Detailed prescription drug information
The MarketScan claims databases contain complete information on outpatient prescriptions. These databases afford distinct advantages over others that track only prescription fills. MarketScan data allow identification of the type of disease (from medical claims) and can be used to determine whether clinical, demographic, and provider characteristics influence prescribing patterns. Individual patients’ prescription fills are recorded so therapies prescribed concurrently (and presumably used in combination) can also be identified. This provides vital information about actual drug use patterns, as opposed to individual drug prescription trends.

The MarketScan Hospital Drug Database provides researchers with inpatient drug utilization data derived from hospital discharge records. These data and a proprietary projection methodology allow researchers to understand drug use in the inpatient and outpatient environment, including hospital use patterns, switching behavior, combination therapy, and patient characteristics. This information is used to help determine if introduction or earlier use of a product would improve clinical and overall cost outcomes and to analyze diagnosis volumes.

High-quality coding
A major advantage of MarketScan claims data involves their comprehensive and high quality coding. Key examples include:
- Diagnosis coded on 99 percent of all claims
- Procedure coded on 85 percent of physician claims
- Fully paid and adjudicated claims
- Complete payment/charge information, including amount of patient responsibility
- Complete outpatient prescription drug information, including patient copayments, mail order, injectables, specialty pharmacies, all carve-outs, manual and electronically submitted claims, and plan/formulary summaries

Limitations of the data
As with any data source, MarketScan claims data have limitations. Some of these have to do with the nature of claims data, and others with the nature of the MarketScan sample population. Key limitations include the following:
- The MarketScan claims 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 datasets or be used as benchmarks against them.
- The data come mostly from large employers; medium and small firms are not represented.
- Accessing the data requires data management software. DataProbe®, MarketScan online tools—Sample Select, Sample Select Prevalence, Inpatient View, Outpatient View—and programmer support can facilitate access.

Numerous research applications
In combination, the features of MarketScan claims databases enable analysts to conduct a broad range of health services studies, including:
- Cost-effectiveness and cost-offset studies
- Pharmacoeconomic outcomes evaluations
- Burden of illness analyses
- Surgical and pharmaceutical treatment comparisons
- Forecasting and modeling
- Assessment of best practices and benchmarking against empirical norms or clinical practice guidelines
- Clinical trial planning and support

In the remainder of this white paper, we discuss MarketScan claims data and research applications in greater detail. We begin by describing how the MarketScan databases are built and the elements contained in each database. Following that, we provide highlights of studies conducted with MarketScan data. At the conclusion of the paper, we offer more information about how to obtain MarketScan data.
OVERVIEW OF MARKETSCAN CLAIMS DATA

How the MarketScan datasets are constructed
Thomson Reuters constructs the MarketScan warehouse by collecting data from employers, health plans, and state Medicaid agencies and placing them into databases. Data comprise service-level claims for inpatient and outpatient services, and outpatient prescription drugs. All claims have been paid and adjudicated. We standardize financial, clinical, and demographic fields, and then add contributor-specific fields. Drug detail (e.g., therapeutic class, therapeutic group, manufacturer’s average wholesale price, and a generic product identifier) and clinical detail (on disease episode grouper) are also added.

A unique enrollee identifier is assigned to each individual in a MarketScan claims database. This identifier is created by encrypting information provided by data contributors. This information includes the employee identifier, the relationship of the enrollee to the contract holder, the gender of the enrollee, and the enrollee’s date of birth. We then combine the standardized fields of the individual databases and create links between years of data and across all data types. The data are collected for the MarketScan annual databases when nearly 100 percent of claims have been paid, eliminating the need for completion factors and improving the reliability and accuracy of the data.

Protecting the privacy of patient data as well as the privacy of our customers is a core principle of Thomson Reuters, so the MarketScan research databases fully comply with the Health Insurance Portability and Accountability Act of 1996 (HIPAA). The MarketScan databases meet the criteria for a limited-use dataset and contain none of the data elements prohibited by HIPAA for limited-use datasets. In addition, Thomson Reuters has taken steps to go beyond these HIPAA requirements. For example, instead of providing five-digit ZIP codes for employees and providers in the databases, we provide three-digit ZIP codes. Furthermore, the MarketScan databases underwent a statistical analysis by a third party to verify that they met HIPAA requirements for fully de-identified datasets. While meeting these requirements is optional given the current MarketScan licensing process, this additional step further demonstrates the Thomson Reuters commitment to HIPAA compliance and to protecting the confidentiality of patient-level and provider-level data. All patient-level and provider-level data within the MarketScan research databases contain synthetic identifiers to protect the privacy of individuals and data contributors.

