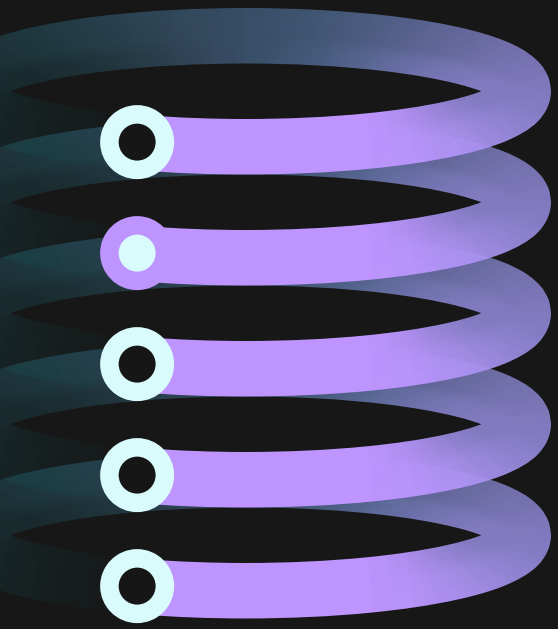
Watson Health

IBM MarketScan Research Databases for life sciences 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 life sciences research. Specifically, the paper illustrates the different attributes of individual MarketScan Research Databases, 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 life sciences 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).

The MarketScan Research Databases were created 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 collections 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, hospital and mortality data.

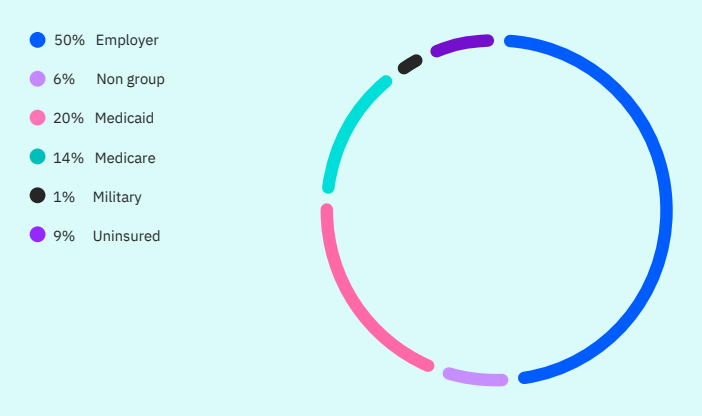


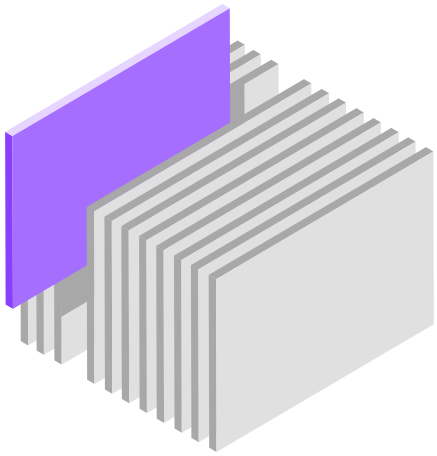
Figure 1. US population distribution by insurance status – 2018
Source: Kaiser Family Foundation estimates.
https://www.kff.org/other/state-indicator/total-population. Accessed May 2020

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 273 million unique patients since 1995. In the most recent full data year, MarketScan databases contain healthcare data for more than 39.7 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.



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.

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.
- Linked mortality data can be essential for research in disease areas such as oncology.
- Medical data linked to dental claims, allowing research in the area of cardiovascular disease and its relationship to oral health, for example.
- Life sciences 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.

Data completeness

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

- Fully paid and adjudicated claims
- Payment and charge 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 2,650 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:

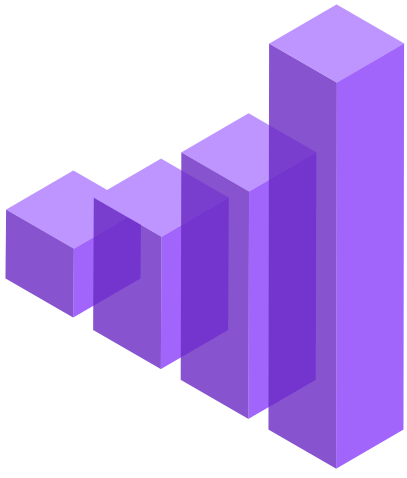
- MarketScan databases are based on a large convenience sample. Because the sample is not random, it may contain biases or fail to generalize well to other populations. However, these data can complement other data sets or be used as benchmarks against them.
- Data come mostly from large employers; medium 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 **37 billion** service records
- More than **273 million** covered individuals
- More than **120** 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.

