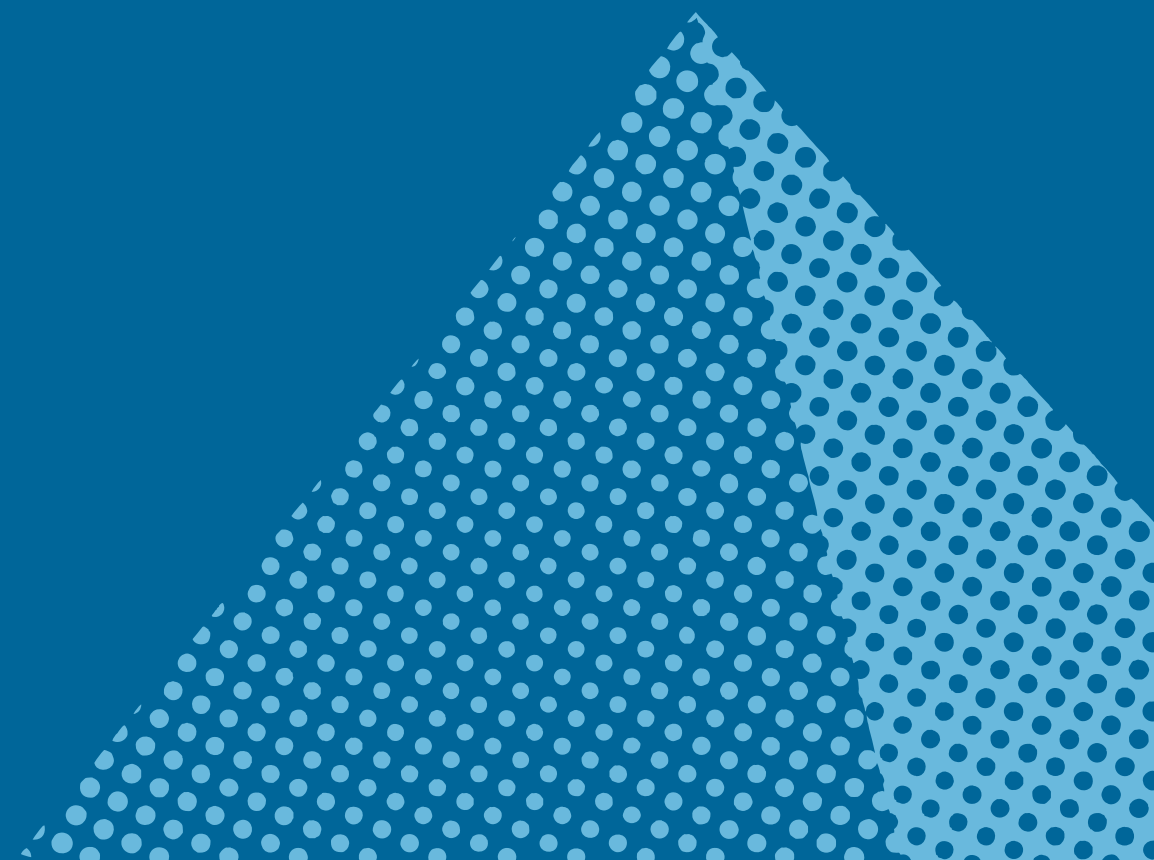
Clarify

Bringing Clarity to Healthcare Price Transparency Data

Methodology for ingestion, enrichment, and conversion into easy-to-query rate intelligence



MARCH 2023

Table of contents

Executive summary

After decades of price negotiations between health insurance companies and healthcare providers being cloaked in secrecy, payers and hospitals are now required to publicly share negotiated rates. However, the public rate data that has been released is complex and challenging to work with in its raw state. Raw rates data released by payers contains trillions of records, is not consistently formatted, and includes payer-specific gaps in reporting, creating a web of convoluted healthcare pricing in the commercial market. Hospital-reported rates data is beset by lower coverage, even less standardization, and is reported for a smaller set of inpatient services. ***However, for those with the infrastructure to parse through the noise, there is plenty of value to be uncovered from price transparency data.***

Since 2015, the Clarify Platform® has been managing massive healthcare data sets using big data efficiencies that are commonplace in the consumer and financial industries. Our investments in processing power provide the perfect infrastructure to handle the size and scale of price transparency data. As a result, we can provide price transparency intelligence that helps healthcare organizations stay ahead of the curve by strengthening contract negotiations and improving provider network affordability.

We deliver rate intelligence via a powerful no-code query engine in cloud software. Healthcare organizations can instantly query 500+ terabytes of enriched rates data and generate reports on market prices in seconds. This puts the power in the hands of the end user, reducing the need for a fully staffed analytics team to wade through the vast sea of price transparency data. Aside from its no-code query engine, what makes our solution unique is our proprietary data enrichment process. By enriching raw rates with claims data on 300M+ annual lives, the product displays prices that have been billed instead of millions of meaningless prices.

In this paper, we describe the robust methodology that we have developed to convert price transparency data from multiple sources into easy-to-query rate intelligence. We detail our data ingestion, cleaning, and proprietary enrichment process that is first-of-its-kind in the industry. Our enrichment process combines claims data with trusted clinical informatics methodology to map standard codes and service lines, estimate Medicare payments for those services, and calculate rates as a percentage of Medicare payments. It assesses claims volume to determine utilization and remove 'zombie rates,' rates that are reported for providers who never have and never would provide certain services. By integrating insights from claims alongside rates data, Clarify adds the critical dimensions of service mix and volume. This also enables the display of the weighted percent of Medicare prices (versus straight average) so users can better inform their understanding of the true economics and value delivered.

With a better understanding of our methods, we hope you can determine if our approach is suitable for your technical and business needs.

> *"With Clarify's rate data, our team can now leverage the combined power of Clarify's provider quality performance insights in tandem with rate insights to fully assess provider performance and help our members get the higher quality outcomes they deserve at an affordable price."*
>
> **VP, Market and Provider Analytics**
> **National Health Plan**

The new era of price transparency

As healthcare costs continue to climb in the US, healthcare decision-makers search for strategies to curb those ever-growing costs. After decades of negotiations between health insurance companies and healthcare providers being cloaked in secrecy, in January 2021, the Healthcare Price Transparency Act went into effect, requiring hospitals to publish their cash pay rates and rates negotiated with health insurance companies. In addition, the No Surprises Act (part of the Consolidated Appropriations Act) introduced a number of transparency and disclosure requirements for hospitals and health plans. Beginning on July 1, 2022, health insurers were required to share their negotiated rates for all in-network covered services and items as well as their out-of-network allowed amounts and billed charges. By January 1, 2023, cost-sharing estimates for 500 common shoppable services became available in consumer-friendly formats, and by 2024 the remainder of covered services must be made available.

