

Merative™ MarketScan® Research Databases

Biostatistics and Data Science Core

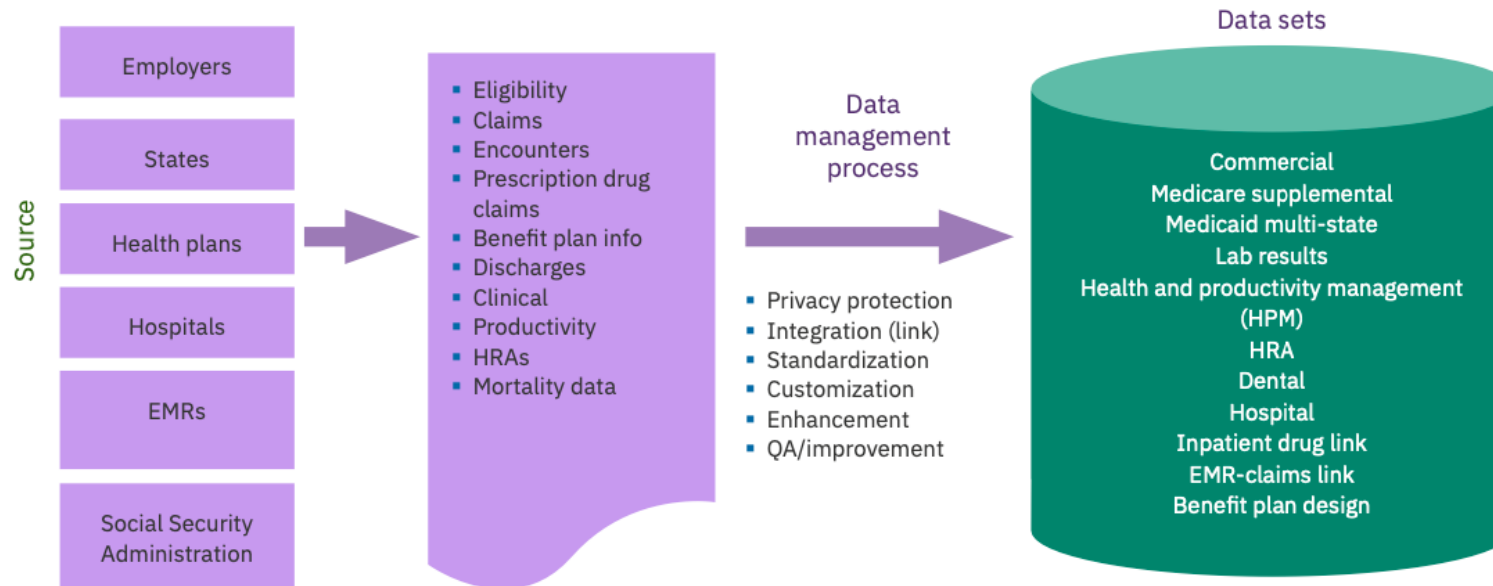
Database version: May 2024

Update date: 6/14/2024

For data consultation, email BDSC@UTSouthwestern.edu

Overview

The Merative™ MarketScan® Research Databases are a collection of de-identified patient-level health data that includes medical, drug, and dental information, as well as productivity data, laboratory results, health risk assessments, hospital discharges, and electronic medical records. The data is contributed by large employers, managed care organizations, hospitals, EMR providers, Medicare, and Medicaid.



From IBM MarketScan Research Databases for life sciences researchers(2018)

Overview

MarketScan databases offer some of the largest convenience samples available in proprietary US databases—with over 245 million unique patients since 1995.



Not included in the UTSW purchase

From IBM MarketScan Research Databases for life sciences researchers(2018)

Using MarketScan at UT Southwestern

Current data coverage:

- Commercial Claims & Encounters data
- Medicare Supplemental & COB data
- Data time span: 2019 to 2022
 - Currently, UTSW has access to a 4-year span of MarketScan data

Access:

Users can request to receive a derived subset of MarketScan data tailored to the scope of project.

Biostatistics and Data Science Core (BDSC) provides support on MarketScan data query, curation, and general consultations:

- Service is provided on hourly for-fee-service basis (\$105/hour)
- Downstream analysis service can be requested separately (hourly-based or FTE-based)
- Data Request: BDSC@UTSouthwestern.edu

Important: Users are required to follow the Usage Rules (see next slide).

