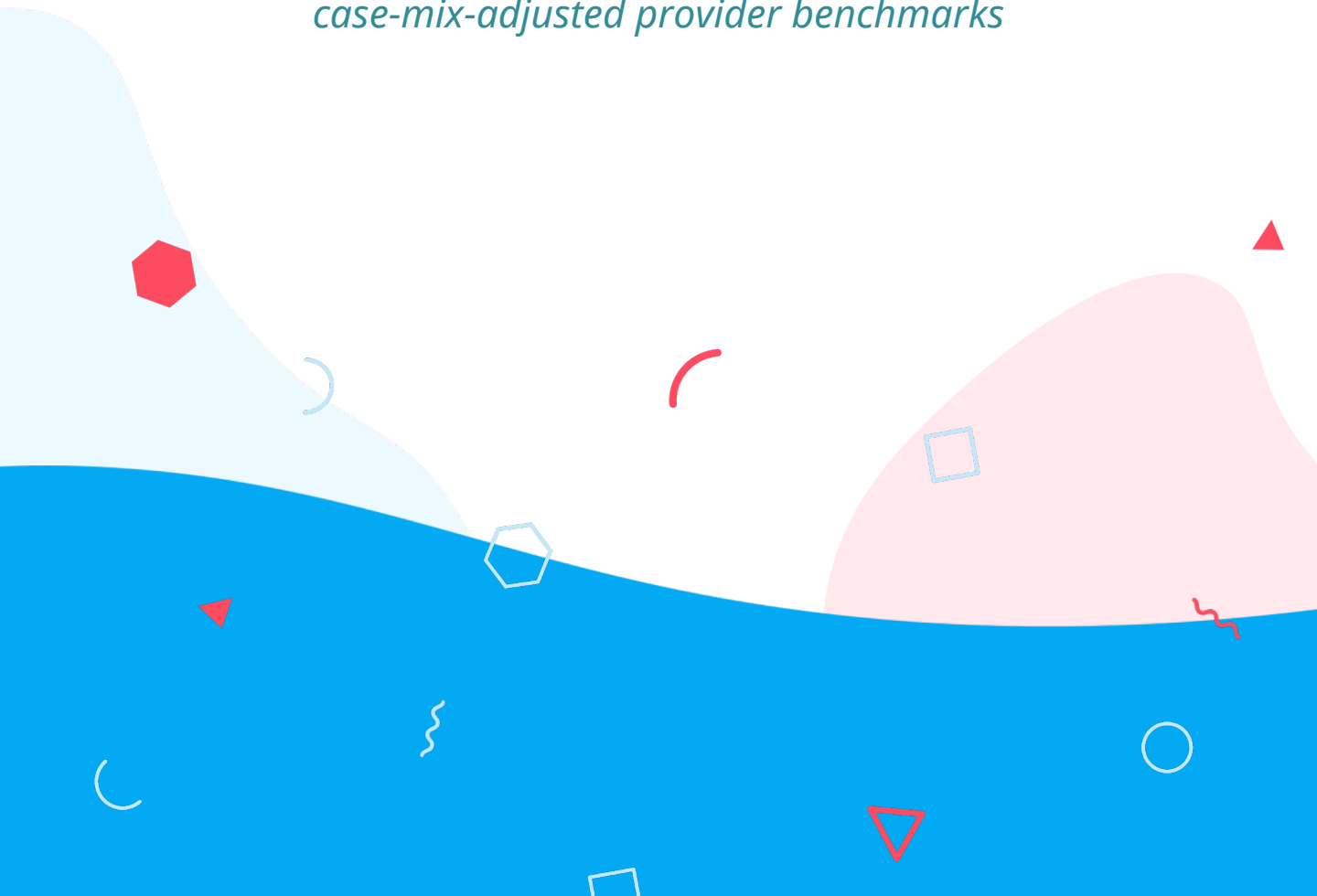


# The Future of Benchmarking

*Transparent and fair  
case-mix-adjusted provider benchmarks*



# Table of Contents

Background	3
The role of big data in fair benchmarking	3
Dated incumbent case-mix adjustment methodologies	4
Next-generation dynamic provider benchmarking	5
External validation of Clarify's methodology	6
HEAD-TO-HEAD STUDY VS. INDUSTRY INCUMBENT'S METHODOLOGY	6
RESULTS	6
SAMPLE BLINDED COMPARISONS BETWEEN CLARIFY VS. INCUMBENT VENDOR	7
EXAMPLES OF CLARIFY'S UTILIZATION BENCHMARKING:	8
EXAMPLES OF CLARIFY'S OUTCOME MODELING & BENCHMARKING	8
Conclusion	9
About Clarify	9

# Background

Accurately measuring and reporting on provider performance is challenging. The analytics required to produce valid and reliable metrics is complex and, as an industry, healthcare has often missed the mark. Many healthcare organizations have focused their provider performance analytics on cost, utilization, quality, and operations-based metrics, and have delivered static reports to providers in an attempt to highlight opportunities for improvement. Though these analytics have solid underpinnings, they often do not accurately and transparently account for the differences in complexity of cases under the physician's care, rendering the reports unactionable.