Additional enhancements to the data during database creation include:
• Comparing diagnosis and procedure codes to codes that were in effect when the raw data were collected; and editing of the diagnosis and procedure codes, if necessary
• Adding the care provider’s Metropolitan Statistical Area (MSA) to claims integration of benefit plan characteristics, enrollment, outpatient prescription drug, and medical/surgical data
• Adding Major Diagnostic Categories (MDCs) and Diagnosis Related Groups (DRGs) to claims, plus application of other classification systems, such as Outpatient Treatment Groups and Medstat Disease Staging®
• Identifying the type of plan, such as health maintenance organizations (HMOs), preferred provider organizations (PPOs), and point-of-service (POS) or comprehensive plans
• Verifying that both the experience (claims) and the denominator populations (eligible enrollees) exist for all sets of data contributed to the database

THE MARKETSCAN WAREHOUSE: EIGHT FULLY INTEGRATED CLAIMS DATABASES

The end product is one of the nation’s largest collections of patient data, featuring:
• An opportunity sample from multiple sources (employers, states, health plans)
• Over four billion patient records
• 77 million covered lives
• 77 contributing employers; 12 contributing health plans
• Representation from over 126 unique carriers
The MarketScan warehouse consists of eight claims databases (see Figure 2). These are described in further detail below.

1. The MarketScan Commercial Claims and Encounters Database consists of employer- and health plan-sponsored data containing medical and drug data for several million individuals annually. Nearly 18 million individuals are included in the 2006 database, encompassing employees, their spouses, and dependents who are covered by employer-sponsored private health insurance. Healthcare for these individuals is provided under a variety of fee-for-service (FFS), fully capitated, and partially capitated health plans, including preferred and exclusive provider organizations (PPOs and EPOs), point of service plans, indemnity plans, health maintenance organizations (HMOs), and consumer-directed health plans. Medical claims are linked to outpatient prescription drug claims and person-level enrollment information.

The Commercial database is constructed by combining, standardizing, and enhancing the databases Thomson Reuters builds on behalf of large employers and health plans nationwide. Sample data elements are shown in Table 1.

Understanding the total health cost of a particular medication—as opposed to the direct cost of medication alone—is critical to assessing cost-effectiveness. Why? Because a more expensive drug therapy may also produce better outcomes, ultimately reducing long-term medical costs. For example, a study compared the cost of two different prescription therapies for treating depression: selective serotonin reuptake inhibitors (SSRIs) and tricyclic antidepressants (TCAs). MarketScan data revealed that the two-year average prescription cost of TCAs was lower, but the overall cost of treatment using TCAs was higher (see Figure 3).
The Commercial database offers 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 pharmacy are included to fully capture prescription fills from all locations. Comprehensiveness of drug data from all sources is particularly important for adherence studies and those looking at injectable drugs.

2. The MarketScan Medicare Supplemental and Coordination of Benefits (Medicare Supplemental) Database is the first in the United States to profile the healthcare experience of retirees with Medicare supplemental insurance paid for by employers. In 2007, 23 percent of the 44 million Medicare beneficiaries received their drug benefits through an employer or union-sponsored health plan. The database includes the Medicare-covered portion of payment (represented as Coordination of Benefits Amount, or COB), the employer-paid portion, and any out-of-pocket patient expenses.

The 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 valuable information. This feature makes the database a powerful tool for pharmacoeconomic and outcomes research and provides valuable insight into the drug use and spending patterns of older Americans. In addition, this drug data feature can address the same set of questions as the Commercial database.

The data elements in this database are the same as those appearing in the Commercial database, but pertain to patients with Medicare supplemental insurance.

3. The MarketScan Health and Productivity Management (HPM) Database provides the opportunity to combine data on workplace absence, short-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 this dataset, researchers can:
- Assess the direct and indirect costs associated with a clinical condition
- Measure the impact of diseases on absenteeism, short-term disability, and workers’ compensation
- Track total healthcare costs across both medical and workers’ compensation systems
- Estimate the potential return on investments in wellness or disease management programs
- Assess the impact of a child’s or spouse’s illness on employee absence
- Determine the relative costs of alternative pharmaceutical and medical device interventions, considering both group medical costs and absenteeism costs
- Develop predictive models that help define relationships between demographic factors and HPM outcomes
Sample data elements for the MarketScan Health and Productivity Management database are presented in Table 2.