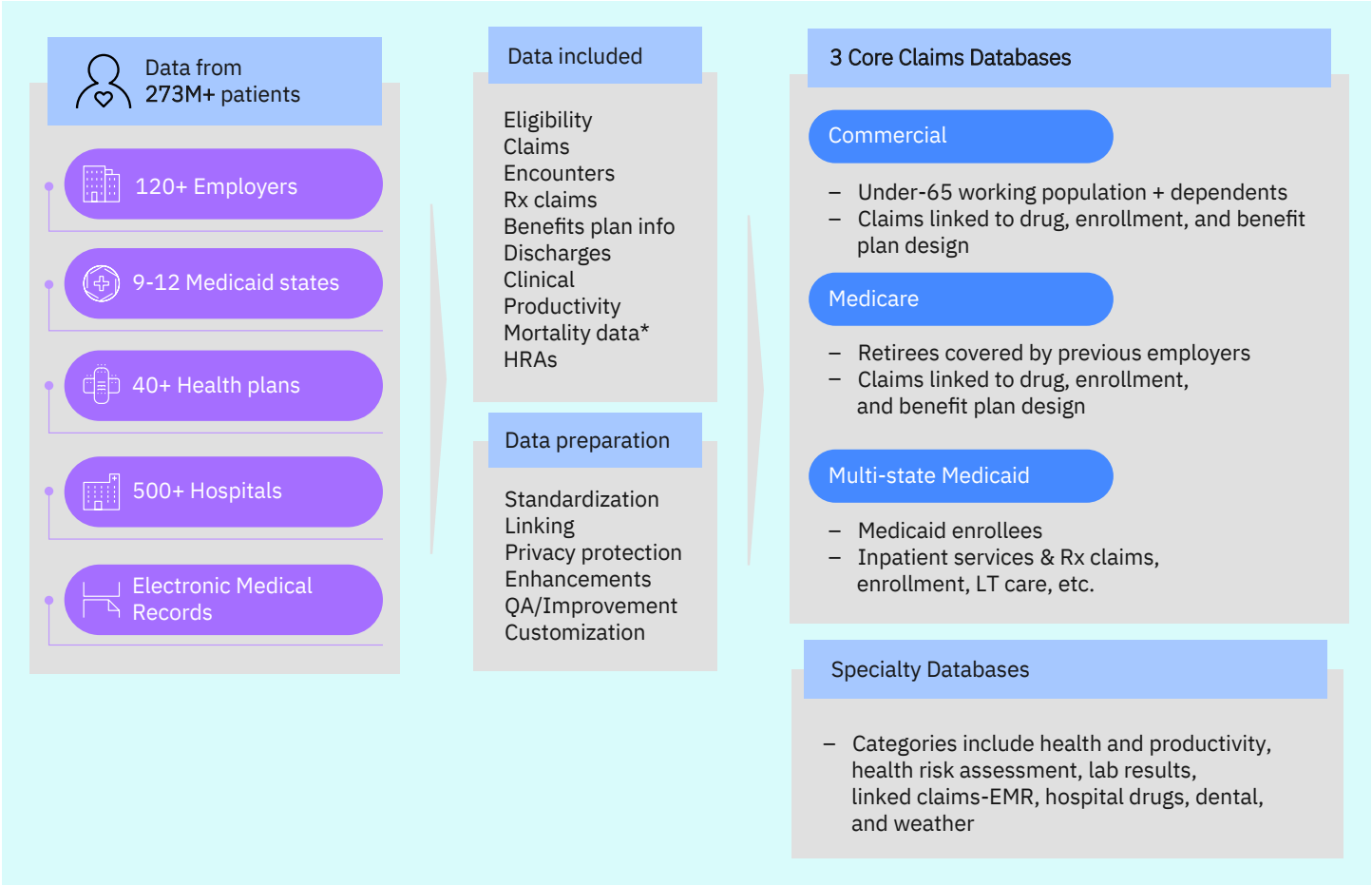
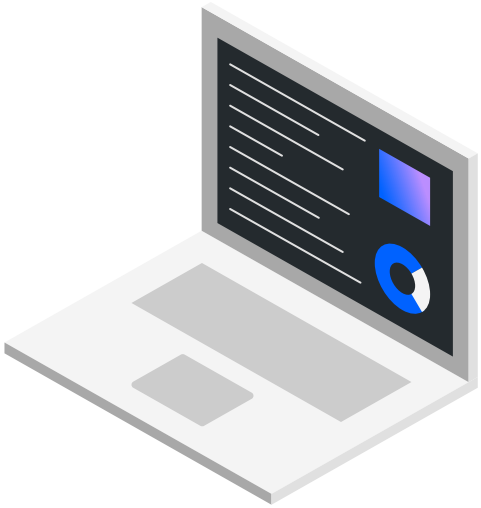


