



## **GoodRx Research Data Description**

Jeroen van Meijgaard, PhD

Diane Li,

Tori Marsh, MPH,

Amanda Nguyen, PhD

**GoodRx Research**

April, 2020

## **U.S. Claims Data**

GoodRx aims to provide the highest level of quality research about drugs, their availability and their cost. To fulfill this aim, the company supplements the data generated internally with data from external sources. The GoodRx Research team leverages these sources to track drug utilization, drug pricing, and drug shortages, and inform patients that may be impacted. Some of these sources are in the public domain, provided by the government generally, while other data sources are licensed. This document describes the coverage of licensed data and potential use to describe drug trends and utilization patterns in the United States.

GoodRx uses a representative sample of de-identified pharmacy claims across the United States, representing a wide range of channels and providers. Data includes the drug and quantity that was dispensed, the location and date of the transaction, and the usual and customary cost (U&C) of the product that was dispensed. The U&C cost is what would be paid by the patient in the absence of a contracted rate, or insurance price.

These data do not include any personal identifiers, and do not include information on what was paid for the medication.

The GoodRx Research team combines these claims with information on drugs from the GoodRx proprietary drug database and data on pharmacies to analyze trends in utilization and variation across geographic locales for specific drugs, classes of drugs, and medical conditions for which different drugs are prescribed.

## **Coverage of the Data**

### *Payers and plan types*

The claims cover both private and public insurance, including government payers like Medicare and Medicaid, and insurance products with different reimbursement structures like HMOs and Medicare Advantage.

### *Pharmacies*

More than 70,000 different pharmacies are represented across the six years reviewed (2014-2019), with just over 60,000 pharmacies identified as retail community pharmacies (RCPs) following the CMS definition (open to the public, operating in the community and not generally a mail order pharmacy, or servicing a member-only population such as an HMO pharmacy). The distribution of claims between chain and

independent pharmacies is shown in table 1. Additionally, just under 1% of all pharmacies provide mail order services, with a total claim share of 6.8% (2014-2019).

### Table 1. Share by Pharmacy Type (% , 2014-2019)

Pharmacy Type	Claims Share
Chain	62.2%
Independent	37.8%

GoodRx

#### Geography

The data represent claims across the country, with the breakdown shown in table 2. Due to the nature of the source of the data the claims are clustered geographically with 80% of the claims dispensed at about 20% of the pharmacies.<sup>1</sup>

### Table 2. Claims by Region (% , 2014-2019)

Region	Claims Share
Northeast	17.4%
Midwest	21.1%
South	37.9%
West	23.7%

Note: Adjusted estimates using claim weights

GoodRx

<sup>1</sup> By comparison, the top 15 largest U.S. pharmacy chains accounted for 76.1% of prescription drug revenues in 2018. Adam J. Fein. (2019). "The Top 15 U.S. Pharmacies of 2018: M&A Reshapes the dsMarket." Drug Channels Institute. Retrieved 4/23/2020 from <https://www.drugchannels.net/2019/02/the-top-15-us-pharmacies-of-2018-m.html>.

### *Drugs and manufacturers*

The data is not restricted to any particular set of drugs or manufacturers.

### **Adjusting, Weighting and Limitations of the Data**

Since the data is not a random sample, there is some clustering, both geographically and by patient population. Depending on the type of analysis, potential bias caused by this clustering is addressed in various ways. These include weighting the data, for example using geographic weights, to scale to a nationally representative sample, indexing trends to highlight changes in utilization rather than level of utilization, and adjusting and updating the product mix for indexing composite measures.

The data set is especially useful for changes that happen over a short time horizon, when the patient population is relatively steady, such as the uptake of new generic medications, shifts in prescriptions during shortages of specific products, or claim patterns during large events that impact large areas of the country, such as the COVID-19 pandemic.

### **Representativeness of the Data**

To assess how representative this data set is, we compare the share of claims for different products with publicly available data sets. First, we compare the share of claims to the prescription data from the Medical Expenditure Panel Survey (MEPS) conducted by the Agency for Healthcare Research and Quality (AHRQ)<sup>2</sup>, which is a nationally representative dataset of health care utilization and expenditures. We also compare the share of claims with data from Medicare (the prescriber data)<sup>3</sup>, and Medicaid (State Utilization data)<sup>4</sup>.

Each of these sources group drugs slightly differently. For the purposes of this comparison, we group drugs along the lines of MEPS, where drugs are generally identified using the generic compound name. The top 10 most commonly prescribed drugs in our data set are shown in table 3, with the share of claims shown for each of

---

<sup>2</sup> Agency for Healthcare Research and Quality. (2017). Medical Expenditure Panel Survey (MEPS). Retrieved 4/23/2020 from <https://www.ahrq.gov/data/meps.html>.

<sup>3</sup> Centers for Medicare and Medicaid Services. (2017). Medicare Provider Utilization and Payment Data: Part D Prescriber. Retrieved 4/23/2020 from <https://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/Medicare-Provider-Charge-Data/Part-D-Prescriber>.

<sup>4</sup> Medicaid. (2017). State Drug Utilization Data. Retrieved 4/23/2020 from <https://www.medicaid.gov/medicaid/prescription-drugs/state-drug-utilization-data/index.html>.

the comparison data sets. Note that the top 4 drugs in our list are also shown as the top 4 drugs in the MEPS data. The Medicare shares are slightly inflated due to reporting limitations where drugs are not counted when claims for a specific provider are less than a certain threshold. The claim patterns are markedly different for Medicare and Medicaid due to makeup of the population.

**Table 3. Drugs Dispensed, Claims Compared to Other Sources (% of total, 2017)**

Drug	GoodRx Sample	MEPS (95% CI)	Medicare (prescriber file)	Medicaid (utilization file)
Lisinopril	2.28%	3.14% - 3.56%	3.74%	1.7%
Atorvastatin	2.20%	3.15% - 3.55%	4.79%	1.61%
Levothyroxine	2.19%	3.03% - 3.48%	4.29%	1.33%
Metformin*	1.96%	2.34% - 2.68%	2.92%	1.6%
Hydrocodone/Acetaminophen	1.79%	1.14% - 1.42%	1.29%	1.7%
Amlodipine	1.74%	2.15% - 2.49%	3.76%	1.29%
Metoprolol*	1.72%	2.01% - 2.35%	3.86%	1.07%
Albuterol*	1.72%	1.46% - 1.75%	0.67%	3.27%
Omeprazole	1.51%	1.73% - 2.03%	2.63%	1.53%
Gabapentin	1.43%	1.33% - 1.61%	1.99%	1.84%

Note: Adjusted estimates using claim weights; \*Products grouped to align with MEPS data: metformin includes metformin ER; albuterol includes all albuterol inhalers; metoprolol includes metoprolol ER