### TABLE 2: Sample Data Elements in MarketScan Health and Productivity Management Database

<table>
<thead>
<tr>
<th>Demographic Information</th>
<th>Short-Term Disability</th>
<th>Workers' Compensation</th>
<th>Workplace Absence</th>
<th>Health Plan Features</th>
<th>Financial Information</th>
<th>Inpatient &amp; Outpatient Medical Information</th>
<th>Drug Information</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patient ID</td>
<td>Case days</td>
<td>Body part injured</td>
<td>Dates and hours of absence</td>
<td>Coordination of benefits amount</td>
<td>Total payments</td>
<td>Admission date and type</td>
<td>Generic product ID</td>
</tr>
<tr>
<td>Age</td>
<td>Disability type</td>
<td>Case diagnosis</td>
<td>Absence type (sick, disability, vacation, etc.)</td>
<td>Deductible amount</td>
<td>Net payments</td>
<td>Principal diagnosis code</td>
<td>Average wholesale price</td>
</tr>
<tr>
<td>Gender</td>
<td>Case diagnosis</td>
<td>Indemnity payments</td>
<td>Copayment amount</td>
<td>Plan type</td>
<td>Payments to physician</td>
<td>Discharge status</td>
<td>Prescription drug payment</td>
</tr>
<tr>
<td>Employment status and classification (hourly, etc.)</td>
<td>Total payments</td>
<td>Case days</td>
<td></td>
<td></td>
<td>Payments to hospital</td>
<td>Major diagnostic category</td>
<td>Therapeutic class</td>
</tr>
<tr>
<td>Relationship of patient to beneficiary</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Payments total admission</td>
<td>Principal procedure code</td>
<td></td>
</tr>
<tr>
<td>Geographic location (state, ZIP Code)</td>
<td></td>
<td>Medical payments</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Industry</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

4. **The MarketScan Medicaid Database** contains the medical, surgical, and prescription drug experience of nearly 22 million Medicaid enrollees from multiple states. It includes records of inpatient services, inpatient admissions, outpatient services, and prescription drug claims, as well as information on long-term care and other medical care. Data on eligibility (by month) and service and provider type are also included. In addition to standard demographic variables such as age and gender, the database includes variables of particular value to researchers investigating Medicaid populations, such as aid category (blind/disabled, Medicare eligible) and race.

Using this database alone or in conjunction with other MarketScan research databases, researchers can:
- Analyze disease conditions that are especially prevalent among Medicaid populations, such as HIV/AIDS, schizophrenia, and diseases of the elderly
- Assess trends in healthcare cost, utilization, and outcomes for diseases that strike broadly across all populations, such as asthma, cancer, and cardiovascular conditions
- Incorporate variables not available in other claims databases, such as race and aid category
- Determine the cost burden of particular diseases or conditions in Medicaid populations

5. **The MarketScan Benefit Plan Design (BPD) Database** contains detailed information about benefit plan characteristics for a subset of the health plans represented in the Commercial and Medicare Supplemental databases. This information, which is abstracted from summary plan description booklets, includes financial provisions, health service benefits, managed care features, and health coverage types.

The Benefit Plan Design database allows researchers to:
- 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
- Measure changes in plan design and benefit characteristics from 1995 onward
Sample data elements for the Benefit Plan Design database are presented in Table 3.

### TABLE 3: Sample Data Elements in the MarketScan Benefit Plan Design Database

<table>
<thead>
<tr>
<th>Financial Provisions</th>
<th>Health Service Benefits</th>
<th>Managed Care Features</th>
<th>Health Coverage Types</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maximum out-of-pocket (family and individual)</td>
<td>Home healthcare</td>
<td>Precertification requirements/Penalties</td>
<td>Point of Service (capitated and non-capitated)</td>
</tr>
<tr>
<td>Annual and lifetime limits (family and individual)</td>
<td>Extended care/Skilled nursing facility</td>
<td>Utilization review requirements</td>
<td>Health Maintenance Organization</td>
</tr>
<tr>
<td>Coinurance levels</td>
<td>Hospice</td>
<td>Second surgical opinion requirements</td>
<td>Basic/Major Medical</td>
</tr>
<tr>
<td>Copayments</td>
<td>Prescription drug</td>
<td>Case management requirements</td>
<td>Comprehensive</td>
</tr>
<tr>
<td>Deductibles (family and individual)</td>
<td>Mental health/Substance abuse</td>
<td>Formulary utilization</td>
<td>Preferred Provider Organization</td>
</tr>
<tr>
<td></td>
<td>Physical therapy</td>
<td>Mental health and substance abuse carve-out provisions</td>
<td>Consumer Directed Health Plan</td>
</tr>
<tr>
<td></td>
<td></td>
<td>In-network incentives</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Out-of-network penalties</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Preventive care</td>
<td></td>
</tr>
</tbody>
</table>

