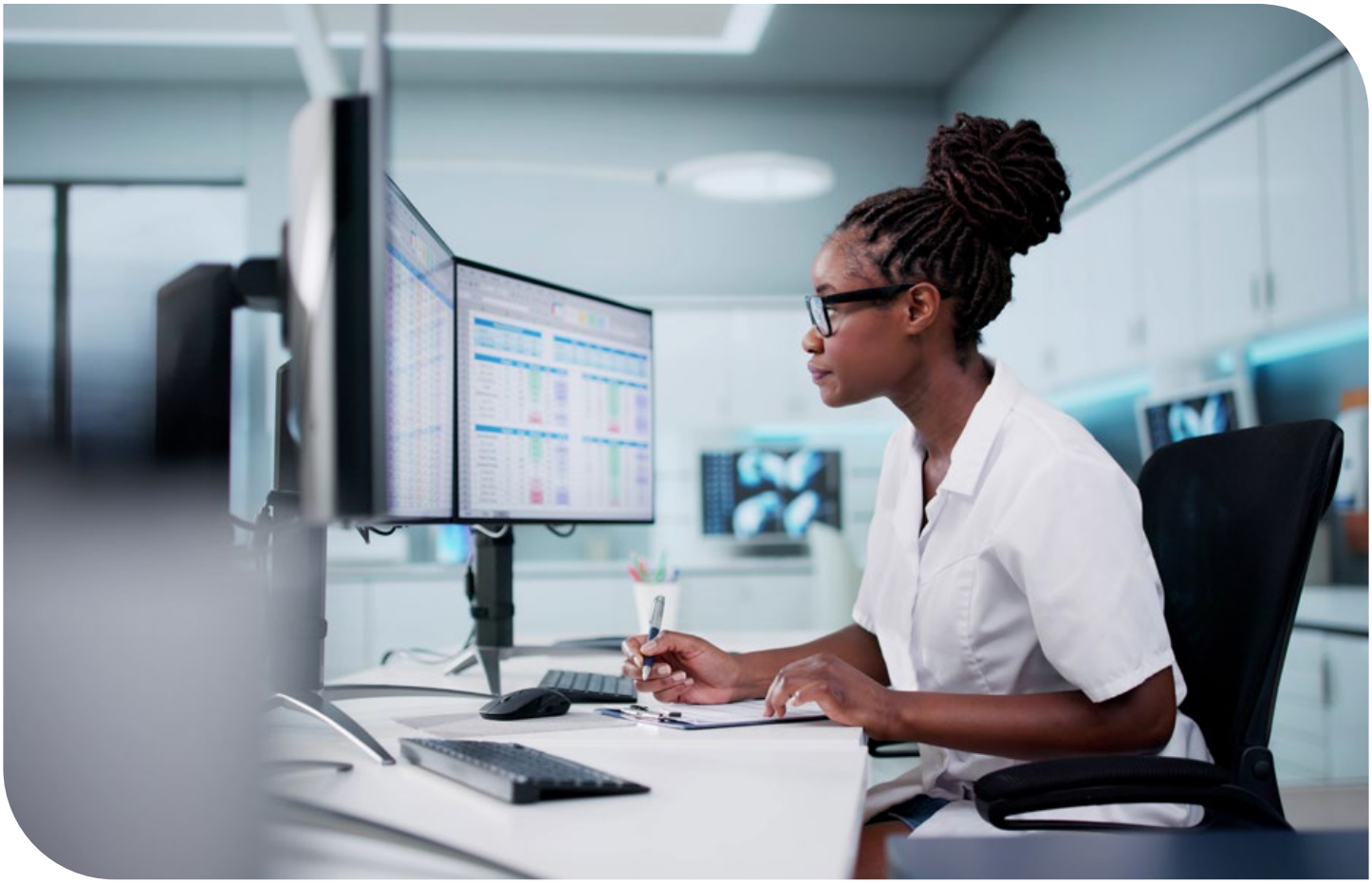

August 2025

Creating a Data Framework

by Combining Health Care Cost Transparency and Quality Data for Purchasers

Purchaser Business Group on Health

Introduction

Health care in the U.S. is unaffordable for workers and their families[1] with the total cost of health care continuing to rise at a pace consistently exceeding overall inflation — rising to $33,214 in 2025 for a family of four.[2] The national health expenditure is projected to grow from $4.9 trillion in 2023[3] to over $7 trillion by 2032.[4] Despite spending more on health care per capita compared to other peer nations,[5] the value of health care in the U.S. is comparatively low as quality remains variable,[6] access is increasingly limited,[7,8] and outcomes are poor compared to peer nations.[9]

The U.S. is uniquely reliant on employers* as a source of health insurance. Roughly 164.7 million people living in America — or 60.4% of the population — receive health coverage through an employer-sponsored plan.[10] Purchasers that self-insure are in most cases** legally obligated to act as fiduciaries and are responsible for securing high-quality, cost-effective health care for their employees. However, these employers often lack the necessary data and resources to fulfill this role effectively. Despite their responsibility as prudent fiduciaries, self-insured employers face significant barriers to accessing and evaluating the cost, quality, and value of health care services. A key challenge is the persistent opacity of healthcare pricing and institutional resistance to transparency from both insurers and health care providers.[11,12] In response to these barriers and in support of purchasers, the Purchaser Business Group on Health's ("PBGH") advocacy work has focused on working with policymakers to support transparency and competition as strategies to contain health care costs through market-based reform.[13,14]

* Throughout this whitepaper, "purchasers" will be used to refer collectively to public and private employers as well as non-employer purchasers of health care (e.g., Taft-Hartly benefit funds, state health plans, state exchanges).

