UTSouthwestern
Medical Center

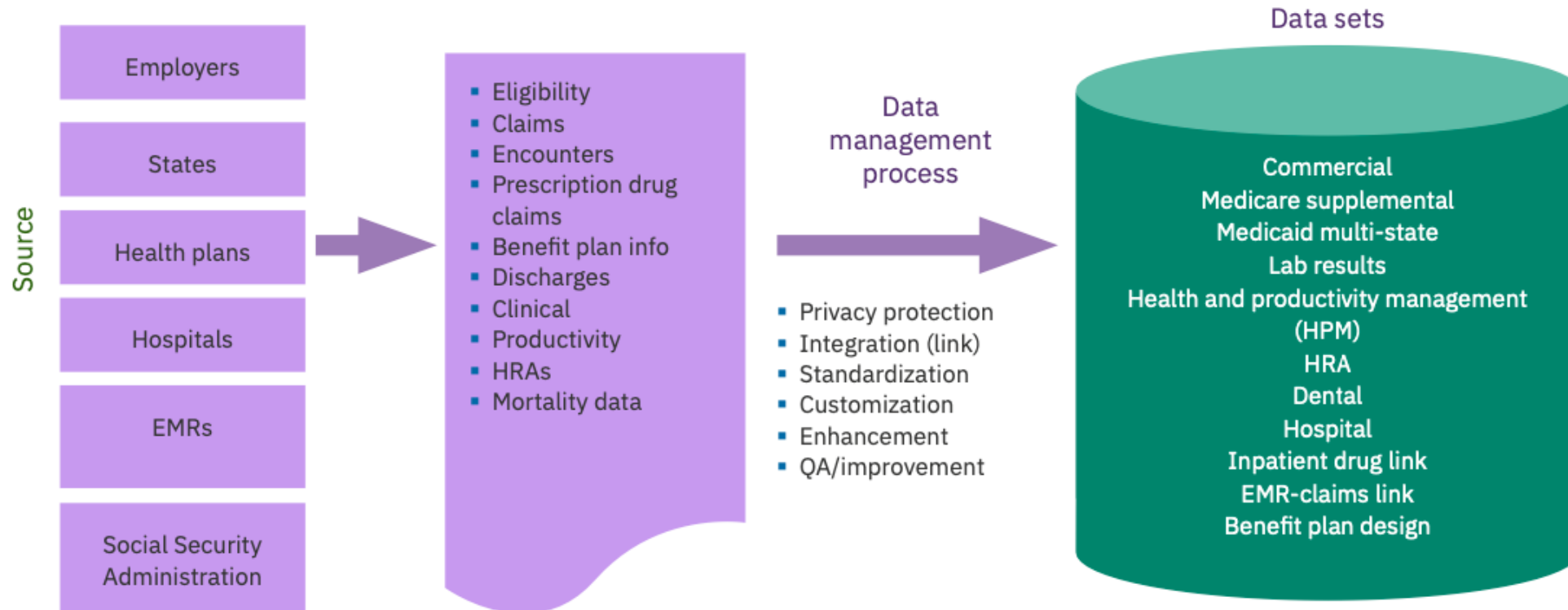# 2023 Merative™ MarketScan® Research Databases

Biostatistics and Data Science Core

File version: October 2023

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

Using MarketScan at UT Southwestern

Current data coverage:
- Commercial Claims & Encounters data
- Medicare Supplemental & COB data
- Data time span: 2019 to 2021
    - Currently, UTSW has access to a **3-year span** of MarketScan data
    - If/when the contract is renewed, the 3-year window will move forward by 1 year.

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.