Figure 3. MarketScan Research Databases: Integrated at the patient level

*Elements only available for IBM-conducted analyses and are not included in data licenses



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: Maternal Immunization in the United States: A Nationwide Cohort Study of Pregnant Women.

Pregnancies that ended between January 1, 2010 and April 30, 2017 were identified in the IBM MarketScan Research Databases (Commercial and Multi-State Medicaid Databases) to assess maternal coverage of tetanus toxoid, reduced diphtheria toxoid, and acellular pertussis (Tdap) and influenza vaccines and gestational age at vaccination. Pregnancies were identified in each database and gestational age for each pregnancy outcome was estimated using a standardized algorithm.

Pregnancies were considered vaccinated if a Tdap or influenza vaccination was received between the estimated date of LMP and the pregnancy end date. Between 2010 and 2017 there were marked increases in vaccination coverage: from 1% to 56% (Commercial) and from 0.5% to 31% (Medicaid) for Tdap; from 15% to 31% (Commercial) and from 10% to 18% (Medicaid) for influenza. Among pregnancies vaccinated against Tdap, 89% of Commercial and Medicaid pregnancies received the vaccination in the time period of 27-36 weeks gestation, the recommended time period for Tdap vaccination with a mean gestational age of 31 weeks (SD: 5) at Tdap administration in both cohorts. Influenza vaccination can be administered at any point during pregnancy. In this study, the mean gestational age at influenza vaccination was 21 weeks (SD: 11) in Commercial pregnancies and 23 weeks (SD: 10) in Medicaid pregnancies. In multivariable analysis, the factors most strongly associated with likelihood of vaccination were receipt of another vaccination (ORs: 3.39-5.19) or having no pregnancy-related outpatient visits (ORs: 0.23-0.40). Although maternal Tdap and influenza vaccination coverage increased substantially from 2010 to 2017 among large, geographically diverse US cohorts, coverage remained suboptimal, potentially putting newborns at risk of pertussis and influenza infection before they receive their infant vaccinations. In addition, a beneficial impact was observed in the relationship between the receipt of one vaccination on the receipt of the other, emphasizing the need for providers to offer vaccination at every opportunity.

Ghaswalla P, Poirrier JE, Packnett E, Irwin DE, Gray S, Buck P. Maternal Immunization in the United States: A Nationwide Cohort Study of Pregnant Women. American Journal of Preventive Medicine. 2019;57(3):e87-e93. doi: https://doi.org/10.1016/j.amepre.2019.04.013

The IBM® MarketScan® Medicare Database consists of linked longitudinal claims data from Medicare Advantage and Medicare Supplemental health insurance plans. The database includes the employer-paid portion and out-of-pocket patient expenses for both Advantage and Supplemental plans, as well as the Medicare-covered portion of payment (represented as Coordination of Benefits Amount or COB) for Supplemental plans. It 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 Database have prescription drug coverage; therefore, dispensed 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 inclusion of both Advantage and Supplemental data into one database helps provide a more representative, complete, and longitudinal view of the commercially-insured age 65+ U.S. population, and can help open up opportunities for unique analyses, such as comparing costs and utilization between Advantage and Supplemental enrollees.