Usage rules

- Users can keep the project dataset (a derived subset of MarketScan data tailored to the scope of project) until study completion (and as long as UTSW maintains an agreement with MarketScan).
- Users will notify BDSC@UTSouthwestern.edu with any publications or grant applications using MarketScan.
- Users must acknowledge the UTSW NIH CTSA grant 1U54TR002361 in publications
- The current data use agreement is not applicable to externally funded projects. For research projects supported by foundations, state, federal government, or commercial companies, an additional fee and approval are necessary.
- PI cannot be a student.
- Data must be stored on a UTSW owned device or server.

What's new in this release

This year, the Merative MarketScan Databases have been migrated to Snowflake, providing a more secure and user-friendly platform for accessing, extracting, and collaborating with the data.

The screenshot displays the Snowflake web interface. On the left is a navigation sidebar with options like 'Create', 'Search', 'Projects', 'Data', 'Databases', 'Add Data', 'Data Products', 'AI & ML', 'Monitoring', and 'Admin'. The main content area is divided into two panels. The left panel, titled 'Databases', shows a table of three databases: CLIENT_ACCOUNT, MARKETSCAN, and SNOWFLAKE. The right panel, titled 'MARKETSCAN', shows a list of views under the 'COMMERCIAL_SET_A' schema, including CCAE_A, CCAE_D, CCAE_F, CCAE_I, CCAE_O, CCAE_R, CCAE_S, and CCAE_T.

NAME	SOURCE	OWNER	CREATED
CLIENT_ACCOUNT	Local	ACCOUNTADMIN	1 month ...
MARKETSCAN	Share	ACCOUNTADMIN	1 month ...
SNOWFLAKE	Share	—	1 month ...

MARKETSCAN

COMMERCIAL_SET_A

Views

- CCAE_A
- CCAE_D
- CCAE_F
- CCAE_I
- CCAE_O
- CCAE_R
- CCAE_S
- CCAE_T

What's new in this release

- MARKETSCAN
 - COMMERCIAL_SET_A
 - Views
 - CCAIE_A
 - CCAIE_D
 - CCAIE_F
 - CCAIE_I
 - CCAIE_O
 - CCAIE_R
 - CCAIE_S
 - CCAIE_T

- MEDICARE_SET_A
 - Views
 - MDCR_A
 - MDCR_D
 - MDCR_F
 - MDCR_I
 - MDCR_O
 - MDCR_R
 - MDCR_S
 - MDCR_T

In comparison to the 2023 version, the 2024 version includes lab data from 2019 to 2022.

Data Coverage

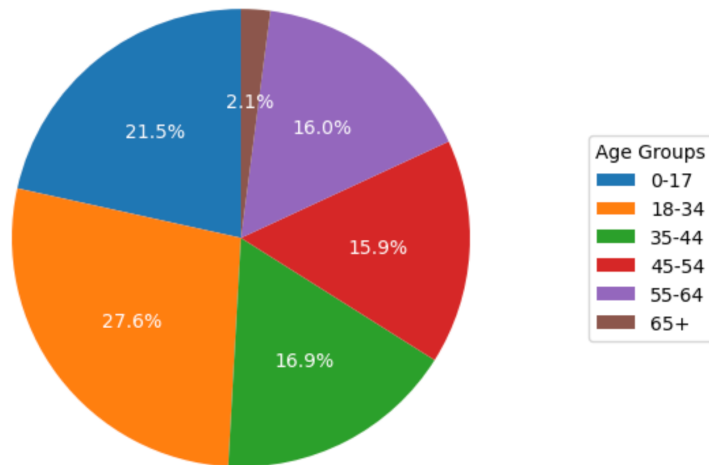
Butler, A.M., Nickel, K.B., Overman, R.A., Brookhart, M.A., 2021. IBM MarketScan Research Databases, in: An Introduction to Epidemiology for Health Professionals. An Introduction to Epidemiology for Health Professionals, pp. 243–251..
https://doi.org/10.1007/978-3-030-51455-6_20

Overall dataset size at UT Southwestern

Size	Commercial Database	Medicare Database
# of unique enrollees	41,726,154	2,748,578
# of patients with outpatient drug orders	29,595,063	2,551,460
# of unique outpatient visits	35,028,463	2,620,780
# of unique inpatient admissions	2,991,753	635,711