** In rare cases, municipalities and non-federal governmental entities have not codified fiduciary responsibility for those responsible for the management and administration of their health plans, but the vast majority of self-insured employers are fiduciaries under ERISA (or a similarly worded state law statute, for non-ERISA plans). Fully-insured health plans are also subject to fiduciary responsibilities under ERISA, however, the focus of this project is self-insured health plans.

Health Care Price Transparency Reforms

Through PBGH's effective policy advocacy, much progress was made in 2019 and 2020 for transparency-based policies, both legislatively and regulatorily. During this time, Congress actively debated a broad piece of legislation entitled the "Lower Health Care Costs Act," which contained several transparency provisions that ultimately became law as part of the Consolidated Appropriations Act of 2021 ("CAA"). Among such provisions was a ban on contractual gag clauses that prevent employers from accessing their health care cost and quality information.[15]

On the regulatory front, in June 2019, President Trump issued Executive Order 13877, entitled "Improving Price and Quality Transparency in American Healthcare to Put Patients First," which directed administrative agencies to improve transparency in health care and empower patients to make informed decisions about their health care.[16] The Departments of Health and Human Services, Labor, and the Treasury ("Departments") responded to the President's Order twofold:

1. Using existing statutory authority under the Public Health Service Act and the Affordable Care Act to enact the Hospital Price Transparency ("HPT") Rule, which mandated hospitals to disclose their prices to the public including the disclosure of plan-specific "negotiated rates" annually through machine-readable files ("MRFs").[17]

2. Enacting the Transparency in Coverage ("TiC") Rule, which mandated that insurers and employers disclose their plan-specific negotiated rates for in-network and out-of-network health care services monthly through MRFs.[*,18]

* The Departments deferred enforcing TiC's third MRF requirement containing prescription drug pricing on August 20, 2021 and, two years later, rescinded the delay on September 27, 2023. Despite this recission, CMS has yet to specify a prescription drug MRF schema and only took further action on this requirement on May 22, 2025, when the agency issued a Request for Information ("RFI") to the public on how to implement the prescription drug MRF. PBGH leveraged insights from the data demonstration project in its July 2, 2025 response to this RFI.

These data sources provide purchasers, researchers, and consumers the first publicly available, independent information on health care pricing in the United States, holding the potential to disrupt how purchasers have traditionally analyzed insurance networks and health care costs. This type of information is crucial for employers to fulfill their fiduciary obligations under ERISA, which were dramatically enhanced with new requirements under the CAA. The core fiduciary obligation required by ERISA is to act "solely in the interests" of plan participants and beneficiaries by using plan assets for the "exclusive purpose" of purchasing high quality health care benefits while paying "only reasonable expenses."[19]

Forward thinking purchasers that understand the essence of their fiduciary duties, despite dozens of pending court cases litigating the exact scope and nature of those duties, are nonetheless proactively seeking ways to demonstrate fiduciary prudence.[20] These purchasers also highly value the quality of the health care they are purchasing on behalf of employees, a key component of fiduciary responsibility.*[,21,22]

The newly available price transparency data holds vast potential to inform health care purchasing strategies and enable use cases to lower health care costs. Despite apparent gaps in the MRFs that limit the usefulness of the data — and the substantial technical expertise, actuarial rigor, and data infrastructure to analyze it — the data can still yield actionable insights if it is supplemented with the appropriate datasets. PBGH's data demonstration project was specifically designed to bring the necessary datasets together to test the utility of the price transparency data and has demonstrated that such aggregated dataset is uniquely valuable to health care purchasers.

* The Department of Labor ("DOL") has specifically stated its opinion that it is important for health plan fiduciaries to not only consider the cost of services, but the quality as well. *("In selecting a health care provider [], as with the selection of any service provider under ERISA, the responsible plan fiduciary must engage in an objective process designed to elicit information necessary to assess the qualifications of the provider,* ***the quality of services offered,*** *and the reasonableness of the fees charged" . . . "failure to take quality of services into account . . . would constitute a breach of the fiduciary's duty")*

The PBGH Health Care Data Demonstration Project

PBGH spent over a year beginning in 2023 assessing and defining the unmet needs of large purchasers and the various use cases for which they wanted to use the newly available information. As these needs and their requirements were documented, PBGH assessed the market to understand who was developing resources to meet these needs. PBGH ultimately found no viable solutions; at the time, few if any vendors were collecting and integrating needed data - an unsurprising discovery given the complexity of the data as well as the fragmentation of data across multiple entities. As such, PBGH decided to pilot a Health Care Data Demonstration Project with our membership to test the feasibility of productively using the data and to identify opportunities and use cases for enabling effective health care purchasing decisions.

PBGH is uniquely qualified to lead a multi-employer initiative. PBGH's 35-year track record of designing and implementing innovative purchasing strategies with and for large employers and public purchasers is singular in the industry. PBGH has enabled multi-employer purchaser initiatives from group purchasing and direct contracting to a Centers of Excellence network. PBGH staff and partners have needed technical capabilities and deep subject matter expertise* to design and implement initiatives that benefit self insured employers and their workforce with critical independence from the health care industry that avoids conflicting incentives and interests.** Leveraging our credibility and strong relationships with our members, we enlisted five large purchaser members to join the pilot and contribute their historical claims and demographic data to explore what actionable insights and results could come from this project. This initial project focused on 10 regions across the U.S. where the participating purchasers had significant headcount:

- Atlanta, Georgia
- Chicago, Illinois
- Dallas, Texas
- Denver, Colorado
- Northern California
- Northern New Jersey/New York City
- Oregon
- Phoenix, Arizona
- Seattle/Puget Sound, Washington
- and Southern California.

* PBGH's unique expertise spans a wide range of health care issues, including policy advocacy, fiduciary and CAA knowledge, quality measurement, payment reform, and contracting principles development and implementation.

** PBGH is the only group representing large employers and public purchasers that has strict limitations for its membership which prevent any health care companies or affiliates from becoming a member.