6. **The MarketScan Lab Database** helps researchers understand:
   - How well a drug is performing in the real-world clinical setting
   - Diagnostic test results administered prior to initiation of drug therapy
   - Lab test results as indicators of results of drug therapy
   - Frequency of performing safety monitoring lab tests while a patient is on drug therapy
   - Differences in treatment patterns between patients whose disease is under control compared to patients whose disease is not

MarketScan Lab includes inpatient, outpatient, drug, enrollment, and lab test results for patients during the time period from 2003 through 2006.

7. **The MarketScan Hospital Drug Database**, acquired as result of the acquisition of Solucient, is derived primarily from hospital billing systems from over 586 hospitals. This database provides the most detailed and comprehensive data available for understanding hospital care, including drug utilization in the inpatient setting. The database includes over 38 million hospital discharges between 2002 and March, 2008. Data elements include admission and discharge dates, primary and secondary diagnoses, primary and secondary procedures, patient demographics, drug utilization, and supplies.

Data are updated monthly with a 45-day lag after the close of the month. A unique feature of this database is the proprietary projection methodology allowing data users to forecast market potential, monitor product uptake and share, determine if earlier use of a product would improve clinical and overall cost outcomes, and analyze diagnosis volume and drug ordering by physician specialty. A full spectrum of semi-custom data reports are supported by the MarketScan Hospital Drug Database including those used to understand hospital market share, combination therapy, switching behavior, and for profiling patients.
8. The MarketScan Health Risk Assessment (HRA) Database provides specialized data allowing researchers to understand the contribution of patient behaviors to health outcomes. Health Risk Assessments can also be invaluable for researchers as they provide self-reported data on clinical variables that may be otherwise unavailable. Like other MarketScan databases, MarketScan HRA standardizes and links HRA data with the claims experience of patients presenting 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 HRA and HPM databases enriching health and productivity management studies. Researchers looking at diabetes, cardiovascular disease, insomnia, and smoking cessation find these data uniquely valuable.

Sample data elements are presented in Table 4.

<table>
<thead>
<tr>
<th>TABLE 4: Sample Data Elements in the Health Risk Assessment Database</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Biometric information</strong></td>
</tr>
<tr>
<td>Height</td>
</tr>
<tr>
<td>Total cholesterol</td>
</tr>
<tr>
<td>Glucose</td>
</tr>
<tr>
<td><strong>Health Behaviors</strong></td>
</tr>
<tr>
<td>Smoking status</td>
</tr>
<tr>
<td>Seatbelt use</td>
</tr>
<tr>
<td><strong>Exercise</strong></td>
</tr>
<tr>
<td><strong>Intent to Change Behavior</strong></td>
</tr>
<tr>
<td>Smoking</td>
</tr>
<tr>
<td>Diet/Weight</td>
</tr>
<tr>
<td><strong>Self-Assessed Health</strong></td>
</tr>
<tr>
<td>Physical health</td>
</tr>
<tr>
<td>Family history</td>
</tr>
</tbody>
</table>

9. The MarketScan Inpatient Drug Link Database, nicknamed the Hospital Spillover Database, helps to answer research questions regarding the effect of an inpatient stay on drug utilization. The database contains matched patients from MarketScan claims data and hospital discharge records. These data help researchers understand:
• 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 pre-, during, and post-hospitalization providing rich cross-sectional and longitudinal details about patient treatment patterns. Hospital discharge data provide the inpatient drug component resulting in enriched insights into the transition between inpatient and outpatient treatment.
ADDITIONAL DATA TOOLS

Software tools can be used in conjunction with MarketScan to increase analytic power and gain quick access to summarized data. These include Disease Staging, Medical Episode Grouper, MarketScan Sample Select, MarketScan Sample Select Prevalence, Disease Profiler™, DataProbe, MarketScan Inpatient View, and MarketScan Outpatient View, described in more detail below.

Disease Staging
Disease Staging is a clinically validated risk-adjustment methodology developed by Thomson Reuters for studying the impact of illness severity on care complications, treatment patterns, resource consumption, and costs.

Medical Episode Grouper
Medical Episode Grouper is a methodology that enables researchers to analyze a patient’s inpatient, outpatient, and pharmaceutical costs by episode of care.

