How To Analyze And Improve The Usability Of Insurer Price Transparency Data

Yang Wang, Yuchen Wang, Angela Liu, Gerard F. Anderson, Ge Bai

MAY 25, 2023 **DOI:** 10.1377/forefront.20230524.904150

Editor's Note

This article is the latest in the Health Affairs Forefront *series, Provider Prices in the Commercial Sector, featuring analysis and discussion of physician, hospital, and other health care provider prices in the private-sector markets and their contribution to overall spending therein. Additional articles will be published throughout 2023. Readers are encouraged to review the Call for Submissions for this series. We are grateful to Arnold Ventures <https://www.arnoldventures.org/> for their support of this work.*

The Transparency in Coverage Final Rule <https://www.cms.gov/CCIIO/Resources/Regulations-and-Guidance/Downloads/CMS-Transparency-in-Coverage-9915F.pdf> , implemented on July 1, 2022, required all commercial insurers offering individual and group health plans to publicly disclose their in-network negotiated rates for specific procedures and providers. This regulation is designed <https://www.cms.gov/healthplan-price-transparency> to use price transparency to empower payers and patients, and to further promote price competition in health care markets to contain health care spending.

One *Health Affairs Forefront* article suggests that the insurer price transparency data may provide an enormous and comprehensive repository of health care pricing information, with identifiable insurers, plans, procedures, and providers, for the US commercial insurance market. The data demonstrate strength compared to other existing pricing data, such as the hospital price transparency data, state all-payer claim databases, and proprietary commercial claims databases.

Another *Health Affairs Forefront* article published today shares an overview of the data and provides a preliminary summary of in-network rates for a sample of procedures disclosed by four national health insurers. That piece notes that the massive size of the files will present a challenge to many researchers. Meanwhile, other researchers have also expressed concerns <https://georgetown.app.box.com/s/1ezsggz1c7smsaexkr8rght15sokgusl> about the data, given the large size, complicated structure, and data redundancy that could potentially undermine its usefulness for research and policy efforts.

In this article, based on our experience working with the data, we provide a simple guide on how to query and compile the insurer-disclosed data into a standardized pricing data set that can be used for empirical health policy research. We also identify a list of issues that insurers and policy makers could address to improve the usability of the data for the purpose of containing health care spending.

How To Analyze Insurer Price Transparency Data

The critical first step in analyzing these insurer-disclosed data is to generate a standardized pricing data set with insurer, plan, procedure, provider, and price information. To illustrate this process, we used the price transparency data disclosed for the month of January 2023 by the six largest commercial insurers <https://www.ama-assn.org/system/files/competition-health-insurance-us-markets.pdf> —United

HealthGroup, Elevance Health (formerly Anthem), CVS Health (formerly Aetna), Cigna, Kaiser Permanente, and Health Care Service Corporation (HCSC). These six insurers accounted for 61 percent <https://www.ama-assn.org/system/files/competition-health-insurance-us-markets.pdf> of the US commercial market in 2021.

We used the insurer price transparency compiled by Turquoise Health as of March 14, 2023. Turquoise Health is a health care data company whose nationwide hospital price transparency data set has been widely used in published academic studies. We accessed the Turquoise data via Trino platform, using SQL and R programming software. If researchers access the insurer price transparency data directly or through other vendors, the query interface might differ, but the data structure for each insurer will remain the same as originally disclosed by the insurer. Therefore, the discussion below has broad generalizability.

Larger insurers typically disclose their pricing information in multiple data files, where each data file is identified by a unique "data source name." Exhibit 1 shows a total of 2,926 data files from the six major insurers. For some insurers, data files with different data source names actually contain identical pricing information. We consider data files to have duplicated pricing information if their price tables include both the same number of total observations and the same total price amounts at the unit of cents. Notably, 83 percent, 43 percent, and 32 percent of data files disclosed by UnitedHealth Group, CVS Health, and Elevance Health, respectively, appear to be duplicative (exhibit 1). After dropping these redundant files, there are 1,554 (53 percent) files with unique pricing information.