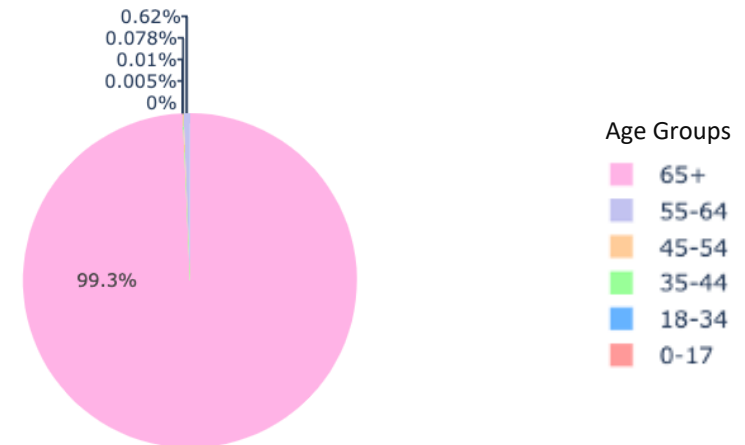
Distribution of Age

Commercial Database



Based on a random sample of n=100,000 records

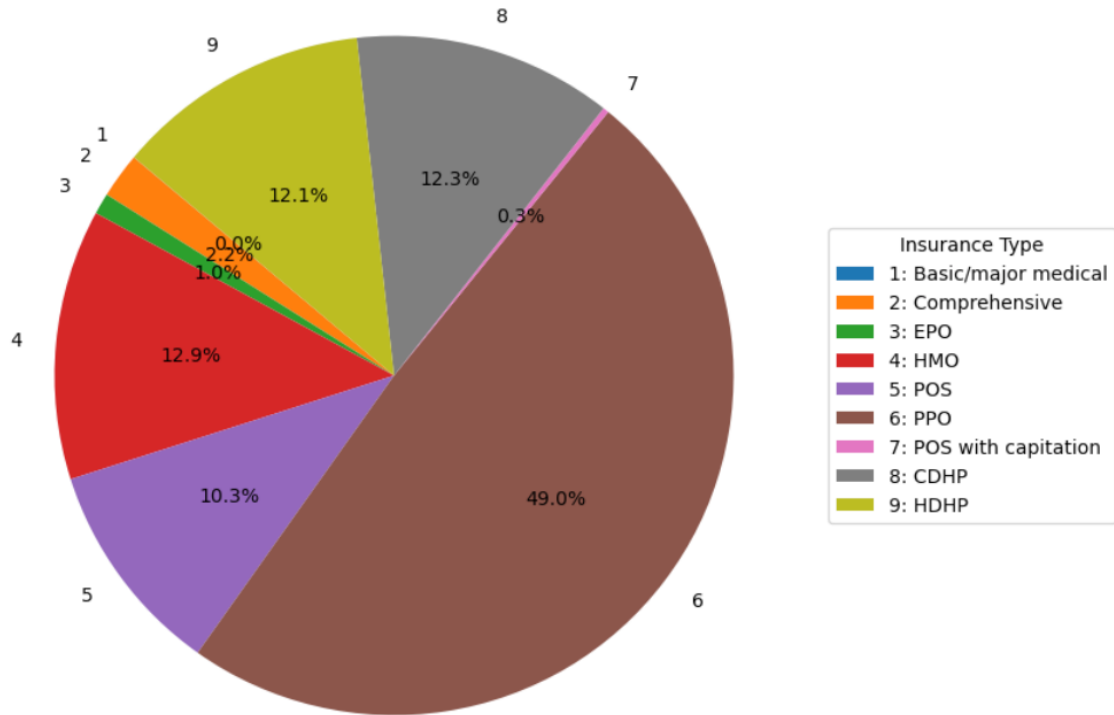
Medicare Database



Based on a random sample of n=100,000 records

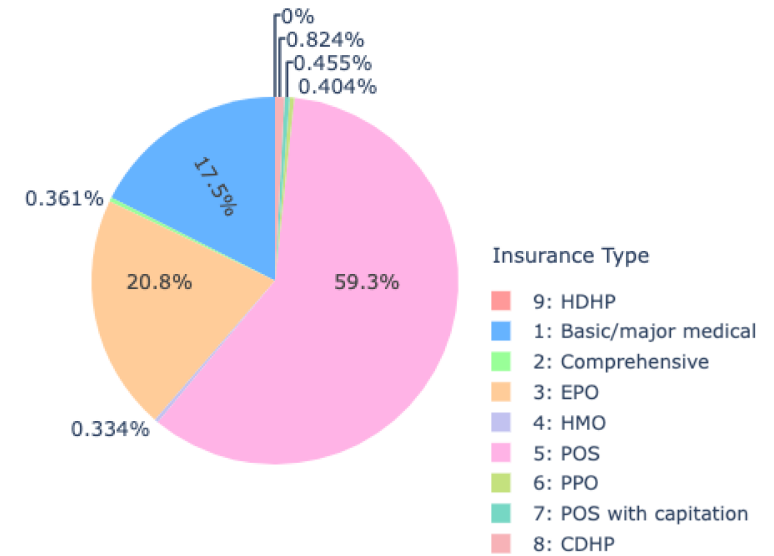
Distribution of Insurance

Commercial Database



Based on a random sample of n=100,000 records

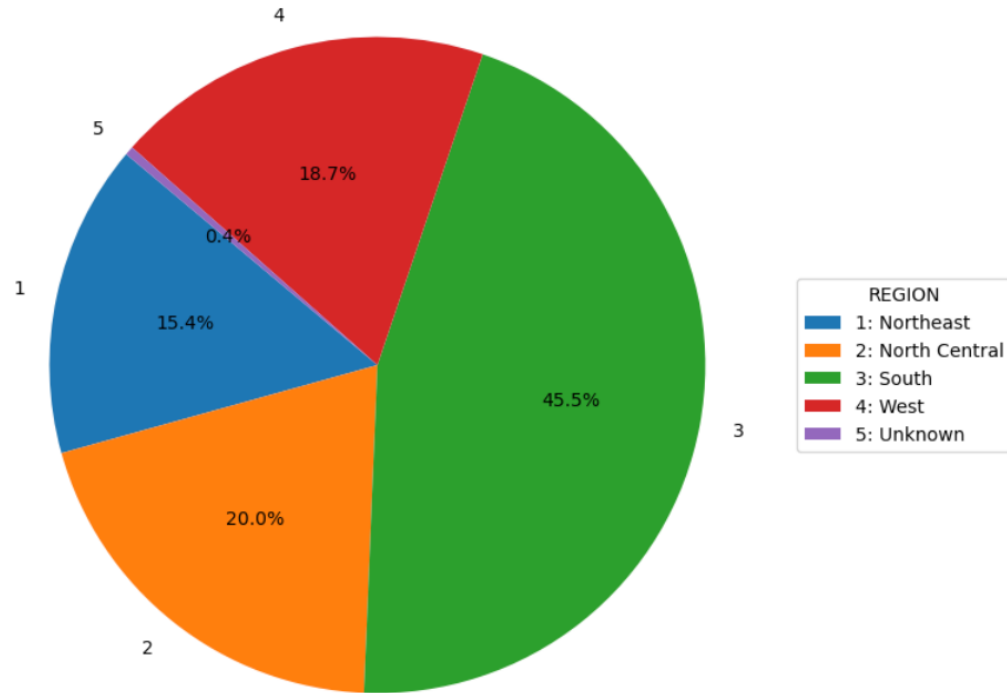
Medicare Database



Based on a random sample of n=100,000 records

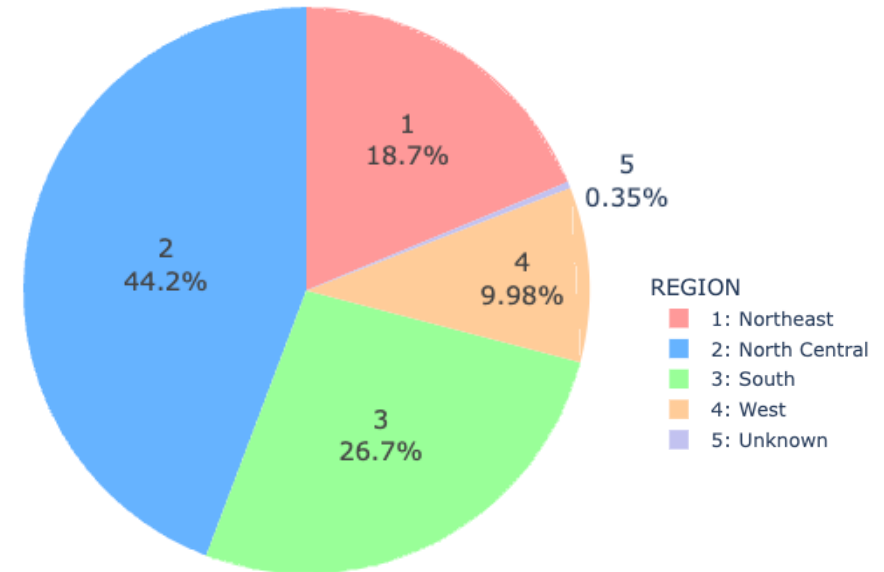
Distribution of Region

Commercial Database



Based on a random sample of n=100,000 records

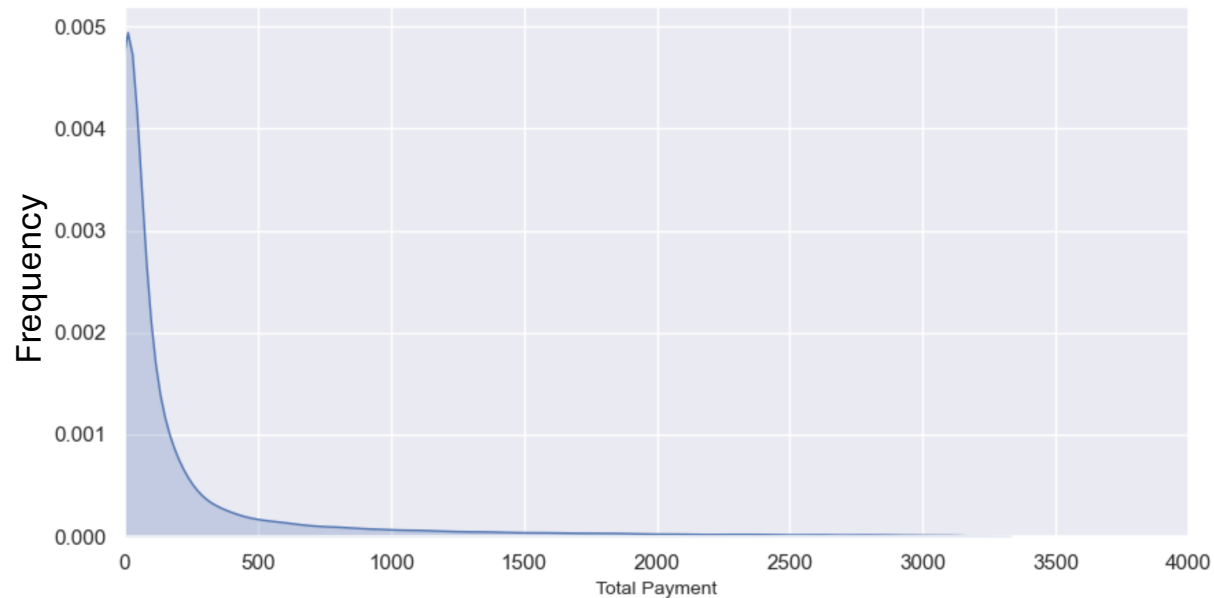
Medicare Database



Based on a random sample of n=100,000 records

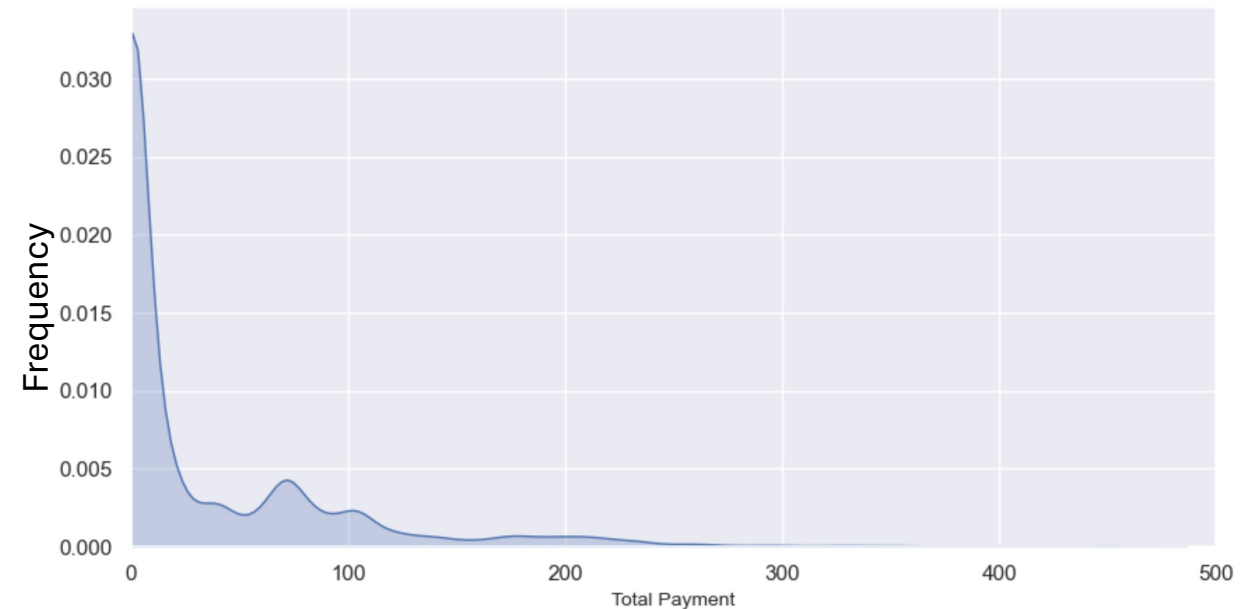
Histogram of Total Payment – Outpatient Drug Claims

Commercial Database