PBGH conducted a request for proposal ("RFP") to find a partner with deep technical expertise and capabilities as well as a flexible platform needed to serve jumbo employers and process large, complex datasets. PBGH selected a data analytics partner and health care quality data suppliers to synthesize cost and quality variations by marrying the HPT and TiC data, provider quality metrics, and purchaser demographics and claims data. The combination of this information allowed for the development of commercial comparative benchmarks* and actionable insights for purchasers. PBGH selected Milliman as the data analytics and actuarial partner to enrich the data and develop reporting for each employer group using the publicly posted HPT and TiC transparency data (licensed to Milliman by Turquoise Health for use in the Milliman Transparent product) and provider quality and hospital safety data supplied by Embold Health ("Embold") and The Leapfrog Group ("Leapfrog"), respectively.

With the support of these partners and the financial support of the Peterson Center on Healthcare, PBGH's Health Care Data Demonstration Project findings and insights are meaningful and actionable for purchasers. The discourse about the usability of the transparency files has centered around how the files have been unusable for meaningful action.[23] However, this whitepaper demonstrates how the information can be used effectively, the methodology we used to create our data framework, and what can be learned about the data in its current form. This report does not necessarily reflect the views or positions of Milliman, Turquoise Health, Embold, or Leapfrog.

* The commercial market has long lacked independent reference points for prices, and has therefore traditionally relied on Medicare's payment rate as a stand-in, *de facto* benchmark for comparison purposes. Through this Health Care Data Demonstration Project, PBGH has made inroads in developing a benchmark that can be used to compare and evaluate the reasonableness of health care prices in the commercial market: A commercial comparative benchmark.

Methodology

Health care price transparency data is a powerful new resource that employers can use to guide strategic benefit design, plan management and vendor accountability, and support fiduciary compliance. The HPT Rule[24] and TiC Rule[25] require hospitals and payers, respectively, to publish negotiated payment rates by service code, e.g., Healthcare Common Procedure Coding System ("HCPCS") or diagnosis-related group ("DRG"), in publicly available MRFs.

Price transparency data provides the opportunity to perform unblinded, employer-specific comparisons of payment relationships between various payer networks and providers with current price information. However, developing meaningful comparisons with the price transparency data is very complex as the files are massive in size — encompassing petabytes* of raw data when all data is collected and maintained — and structured in a way that makes it difficult for benefit consultants, brokers, and employers let alone patients to work with and interpret the data.

Transparency data alone is inadequate to inform purchasing strategies. For meaningful and actionable analytics leveraging price transparency data, it is necessary to have deep subject matter expertise related to healthcare contracting and in-depth knowledge of employer-sponsored health care benefit plans. It also requires the combination of multiple data sources and actuarial and analytical rigor to assess data completeness, usability, and reliability of comparisons. Purchasers looking to work with the transparency files should have combined data with trusted analytic, actuarial, and expert advisory support to guard against the significant risk of misinterpreting the data.

For the purposes of the demonstration project, the following data was acquired:

- TiC data (also referred to as payer transparency data)
- HPT data (also referred to as hospital transparency data)
- Claims and demographic data from the five participating purchasers
- Leapfrog hospital safety data
- Embold provider quality scores

PBGH will publish results and comparative commercial benchmarks developed from the analyses separately, but this paper provides detailed information about each data source and describes how it was utilized for the PBGH Health Care Data Demonstration Project.

* A petabyte is equivalent to 1,000 terabytes or 1 million gigabytes. It is "commonly used to measure the capacity of hard drives, data centers, and cloud storage systems." A single petabyte can hold about "500 billion pages of text" or "256 million photos." Sources: [Link] / [Link]

TiC and HPT Data

Overview of Data Sources

Although the HPT and TiC data are often discussed interchangeably as "price transparency data," the two data sources have distinct elements and are subject to different sets of regulation. Each dataset contains important information that supports the goal of extracting meaningful and actionable insights to identify and evaluate the cost and relative value of health care services.

HPT Rule was effective January 1, 2021, and required hospitals to publish MRFs with contractual payment rates, to be updated at least annually. Hospitals must report all negotiated arrangements, so the data includes the commercial lines of business reported by payers, as well as Medicare Advantage and managed Medicaid. HPT data includes the negotiated rates, billed charges (i.e., list price or chargemaster amount), and cash prices.

The TiC Rule was effective July 1, 2022, and required payers to publish MRFs with contractual payment rates, to be updated monthly. Data is limited to commercial group and individual lines of business but includes health care pricing data far beyond the hospital setting — including professional and institutional rates and rates for all non-hospital facilities. Although the TiC MRFs will include full prescription drug pricing information in the near future, only medical-drug data was available during this initial iteration of the project.

In November 2023, CMS published changes to the HPT regulations as part of the CY 2024 OPPS Final Rule.[26] The Final Rule included changes to the HPT data that were effective July 1, 2024, and January 1, 2025. According to CMS, these changes were designed to improve monitoring and enforcement capabilities, reduce the compliance burden on hospitals, and align certain provisions with the TiC data.[27]

A high-level overview of the key requirements and differences in all the available data to date are summarized in Figure 1. How much of this data is needed by each individual stakeholder depends on the scope and purpose of proposed work.