Hospital Price Transparency Rule (HPTR)

Hospitals, as of January 1, 2021, are required to publish not only their chargemaster of list prices but also their allowed amounts for each contracted payer network, their cash discount rate, and the minimum and maximum payments for those services. CMS requires these to be published in a machine-readable file (MRF). Hospitals must also publish the figures for the most commonly scheduled procedures to an interactive dashboard that prospective patients could peruse before choosing a provider. CMS has recommended hospitals report on up to 500 shoppable services to aid consumers. Penalties for hospitals failing to comply will incur a minimum Civil Monetary Penalty (CMP) of $300 a day. Despite these stipulated penalties, both compliance and enforcement of the HPTR have been poor. Not all hospitals, including many regional health systems, have complied with the requirements to make rates available. Current estimates of hospital compliance range from 25 to 70 percent, leading to calls by Congress for more federal oversight and enforcement to increase hospital compliance.

Transparency in Coverage (TiC) Rule

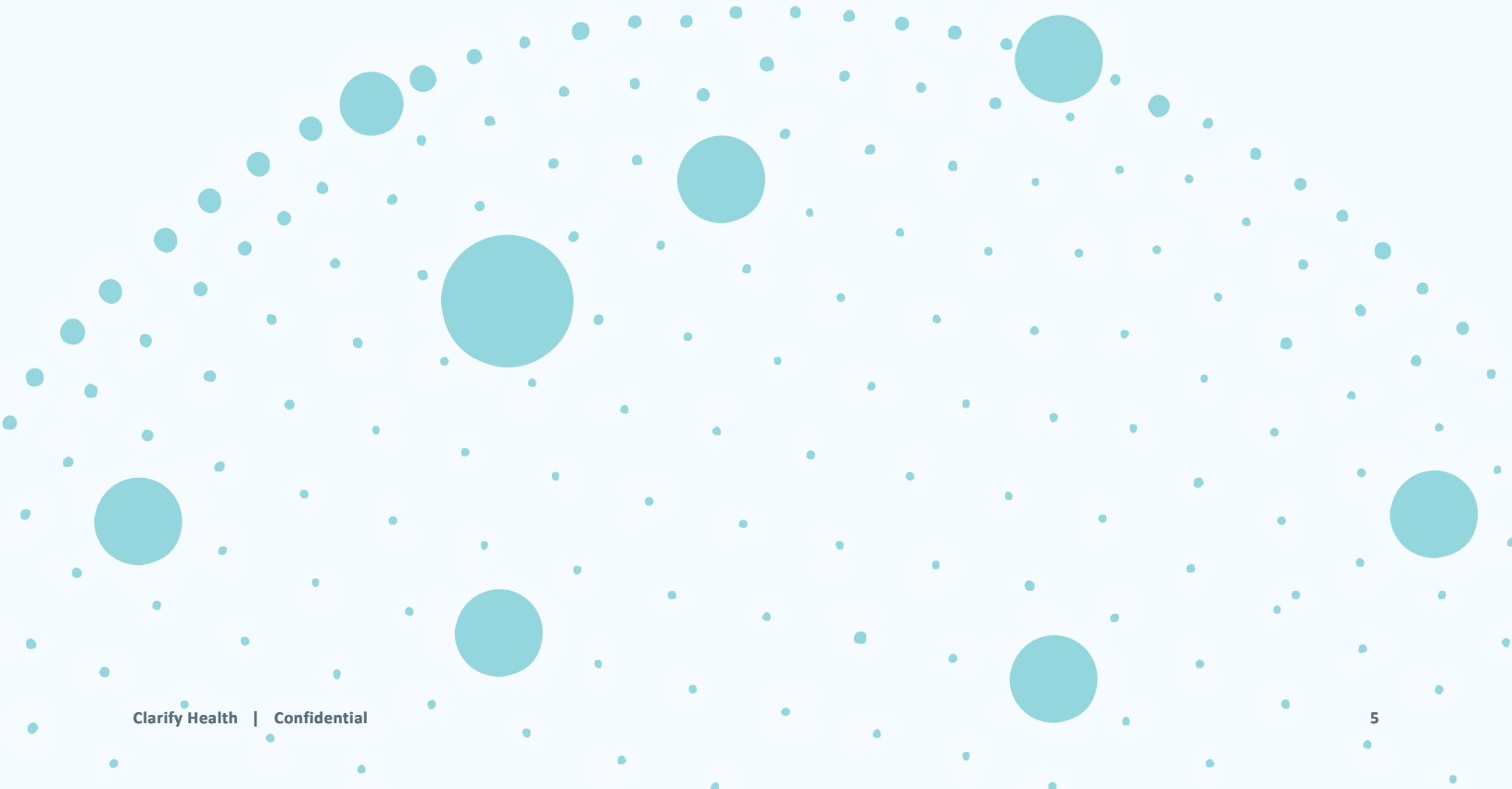
Health plans and self-insured employers, as of July 1, 2022, are required to disclose their contracted in-network rates with all provider sites in their network, not only for hospitals but also organizations such as physician groups and ambulatory surgical centers.

Like the HPTR, the TiC rule requires payers to publish their pricing data as MRFs containing the following sets of costs for items and services:

1. In-Network Rate File: Rates for all covered items and services between the plan or issuer and in-network providers.

2. Allowed Amount File: Allowed amounts for, and billed charges from, out-of-network providers.

CMS penalties for payers failing to comply were set out in existing law (Public Health Services Act and ERISA), which is $100 a day per impacted individual. Relative to HPTR penalties for hospitals, $100 per insured member per day is a far larger amount and, likewise, compliance by payers has been high. While some gaps are evident in Clarify's review of the payer price transparency data, as discussed below, nearly 100% of all health insurance plans in the US are in at least partial compliance with TiC regulations.