The IBM® MarketScan® Multi-State Medicaid Database contains the medical, surgical and prescription drug experience of more than 47 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

Case Study: Opioid prescribing rates in the emergency department.

Data on 19.1 million emergency department (ED) visits contained in the MarketScan Research Databases (Commercial and Medicaid Databases) were examined among patients aged 12-64 over an 11-year period. Rates of opioid prescriptions at the time of discharge were reported.

Up to a fifth (15-20%) of all ED visits resulted in an opioid prescription. Patients aged 25-54 had the highest rates of opioid prescriptions, with hydrocodone being the most commonly prescribed. The study examined an upward trend in opioid prescribing in the first half of the study (2005-2010), which declined steadily through more recent years.

Although rates of opioid prescriptions in the ED declined steadily through 2010, upwards of 15% of patients still received a prescription during the latter years of the study (through 2016), despite the national opioid epidemic.

Ali MM, Cutler E, Mutter R, Henke RM, Mazer-Amirshahi M, Pines JM, Cummings N. Opioid prescribing rates from the emergency department: Down but not out. Drug and Alcohol Dependence 2019; 205:107636, https://www.sciencedirect.com/science/article/piiS0376871619304132?via%3Dihub

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 medical costs and absenteeism costs
- Develop predictive models that help define relationships between demographic factors and HPM outcomes

Case study: Productivity loss and indirect costs among patients with cardiovascular events.

Using the MarketScan HPM Database, researchers were able to estimate the societal burden of cardiovascular disease and related clinical procedures (CVERP) disease beyond direct costs, by examining workplace absenteeism and short-term disability. Adult patients who were considered full-time employees, with evidence of hyperlipidemia and at least one prescription for lipid-lowering therapy were selected for the study and further stratified into two risk groups based on additional clinical procedures (CVERP vs. non-CVERP).

Patients with CVERP had significantly higher absenteeism hours and related costs as compared to patients without CVERP, resulting in 43 excess absenteeism hours and $1,267 in costs during the first month of the study. Short-term disability and associated costs were similarly higher in the CVERP group, with 57.7 excess hours and $996 in costs during the first month.

Cardiovascular disease and other clinical procedures were associated with substantial work loss and indirect costs, and programs aimed at reducing or preventing CVERP may result in cost savings to employers.

Song X, Quek RGW, Gandra SR, Cappell KA, Fowler R, Cong Z. Productivity loss and indirect costs associated with cardiovascular events and related clinical procedures. BMC Health Services Research 2015; 15(245), https://bmchealthservres.biomedcentral.com/articles/10.1186/s12913-015-0925-x

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 Databases. This data asset has undergone a major transformation. Previously the data set 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 data set, 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 health care utilization
- Assess the relative performance of plan types with varying managed care features
- Include detailed plan provisions, such as co- payments, deductibles and coverage options in analysis of healthcare cost and use

The IBM® MarketScan® Lab Results Database, a linked claims-lab results database, includes 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: Utilizing Laboratory Data to Measure A1c Outcomes among Type 2 Diabetes Patients

For management of type 2 diabetes, the American Diabetes Association recommends regular A1c testing and a treatment goal of A1c below 7% for non-pregnant adults.[5] Researchers were interested in utilizing the A1c <7% treatment target to compare clinical effectiveness of two types of type 2 diabetes therapy: fixed-dose combination (FDC) (i.e. two active ingredients combined into one tablet), versus loose-dose combination (LDC) therapy. The MarketScan Lab Results Database, which includes patients covered by both commercial and Medicaid insurance, provided the necessary A1c laboratory results and administrative claims data. A1c laboratory values were assessed in the 12-months prior to and following treatment initiation. The results showed that a significantly higher proportion of patients initiating FDC therapy achieved the A1c treatment goal (37.2%) compared to patients initiating LDC therapy (29.6%). In addition to improved A1c outcomes, FDC therapy was also associated with better medication adherence and lower type 2 diabetes medical costs in the 12 months following therapy initiation.