Figure 1: HPT And TiC Data Attributes

	HPT (Hospital)	TiC (Payer)
Component	**(HPT regulation)**	**(TiC Final Rules)**
Data Update Frequency	Annually	Monthly
Relative Size of a Single MRF	Megabytes	Terabytes
Lines of Business Included	Commercial (Individual and Group), Medicare Advantage, Managed Medicaid	Commercial (Individual and Group)
Lines of Business Identified	No	Yes
Attestation Required	Yes (by an “authorized hospital executive”)	No
Data Elements Included		
Institutional Providers	Yes	Yes
ASC Providers	No	Yes
Professional Providers	No	Yes
Billed Charges	Yes	No
In-Network Negotiated Rates	Yes	Yes
Out-of-Network Rates	No	Separate file
Capitation / Bundled Payments	No	Yes
Medical Pharmacy Data	Yes	Yes

Overcoming Challenges with the Price Transparency Data

Although the transparency data is publicly available, both datasets present significant challenges that make it difficult to ingest and interpret the data directly from the MRFs. Therefore, it is critical to either invest in the resources or work with a data partner that understands the schemas and has developed processes and methodologies to efficiently parse and transform the data into usable formats.

One challenge is that the transparency data files are posted on different websites for each hospital and payer, and they must be located before they can be ingested. There are hundreds of thousands of MRFs that are all located in different locations. To overcome the challenges with data collection, ingestion, parsing, and aggregation, Milliman licenses the data from Turquoise Health.

Collecting and normalizing the raw transparency data forms an essential foundation, which enables further data enrichment and analysis to maximize its utility for purchasers. Once the desired files are ingested and parsed, the data must be reviewed and classified to identify the appropriate rates for the desired product (e.g., PPO, HMO), line of business (e.g., group, individual, Medicare, Medicaid), and network (e.g., Aetna Choice POS II, UHC Choice Plus, Cigna OAP).

After the data is appropriately classified, significant data cleaning and analytic expertise is required to gain meaningful insights. There is wide variation in the overall data completeness and quality, as the data is frequently incomplete and contains many different formats and contract structures that are not easy to compare without additional analysis and subject matter expertise. A February 2025 brief by the Peterson-KFF Health System Tracker[28] documented similar challenges to those faced in the PBGH Data Demonstration Project, such as: (1) "ghost rates" in the data (rates for services not provided by a certain provider, e.g., colonoscopies performed by a dermatologist); (2) multiple rates posted for the same service and provider; (3) provider mapping ambiguities (e.g., facilities with multiple National Provider Identifiers (NPIs)), and; (4) rate structure differences. To overcome these challenges, Milliman applied notable enhancements to the data to address each of these limitations:

Network and Plan Identification

- Logic is applied to both datasets to determine standardized payer name, line of business, product, and network (where possible).
    - In the HPT data, hospitals are required to provide the payer's name and plan name for each of their contracts. However, there are no standardized values for these fields and hospitals can use whatever naming convention they choose. This leads to wide variation across hospitals. An algorithm based on keywords was developed to parse out clean payer, line of business, and product names. It is often not possible to identify a specific network in the HPT data.
    - In the TiC data, a table of contents file is often provided by each payer. These files give insight into plan IDs that correspond to each payer's set of networks and link to various MRFs. Each payer's posted files, including table of contents and in-network rate files, are manually reviewed to classify networks. This is critical for employers who need to understand the contracted rates for specific payer networks, including broad PPOs, high-performance networks, and HMOs, as the negotiated rates can vary by network.

Utilization Data

- Neither the HPT or TiC data includes utilization or service mix information to allow for aggregations of the code-level data. In order to create actionable insights from the data, nationwide utilization data specific to the line of business (e.g., group commercial), provider type (e.g., short-term acute care hospital), place of service, bill type, and code set was applied from Milliman's research databases of over 70 million commercial lives. The utilization was appended to the transparency data, on a code-level basis, to aggregate the transparency data into meaningful service categories. Adding utilization information is a critical enhancement to the data, as it enables a purchaser to compare whole service lines and contracts (including any contractual differences that may exist).*
- Purchaser-specific claims data will generally not have large enough volumes to make broad assessments of cost differences within and across large geographic regions, which is why a national commercial dataset was used to inform the broad analyses. However, the purchaser-specific data allowed for targeted and in-depth analysis of top providers and services.

Medicare Benchmarks

- Milliman's Global Relative Value Units ("GRVUs") are assigned to the data, and are the basis for the key metric used for comparison analyses (Percent of GRVU Medicare) and gauging reasonableness of individual unit prices.[29]
  - GRVUs are a set of Relative Value Units ("RVUs") that cover the entire range of healthcare services. GRVUs can be thought of as an all-payer version of Medicare as it overcomes common limitations of contract comparisons that use Medicare fee schedules. In particular, GRVUs include adjustments to better account for service categories that are common in commercial populations but are not captured adequately by an age 65+/disabled Medicare population, (e.g., maternity, neonatal intensive care unit, pediatrics). The GRVU Medicare value for many high-volume codes will match Medicare, but some services will vary significantly.
  - While CMS Medicare is a common benchmark,** one key limitation is that Medicare reimbursements vary substantially by geography and provider type, which does not allow for a relative price comparison between two differently situated facilities (e.g., an Academic Medical Center in New York City compared to a short-term acute hospital in rural Pennsylvania).

* In CMS's CY 2026 OPPS Proposed Rule, CMS proposed to require hospital MRFs to include "count[s] of allowed amounts." PBGH's response to the Administration's hospital enforcement RFI voiced support for CMS to finalize this proposal based on this current limitation with the price transparency data. Likewise, in PBGH's TiC RFI response, we advised CMS to require drug frequency and counts in the prescription drug MRFs.

** Medicare's fee-for-service rate is often used as a benchmark for commercial market pricing. For instance, 250% of Medicare would represent a commercial price for an item or service that is 2.5x what Medicare would pay for the same item and service, at the same location.