MarketScan Sample Select
Sample Select provides Internet access to the most recent five years of MarketScan Commercial and Medicare Supplemental, HPM, and HRA data for querying counts of patient cohorts based on disease, diagnosis, and/or procedures. This desktop tool enables researchers to quickly access patient population counts to assess research protocols and to gather quick facts. Summary reports provide demographic, clinical, and utilization details on the selected population.

MarketScan Sample Select Prevalence
Sample Select Prevalence uses the online Sample Select platform to quickly project the estimated prevalence of a treated condition or diagnosis among patients actively engaged in the U.S. healthcare system and covered by employer-sponsored insurance—an estimated 57 percent of Americans. Importantly, these estimates represent those patients seeking care and covered by an important payer in the healthcare arena. Data come from the MarketScan Commercial and Medicare Supplemental databases and projections are based on the proprietary MarketScan Weights.

Disease Profiler
Disease Profiler provides Internet-based reports that present summary statistics on more than 600 disease categories. The annual reports are based on the Commercial and Medicare databases. Overall prevalence rates (per 1,000 lives) and mean annual payments (per patient) are provided. Disease Profiler also includes the metrics indicated below, by database and year. Currently, reports are available for 1998 through 2006 on a range of topics, for individual diseases including:

- Prevalence of condition by age group and gender
- Healthcare utilization and payments by place of service
- Mean annual healthcare payment percentages
- Top five therapeutic classes
- Top 10 drugs prescribed
- Top 10 comorbidities

DataProbe
DataProbe is a PC-based software package designed for data analysis to support research and decision making. DataProbe facilitates analyses of multiple data sources, including the MarketScan databases. DataProbe offers users access to healthcare data through applications that guide them in creating customized queries and reports. DataProbe is not data-specific, nor does it require particular variables or data formats. It imports any flat-file dataset and provides several tools for analyzing, combining, and aggregating data from multiple sources. Examples of such sources include public-use files, Medicare, hospital discharge files, patient survey data, vital statistics, and the MarketScan databases.

MarketScan Inpatient View
Inpatient View, another online tool, provides current and five-year projections of diagnosis and procedure volumes for hospital-based inpatient care in the United States. Data are derived from more than 20 million inpatient records representing one of every two discharges from U.S. hospitals annually. For each ICD-9 code, statistics are provided including detailed patient demographics, admissions data, length-of-stay distributions, cost, regional and facility distributions, patient disposition, and payer mix information.
Outpatient View details the total annual volume for diagnoses and procedures by outpatient setting: ambulatory surgery, outpatient hospital, and physician offices. Volumes are detailed by region, age, gender, outpatient setting, and payer for the most recent year. Like Inpatient View, Outpatient View also provides a five-year forecast.

MARKETSCAN DATA IN ACTION: HIGHLIGHTS OF STUDIES

With more than 145 publications appearing in major peer-reviewed journals in the last five years, MarketScan databases are the most published proprietary databases in the United States, used by government, university, and private-sector researchers. MarketScan data have supported a range of health services research and analysis, including studies of:
- Economic burden of illness: Costs and resource utilization
- Evaluation of treatment and interventions
- Effects of healthcare coverage and managed care
- Benchmarking and performance measurement
- Projecting disease prevalence
- Health and workforce productivity
- Pharmacoeconomic analysis
- Methodological studies

Each of these areas is discussed in more detail and exemplified below:

**Economic Burden of Illness: Costs and Resource Utilization**

Using the MarketScan databases, researchers can track complete episodes of care for patients and their families. Therefore, the data enable analysis of the direct and indirect costs of specific diseases—cancer, depression, or diabetes—as well as other health conditions, such as heroin addiction (discussed below). Two recent examples of research drawing on MarketScan data have studied the economic burden of particular conditions.

Analysts used MarketScan data from the Commercial, Medicare, and HPM databases to estimate the cost burden of cancer. Prior to the study, most estimates for cancer costs had been based on data sources with significant limitations for generalizing to the broader U.S. population, such as Medicare data or survey results that included only inpatient spending. This study focused on newly diagnosed patients with one of seven types of tumors in 1999 and 2000 and compared their direct and indirect costs with those of a matched control group. Results revealed that the total cost burden of cancer during the study years was substantial—ranging from $2,187 for prostate cancer to $7,616 for pancreatic cancer per month. By comparison, costs for the control group averaged $329 per month. The study also demonstrated the utility of using administrative data for deriving tumor-specific estimates of cancer costs, suggesting that policy makers charged with creating a national estimate of cancer costs should broaden the customary range of data sources to obtain a more comprehensive picture.