Reference: Vlahiotis A, Wittbrodt E, Patel J, Bell KF, Riehle E. Clinical and Economic Outcomes in Type 2 Diabetes Patients Treated with Fixed-Dose versus Loose-Dose Combination Anti-hyperglycaemic Therapies. Poster presented at European Association for the Study of Diabetes (EASD) 2017. American Diabetes Association. 6. Glycemic Targets: Standards of Medical Care in Diabetes-2020. Diabetes Care. 2020 Jan;43 (Suppl 1):S66-S76. doi: 10.2337/dc20-S006.

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.

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: Antibiotic Prophylaxis for Dental Procedures and the Incidence of Infective Endocarditis

Patients with underlying cardiac conditions are predisposed to developing infective endocarditis following invasive medical or dental procedures and may be prescribed prophylactic antibiotics. In 2007, The American Heart Association updated recommendations, advising that antibiotic prophylaxis was only appropriate for high-risk patients and not those deemed moderate or low-risk.

In light of this update, researchers sought to quantify the change in antibiotic prophylaxis prescribing patterns and whether reducing prophylactic antibiotic utilization led to increases in incidence of infective endocarditis. Patient medical claims data in the MarketScan Commercial and Medicare Databases were used to characterize patients as high-risk, moderate-risk, and unknown/low-risk of infective endocarditis. The results showed that by 2015, the 2007 recommendation change was associated with a significant decrease in antibiotic prophylaxis prescribing (64%) for moderate-risk individuals and a 20% fall for those at high-risk. While antibiotic prophylaxis prescribing fell among all risk groups, there was a significant increase in infective endocarditis incidence among high-risk individuals. Although these data do not establish a cause-effect relationship between the reduced utilization of antibiotic prophylaxis and the increase in infective endocarditis, the association warrants further investigation.

Reference: Thornhill MH, Gibson TB, Cutler E, Dayer MJ, Chu VH, Lockhart PB, O'Gara PT, Baddour LM. Antibiotic prophylaxis and incidence of endocarditis before and after the 2007 AHA recommendations [published online November 5, 2018]. J Am Coll Cardiol. 2018;72(20):2443-2454. doi:10.1016/j.jacc.2018.08.2178

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.

Case Study: Trends in Antihyperglycemic Agent Use Around Hospitalization in Diabetes

Management of type 2 diabetes is focused on control of blood glucose levels. For many patients this includes the utilization of antihyperglycemic agents. Low adherence to diabetes treatment regimens can result in poor blood glucose regulation which increases the risk of long-term complications. Therefore, adherence to treatment is a critical component of care. Hospitalizations can disrupt regular medication routines and have been associated with adverse clinical outcomes in diabetes. In this analysis, researchers used the MarketScan Hospital Drug Database linked to the MarketScan Commercial and Medicare Database to investigate antihyperglycemic agent use around transitions in care across the inpatient and outpatient settings. Hospitalized patients with a diagnosis of diabetes were identified in the database and AHA use was assessed in both the inpatient (hospital administered drugs) and outpatient (pharmacy fills) settings. Eighty-three percent of hospitalized diabetes patients received AHAs in the inpatient setting. Despite the high rates of inpatient AHA administration, only 40% of patients had an AHA pharmacy fill in the 30 days following hospitalization. Among patients with no evidence of outpatient AHA fills prior to hospitalization, 70% continued to lack of AHA fills post-hospitalization. Further, approximately half the patients with AHA fills prior to hospitalization had no evidence of AHA fills following, suggesting they ceased use of their AHA. These findings highlight how hospitalization can disrupt regular medication routines for diabetes patients. Results suggest that institution of interventions to ensure patients continue self-management practices after hospitalization could help improve management in diabetes.

Montejano L, Vo L, McMorrow D. Transitions of care for people with type 2 diabetes: Utilization of antihyperglycemic agents pre- and post-hospitalization. Diabetes Therapy. 2016; 7(1): 91-103. doi: 10.1007/s13300-015-0148-5.

The IBM® MarketScan® Weather Database helps researchers analyze the impact of weather-related triggers on specific diseases and how weather can exacerbate a patient's condition. This database offer US claims data linked at the patient level with geographically segmented weather data in a HIPAA compliant manner. Researchers can study several diseases including asthma, COPD, allergies, migraines, cold and flu, autoimmune and diabetes.

This database provides access to previously unavailable weather data linked to claims. Some of the key data elements include pollen counts, barometric pressure, air quality index, relative humidity and precipitation amounts.