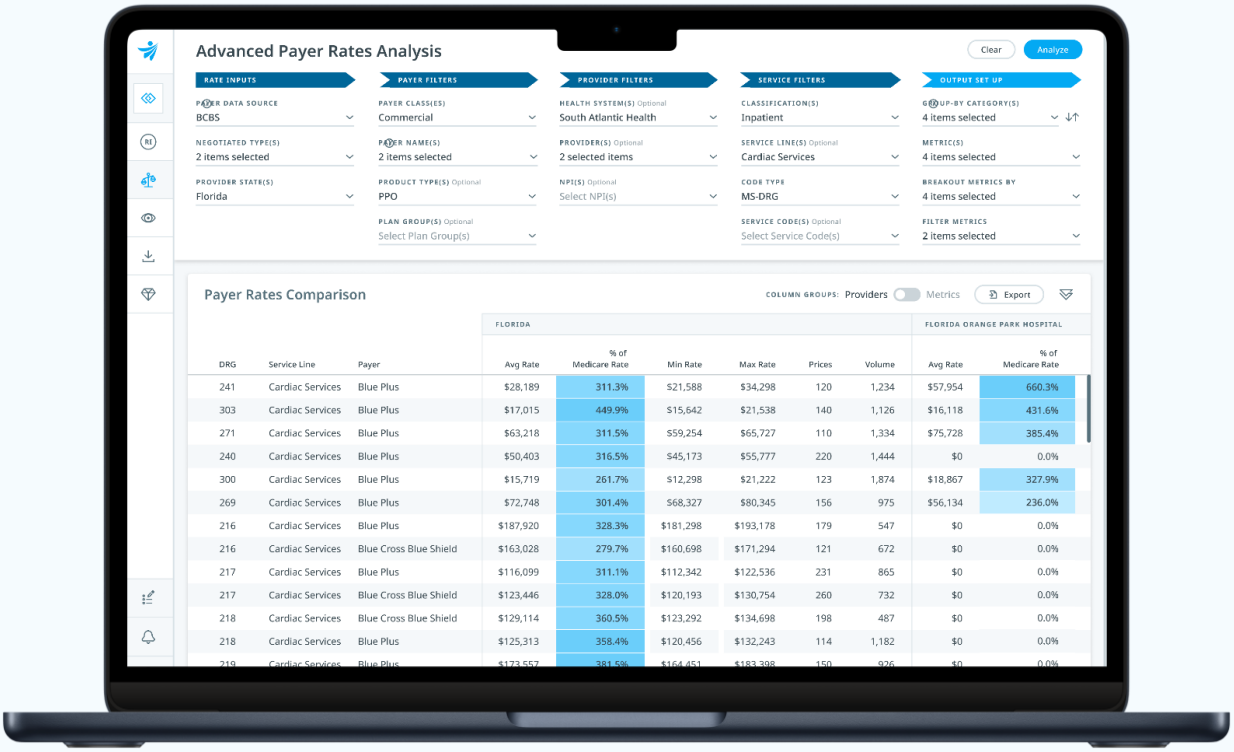
Clarify's price transparency solution

Clarify Rates software

Clarify Rates, a product powered by the Clarify Atlas Platform, is an on-demand software for rich intelligence on healthcare rates negotiated between health insurance companies and providers. It offers a competitive edge in contract negotiations and helps build more affordable provider networks.

With 500+ terabytes of price transparency data that has been enriched with Clarify's claims data, consisting of 300M+ annual lives, users can download hundreds of prices that have actually been billed (instead of viewing millions of meaningless prices). Additionally, by integrating insights from claims alongside rates data, Clarify adds the critical dimensions of service mix and volume. This also enables the display of the weighted percent of Medicare prices (versus straight average) so users can better inform their understanding of the true economics and value delivered.

The software has a no-code query engine that delivers insights in seconds instead of hours. It can instantly run queries on our massive data set, dynamically filter prices by over 15 categories, benchmark rates to Medicare and regional averages, and even create executive-level dashboards. It can compare inpatient, outpatient, and professional rates to the market, shown as a percent of Medicare and drill down at the individual payer, provider, DRG and CPT code levels.



Clarify Atlas Platform

The Clarify Atlas Platform houses the healthcare industry's largest collection of patient journeys, delivering the most precise and actionable insights to payers, providers, and life sciences companies. It maps over four billion patient journeys across over 300 million lives, leverages best-in-class grouper technology, and drives 18B+ AI-powered predictions, answering healthcare's most complex business questions in an instant. Leveraging patient journeys to enrich price transparency data provides a cleaner data set for unparalleled access to rate intelligence that is truly meaningful.

Trusted healthcare data

Atlas brings together 15+ billion government and commercial claim records and Rx data covering 300+ million lives with 400+ factors on social determinants of health and 500+ terabytes of price transparency data in secure, HIPAA-compliant data lakes.

Fast processing

The tech stack behind our analytics platform can load and process billions of claims in hours and enable the visualization of 15 million episodes of care in a secure, cloud-based SaaS platform in seconds.

Transparent insights

We train thousands of models, cut data in countless ways, and ensure optimal case-mix adjusted predictive values to deliver clear and actionable insights.

A single platform to power insights for all

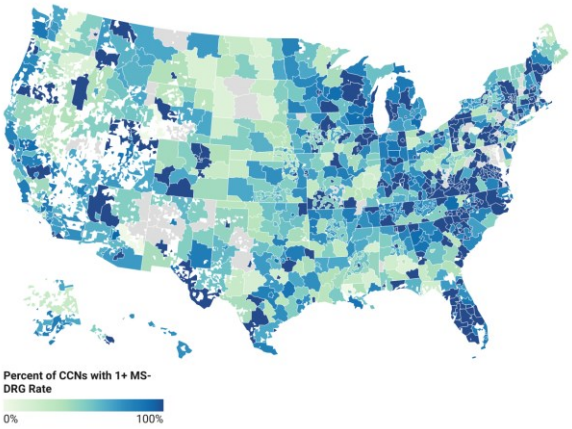
All business applications for payers, providers, and life sciences are powered by a single, modular healthcare analytics platform relying on the same core healthcare data and architecture, data science models, and delivery mechanisms.