An essential challenge in provider performance analytics is creating the right comparisons in order to fairly and accurately evaluate a provider. There are so many factors that may influence a physician's care decisions. Any solution that aspires to fairly benchmark provider performance will need to first address the physician's inherent (and valid) argument, "my patients are different". Provider performance analytics can only spark changes in behavior if clinicians trust that the case-mix adjustment methodologies underpinning any report account for their patients' demographic, clinical, social, and behavioral determinants of health.

## The role of big data in fair benchmarking

Access to a comprehensive data set is critical to fair and accurate benchmarking. However, large-scale machine learning requires a high volume of clean, rich data. Unfortunately, the healthcare industry has been plagued by messy, fractured data sources that make it challenging to leverage the full power of data science, which has transformed other industries. Additionally, most healthcare organizations and vendors do not have access to large, independent clinical data sources. Without an independent data set for developing the benchmarks, performance variation analyses only identify differences within a population, rather than comparing the population to an external view of performance.

Clarify's proprietary approach to cleaning and enriching vast, disparate sources of healthcare data delivers unmatched benchmarking to health plans, health systems, and life sciences companies. The Clarify Platform brings together one of the largest, longitudinal, patient-level datasets in the industry, aggregating traditionally siloed claims, electronic health records (EHRs), prescription, and social and behavioral data from over 300 million lives. Clarify is one of only a handful of for-profit companies with CMS Qualified Entity (QE) status, which feeds the Platform with 100 percent of Medicare fee-for-service (FFS) data sets. These are richer than the more widely used Medicare Limited Data Set (LDS) provided to other vendors. The Platform ingests refreshed data sets and new data sources continuously via an automated data cleaning process that identifies and corrects outliers, distinguishes between unmarked screening and treatment diagnoses, attributes physicians to patients, and sorts claims into appropriate specialty- or disease-related categories.

By linking CMS claims data with commercial claims, EHR, prescription, and socioeconomic data, our models are trained on large cohorts and a more complete picture of each patient's longitudinal healthcare journey. The enrichment of traditional healthcare data with data that are not typically accessible or usable optimizes our Platform's case-mix adjustments and allows our predictive models to uncover patterns that would otherwise be hidden from models built with claims data alone. This allows the Platform to generate patient-level predictive models that more precisely benchmark prior performance and predict future outcomes.

# Dated incumbent case-mix adjustment methodologies

Analytics companies, such as 3M, Milliman, MCG, and Optum, have attempted to make case-mix / severity adjustments in their assessments of provider or clinician performance. Based on our analysis and collaborations with industry partners, we've uncovered several drawbacks and inconsistencies in the methodologies most commonly offered on the market, ultimately leading to less effective performance management and collaboration with providers. Those types of case-mix adjustment include:

Case-mix adjustment type	Details	Drawbacks
<b>Inpatient DRG adjustments</b>	<ul style="list-style-type: none"> <li>Primarily used for making payments in inpatient settings (corollary exists for outpatient) in a way that adjusts for the severity of patients</li> <li>Goes beyond CMS MS-DRG (3 levels) to risk adjust at 4 levels based on secondary diagnosis</li> <li>Option 1: Severity of Illness (Minor, Moderate, Major, Extreme)</li> <li>Option 2: Risk of Mortality (Minor, Moderate, Major, Extreme)</li> </ul>	<ul style="list-style-type: none"> <li>Blackbox methodology</li> <li>Limited risk adjustment</li> <li>Primarily used for payment not benchmarking</li> <li>Susceptible to upcoding by providers</li> </ul>
<b>“Representative” unit</b>	<ul style="list-style-type: none"> <li>Benchmarks created represent average healthcare costs and changes in those costs for a hypothetical ‘typical American family of four’</li> <li>Two values generated for every metric:</li> <li>Standard: MMI for a family of four with employer-sponsored insurance</li> <li>Effective use of managed care: MMI under a managed care setting (about 25% lower)</li> </ul>	<ul style="list-style-type: none"> <li>Only two values for each benchmarking metric</li> <li>Blackbox for the “standard” and “effective use” cohorts</li> <li>National in scale, so does not pick up regional effects</li> </ul>
<b>Cohort-based approach</b>	<ul style="list-style-type: none"> <li>Metrics are benchmarked with cohorts that are meant to create “apples-to-apples” comparison sets between the vendor’s data and customer’s data</li> <li>MCG creates cohorts based on age and gender only and the vendor’s data set is based on 7 million patient claims nationally</li> <li>Metric benchmarks can include multiple values including quartile or decile performance for that cohort</li> </ul>	<ul style="list-style-type: none"> <li>Limited data set and limited case-mix adjustment</li> <li>Limited if sample size of benchmarking data set is insufficient for small cohorts</li> </ul>
<b>Severity-adjustment method</b>	<ul style="list-style-type: none"> <li>Severity adjustment relies upon linear regression models to quantify the impact of both demographic and clinical factors that may influence total episode cost</li> <li>Demographic factors include age and gender; clinical factors include condition status (sets of diagnosis codes that indicate clinically distinct subsets within an ETG base class) and comorbidity markers (to quantify the influence of concurrent medical diagnoses across ETG base classes) associated with an episode</li> </ul>	<ul style="list-style-type: none"> <li>Works for episodic evaluation only</li> <li>Does not adjust or predict the outcome (e.g., cost, LOS). Instead, this method provides a second value that must be used to compare providers</li> </ul>

# Next-generation dynamic provider benchmarking

Leap-frogging incumbent, dated methodologies, Clarify solves for case-mix adjustment with predictive models to (1) retrospectively compare observed patient care to predicted clinical and financial performance, and (2) prospectively predict patient risk and care utilization in real-time at the point of care. The Clarify Platform applies an efficient machine learning data pipeline to generate accurate predictive models that yield decile-based and proprietary predicted benchmarks and values. These benchmarks on expected performance are displayed within an analysis as our signature “Blue Diamonds.” Our Blue Diamonds are completely transparent – there is no blackbox – and can be explored at a granular level to identify the key drivers of actual performance compared to expected.

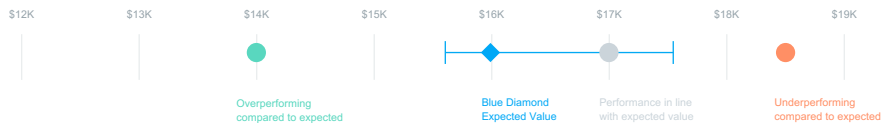


Figure 1: This legend shows a sample of performance benchmark definitions used in Clarify analyses

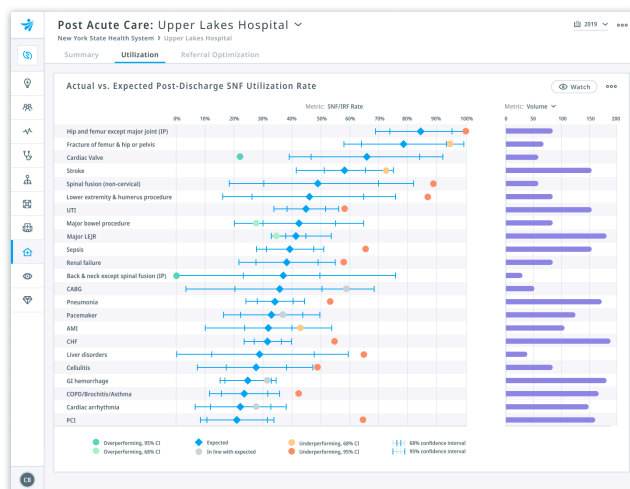


Figure 2: This is a sample analysis of post acute care utilization by service line for a New York hospital

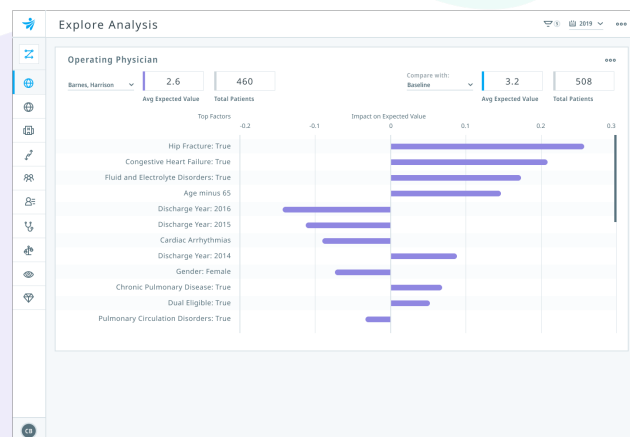


Figure 3: This is an example of a drill-down on provider performance showing the metrics that influenced the benchmark

Furthermore, the Clarify software platform delivers these benchmarks dynamically, on the fly, creating time savings and operational efficiencies heretofore unseen.