Another study examined the economic costs of heroin addiction in the United States. Although heroin use represents only a small proportion of total illicit drug use, its highly addictive nature suggests that its economic impact could surpass a less intensely addicting substance like marijuana. Using an array of data sources, including MarketScan data, the analysis estimated the total cost of heroin addiction in the U.S. in 1996. The results showed that estimated total costs of heroin addiction during the study period were $21.9 billion. Of these costs, productivity losses accounted for the greatest share—53 percent—followed by criminal activities (24%), medical care (23%), and social welfare spending (0.5%). The large economic burden resulting from heroin addiction suggests the importance of investing in prevention and treatment.

Evaluation of Treatment and Outcomes

MarketScan data enable analysts to examine shifts in treatment patterns and assess their impact on a range of outcomes. These data also allow comparative assessment of different interventions.

For example, there are concerns that providers are shifting mental health patients from inpatient to outpatient care as a strategy to contain costs. If this is occurring, the shift could have implications for the quality of care such patients receive.

A study used mental health claims data from 1993 to 1995 to examine whether this shift has occurred, the magnitude of any cost savings, and what components of care might account for savings. The study found that both inpatient and outpatient costs and utilization had fallen during the study period. Inpatient mental health costs fell by 30 percent, driven primarily by decreases in hospital days. Outpatient costs also fell (15%), as did costs for patients using both kinds of care (14%). The study concluded that there has not been a significant shift in the pattern of mental healthcare; rather, an overall reduction of care for mental health patients has occurred.


Another study examined the impact of new treatments on patients with type 2 diabetes. In the 1990s, two new medications for treating hyperglycemia (high blood sugar associated with type 2 diabetes) appeared on the market.

The two new medications could be prescribed simultaneously for patients. A study used data from MarketScan sources to examine whether prescribing patterns for hyperglycemia changed concurrently with the introduction of these new medications and whether these changes were associated with changes in the underlying patient population treated for this condition. The results suggested that from 1997 through 2000, prescription patterns for patients with hyperglycemia shifted in a manner parallel with introduction of the new treatments.


Effects of Health Coverage and Managed Care

MarketScan data contain information on health plan types and features. These details allow researchers to examine utilization patterns in different types of plans, including health maintenance organizations (HMOs), preferred provider organizations (PPOs), and fee-for-service (FFS). The data also contain benefit information, which can be analyzed for their impact on utilization and cost.

For example, managed care strategies that seek to contain costs by reducing hospital length of stay (LOS) have come under criticism. Are managed care patients experiencing shorter hospital stays and does this raise concerns about quality of care under such plans?

To examine this issue, researchers used MarketScan data to analyze the impact of different health insurance plans on length of stay for hospitalizations related to one specific condition: cervical cancer. Because cervical cancer is more likely to affect younger women, the MarketScan data were particularly useful because they reflect a population slightly younger than that of the U.S. overall. The study found that after controlling for other variables such as severity of illness, no significant differences emerged in LOS for patients covered by comprehensive FFS plans versus those covered by other types of plans, including managed care.

MarketScan data also allow analysis of the impact of copayments. One increasingly common feature of drug benefit packages is a three-tiered copayment: the first tier includes generic medications, the second includes specific branded medicines designated as preferred-formulary, and the third includes other non-preferred brands. Patients have an incentive to choose a generic drug, which costs less than the brand-name products. Drugs in the third tier are non-preferred and therefore more expensive. Evidence has shown that this arrangement can reduce drug spending. Less well understood, however, is whether three-tiered copayments influence the choice of medication or restrict the access of high-risk populations to recommended therapies.

This study examined the impact of a three-tiered copayment arrangement on the use of non-steroidal, anti-inflammatory drugs (NSAIDs) among patients with arthritis. In particular, the study focused on the use of recently introduced COX-2 selective inhibitors, which several consensus guidelines recommend for arthritic patients at greatest risk of gastrointestinal complications, but which typically cost more than the older drug treatments. The study employed data from the 2000 MarketScan databases (specifically, Commercial Claims and Encounters, Medicare, and Benefit Plan Design). The findings suggested that the three-tiered arrangement was influencing the selection of medications, and that patients may have been unaware that the more expensive alternative had therapeutic or safety advantages. The analysts concluded that more information about the relative health and safety benefits of different drugs should be made available both to doctors and patients in the context of tiered copayment schemes.