Case Study: Weather Conditions as Predictors of Severe Migraines

While there is anecdotal evidence of weather-related migraine "triggers", the association between weather conditions and migraine occurrence has not been thoroughly investigated in the scientific literature. Researchers used the IBM® MarketScan® Weather Database, which links location-specific meteorological data and patient claims data, to examine the association between weather daily conditions and the occurrence of a severe migraine events (i.e. a migraine diagnosis on an ER or hospital claim). A case-crossover study design compared the meteorological conditions on the date of a migraine event to the conditions on dates when the patient did not have a migraine event. The study revealed that for each 1 inHg unit increase in the day's maximum barometric pressure, the odds of a severe migraine event increase by 19%. This unique, linked weather and healthcare data resource provided valuable real-world evidence of the impact of various meteorological conditions on acute health events.

Reference: Thiel E, Irwin DE. Weather Conditions as Predictors of Severe Migraines: A Case-Crossover Study Using Linked Weather and Claims Data. Poster presented at ISPOR EU 2019. Poster was awarded a finalist prize.

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 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.

The IBM® MarketScan® Explorys® Claims-EMR Data Set integrates claims data from the MarketScan Commercial and Medicare Databases with data for the same patients found in IBM® Explorys® EMRs. The Claims-EMR Data Set allows researchers to investigate situations where explanatory variables are not available in either data set alone. EMR data provide a rich clinical context for interpreting utilization and costs observed in claims data, and claims data provide more complete documentation of medical services that may influence the clinical observations in the EMR. Our data-linking methods provide researchers with statistical information that linked records represent the same patient. The linkage of the two data sets is certified for compliance with HIPAA and regularly updated.[8]

Case Study: Understanding Progression of Non-Alcoholic Fatty Liver Disease

There are a number of risk factors for non-alcoholic fatty liver disease (NAFLD) and development of metabolic syndrome (METS), which is characterized by the confluence of multiple conditions, is one of these factors. Since presentation of METS can vary from person to person, researchers used the Claims and Electronic Health Record Database (CED) to investigate the relationship between METS and the development of NAFLD based on a variety of clinical factors. Patients with metabolic syndrome without evidence of liver disease were followed over a two-year period and the rate of progression and time to NAFLD was examined in conjunction with healthcare service utilization and cost outcomes. Use of the CED database allowed researchers to examine clinical records (laboratory test results, vital signs, etc.) in addition to healthcare claims which provided diagnosis, service utilization, and cost data. Based diagnoses, laboratory test results, and procedures occurring over the six-month baseline period patients were classified as being at high or low risk of NAFLD progression. Over the two-year follow up a greater proportion of high-risk patients progressed to NALFD compared to low risk patients (6.9% vs. 3.9%); high risk patients also evidenced a shorter time to NAFLD progression. Patients who progressed to NAFLD were found to have increased baseline liver enzymes (ALT and AST) and low HDL based on laboratory test result data; height and weight measurements also revealed patients who progressed to NAFLD were more likely to be obese. Analysis of healthcare costs derived from the administrative claims revealed that patients who progressed incurred healthcare costs that were 3-fold higher than patients who did not progress over the study period. This analysis provides real-world insight into the development of NAFLD in patients with METs. The combination of claims and clinical data from EHR records allowed researches to identify risk factors of progression while simultaneously investigating the economic impacts of developing NAFLD.

Brady B, Irwin D. Progression from metabolic syndrome to non-alcoholic fatty liver disease-analyses in an EMR-Claims Database. Value in Health. 2019 May; 22(S1).

The MarketScan Mortality File integrates death information with the enrollment and claims data for a subset of individuals who are found in the MarketScan Commercial and MarketScan Medicare Databases. The subset consists of individuals for whom a death could be ascertained. Individuals who died would have a month and year of death reported in this database; those who were alive during the time period of the database would have a missing date value. Death information allows IBM researchers to evaluate mortality as a potential outcome in various studies, including:

- Safety and effectiveness research
- Evaluation of the healthcare delivery system and access to care
- Effectiveness of medical, drug and surgical interventions
- Economics of end-of-life care
- Population health trends and inequalities in care over time

Mortality information—in conjunction with other data elements available within the MarketScan claims data files, such as age, gender and diagnoses—provides IBM researchers with a robust data asset that supports for age- and risk-adjusted mortality studies. The MarketScan Mortality File is only available for IBM-conducted analyses and is not available through data licenses.