The predicted value the platform generates provides the most accurate case-mix-adjustment on the market, accounting for provider factors (e.g., hospital size, academic vs. non-academic), patient characteristics (e.g., co-morbidities, DRGs, procedure types, age, etc.), and regional factors (e.g., rural versus urban) to derive how an average provider would be expected to perform on that metric on the exact same set of patients.

### Clarify's approach to provider performance benchmarking is differentiated in the following ways:

1. **Comprehensive case-mix / severity-adjustment factors** including demographics, clinical diagnoses and social determinants of health drawn from the largest patient-level dataset on the market.
2. **Precise modeling using machine learning.** Instead of cohort-based approaches, we used generalized linear models which look at each individual provider's panel and the values associated with their unique patients. This is basically the difference between using calculus to calculate (our approach) vs. summing rectangles to approximate (cohort) area under the curve.
3. **Comparison of the benchmark value to actual value**, which means that you do not require any additional "severity score" when comparing outcomes.
4. **Not a black box.** Clarify can highlight how each variable contributes to the case-mix / severity-adjustment, not just provide a black box score.
5. **Dynamic benchmarking and custom care groupings.** Because of the speed and precision of the machine learning platform, Clarify allows for flexible interrogation and drill-down within any category, including the ability to change and the comparator dataset on the fly.

## External validation of Clarify's methodology

### HEAD-TO-HEAD STUDY VS. INDUSTRY INCUMBENT'S METHODOLOGY

A leading multi-state health insurer and Integrated Delivery Network based in Pittsburgh, Pennsylvania, partnered with Clarify because they wanted an innovative, precise, and software-based approach to understanding provider performance. The first step in the partnership was to validate Clarify's benchmarks versus the incumbent's.

### RESULTS

The results revealed that Clarify's precise benchmarking methodologies, driven by granular case-mix adjustment and specialty-specific total cost and episodic modeling, corrected multiple issues in the incumbent vendor's methodology.

For example, there were clear patterns of over-predicting and under-predicting across different specialties. For lower per-unit cost specialties like Orthopedics and Sports Medicine, Clarify found that the existing models were under-predicting in over 20% of specialists. That means that the customer was failing to identify **potential performance issues in 20% of their specialist providers**, leading to incorrect decisions on network, referral, and contracting decisions.

Additionally, for higher per-unit cost specialties like Interventional Cardiology, Cardiac Surgery, and Neurosurgery, the customer determined that the incumbent vendor was significantly over-predicting costs and performance. **This issue was found in 50% of Cardiologists / Cardiac Surgeons and 25% of Neurosurgeons.**

This meant that the customer was significantly over-flagging these specialists as having performance issues, whereas Clarify's models accurately found that the higher costs were justified given more complex case-mix and severity of the patient panel or procedures.

These differences in Clarify's benchmarking methodology compared to the customer's existing methodology demonstrated the power of Clarify's models to identify performance accurately at both ends of the spectrum - higher volume, lower cost specialties as well as lower volume higher cost specialties.

### SAMPLE BLINDED COMPARISONS BETWEEN CLARIFY VS. INCUMBENT VENDOR

SURGEON	INCUMBENT			CLARIFY			DIFFERENCE	
	EPISODES	OE	PREDICTED	EPISODES	OE	PREDICTED	OE	PREDICTED
Orthopedic	110	1.05	\$ 24,377	53	1.10	\$ 33,030	(0.04)	\$ (8,653)
Orthopedic	30	1.38	\$ 36,020	32	1.31	\$ 79,240	0.07	\$ (43,219)
Cardiac	81	1.10	\$ 54,081	100	0.92	\$ 59,599	0.17	\$ (5,518)
Orthopedic	22	0.72	\$ 10,095	29	1.05	\$ 44,397	(0.33)	\$ (34,302)
Neuro	17	1.24	\$ 33,588	20	0.76	\$ 35,434	0.48	\$ (1,846)
Neuro	19	1.34	\$ 35,860	45	1.06	\$ 65,411	0.28	\$ (29,552)
Orthopedic	48	0.91	\$ 13,839	87	0.97	\$ 26,488	(0.06)	\$ (12,649)
Orthopedic	16	0.88	\$ 23,839	17	1.08	\$ 56,887	(0.19)	\$ (33,189)
Cardiac	144	1.04	\$ 18,448	17	0.94	\$ 19,992	0.10	\$ (1,545)