Based on a sample size (n=100,000 records). The x-axis shows the gross payments to a provider for a service. Payment equals the amount eligible for payment under the medical plan terms after applying rules such as discounts, but before applying COB, Copayments, and Deductibles.

Medicare Database



Based on a sample size (n=100,000 records). The x-axis shows the gross payments to a provider for a service. Payment equals the amount eligible for payment under the medical plan terms after applying rules such as discounts, but before applying COB, Copayments, and Deductibles.

Demographics and geographics

Domain	Data Elements	Commercial Database	Medicare Database
Demographics	<ul style="list-style-type: none"> Age, birth year, sex Employment status and classification Geographic region, zip code of employee residence Industry classification of the employer State of hospital 	Yes	Yes
	<ul style="list-style-type: none"> Race, ethnicity Lifestyle factors (SES, tobacco/substance use, income, education) 	No	No
Insurance and financials	<ul style="list-style-type: none"> Plan type (e.g., HMP, POS, PPO) Date enrollment start, date enrollment end Payment info for each service/hospitalization: <ul style="list-style-type: none"> Total pay, copay, deductible, coinsurance, COB and other savings, net pay 	Yes	Yes

Inpatient and outpatient care records

Domain	Data Elements	Commercial Database	Medicare Database
Encounter	<ul style="list-style-type: none"> Visit type/department Dates of admission, discharge Date service start, end, payment Length of stay, place/quantity of service 	Yes	Yes
Clinical	<ul style="list-style-type: none"> Principal and secondary diagnosis codes <ul style="list-style-type: none"> ICD-9/10 and others Principal and secondary procedure codes <ul style="list-style-type: none"> CPT and others Place/quantity of service Clinical measures (e.g., REFHEIGH, REFLOW, RESLTCAT) 	Yes	Yes
	<ul style="list-style-type: none"> Clinical measures (e.g., BP, BMI, lab results) Physical examination results Patient-reported measures 	No	No
Medication	<ul style="list-style-type: none"> Drug name, ID, and codes <ul style="list-style-type: none"> National Drug Code and others Days Supply, Quantity, Refill Number, Pharmacy ID 	Yes	Yes

Example Use Cases

Example use cases

Typical questions the MarketScan could address:

- Cost of care analyses
- Trends or patterns of medication use or procedures
- Healthcare utilization analyses
- Associations between an outcome and clinical (e.g. diagnoses)/demographic covariates
- Lab data

Limitations to MarketScan:

- No death data – outcomes studies for mortality not possible, difficult to account for competing risk of death for certain outcomes.
- High “Churn”, especially in commercial database limits longitudinality of data (when people switch insurance they may drop out of the database)
- OTC meds not captured

Contact

For MarketScan data inquiries, email BDSC@UTSouthwestern.edu

Consultants:

- Pamela Chan
- Donghan (Mo) Yang