A comparison of hospital data to health plan data

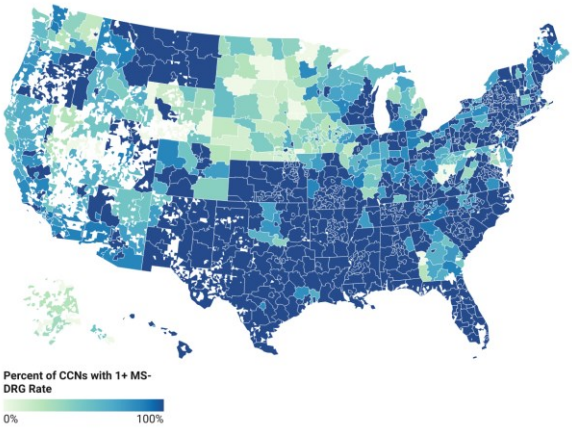
Clarify processed, analyzed, and compared data from approximately 5,600 hospitals and all large national payers since the inception of each data set to assess quality differences between the data sets. We observed that health plans are more compliant with price transparency regulations than hospitals, resulting in higher data coverage across payer-sourced data.

Figure 1. Payer-sourced rate data have higher coverage & compliance.

Hospitals (identified by CMS CCN IDs) reporting at least one MS-DRG hospital rate:



CCNs with at least one MS-DRG payer rate reported by Aetna/CVS, BCBS, Centene, Cigna, United Healthcare:



Relative to payer reporting, hospital price transparency reporting has been poor due to a weaker regulatory framework (e.g., no initial requirements for standardization), resulting in worse compliance by hospitals and, therefore, lower coverage across hospital-sourced rates compared to payer-sourced rates. However, we recognize that compliance with price transparency efforts has been mixed for both hospitals and payers. So, in addition to assessing coverage, we have quantitatively and qualitatively assessed the quality of payer and hospital-reported rates data.

We observed:

- Hospital-sourced rates data is limited to hospital services (e.g., it is missing professional and ambulatory services), whereas over 90% of unique negotiated rates reported by health plans reflect rates negotiated with professionals.
- Notably, CMS also did not require a standard schema for hospital MRF formatting, whereas health plans must follow a defined schema. This has led to a wider variation in hospital reporting format and quality.

Figure 2. Clarify's analysis of provider-source and payer-sourced rates data quality

		Hospital Data	Payer Data
Data Insights	Institutional data	✓	✓
	Professional data		✓
	Employer and TPA plan data		✓
	Medicare Advantage and Managed Medicaid plan data	✓	
	Larger reported code & service basket		✓
	Signal on place of service	✓	✓
	Negotiation arrangement information		✓
Data Fidelity	Standard schema defined by CMS		✓
	Higher refresh cadence		✓
	Higher compliance rate		✓
	Higher penalties for non-compliance		✓

While payer rate files are many times larger and more complex to manage, our findings show they are a far better data source. Because payer-sourced rates data are higher coverage and higher quality, the Clarify Rates product leverages payer-sourced negotiated rates as its primary source of data.

Data preparation

We have fully extracted, cleaned, and enriched these data for over 65 national and regional payers as of February 2023, including United Healthcare, Cigna, Aetna CVS, Humana, and multiple Blue Cross Blue Shield (BCBS) payers. Clarify has also ingested MRFs from 5,570 hospital facilities.

Acquiring, scraping, and processing these MRFs can be challenging, given their sheer magnitude and complexities created by some payers that make their full set of files more difficult to download or process. A high-level overview across twelve payers (including all BCBS payers) is provided below, covering both file counts and coverage in terms of unique service codes and national provider identifiers (NPIs) as of February 2023.

500 TB

Since the release of payer MRFs in July 2022, Clarify has ingested over 500 terabytes of raw, compressed data.

Figure 3. Payer MRF Quality Stats

Issuer	# Files Downloaded	Size of Files Downloaded	Total Unique MS-DRGs Reported	Total Unique HCPCS Reported	Total Unique CPTs Reported	Total Unique NPIs Reported
Aetna	59,924	515 TB	767	6,087	10,917	1,159,059
BCBS	20,190	55 TB	707	18,106	18,119	110,975
Centene	45	599 MB	638	2,257	9,042	830,540
Cigna	84,509	683 GB	767	6,847	15,050	1,253,402
Humana	9,130,896	414 TB	794	6,924	10,900	31,396
Kaiser	413	758 GB	801	6,960	10,984	781,389
Medica	235	270.4 GB	693	6,615	17,637	49,056
Molina	6,547	168 GB	-	-	18,102	360,235
Mountain Health Coop	492	300 MB	767	6,188	10,689	1,220
Oscar	117	24.8 GB	-	14,485	10,142	531,980
UHC	54,778	31 TB	693	6,254	10,934	1,283,980
UPMC	4,549	15 GB	767	3,221	9,430	73,596