Benchmarking and Performance Measurement
MarketScan data resources allow benchmarking analysis to compare the relative health of different populations or the quality of treatment for specific conditions.

One study used the MarketScan databases along with data from the Department of Veterans Affairs (VA) to compare the quality of drug treatment for patients with a specific mental health condition: schizophrenia. The analysis focused on the use of antipsychotic medications and whether dosing adhered to established treatment recommendations. Results revealed no statistical difference between the two populations in terms of the percentage of patients receiving recommended dosages (60% for the VA; 58% among privately insured patients). However, VA patients were more likely either to be dosed above or below guideline-recommended amounts. The study noted that there was room for improvement in the treatment of both populations in terms of adhering to guidelines.


MarketScan data can also support performance measurement analyses. For example, one study measured the short-term cost of adhering to clinical guidelines. The analysis focused on relationships between adherence to seven guidelines for treating diabetes and medical spending based on 1996 data. The study found that overall expenditures during the year were $713 higher if diabetes patients received all guideline-based treatments. There were two important exceptions: adherence to the suggested frequency of hemoglobin A1c blood sugar tests and to eye exams were associated with significantly lower expenditures. One implication of the study was that further research was needed to assess the potential for longer-term payoffs of guideline adherence.

Projecting Disease Prevalence
Researchers at Thomson Reuters have developed the MarketScan Weights that allow MarketScan Commercial and Medicare Supplemental data to provide reliable projections for the number of patients in the employer-insured (ESI) population seeking treatment for a disease or medical procedure of interest. This population represents almost 57 percent of Americans and is a large healthcare payer segment. The MarketScan Weights combine the power of a large claims database, a convenience sample, with the representativeness of a probability sample such as the Medical Expenditure Panel Survey (www.meps.ahrq.gov).

Thomson Reuters researchers looked at the “The prevalence of antipsychotic use among privately insured patients ages 0-64” to understand the implications for education and safety studies. In the 1990s, second generation antipsychotics (atypicals) came on the market presenting a better side effect profile than older typical antipsychotics. Recently, atypical medications have been shown to have other potential liabilities and the FDA issued warnings regarding the cardiac, metabolic, cerebrovascular, and mortality risks associated with the use of antipsychotic medication in elderly patients with dementia. The study used Sample Select Prevalence to calculate antipsychotic medication use by age and diagnosis. Results estimated that one child in every 185 and one in every 100 adults aged 45-64 in the ESI population filled a prescription for an antipsychotic medication in 2005. The most commonly indicated diagnosis was bipolar disorder.

Reference: Mark T, Chang S, Hansen L. The prevalence of antipsychotic use among privately insured patients ages 0-64: Implications for education and safety studies. 23rd International Conference on Pharmacoepidemiology and Therapeutic Risk Management; August 19-22, 2007; Quebec City, Quebec, CANADA.

Health and Workforce Productivity
MarketScan workforce data allow researchers to assess the impact of specific conditions and particular courses of treatment on health maintenance and job productivity.

A team of Thomson Reuters researchers examined the health and productivity costs of the “Top 10” physical and mental health conditions affecting six large U.S. employers. The analysis was intended as a first step in identifying high-cost conditions that are most amenable to intervention. By far the most expensive condition for U.S. employers in the study year (1999) was angina pectoris (chronic chest pain). Costs for this condition averaged $235 per month across all eligible employees. The next closest was hypertension, which costs $160 per employee. The most expensive mental health condition was bipolar disorder, which costs $64 across all eligible employees.


Another study examined the impact of treatment for depression on workplace productivity. One goal of antidepressant therapy is to return patients to normal functioning, so an important measure of effectiveness is restoring normal work capability, as measured in terms of days present on the job. The study analyzed a population of employees diagnosed with depression. It recorded absences for the six-month period prior to treatment and the six-month period following for two groups of patients: those treated with tricyclic medication and those treated with selective serotonin reuptake inhibitors (SSRIs). The study found that anti-depressant therapy was effective in reducing absenteeism for both groups.

Pharmacoeconomic Analysis
Pharmacoeconomic analysis focuses on understanding the impact of pharmaceutical use on treatment costs. In most cases, a treatment followed consistently over a long time will cost more than one followed sporadically or for a short period. However, the longer, more consistent treatment may produce better clinical outcomes, thus reducing the long-term utilization of medical resources.