Exhibit 1: Duplication of price data files released by six large insurers, January 2023

Major Insurers	2021 Market Share	Number of Original Data Files	Number of De-duplicated Data Files (percent)
UnitedHealth Group	15%	436	72 (17%)
Elevance Health (Anthem)	12%	243	166 (68%)
CVS Health (Aetna)	11%	2,134	1,208 (57%)
Cigna	10%	40	36 (90%)
Kaiser Permanente	7%	30	30 (100%)
Health Care Service Corporation (HCSC)*	6%	43	42 (98%)
Total	**61%**	**2,926**	**1,554 (53%)**

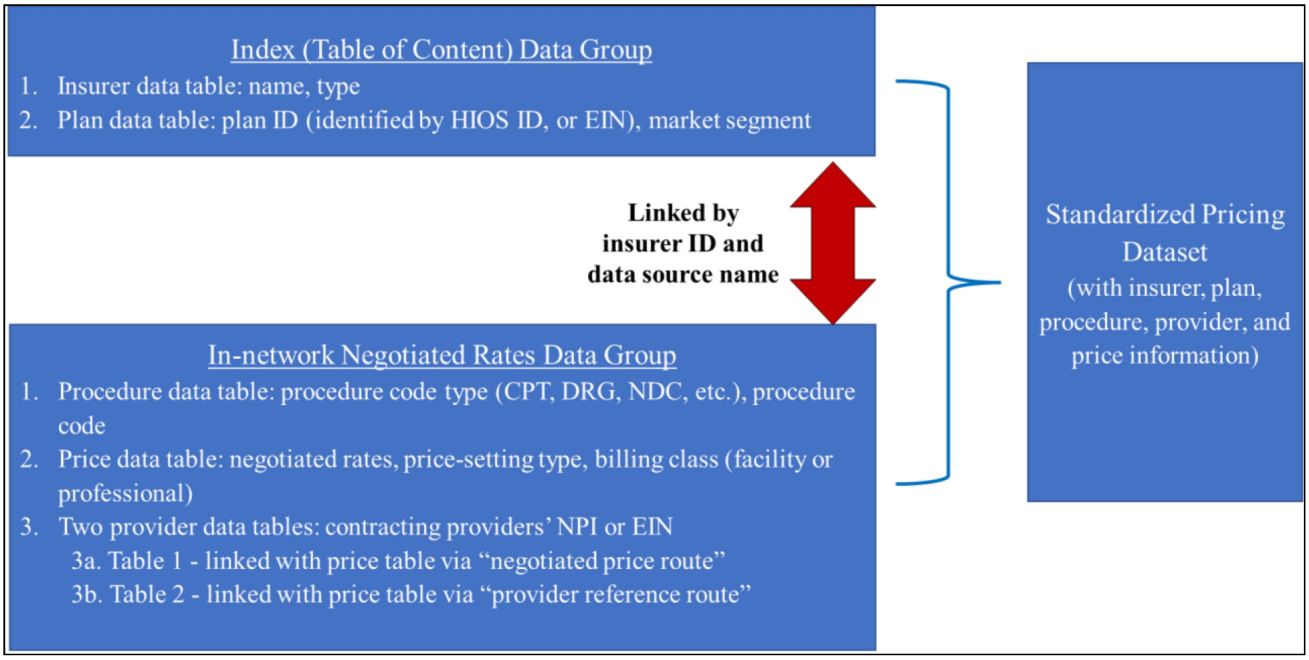
*Source: Authors' analysis of data released under the Transparency in Coverage rule <https://www.cms.gov/CCIIO/Resources/Regulations-and-Guidance/Downloads/CMS-Transparency-in-Coverage-9915F.pdf> . *Note: HCSC include Blue Cross Blue Shield of Texas, Oklahoma, Illinois, Montana, and New Mexico.*

We created one standardized pricing data set using information from these 1,554 data files. As illustrated in exhibit 2, we first queried tables from the index data group (table of contents) to extract basic insurer information (for example, name and type, such as insurance issuer, third-party administrator, and so forth) and plan information (for example, Health Insurance Oversight System identification number for Affordable Care Act [ACA] Marketplace plans or employer identification number [EIN] for employer-sponsored health plans).

We then queried tables from the in-network negotiated rates group to obtain corresponding information from the procedure table (procedure type and procedure code), price table (in-network rates, price-setting types, facility or professional setting), and provider table (contracting providers identified by either EIN or national provider identifier [NPI]). One challenge is that insurers uploaded provider information via two different pathways (the negotiated price route and provider reference route). As a result, there are two corresponding provider tables that both need to be included for data querying.

We then merged tables between the index data groups and in-network negotiated rates group using insurer identification and data source name to finish the construction of the standardized pricing data set. Due to the large data size and complex structure, it is computationally very demanding to query the entire 1,554 data files from the 6 largest national insurers at one time. Instead, we constructed loops in R to query through these data tables for each individual data file at a time.

Exhibit 2: Data tables needed to create the standardized pricing data set from insurer files released under the Transparency in Coverage rule



Source: Authors' analysis.

Notably, because we focused on the six major insurers with data parsed and compiled by Turquoise as of March 14, 2023, our results might not be extrapolated to data disclosed by other insurers. In addition, insurers upload their data on a monthly basis. Since the quantity and quality of data disclosed by insurers might vary from time to time, our results might not be generalizable to other time periods.