Addressing gaps and missingness

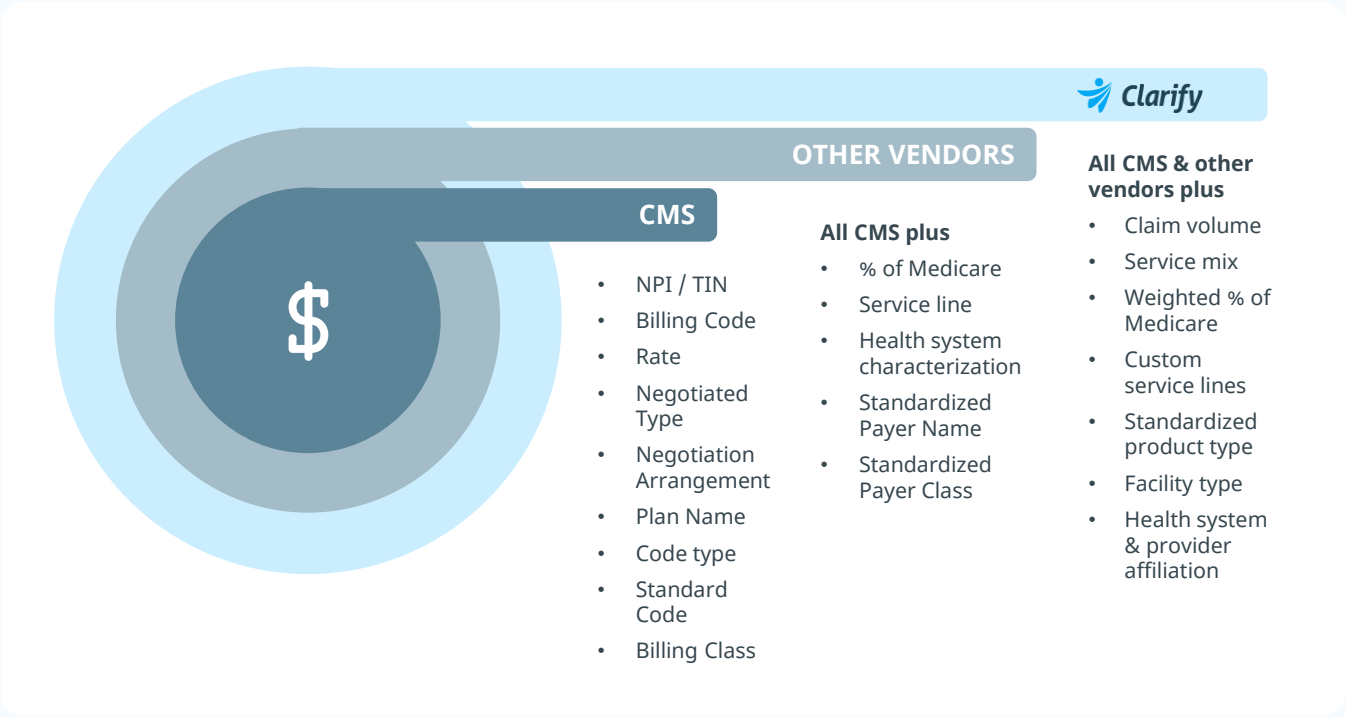
We comprehensively review each payer's MRF during our cleaning and enrichment processes. Through this process, we identify gaps in reporting by some payers. We communicate with payers about these gaps and expect unintended non-compliance with the price transparency regulations to improve over time. Examples of identified gaps or complications in processing from the above set of payers include:

- BCBS rates (compiled from over 40 individual BCBS payers) contain substantial duplication due to BCBS payers' reciprocity agreements, such as the BlueCard program, and the payers' interpreted compliance with regulations requiring reporting of all reciprocal rates by the specific BCBS payer.
- Centene's overall MRF size is observed to be smaller than other national payers, resulting in various gaps in specific rates and service codes.
- Cigna institutional rates are typically reported as individual rather than organizational NPIs (over 90%). In its 2022 releases, Cigna institutional rates only included around 1,400 organizational NPIs. While over 90% of NPIs continue to belong to individual NPIs in February 2023, overall NPI count increased by 61% compared to NPI counts from November 2022, and there are now 5,900 organizational NPIs reported for Cigna institutional rates.
- Humana's MRFs are massive, and their posted files total over 400 terabytes. Humana was the only national insurer to initially post their data in a text format (.CSV). While JSON formats have been posted recently, Humana's MRFs remain the largest, which results in slower processing speed.
- Molina has only posted institutional rates related to revenue center (RC) and CPT code types. We would expect institutional inpatient rates to include negotiated rates for MS-DRGs.
- UHC's rates data underreports institutional outpatient rates, with between 25-50% of unique HCPCS for outpatient institutional rates missing in some markets. Clarify also observed that UHC initially reported uniform negotiated rates across all DRGs for some hospitals, which is likely a reporting error.

Enrichment methodologies

Clarify's all-payer claims data is used to highlight the rates that actually matter for payers and providers while filtering rates that will never be utilized. Clarify incorporates provider and health system characteristics and identifiers, identifying plan contracting groups, plan types, and employer affiliations, adding service line categories, estimating rates as a percentage of Medicare payments, and combining with national commercial claims samples to assess utilization. Clarify then uses a distinct approach leveraging our proprietary data sources to meaningfully assess market prices of specific services, which often number in the thousands for a single service code within a given market. Importantly, we limit rate analytics to providers with documented volume for that code. This eliminates 'zombie rates,' which are rates that are reported for providers who never have and would never, provide certain services. We can also calculate volume-weighted averages and case mix adjustments across baskets of services of most interest to specific users.

Figure 4. Clarify has the most enriched rates fields in the industry.



Rates enrichment

A variety of steps are taken to enrich negotiated rates, including mapping standard codes to service lines, estimating Medicare payments for those services, calculating rates as a percentage of the Medicare payment, and assessing claims volume to assess utilization and remove 'zombie rates.'

Service code types, service classification, and service line mapping

Rates are provided across a range of standard codes. These standard code types are identified by the payer and validated by Clarify using standardized logic. Standardization includes re-defining or excluding codes that do not align with coding conventions, such as codes reported to be MS-DRGs that are more than three digits long. In addition to the most common codes, including MS-DRGs, HCPCS, and CPT codes, a wide range of rates associated with other code types are available in the Clarify Rates product, as needed for specific use cases. These include: [APC, AP-DRG, APR-DRG, CDT, CPT, CSTM-ALL, EAPG, HCPCS, HIPPS, ICD, LOCAL, MS-DRG, NDC, PROC, RC, and UNKNOWN code types.

For rates data corresponding to an MS-DRG, CPT, HCPCS, APC, EAPG, CDT, or revenue center code, services are mapped by the standard code to one of four service classifications. These include dental, inpatient, outpatient, and unknown service classifications.

Similarly, for rates corresponding to an MS-DRG, CPT, or HCPCS, services are mapped by the standard code to one of 41 service lines. A complete list of service line mappings is available in the Appendix.

Medicare rate logic

Medicare rates are calculated for each standard service code in Clarify Rates, primarily to calculate a percentage of Medicare benchmark for each rate. Separate methodologies are applied for inpatient, outpatient, and professional services, respectively, consistent with differences in payment by Medicare.

Inpatient and outpatient services

Medicare payment rates for inpatient services are calculated from the Inpatient Prospective Payment System (IPPS), while rates for outpatient services are based on the outpatient prospective payment system (OPPS).

Clarify first calculates Medicare payment amounts for inpatient and outpatient codes specific to the hospital's CMS Certification Number (CCN) of the hospital. The CCN-specific inpatient rates adjust for CCN-specific wage indices and cost of living adjustments. In contrast, the DRG weights, labor share, labor rate, and non-labor rate are applied uniformly to all CCNs in a given year consistent with Medicare payments. Similarly, calculated outpatient rates adjust for CCN-specific wage indices and the relative weights and conversion factors applied to all CCNs in a given year.