Data Quality and Usability

- The quality of data for each unique price record is addressed by assessing the reasonableness of the negotiated amount and service code. Negotiated amount outliers are identified as well as duplicates and invalid codes. If there are multiple rates (or duplicate rates) for a specific provider / plan / code / place of service / modifier combination, the data record with the median negotiated rate for that combination was included and the rest was flagged as unusable. Records flagged for exclusion are excluded from the aggregated relativities.
- To aid with analytics, metrics were developed to quantify the usability of the data. The most important data usability metric is a *Percent of Expected* value. *Percent of Expected* measures the total RVUs associated with transparency data where negotiated rates are available (or can be derived) and deemed reasonable, expressed as a percentage of the total RVUs we would expect for the given provider type and line of business ("LOB") if all negotiated rates were available. This metric helps quantify the comprehensiveness and usability of the posted transparency data.*
- For example, a *Percent of Expected* value of 90% indicates that the posted data reflects 90% of the total volume of services one would expect to be present, which represents a comprehensive data point. *Percent of Expected* values between 50 – 75% are considered generally reliable, 25 – 50% are moderately reliable and additional investigation is recommended, and less than 25% implies the need for code-level data investigation.

Service Type Grouping

- Milliman's *Health Cost Guidelines* categories are assigned to code-level data, allowing rolled up price relativities to be reviewed by service type, setting, and in total.

Provider Group Mapping

- Facilities were mapped in the transparency data by linking organization National Provider Identifier ("NPI") to CMS Certification Number ("CCN"). Professional providers are mapped into meaningful provider groups based on Milliman's own database of NPIs and Taxpayer Identification Numbers ("TINs").

Rate Methodology Type

- An appropriate rate methodology (e.g., per case, per day) is determined for each data record as not all negotiated rates in the transparency data are reported on a consistent reimbursement basis.

Percent of Charge Records

- In the HPT data, both billed and allowed amounts are present in the data. In the TiC data, billed charges are not a required field and, therefore, not present. However, payers are permitted to report the percentage value that applies to each code for services reimbursed as a percentage of billed charges. Milliman crosswalks between the HPT and TiC datasets to utilize billed charges from the hospital data to estimate negotiated amounts based on the reported percentage of charge when the code level data overlaps.

* However, it is important to note that the *Percent of Expected* metric is not intended to be a measure of compliance with the HPT or TiC requirements.

Comparing HPT and TiC

- Since the HPT and TiC data are not always fully populated or complete, the two datasets can complement each other and help fill in gaps in the data or corroborate findings. While posted negotiated rates do not always match adjudicated claims (discussed below), when prices in the two datasets do align, it increases confidence in the accuracy of the results. In cases where results from the payer and hospital data do not align, utilizing purchaser claims data (where available) to validate which data source more closely aligns with how the claims are adjudicated can help identify which dataset is more reliable.

Key Learnings Regarding the HPT and TiC Data

There were several key learnings from the project that will help inform future use of transparency data for purchasers:

- **Supplemental data sources are required:** Limitations of the publicly posted price transparency data require other data assets, such as utilization data, purchaser claims data, and a benchmark like GRVUs to help make the data comparable. Standalone transparency files without supplemental data are not adequately reliable for large scale health care purchasing decisions. Code-by-code level analysis is insufficient to provide the amount of context needed for a purchaser to evaluate the total cost of care.
- **Recognition of differences between adjudicated claims and negotiated rates:** Sometimes an adjudicated claim will match a negotiated rate, but not always. The transparency data does not capture all the nuances associated with an adjudicated health care claim and is therefore wholly insufficient to understand pricing without purchaser claims data. For example:
  - Some claims are subject to payment terms and adjustments not reflected in the posted rates including outlier claims / stop-loss provisions.
  - Some types of services are coded / priced as a bundle or set of services receiving one payment and bundling is not always clear in the posted rates.
  - Some third-party administrators ("TPAs") contract differently than others or report prices with different code sets or reimbursement methods (i.e., percent of charge vs. case rate vs. fee for service).
  - Claim payment hierarchies are not clear in the posted data and may appear as two prices for the same service (e.g., maternity reimbursement structured as case rate for 0 – 2 days, then as *per diem* for days 3+).

Because the price transparency data does not always explicitly match claims data, knowledge of healthcare contracting is important to understand and interpret the posted data.* Mismatches between the adjudicated claims and the negotiated rates *may* also indicate that there are provider network agreement terms that are affecting what adjudicated price ultimately results, or *may* indicate there are third-party entities adding costs / fees on top of negotiated rates.** Specifically, these terms could relate to claims adjudication logic like outliers, "lesser of," multiple procedures, value-based care, etc. Knowing this information is vital for a purchaser to understand their health care costs.

- **Significant attention is needed to identify the correct network that is most appropriate to use for comparison:** Since HPT and TiC each present specific contracts in different ways, detailed research into the MRFs is required to ensure any analysis reflects the posted contractual rates for both the purchaser and the desired peer networks for comparison. Otherwise, analyses reflect only medians or averages across various networks, which is not actionable for specific purchasers.
- **Recognition of significant, large gaps in the reported data:** The transparency data can be sparsely populated, resulting in significant gaps and poor data quality. When this occurs, it tends to be consistent across either (1) the same payer, or (2) the same geographic region. In addition, some of the data must be excluded from analysis when it clearly falls outside the bounds of a reasonable range of costs. For example, when posted negotiated rates are less than half of the known Medicare reimbursements, those rates are considered to be unreasonably low.
- **Incomplete data can still allow for valuable insights:** The gaps in the data do not render the data unusable, as there are still valuable insights to mine at the code level or service line level. To reliably use this data despite its limitations — the key is understanding how much data supports a specific comparison.
- **Imputing data can lead to inaccurate conclusions.** Given the numerous challenges in interpreting the posted data, including the need to assess completeness and reliability, we caution purchasers against extrapolating findings or forming broad generalizations. For example, when assessing which carrier network has the best contracts in a given geographic region, it is important to accept the data as it is and not draw conclusions through imputation or extrapolation of data that is incomplete or inaccurate.