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

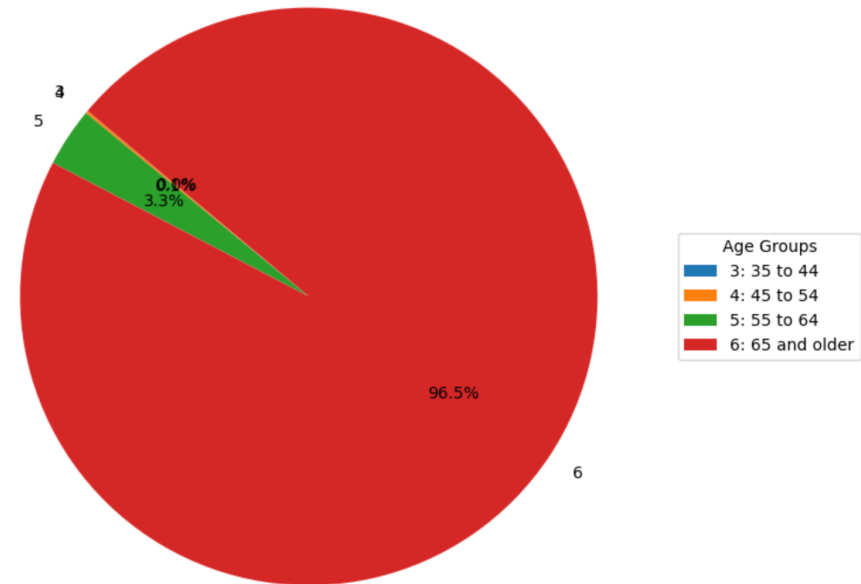| Size	Commercial Database	Medicare Database
# of unique enrollees	36,106,309	2,295,821
# of patients with outpatient drug orders	25,668,551	2,142,471
# of unique outpatient visits	1,586,266,747	279,534,108
# of unique inpatient admissions	3,152,322	1,177,405

Distribution of Age

Commercial Database

Based on a random sample of n=361,063 enrollees

Medicare Database

Based on a random sample of n=22,958 enrollees

Distribution of Insurance

Commercial Database

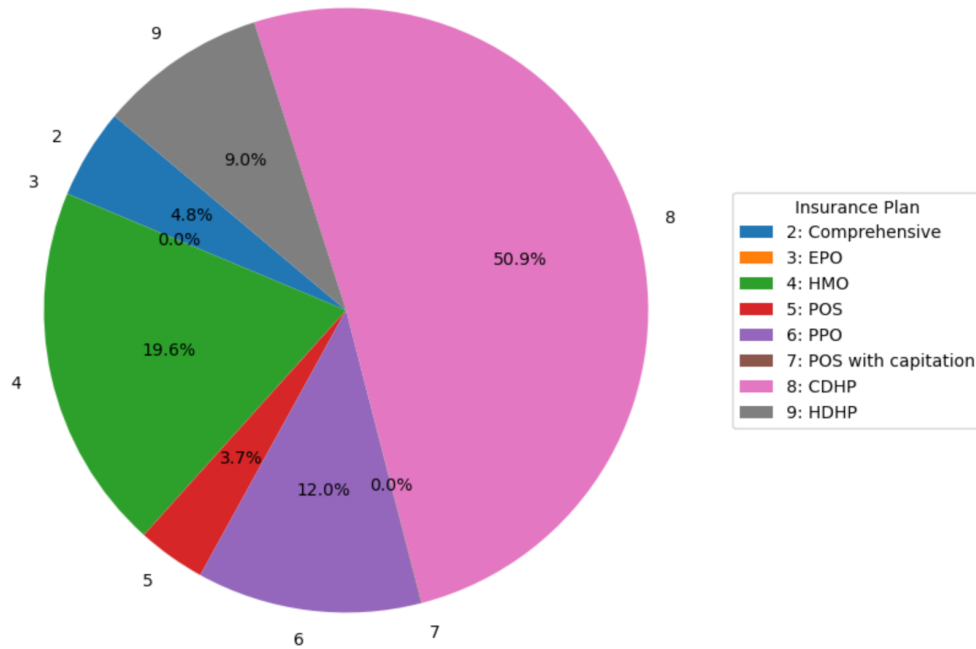Based on a random sample of n=361,063 enrollees
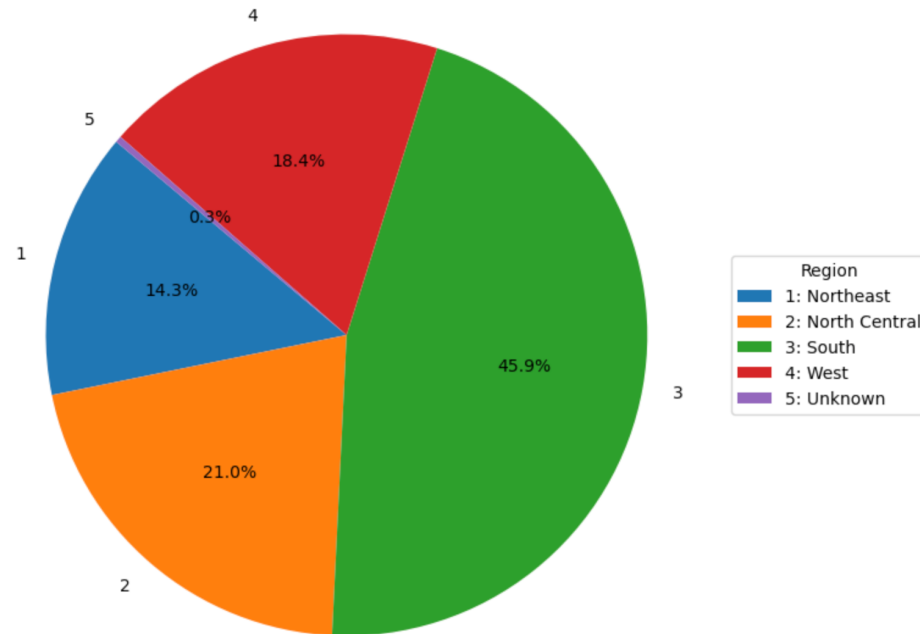
Medicare Database