CCN provider identifiers are not supplied in publicly released rates files. To assign CCNs, Clarify maps the provided NPIs to CCNs using reference data sources. In some cases, we cannot match NPIs to CCNs necessary to calculate CCN-specific payment amounts. Because the IPPS and OPPS rates are tied to CCNs, we cannot always assign the CCN-specific calculated rates from our reference data to the enriched payer-sourced data. When NPIs cannot be assigned to a CCN, payment rates are set at the zip-code level (either a five or three-digit zip code level). The zip code associated with a rate is sourced from Clarify reference data based on the rate's specific NPI. These estimates aggregate all CCN-specific payment amounts in a five-digit zip code and three-digit zip code and calculate the median rate for all codes in that zip code in a year. A final Medicare rate is assigned for all relevant service codes (MS-DRG, HCPCS, or CPT) for all providers using the following hierarchy:

- CCN-specific rate for a code from most recent year available
- Five-digit zip code's median rate from most recent year
- Three-digit zip code's median rate from most recent year
- For outpatient codes only, unadjusted outpatient rates from most recent year

Examples of this hierarchy:

- An NPI-DRG combination has a CCN identified in Clarify reference data. A CCN-specific rate for the DRG is available from 2021, and a five-digit zip code median rate for the DRG is available from 2022. The CCN-specific rate from 2021 will be selected as the Medicare rate.

- An NPI-DRG combination does not have a CCN populated. A five-digit zip code median rate for the DRG is available from 2021, and a three-digit zip code median rate for the DRG is available from 2022. The five-digit zip code median rate from 2021 will be selected as the Medicare rate.

This approach is not without potential limitations, with certain Medicare payment policies not applied to allow for equitable comparison of payment rates between local providers providing the same healthcare services. Specific to inpatient payment amounts, the IPPS rate calculation used by Clarify does not account for disproportionate share hospital (DSH) adjustments or indirect medical education (IME) adjustments. Hospitals that treat a high percentage of low-income patients may qualify for DSH adjustments, and approved teaching hospitals may qualify for IME adjustments. Regarding outpatient payment amounts, the OPPS rate is always calculated using each year's full conversion factor. Under the OPPS, CMS subjects hospitals that do not submit data on standardized quality measures to a reduced conversion factor; we do not apply this reduced factor to any provider to improve cross-provider rate comparisons. In addition, the OPPS applies payment adjustments for certain qualified hospitals (i.e., rural sole community hospitals, hold-harmless cancer centers, and pediatric hospitals) intended to promote specific CMS policy goals: Clarify's calculations also exclude these adjustments.

Professional services

Medicare rate estimates for professional services codes are based on CMS, local area-specific (i.e., carrier-locality specific) Physician Fee Schedule (PFS) pricing inputs available from CMS. The carrier-specific PFS rates are an additional source of reference data maintained by Clarify for use across the Clarify Platform. No further adjustments are made to the data, with PFS rates available from 2013 to 2023 and the most recently available year applied.

Specifically, the most recent year's non-facility PFS amount is used as a code's Medicare rate based on matching the code, carrier, and locality to Clarify's reference data. The carrier and locality associated with a rate is determined based on the rate's zip code. The PFS non-facility fee schedule amounts can vary for a code by HCPCS modifier for a year, carrier, and locality combination. Only amounts without HCPCS modifiers are selected as the Medicare rate, regardless of the modifier(s) present on a payer-sourced rate. Note that modifiers are not required to be included under TiC regulations.

Percent of Medicare calculation

Following the calculation of a Medicare payment amount for a DRG, CPT, or HCPCS code's rate, the negotiated rate amount is divided by the Medicare rate to determine the percent of Medicare. This calculation is not performed if the record is a per diem or percentage rate as indicated by its negotiated type.

Percent of Medicare rates have been used in health services research in several applications to normalize for differences in price levels and service intensity across fields.

Claims volume

Despite the trillions of rows of rates data released by payers, much of this data is not associated with legitimate service volume, and so it is not meaningful for either payers or providers. Comparing rates for specific service codes by NPI with Clarify's all-payer claims data, we find substantial over-reporting of rates for provider-service code pairs with no documented utilization in claims. In some cases, these findings are obvious: dialysis clinics and clinical laboratories are not actual providers of elective surgeries despite payer-reported rates for such services. Others are less clear and show the benefit of a claims-based approach to review, such as rates negotiated with hospitals for specific services they do not actively provide.

Clarify considers multi-payer service volume, including both traditional Medicare service volume and commercial volume. Traditional Medicare service volume is estimated from 100% of Medicare claims available to Clarify, a CMS Qualified Entity; commercial claims volume is sourced from multiple proprietary commercial claims samples licensed by Clarify and applied across the Atlas platform.

Payer plan enrichment

Issuer

Every health insurance company identified by Clarify for rates sourcing is assigned an Issuer ID by its primary name. “Issuer ID” is a field in the user interface that can be used to stratify the data.

Derived payer name

Clarify extracts payer name information from the rates data to assign a unique payer subsidiary for all observations. These are separated into two types of payers: employer plans and all other named payers operating under the same issuer. Additional categorization is then completed by payer, with a uniform approach implemented except where specific deviations are warranted. For example, as noted above Clarify has highlighted substantial overlap in negotiated rates within states across BCBS payers. For this reason, we have chosen to compile BCBS rates into a single database. However, we undertake steps using the payer name field to distinguish different BCBS payers’ (e.g., Anthem, Regence, BCBS Texas) rates as needed.

Product type

Product type is derived from two data sources:

- CMS HIOS product level data available through proprietary Clarify reference data sources
- Natural-language text analysis of raw plan name values provided by hospitals and payers in their rates files

The possible output values include:

- HMO: Health Maintenance Organization plans
- PPO: Preferred Provider Organization plans
- EPO: Exclusive Provider Organization plans
- POS: Point-of-Service plans
- Indemnity: Indemnity-type plans
- Dental: Dental coverage only plans
- Other/Describe: Other plan types (including ACOs and other non-traditional plan types)
- Multiple products: Plans containing references to multiple products

An additional description of product type reference data sources is helpful to understand the latency and quality of the product type field. HIOS data can only be used to determine product type in payer-sourced rates data, as hospital-sourced rates data does not provide any HIOS information. Further, it is only possible for plans that are reported using a 10-digit (or longer) HIOS ID. Some issuers only report a 5-digit HIOS ID, and many plans (primarily employer plans) are reported using an EIN. The number of plans whose product type can be identified using HIOS data will vary from issuer to issuer.