* Equally as important are the provider network agreements themselves. Section 201 of the CAA entitles purchasers to the "claims-related financial obligations" that are included in the provider contract, but not access to the provider contract itself. This is a severe limitation that presents an obstacle to purchasers understanding why their adjudicated claims do not match negotiated rates. See PBGH's recent RFI response beginning on p. 13.

** Last year, the U.S. Senate questioned the role of TPA intermediaries in driving up health care costs following a New York Times investigation of MultiPlan. Georgetown CHIR published a May 2025 article in *Health Affairs* describing the state of knowledge on these intermediaries, the conflicts of interest at play, and the need for greater scrutiny into these undisclosed fees — all of which may finally be illuminated from future iterations of this data project.

Overview of Data Sources

Detailed medical claims and enrollment data were collected from each participating purchaser to determine which geographic markets, providers, and services were most relevant for each purchaser. As noted earlier, purchaser-specific claims were not used to aggregate code-level data since claims for a single employer may not be credible when making broad market assessments. Claims data also served as a validation source for price transparency information when discrepancies emerged. Each purchaser was asked to provide claim line-level data for the most recent 36 months available. The data request included essential fields to facilitate alignment with the price transparency data. In particular, the request asked for all relevant service codes (CPT, HCPCS, Modifiers, DRGs), rendering provider information including both NPIs and TINs, place of service classification, allowed amounts, paid amounts, claimant demographic information, and dates of service.

The quality of the data received varied between purchasers. In general, the data from each purchaser demonstrated a high level of completeness and validity across the critical coding fields. This allowed for clean mapping to Milliman's *Health Cost Guidelines* service categories to enable reliable cost and utilization analysis. However, there were certain flaws with some data that required extra processing steps. Several examples of which purchasers should be aware are included below:

- Data translation mappings* are necessary, as certain fields like place of service contained values that were not consistent with expected values.
- Unexpected and inconstant placeholder values, such as "~" and "~Missing", must be removed.
- Fields with unprintable characters and excess whitespace must be scrubbed.
- There are missing or unavailable data fields, such as billed amounts,** NPI or TIN, or modifier codes.

* Data translation mappings refer to information carriers / TPAs provide that maps unexpected values to expected values.

** For example, when the TPA or data warehouse vendor only provided allowed amounts for the purchaser's claims.

Of note, none of the TPAs or data warehouse vendors met the file specifications. Many provided an existing or general data extract, which required additional time to map the data into Milliman's required format. It is important for purchasers to account for the lead time needed to transcribe and clean data for any analysis project, so it can appropriately inform purchasing decisions.

Process to Obtain Purchaser Claims Data

The process to obtain the data varied significantly for each purchaser. All purchasers were provided with the same data request to use with their vendors. This request included the required data elements, desired format, and descriptions of each data field. Purchasers generally requested that we work directly with their data warehouse vendor or TPA to obtain the requested information. In some cases, TPAs expressed concern with providing the specific data elements of the purchasers' claims data, as they considered it proprietary. For example, some TPAs objected to including both billed amount and allowed amount information. These concerns were overcome because the contractual payment rates are already publicly available in the HPT and TiC data. Furthermore, self-funded purchasers have a right to access and share their own claims data under Section 201 of the CAA. Despite having clear data rights, at times purchasers needed to enlist legal counsel to help enforce their rights to their own data. Milliman ultimately received the minimum necessary data for each purchaser to complete the analyses.

The de-identified data was provided via Secure File Transfer Protocol ("SFTP"). Upon receipt, all data was stored in secured, access-controlled environments. Data from each purchaser was segregated into distinct network locations and databases, ensuring strict logical separation. This segregation was enforced using access control lists ("ACLs"), which restricted access to each dataset on a need-to-know basis and only to authorized personnel. This multilayered approach to data handling ensured that each purchaser's information remains isolated, secured, and protected throughout all stages of storage, data processing, and analysis. Adhering to these protocols, especially those related to data use agreements, is critical when handling sensitive claims data to maintain confidentiality, integrity, and compliance with applicable data protection standards.

Key Learnings from Purchaser Claims Data

This project required us to align purchaser claims data with price transparency data to create meaningful cost comparisons across different TPAs, network products, and providers. While the quality of the claims data varied by purchaser, as did the data issues we encountered, PBGH learned several takeaways:

- **Purchasers should ensure they have access to all detailed, de-identified claims data needed for meaningful analysis:** While there were minimal invalid codes, complete and detailed claim elements are necessary to make credible comparisons of top services. As an example, modifiers can significantly impact the allowed amount and should be included.

  The statutory language of the CAA is clear that purchasers have (1) complete, unrestricted access to their deidentified claims data, including billed and allowed amounts and any terms in provider contracts that affect payment, and (2) unrestricted rights to share this data with a HIPAA business associate of their choosing for analysis, interpretation, validation, and any other purposes they deem important — including this demonstration project.[30]

  However, one purchaser in the PBGH Health Care Data Demonstration Project encountered hurdles to accessing their de-identified claims data for the purposes of the project. This

provides further evidence of the need for policymakers to clarify and strengthen the CAA's prohibition on gag clauses, so purchasers have fewer hurdles to accessing their data.*

- **Accurate provider ID fields are essential for 'apples to apples' comparisons:** Although we requested both NPI and TIN values for this project, the data received was often incomplete or inaccurate. We expected TINs or organizational NPIs for facility and Ambulatory Surgical Center (ASC) claims, and individual NPIs for professional claims. However, this is not what we received.
  - Some provider ID issues inhibited our ability to map a facility or medical group name to each claim. For example:
    - Claims missing provider IDs could not be classified.
    - We observed inpatient / outpatient facility claims with individual NPIs only. This limited the ability to map claims to the rendering facility and contracted rate from the transparency data.
    - We received claims for individual professional services with TIN or organizational NPI, which limited the ability to map quality scores for individual physicians.
- **Raw data without any prior manipulation is preferable:** Data warehouses can transform data in ways that hinder analysis based on the information's prior interpretation and synthesis. Each purchaser has existing, agreed-upon data feeds, which may not contain all the necessary information for this type of project. When considering data warehouses, purchasers should carefully vet any processing steps to ensure flexibility for future data analysis.