Additional data tools

Software tools and methodologies can be used with MarketScan data to help increase analytic power or gain quick access to information. The MarketScan portal includes the following tools:

IBM® MarketScan® Sample Select

MarketScan Sample Select offers access to MarketScan Commercial, Medicare, and Medicaid data for querying counts of patient cohorts based on disease, diagnosis and/ or procedures. This online tool helps enable researchers to access patient population counts to assess research protocols and gather facts. Summary reports provide demographic, clinical and utilization details on the selected population.

IBM® MarketScan® Inpatient View

MarketScan Inpatient View, another online tool, uses a catalog of US hospital-based inpatient care to provide users with diagnosis and procedure volumes and key statistics for the most recent year with trending information. For each diagnosis code, statistics include detailed patient demographics, admissions data, length-of-stay distributions, cost, regional and facility distributions, patient disposition and payer mix information.

Data are derived from all-payer data gathered from 28 million actual inpatient records, representing approximately 50 percent of discharges from US hospitals per year. This detail-rich database is the IBM Projected Inpatient Database (PIDB)—a proprietary database that is one of the largest all- payer inpatient database available. The database supports publications, products and custom studies, the results of which are applicable to short-term, general, nonfederal hospitals in the US. This database combines public and proprietary state data, as well as individual and group hospital contracts.

IBM® MarketScan® Outpatient View

MarketScan Outpatient View details the total annual volume for diagnoses and procedures by outpatient setting: ambulatory surgery centers, outpatient hospitals and physician offices. Volumes are detailed by region, age, gender, outpatient setting and payer for the most recent year. Like MarketScan Inpatient View, MarketScan Outpatient View also offers a five-year forecast.

IBM® MarketScan® Treatment Pathways

IBM MarketScan Treatment Pathways overlays a cloud-based analytic interface onto large patient-level databases such as the IBM MarketScan Research Databases. 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).

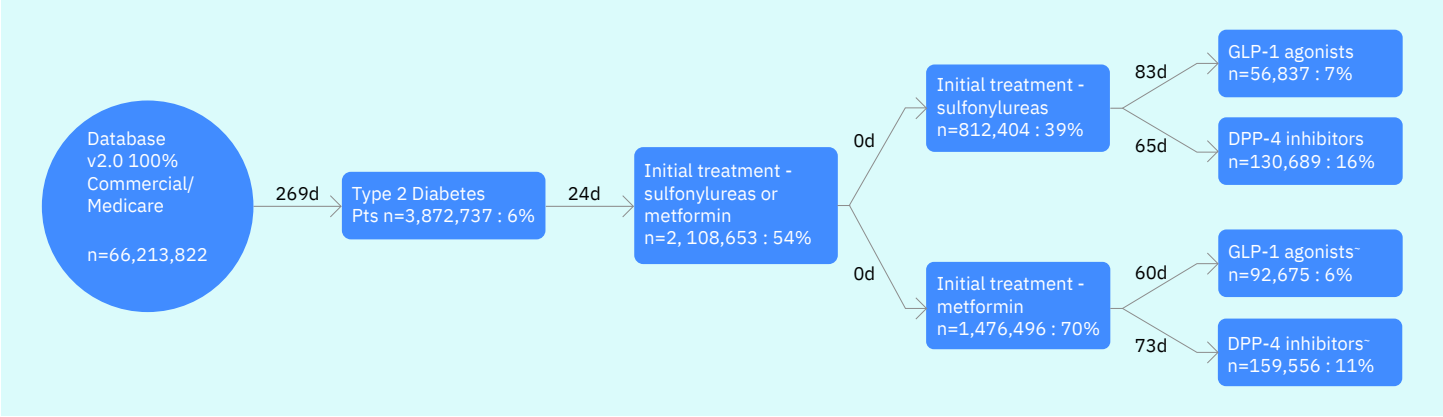


Figure 5. MarketScan Treatment Pathways helps increase the speed of analyzing MarketScan data