In addition to using HIOS data, Clarify has also developed logic that searches the raw plan name value for specific keywords related to the product types, e.g., HMO, PPO, EPO, POS, and Indemnity. If, for example, a plan name matches the keyword search for HMO, the plan is assigned an HMO product type. In some cases, a plan name will contain multiple product type keywords (e.g.,"Cigna HMO/PPO"). We do not give preference to one product type over another in these instances and instead designate the plan as "Multiple products." These plans are likely a rollup of various plans with varying product types that share rates.

In payer-sourced rates data where both methods are possible for identifying product types, preference is given to the HIOS-based product type. That is, if a product type can be derived from both the HIOS data and plan for a plan, the HIOS-based product type is selected.

Limitations exist for the product type categorization as currently implemented. The fill rate for product type will vary from issuer to issuer based on a considerable variance in the raw data. Some issuers seldom report 10-digit HIOS IDs, and some use plan names that cannot convey a product type. We have identified dental plans that indicate an additional product type (e.g., a dental PPO plan). We feel that in these situations, it would be more consistent to classify the plan as a PPO product type. Note that these plans do not qualify for the "Multiple products" designation as the dental identification comes from the HIOS method, while the PPO identification comes from the plan name method.

Payer class

As a last step in our enrichment of payer data, each plan and rate are mapped to a payer class. While commercial plans are a primary interest for most users, a diverse set of payer classes are available in the Clarify Rates data, including Commercial, Medicare Advantage, Exchange, Medicaid payer class types. However, Medicare Advantage and Medicaid plans are not required to be included in payer data per the TiC regulations and most plan rates are classified as either commercial or Exchange plans.

Plan grouping methods

A main technical issue with payer rates data is the sheer size of payer-released files. Close to a petabyte of raw data is available for download, and this is prior to additional enrichment steps taken by Clarify. A main source of data size is the large number of raw plan names corresponding to each unique service-specific rate-NPI combination for a given payer. Many of these plan name differences do not correspond to different rates; they are "duplicates" in the sense that they do not provide any additional value or insight on rate info outside of the specific plan name. Clarify has observed examples where the same rate for the same provider map to over 40,000 unique plans offered by the same payer.

These data may be of use to some users but, as a consequence, fully joined data is extremely large and computationally inefficient, increasing run times and unnecessary for many users. To overcome this feature of the released data and increase computational efficiency, Clarify has prepared a schema to group plan names with the same rate into buckets to reduce data size.

This dramatically improves processing time, particularly when considered within the Clarify Rates product. For example, if we have 40k rows of rate info for United at a single facility for a single MS-DRG, Clarify categorizes the plans into groups based on rate variance. We leverage product type categorizations to divide up groups by HMO, PPO, and other plan types. After reviewing rates and assigning all plans to plan groups, users can leverage two flexible approaches:

- Rates grouped at Plan Group ID level: Allowing queries at the plan group level, which makes insights easier to understand and faster to load.
- Rates grouped at the Raw Plan Name level: Separate tables are maintained for each payer mapping groups to plan name, as it still needs to be discoverable for some users.

Enriched rates data grouping plans results in no loss of information except for specific plan names, while allowing Clarify Rates to combine multiple payers for consolidated, cross-payer review and analysis. It allows queries in Clarify Rates to run more efficiently and provides for flat file delivery with less noise. As necessary, further drill-downs into the exact plan names may then be completed.

Provider enrichment

In addition to rate-specific and payer plan-specific enrichment efforts, enrichment of provider data found in released rates data is one additional way in which Clarify's platform standardizes and improves on publicly available data.

Provider location information

Working from released provider identifiers (NPIs) and rich, proprietary reference data, each rate is matched to a street address, city, state, zip code, locality, carrier, and census division.

This enrichment helps group data into meaningful local market comparisons across a wide variety of use cases.

Facility roll-up, provider group, affiliation logic

Similarly, working from released provider identifiers in the raw rates data, Clarify applies a rich mapping of provider identifiers to the organizations they operate in. This includes mapping NPIs to provider names and health system characteristics.

In addition to provider group characteristics, we apply specialty information and other provider-specific characteristics to each observed rate.

TIN - organization NPI crosswalk (payer rates derived)

Sourcing both from proprietary reference data sources and the amassed rates data itself, Clarify can strengthen the quality of our provider data by creating a TIN to organization NPI crosswalk file to fill in gaps in specific price transparency files. This allows us to allow search by both NPI and TIN across payers.

Health system

Using a proprietary Clarify reference data source, we check all providers for affiliation with a health system. Providers, including hospitals, physician groups, and individual clinicians, are assigned to either a specific health system or are given the value 'No Parent Affiliation.'

Validation and quality review

Clarify assesses data quality and regulatory compliance of payer-reported rates data. We evaluate payers on their compliance, assessing multiple metrics related to their released rates:

- A check to determine if the Payer in question posted any MRFs per the TiC regulations
- The number of Unique Health Systems included in the MRFs
    - Based on NPI
    - The number of unique NPIs included in the MRFs
- The number of unique MS-DRG, APR-DRG, & HCPCS codes included in the data
- With consideration given to the % of possible DRGs and a % of possible HCPCS codes
    - Do the files have all the DRGs that are listed?
    - Do the files have all the CPT/HCPCS Codes listed?
- A check to ensure the files include both inpatient, outpatient, & professional rates
- Number of unique payer names
- Recency of updates (must have at least one a year)

Statistical analysis and trends

Clarify's analytics teams are actively working with the price transparency data to support customer-specific queries and help users better understand the data, given its recent release. Our team is made up of an interdisciplinary team of over 20 analysts with expertise in actuarial methods, data processing, data science, health economics, healthcare finance, health services research, and statistics. We can do trends analysis on our enriched rates data using service line mapping, volume data from commercial and Medicare claims, geographical attributes, and other specific features of a market, across payer plans, or for a set of providers depending on the specific use case. In general, and particularly for any national or regional analysis, there is often too much data to be able to assess meaningful trends in the market unless we use our enrichment fields to inform the relevance of a specific rate. For example, lab providers reporting rates for surgeries but without any recorded experience with such procedures are completely meaningless and should be excluded from analyses. The process of identifying trends:

Combining data to the service code, provider, payer level (including the rate amount, reporting entities, etc.)