* PBGH has seized every opportunity to convey this to the Congress and the Administration and has suggested — in multiple venues — that aside from clarifying and strengthening the CAA as-written, it would be appropriate to institute meaningfully penalties on service providers (i.e., carriers, TPAs) that continue to data block. See: [Link]

Health Care Quality Information

The TiC and HPT transparency data is price data only and does not include any information related to health care quality or safety. As a result, quality and safety information needs to be sourced independently before being ingested and appended. As part of this data demonstration project, we obtained health care quality information from two sources:

1. **Leapfrog Hospital Safety Data**

   Leapfrog supplied its Hospital Safety Grades for all hospitals in the Leapfrog database.

   The Leapfrog Hospital Safety Grade is a nationally recognized measure of patient safety, assigning letter grades to nearly 3,000 general acute-care hospitals twice a year based on their ability to prevent medical errors, injuries, and infections. Using up to 30 evidence-based performance measures from trusted sources like CMS and the Leapfrog Hospital Survey, the grades help patients and health care purchasers make informed decisions about where to seek the safest care.

   As part of the Health Care Data Demonstration Project, the letter grade supplied by Leapfrog was reported alongside the transparency price comparisons.

2. **Embold Provider Quality Scores**

   Embold supplied its proprietary provider quality scores by individual NPI.

   Using aggregated data from public and proprietary sources, Embold identifies quality measures with the highest clinical impact and applies them across a broad dataset. This data allows users to look beyond a single episode of care and follow a patient's complete healthcare journey to reliably evaluate the appropriateness and effectiveness of the care that they receive. Embold can identify appropriate and effective care practices in local communities by applying these measures across our dataset.

   Milliman reported this data by aggregating and averaging individual provider scores at the medical group level, either in total or by specialty. This approach was consistent with the way Milliman Transparent reports price data. Given the importance of health care quality to self-insured employers, this component added unique value to the work. The Health Care Data Demonstration Project did not demonstrate a correlation between higher prices and higher quality, challenging purchasers' assumptions and informing future network design decisions.

Conclusion

PBGH has developed an important health care data framework for its purchaser members that are seeking to purchase high value health care and hold vendors accountable, as prudent fiduciaries. This Health Care Data Demonstration Project will be extended to additional members and non-member purchasers as demand for this type of analysis continues to grow. Expanding the project is also consistent with PBGH's public mission of sharing the process, challenges, and learnings to inform all industry stakeholders and policy makers working with the transparency files to improve health care value.

This data demonstration project also has significant policy implications that align with policymakers' support for transparency as a market enabler. Transparency is necessary for meaningful private market reforms, supports the goals of many self-insured purchasers, and remains a top policy priority of PBGH and our members. The PBGH Health Care Data Demonstration Project provides insights into the usability of the data to inform purchasers' cost-containment goals, which in turn provides insights into whether the implementation of the federal government's transparency-based policies has been successful.

PBGH with the support of its partners has demonstrated that its framework of combining the transparency files (TiC and HPT), health care quality data, and Purchaser claims data yields new insights and approaches to analyze, manage, procure, negotiate, and measure health program value and effectiveness. Additional insights and findings will be shared in upcoming publications by PBGH.

About the Purchaser Business Group on Health (PBGH)

PBGH is a 501(c)(3) non-profit coalition of health care purchasers comprised of members that include the largest public and private purchasers of health care in the United States. Collectively, these organizations spend roughly $350 billion annually buying health care for nearly 21 million employees and their families. PBGH supports its members in implementing innovative solutions to improve health care outcomes and value.

Acknowledgement of the Peterson Center on Healthcare's Financial Support

PBGH is grateful to the Peterson Center on Healthcare for providing funding support and thought partnership for the development and implementation of this data demonstration project and the creation and dissemination of its results and policy implications.

Endnotes

1 Collins et al. (Oct. 26, 2023) "Paying for It: How Health Care Costs and Medical Debt Are Making Americans Sicker and Poorer" *The Commonwealth Fund* [Link]

2 Bell et al. (May 27, 2025) "2025 Milliman Medical Index" *Milliman* [Link] *(The annual figure, inclusive of premiums and all patient out-of-pocket costs, is $35,119 pre-rebate. Post-rebate, the figure is $33,214 as-cited.)*

3 CMS (2024) "National Health Expenditure Data: Historical NHE, 2023" *NHE Fact Sheet* [Link] *(Last modified 6/24/2025)*

4 Fiore et al. (Jun. 12, 2024) "National Health Expenditure Projections, 2023–32: Payer Trends Diverge As Pandemic-Related Policies Fade" *Health Affairs*, Vol. 43, No. 7 [Link]

5 Wager et al. (Apr. 9, 2025) "How Does Health Spending in the U.S. Compare to Other Countries?" *Peterson-KEFF Health Systems Tracker* [Link]

6 Burke et al. (Feb. 2014) "The Complex Relationship between Cost and Quality in US Health Care" *Policy Forum; Virtual Mentor (AMA Journal of Ethics)*, Vol. 16, No. 2 [Link]

7 Sullivan and Jinnett (Dec. 5, 2024). "Barriers Accessing Specialty Care in the United States: A Patient Perspective" *BMC Health Services Research,* Vol. 24, No. 1 [Link]

8 Jabbarpour et al. (Feb. 28, 2024) "The Health of US Primary Care: 2024 Scorecard Report — No One Can See You Now" *Milbank Memorial Fund* [Link]