Based on a random sample of n=22,958 enrollees

Distribution of Region

Commercial Database

Region Distribution in Commericial - Annual Enrollment Summary in the year of 2021
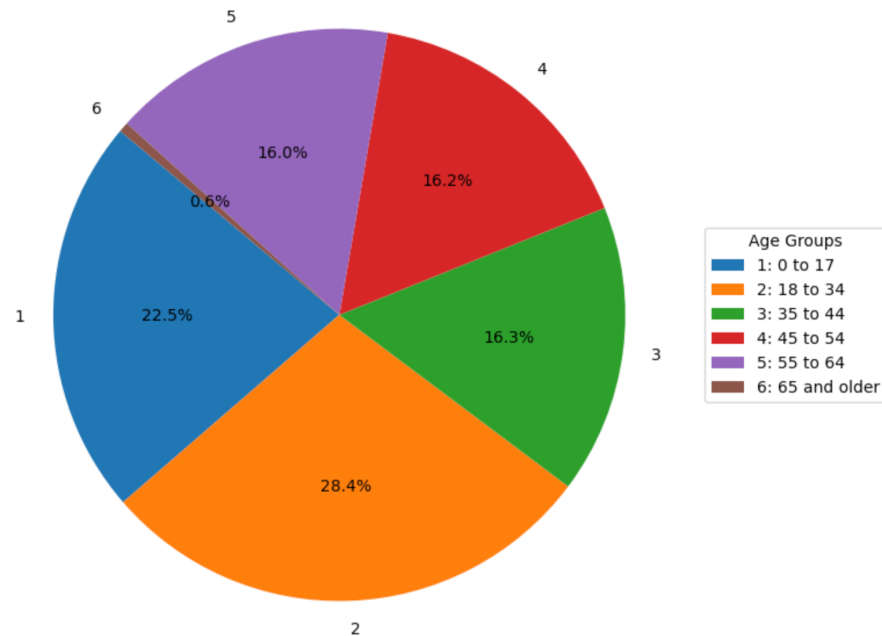
Region
1: Northeast
2: North Central
3: South
4: West
5: Unknown