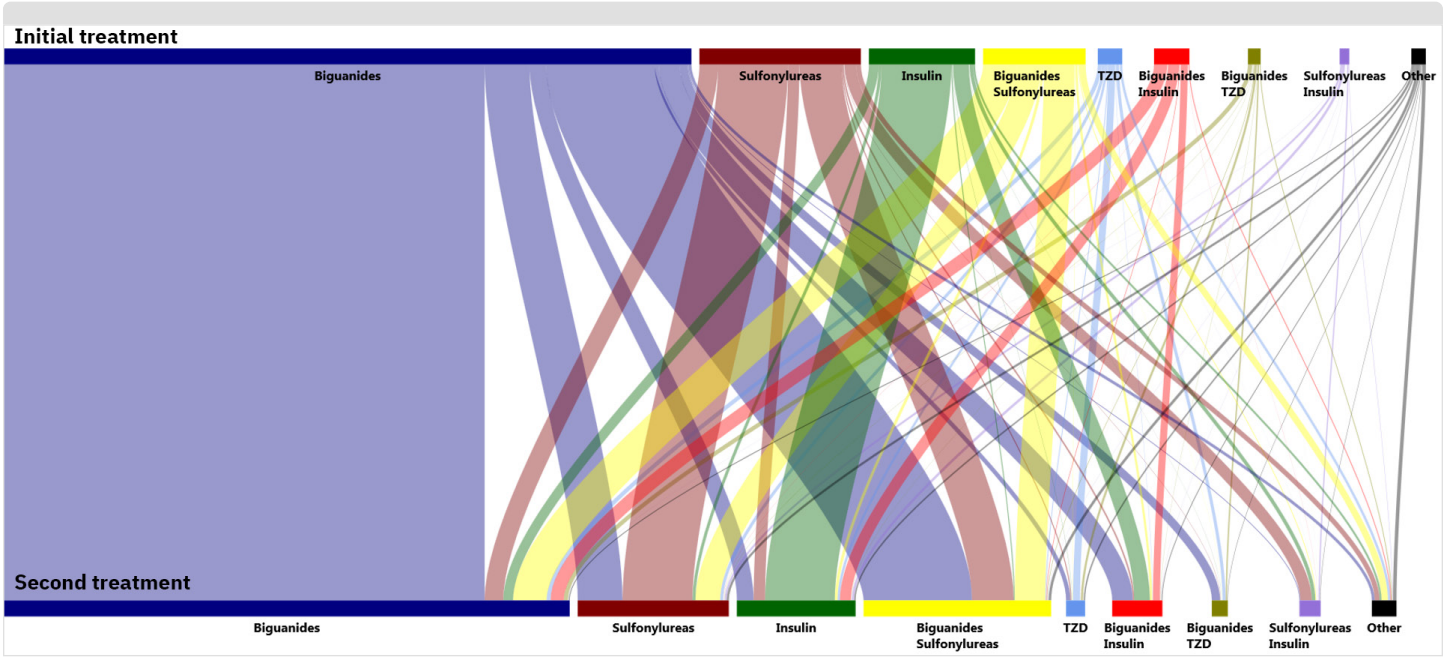


Figure 6. Diagrams show transitions from one treatment pattern to another

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 real-world continuum and cost of healthcare (including the indirect costs)
- Longitudinal tracking of patient data from all sources of care
- Use of MarketScan data in more than 2,600 studies published in peer-reviewed journal articles places the MarketScan Research Databases among the most published in the US

About IBM Watson Health

IBM Watson Health is a data, analytics, and technology partner for the health industry. Supported by the innovation of IBM and intelligence of Watson, we are committed to helping build smarter health ecosystems. Through the combination of our deep industry expertise in health, data and analytics, actionable insights, and reputation for security and trust, Watson Health is working together with its clients and partners to help them achieve simpler processes, better care insights, faster breakthroughs, and improved experiences for people around the world.

For more information

IBM Watson Health, please visit: ibm.com/watson/health.

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; https://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: https://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://www.privacyruleandresearch.nih.gov/pr_08.asp

4 Statistical disclosure review and analyses of 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., https://www.ndei.org/ADA-diabetes-management-guidelines- diagnosis-A1C-testing.aspx.html

Learn more

ibm.com/watsonhealth ©Copyright 2021

IBM Corporation, Route 100 Somers, NY, 10589

Produced in the USA, July 2021