9 Gunja et al. (Jan. 31, 2023) "U.S. Health Care from a Global Perspective, 2022: Accelerating Spending, Worsening Outcomes" *The Commonwealth Fund* [Link]

10 Claxton et al. (May 2024) "Employer-Sponsored Health Insurance 101" *KFF* [Link]

11 See, e.g., OIG (Nov. 8, 2024) "Not All Selected Hospitals Complied With the Hospital Price Transparency Rule" [Link] and Jiang et al. (Dec. 3, 2024) "Enforcing Hospital Price Transparency: Lessons from CMS Actions" *Health Affairs Forefront* [Link]

12 See also, Choi et al. (Jun. 13, 2025) "Technical Challenges with Private Health Insurance Price Transparency Data" *Congressional Research Service, Report No. 48570* [Link]

13 PBGH (Jan. 31, 2025) "Purchaser Policy Priorities in 2025" *Issue Brief* [Link]

14 See also, PBGH (Mar. 31, 2025) "Purchaser Innovation and Policy Engagement Against a Backdrop of Unaffordability and Fiduciary Risk" *Issue Brief* [Link] and PBGH (Jan. 2024) "The 2024 CAA Transparency Data Roadmap" *Fiduciary Discussion Group Webinar Series* [Member-Only Link]

15 Consolidated Appropriations Act, 2021 (Dec. 27, 2020) Pub. L. No. 116-260 [Link] § 201 beginning at **p. 2891**

16 White House (Jun. 24, 2019) "Improving Price and Quality Transparency in American Healthcare to Put Patients First" *Executive Order No. 13877* [Link]

17 CMS (N.D.) "Hospital Price Transparency" *Initiatives* [Link] *(Last modified 5/22/2025)*

18 CMS (N.D.) "Transparency in Coverage" *Initiatives* [Link] *(Last modified 9/10/2024)*

19 U.S. Department of Labor. (Sep. 2019) "Understanding Your Fiduciary Responsibilities Under a Group Health Plan" [Link]

20 See, e.g., *Lewandowski v. Johnson & Johnson* (1/24/24), *S.M.O. v. Mayo Clinic* (4/2/24), *Navarro et al. v. Wells Fargo* (7/30/24), *Stern et al. v. JPMorgan Chase & Co* (3/13/25), and *Barbich and Lindvall v. Northwestern University* (6/20/25). Many more cases are pending litigating the nature of TPA fiduciary responsibilities to purchasers and plan participants.

21 Binder et al. (Jul. 19, 2023) "The CAA and Health Care Quality" *The Leapfrog Group* [Link]

22 Binder and Gelfand (Jul. 14, 2025) "Healthcare Price Transparency Works When Quality is Figured In" *RealClear Health* [Link]

23 Whaley (May 16, 2023) "Health Care Price Transparency: Opportunities to Improve Affordability and Data Effectiveness" *Testimony Before the House Ways and Means Committee* [Link]

24 CMS (Nov. 27, 2019) Medicare and Medicaid programs: CY 2020 Hospital Outpatient PPS Policy Changes and Payment Rates and Ambulatory Surgical Center Payment System Policy Changes and Payment Rates" *Final Rule* [Link]

25 CMS (Nov. 12, 2020) Transparency in Coverage" *Final rule* [Link]

26 CMS (Nov. 22, 2023) "Medicare Program: Hospital Outpatient Prospective Payment and Ambulatory Surgical Center Payment Systems; Quality Reporting Programs; Payment for Intensive Outpatient Services in Hospital Outpatient Departments, Community Mental Health Centers, Rural Health Clinics, Federally Qualified Health Centers, and Opioid Treatment Programs; Hospital Price Transparency; Changes to Community Mental Health Centers Conditions of Participation, Changes to the Inpatient Prospective Payment System Medicare Code Editor; Rural Emergency Hospital Conditions of Participation Technical Correction" *Final Rule* [Link]

27 CMS (N.D) Hospital Price Transparency Frequently Asked Questions (FAQs) [Link] *(Last updated May 8, 2025)*

28 Claxton et al. (Feb. 25, 2025) "Challenges With Effective Price Transparency Analyses" *PetersonKFF Health System Tracker* [Link]

29 For an in-depth explainer of Milliman's GRVUs and how they are applied to the price transparency data, respectively, see: Fox et al. (Jan. 2017) "Milliman GlobalRVUs" *White Paper* [Link] and Fox and McManus (Oct. 9, 2024) "Benchmarking Price Transparency Data: Interpreting Results Using CMS's Medicare Fee Schedules vs. Milliman GlobalRVUs" *White Paper* [Link]

30 CAA § 201, modifying PHSA § 2799A-9 and ERISA § 724. [Link] at **p. 2890** and **2891**

Contributors

Won Andersen
Alliant

Kush Banerjee
The Leapfrog Group

Al Codalbu
Embold Health

Darren Fogarty
PBGH

Mike Gaal
Milliman

Peter Long
Long Strategies

Elizabeth Mitchell
PBGH

David Muhlestein
Simple Healthcare

Caroline Pearson
Peterson Center on Healthcare

Annie Joyce
Peterson Center on Healthcare

Melisa Real
Qualcomm

Erica Reijula
Milliman

David Schleifer
Peterson Center on Healthcare

Julie Selesnick
Health Plan Legal Counsel

Raymond Tsai
PBGH

About the Purchaser Business Group on Health (PBGH)

Purchaser Business Group on Health (PBGH) is a nonprofit coalition representing 40 private employers and public entities across the U.S. that collectively spend $350 billion annually purchasing health care services for more than 21 million Americans and their families. PBGH has a 30-year track record of incubating new, innovative operational programs in partnership with large employers and other health care purchasers. Our initiatives are designed to test innovative methods and scale successful approaches that lower health care costs and increase quality across the U.S.

Purchaser Business Group on Health