Based on a random sample of n=361,063 enrollees

Medicare Database

Region Distribution in Medicare - Annual Enrollment Summary in the year of 2021

Region
1: Northeast
2: North Central
3: South
4: West
5: Unknown
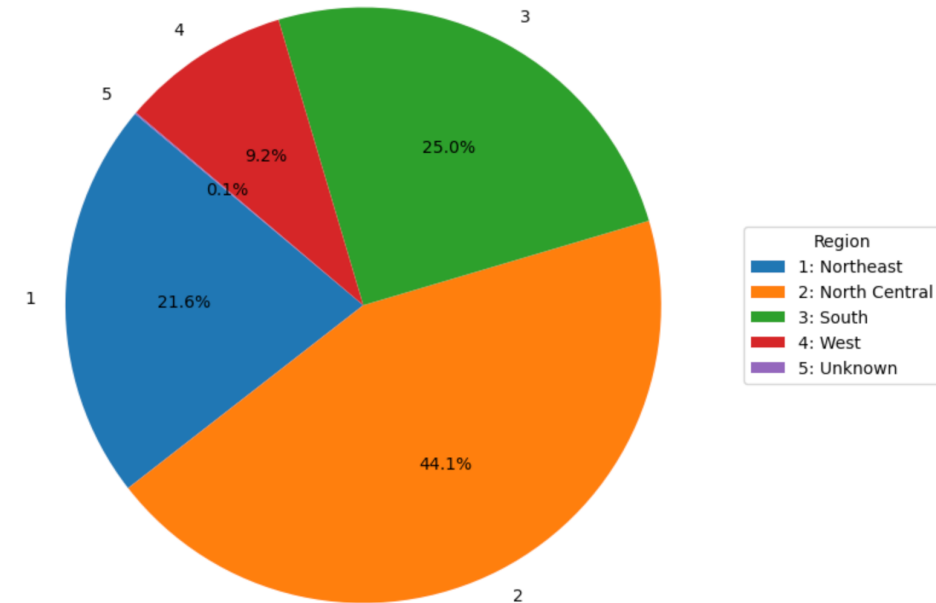
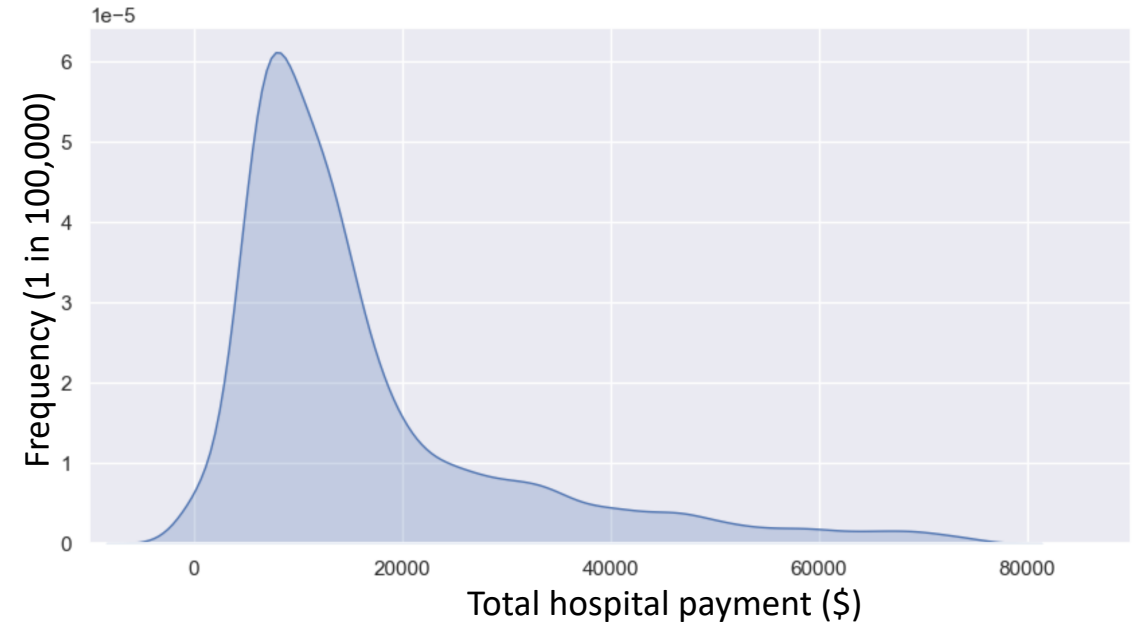Based on a random sample of n=22,958 enrollees