How Insurers Can Improve The Usability Of Insurer Price Transparency Data

The effectiveness of the Transparency in Coverage Final Rule in promoting competition and containing health care spending relies on the usability of the data disclosed by insurers. High usability will reduce processing cost and lower barriers for interested parties to use the data, thus achieving the intended goals of this rule. Based on our experience working with these data, we offer the following recommendations for commercial insurers to consider to improve the usability of the data. These recommendations are also relevant to policy makers interested in motivating insurers to improve. For example, the Department of Health and Human Services could consider issuing guidance regarding these issues.

1. Do Not Upload Duplicate Data Files

Our analysis showed that almost half of the data files uploaded by major insurers include duplicated price data. Redundant information enlarges the data size and complicates data structure. Insurers should upload de-duplicated data files so that each file contains unique pricing information.

2. Upload Provider Information Through One Reference Route And One Type Of Identifier

Currently, all major insurers except for UnitedHealth Group upload their provider information through two reference routes (or pathways for merging with the price table): negotiated price route and provider reference route. In addition, all major insurers used both NPI and EIN to identify providers. These features significantly complicate the data compiling process for external users. Insurers should upload their provider information via a single reference route (preferably the negotiated price route because it is more straightforward) and use one type of provider identification (preferably NPI, because it is more widely used).

3. Upload A Simple And Informative Pricing Data Summary

Such a summary should contain a selection of common procedures, such as the 70 Centers for Medicare and Medicaid Services-designated shoppable procedures as per the Hospital Price Transparency rule <https://www.federalregister.gov/documents/2019/11/27/2019-24931/medicare-and-medicaid-programs-cy-2020-hospital-outpatient-pps-policy-changes-and-payment-rates-and> . Each procedure should have the median, minimum, and maximum prices across all plans for each procedure-provider-network combination or for each procedure-provider-market combination. In this case, “market” refers to the ACA Marketplace, employer-sponsored fully insured market, or employer sponsored self-insured market. These summarized data should be uploaded in CSV format, which is accessible to a broad audience.

4. Disclose Meaningful Pricing Information

Currently, certain key pricing variables are measured inconsistently and ambiguously. For example, insurers disclosed prices under one of five price-setting types: negotiated, fee schedule, percentage, per diem, and derived. Prices disclosed under the percentage type are expressed as values between 0 and 100, most likely indicating the percentage of

the provider's list price. Since the list price is not included in the data, the percentage disclosed is not informative. In addition, prices disclosed under the derived type might not be paid under the fee-for-service mechanism but might be paid as part of a bundled service instead. No explanation was given on how these prices were derived.

5. Improve Clarity For Pharmaceutical Products

This insurer price transparency data include pharmaceutical prices identified by National Drug Code or Healthcare Common Procedure Coding System (for physician-administered drugs). However, it is unclear whether the prices for a given drug are measured by a consistent unit across different insurers and plans, especially for drugs with multiple indications, in which each unique indication may correspond to a unique dosage. This limitation creates a significant barrier to comparing prices across insurers and plans. In addition, the data do not distinguish the prices for the drug product from facility/professional fees incurred when administering or dispending the drug. Finally, it's less clear how the five price-setting types align with different types of drug prices, including negotiated prices, list prices, and average wholesale prices. Therefore, adding drug unit, indication, identification of drug product versus facility/professional fee, and more clarification on type of drug prices will improve the quality of disclosed pricing information for drugs.

Looking Ahead

The insurer price transparency data have an unprecedented and enormous potential to assist various stakeholders in understanding the pricing mechanisms surrounding US commercial health care markets, which will help in designing initiatives to lower health care prices and contain health care spending. Insurers have invested substantial time and effort in disclosing pricing data, but without robust data usability, the full potential of data cannot be achieved. It is time to make simple adjustments in data usability to make a significant impact on data usefulness.

Authors' Note

This work is supported by Arnold Ventures, PatientRightsAdvocate.org, and the Agency for Healthcare Research and Quality (Grant No. T32HS000029).

Article Metrics

<https://www.altmetric.com/details.php?domain=www.healthaffairs.org&citation_id=148956570>

Related

CONTENT

Pharmaceuticals & Medical Technology

TOPICS

Pharmaceuticals
Markets
Data Sets
Costs And Spending
In-Network Providers
Health Care Providers
Researchers
Quality Of Care
Affordable Care Act
Health Insurance Exchanges

Cite As

"How To Analyze And Improve The Usability Of Insurer Price Transparency Data", Health Affairs Forefront, May 25, 2023.

DOI: 10.1377/forefront.20230524.904150