Removing the rates for codes with low materiality (typically identifying a standard market basket - using volume and rate combination)

Rolling up data at different levels of aggregation (service line, code type, NPI, etc.) for a payer using service volumes and/or market case mix of codes within service lines.

Comparing rates and rates as a percent of Medicare between payers and providers. A published example of these types of comparisons was detailed in our December 2022 research brief.

Appendix

Data dictionary

The current version of our data dictionary is available upon request.

Sample SQL queries

Some example SQL queries for use in starting out with our rates flat files are available upon request.

Comparison of hospital and payer price transparency regulations

	Hospital Price Transparency Final Rule	Transparency in Coverage Final Rule
Effective Date	**January 1, 2021**	**July 1, 2022**
Data Requirements	Machine-readable files (MRFs) should include all individual items and services (including service packages) provided by hospital: The gross charge (the charge for an individual item or service that is reflected on a hospital's chargemaster, absent any discounts). The discounted cash price (the charge that applies to an individual who pays cash, or cash equivalent, for a hospital item or service). The payer-specific negotiated charge (the charge that a hospital has negotiated with a third-party payer for an item or service). The de-identified minimum negotiated charge (the lowest charge that a hospital has negotiated with all third-party payers for an item or service). The de-identified maximum negotiated charge (the highest charge that a hospital has negotiated with all third-party payers for an item or service).	MRFs containing the following sets of costs for items and services: **In-Network Rate File: In-network machine-readable files include negotiated rates for all covered items and services at the plan level, for all medical codes,** for contracted rates between the plan or issuer and in-network providers. **Out-of-network Rate File:** Out-of-network machine-readable files include allowed amounts for covered items, services based on billed charges and allowed amounts including historical amounts. Files are not required if there are fewer than 20 claims for a service for a provider.

Refresh Schedule	**Yearly**	**Monthly**
Format	**No schema requirements** from CMS.	Must follow a **defined schema** to allow for data collection, consolidation, and parsing.
Penalties and Enforcement	**$109,500 - $2,007,500** for a full calendar year of noncompliance per hospital. Only two fines issued to date, some warning letters also issued.	**$100 per day** per each affected member. No active enforcement to date.
Compliance	Estimated between 25 – 70 percent in full compliance.	All national and regional payers posting data, with some payers reporting incompletely.
Effective Date	**January 1, 2021**	**July 1, 2022**

Clarify's rate service line categories

Administered Drugs

Ambulance

Ancillary Services

Anesthesia

Behavioral Health

Cardiac Services

Durable Medical Equipment

Emergency Department

Endocrinology

Enteral and Parenteral Therapy

Evaluation Consultation

Gastroenterology

Head Neck

Hematology/Oncology

Home Health Services

Infectious Disease

Lab

Medical and Surgical Supplies

Nephrology

Neuroscience

Obstetrics

Ophthalmology

Orthopedics

Orthotics and Prosthetics

Other

Other Diagnostic Study

Other Therapeutic Procedures

Pathology

Performance Measurement

Physical Therapy

Pulmonology

Rehab Aftercare

Skin and Wound Care

Temporary Codes

Trauma

Unknown

Urology

Vaccination

Vision Hearing and Speech Supplies

Women's Health

Glossary

Atlas Platform: Clarify's analytics platform is designed to provide healthcare price transparency and enriched claims data, insights, and tools to help stakeholders, including both payers and providers make data-driven decisions and better understand healthcare spending and negotiated rates.

Bundled Payment: A single, comprehensive payment made to healthcare providers for a group of related services, such as those associated with a specific medical condition or procedure, to encourage coordinated care and cost containment. Available in price transparency data for select healthcare services.

Fee-for-Service Payment: A traditional healthcare payment model in which providers are reimbursed for each individual service or procedure they perform, rather than receiving a fixed or bundled payment for a patient's overall care. The best, readily available source of pricing information in price transparency data comes from fee-for-service negotiated rates.

Hospital Price Transparency Rule: A regulation implemented by the Centers for Medicare & Medicaid Services (CMS) requiring hospitals to publish their standard charges for items and services, including negotiated rates with payers, in a machine-readable format.

In-network Provider: A healthcare provider who has a contractual agreement with a health insurance plan to provide services at a negotiated rate.

Institutional Services: Healthcare services provided by facilities such as hospitals, nursing homes, or ambulatory surgical centers, encompassing a wide range of treatments, diagnostics, and therapeutic interventions.

Negotiated Rate: The price agreed upon between a healthcare provider and an insurer for specific services, usually lower than the standard charge.

NPI (National Provider Identifier): A unique 10-digit identification number issued by the CMS to healthcare providers for use in administrative and financial transactions, such as billing and claims submission, which primarily identifies payers in price transparency data.

Percent of Charges Payment: A payment methodology in which reimbursement for healthcare services is based on a percent of the provider's standard billed charges. This approach can result in varying payments depending on the provider's charge structure and negotiated rates with insurers.

Per Diem Payment: A fixed daily rate paid to healthcare providers, such as hospitals or long-term care facilities, for each day a patient receives care, regardless of the specific services provided.

Professional Services: Healthcare services provided by licensed practitioners, such as physicians, nurse practitioners, or physician assistants, in the diagnosis, treatment, and management of patients.

TIN (Taxpayer Identification Number): A unique identification number assigned by the Internal Revenue Service for tax reporting purposes, which can be used to identify healthcare organizations and providers in financial transactions and are reported instead of NPIs in the price transparency data by some payers.

Clarify

About Clarify Health

Clarify Health is an enterprise analytics and value-based payments platform company that empowers payers, providers, and life sciences companies to deliver better care, therapies, and outcomes with actionable patient journey insights. Clarify's cloud-based business applications are built on the Clarify Atlas Platform, which maps 300M+ patient journeys to deliver 18B+ AI-powered predictions and surface insights with speed and precision. Clarify's products illuminate actionable opportunities to drive growth, optimize networks, improve care delivery, manage population health, maximize value-based care performance, and bring therapies to market. With Clarify, healthcare organizations can leapfrog from point-solution and manual analytics to self-service, rapid generation of enterprise insights that light the path to better care and outcomes.