Histogram of Total Hospital Payment (Inpatient Admissions)

Commercial Database
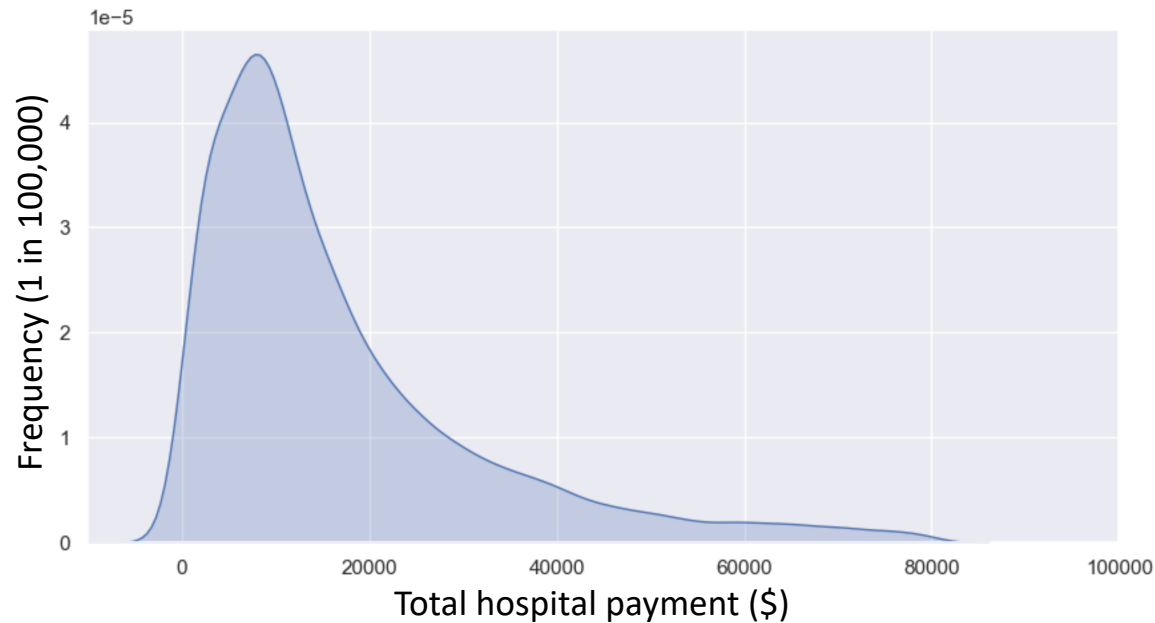
Medicare Database

Based on a random sample of 31,523 enrollees. The x-axis shows the total gross payments to a hospital for covered services provided during an admission.

Based on a random sample of 31,523 enrollees. The x-axis shows the total gross payments to a hospital for covered services provided during an admission.
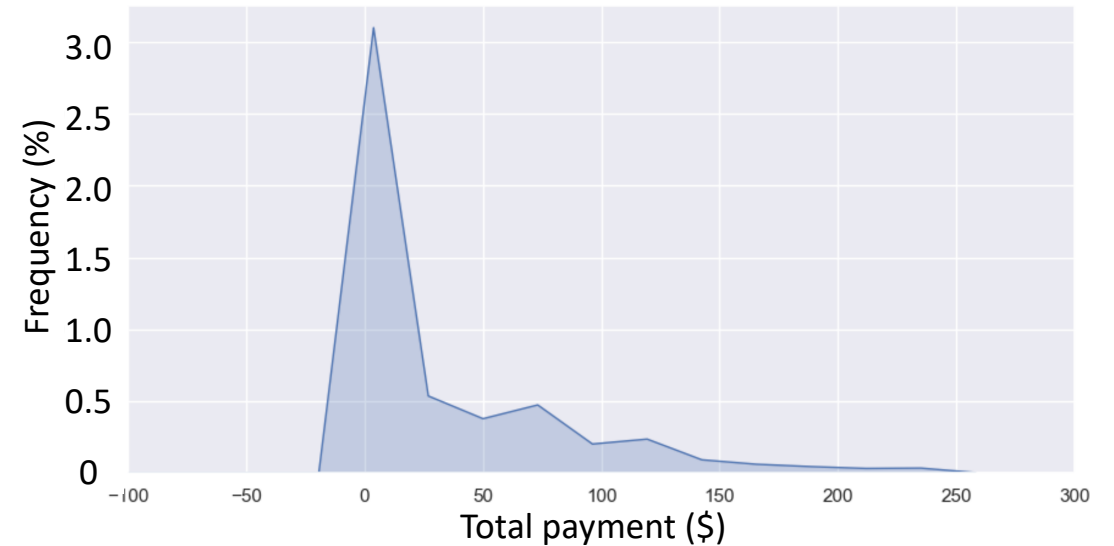
Histogram of Drug Payment – Outpatient Drug Claims