Therefore, a complete understanding of how drug use affects treatment costs requires examining the direct costs of spending on drugs as well as indirect cost offsets that may reduce the overall use of resources. MarketScan data combine clinical encounter and prescription drug data to enable this kind of analysis. One study examined the impact of drug selection and adherence to guidelines in treating depression. The study compared three different SSRI drug treatments for depression. The results showed that patients on one of the three treatments were more likely to be treated according to consensus guidelines; however, adherence to these guidelines did not produce significantly higher costs.


MarketScan data also support analysis of trends in drug costs. In recent years, the price of prescription drugs has increased rapidly as a share of overall medical spending. Researchers sought to determine the drivers of cost growth in drug use. The study employed MarketScan data on prescription drug use in conjunction with other data to study the cost of treatments for seven common health problems. The analysis sought to disentangle the relative contributions of price increases and volume growth in explaining the rising cost of drugs. Results revealed that price increases accounted for only a small part of cost growth and that growth in the volume of prescribing accounted for a much larger share of the increase. The analysis identified several trends driving this increase in volume: increased prevalence of disease as a result of new knowledge and improved clinical practice; demographic shifts toward an older population that requires more treatment; shifts in the mix of existing therapies toward more costly treatments; increasing quantities of medication per patient; and introduction of new therapies for patients who were previously untreated.


Methodological Analysis
MarketScan data can also be used to advance methodological expertise or to support studies that improve analytic capabilities or inform practice.

For example, a team of analysts from Thomson Reuters measured how well current risk-adjustment systems can predict individuals’ future healthcare costs based on current health characteristics. The ideal for such systems is not to overstate future costs, which creates incentives for plans to bar sicker patients, or to understate costs, which may invite overuse of services. It is especially important to understand how well risk-adjustment systems predict the future costs of people with chronic conditions and whether certain kinds of systems work better for certain conditions. The study focused on patients with potentially disabling chronic conditions and compared actual expenditures with those that various risk-adjustment models would have predicted. It found that the use of existing risk-adjustment systems would reduce incentives for health plans to discourage higher-risk enrollees, but not eliminate them. Nevertheless, risk-adjustment would provide a more accurate alternative for predicting future medical costs than simpler formulas based on age, gender, or wages.


Another study set out to define an episode of care. Episode-of-care calculations can be important for informing treatment algorithms or end-point assessments in clinical trials. Using MarketScan claims data from 1993 to 1995, the analysis focused on a specific condition—diabetic foot ulcer, a well-understood and costly complication of diabetes. The study compared two utilization measures: a daily average of patient charges versus a proportion of patients charging. The results showed that the latter measure was a more accurate indicator of an episode of care.

RECENTLY PUBLISHED STUDIES USING MARKETSCAN DATA

2008


Ouyang L, Grosse SD, Kenneson A. Health care utilization and expenditures for children and young adults with muscular dystrophy in a privately insured population. Journal of Child Neurology 2008; Apr 10; [Epub ahead of print].


Sunyecz JA, Mucha L, Basr O, Barr CE, Amonkar MM. Impact of compliance and persistence with bisphosphonate therapy on health care costs and utilization. Osteoporosis International 2008; Mar 20; [Epub ahead of print].


2007


Suh DC, Kahler KH, Choi IS, Shin H, Kralstein J, Shetzline M. Patients with irritable bowel syndrome or constipation have an increased risk for ischaemic colitis. *Alimentary Pharmacology & Therapeutics* 2007; 25(6): 681–92.


Wynn ML, Chang S, Peipins LA. Temporal patterns of conditions and symptoms potentially associated with ovarian cancer. *Journal of Women’s Health* 2007; 16(7): 971–86.


Using and Obtaining the MarketScan Databases

The MarketScan databases offer a powerful, flexible resource for healthcare research. These claims databases have several distinctive features:

- Fully integrated, patient-level data are pooled from diverse points-of-care, reflecting the true continuum and cost of healthcare (including the indirect costs)
- The longitudinal tracking of patient data from all sources of care is the strongest in the industry
- Use in more than 145 studies published in peer-reviewed journal articles during the past five years places these MarketScan healthcare databases among the most published in the United States

For more information on how to obtain the MarketScan research databases for your healthcare research, please contact Thomson Reuters at +1 866 301 5419 or by e-mail at medstat.marketscan@thomson.com. Customized datasets and licensing agreements are available to suit specific data needs.

These papers may be of interest to you:

- MarketScan Bibliography: Lists all known peer-reviewed publications for studies using MarketScan data back to 1990.

To receive these papers, please e-mail your request to medstat.marketscan@thomson.com.