Commercial Database
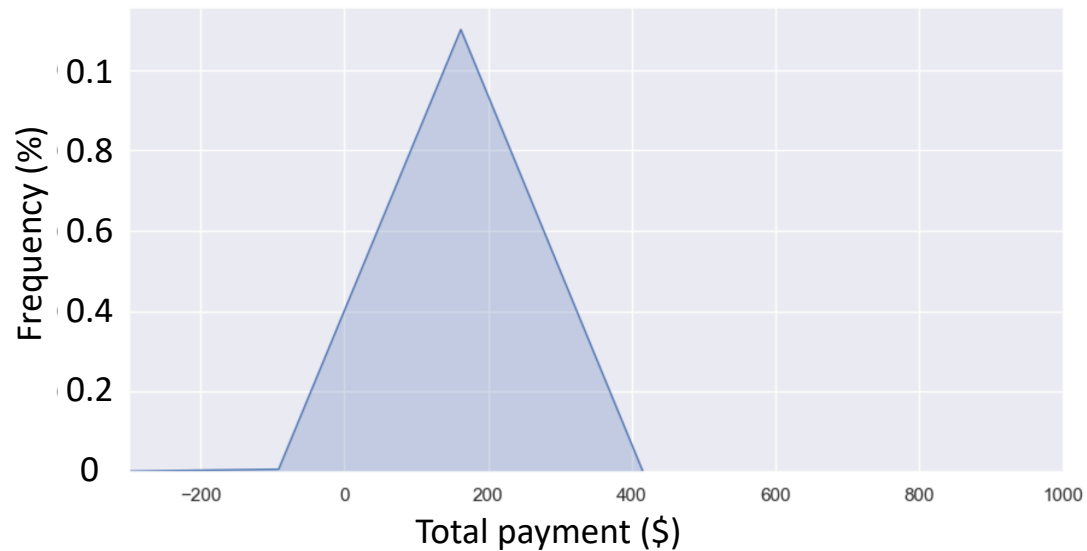
Based on 1% of sample size (n=256,685 patients). The x-axis shows the gross payments to a provider for a service. Payment equals the amount eligible for payment under the medical plan terms after applying rules such as discounts, but before applying COB, Copayments, and Deductibles.

Medicare Database

Based on 1% of sample size (n=279,534 patients). 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	• 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
	• Race, ethnicity • Lifestyle factors (SES, tobacco/substance use, income, education)	**No**	**No**
Insurance and financials	• Plan type (e.g., HMP, POS, PPO) • Date enrollment start, date enrollment end • Payment info for each service/hospitalization: • 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	• Visit type/department • Dates of admission, discharge • Date service start, end, payment • Length of stay, place/quantity of service	Yes	Yes

Clinical	• Principal and secondary diagnosis codes
	  • ICD-9/10 and others
	• Principal and secondary procedure codes
	  • CPT and others
	• Place/quantity of service
		Yes	Yes

	• Clinical measures (e.g., BP, BMI, lab results) • Physical examination results • Patient-reported measures	**No**	**No**
Medication	• Drug name, ID, and codes • 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

Limitations to Marketscan:

- No death data – outcomes studies for mortality not possible, difficult to account for competing risk of death for certain outcomes.
- No lab or vital sign data
- 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 inquires, email Biostatistics and Data Science Core at BDSC@UTSouthwestern.edu

MarketScan data consultants:

- Pamela Chan
- Mo Yang