● SIMILAR PREDICTION   ● INCUMBENT UNDER-PREDICTION   ● INCUMBENT OVER-PREDICTION

## EXAMPLES OF CLARIFY'S UTILIZATION BENCHMARKING:

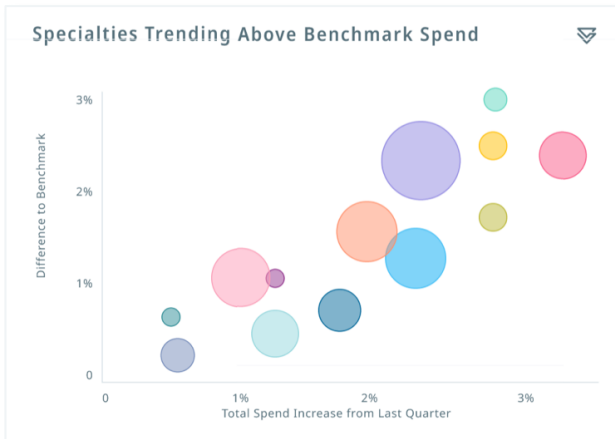


Figure 4: This is an example of an opportunity finder, highlighting the specialties with the largest difference between actual and expected performance

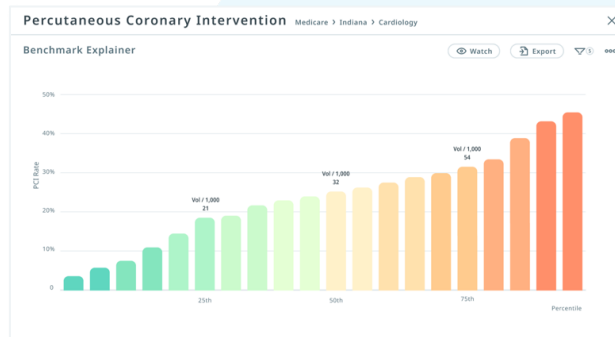


Figure 5: This is an example of a drill-down analysis on provider performance for PCI

## EXAMPLES OF CLARIFY'S OUTCOMES MODELING & BENCHMARKING

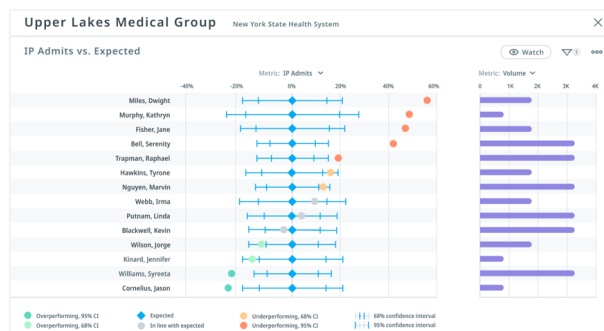


Figure 6: This is an example of an analysis of IP admits by provider



Figure 7: This is an example analysis of length of stay for a cardiology service line

## Conclusion

Realizing the promise of big data in healthcare and life sciences rests on the ability to effectively distill meaning from traditionally fractured, unconsumable data sources. Clarify brought together experts to address the deficiencies of current analytics models used in healthcare and to build the future of benchmarking in dynamic, self-serve software. The methodology described in this paper showcases how Clarify can transparently and precisely uncover previously unrecognized insights about clinicians, service lines, facilities, and care networks to improve care. It is only with fair and trusted performance benchmarking that provider organizations and health plans will truly move the needle in achieving their goals for cost and quality optimization.

## About Clarify Health

At Clarify Health, we turn data into insights, so our provider, payer, and life sciences partners can turn insights into impact. The Clarify Health Platform is the only enterprise analytics platform to power stakeholders across the industry. Our growing data set – the most comprehensive longitudinal data set in the US – links clinical, claims, prescription, lab, and social determinant of health data on over 300 million lives. Our industry-leading analytics platform applies externally validated statistical modeling, machine learning, and AI to process and harness the power of this data to improve patient care and optimize drug commercialization.

To find out how Clarify's cloud-based analytics solutions can help you answer your most important business questions and deliver better patient care, visit [www.clarifyhealth.com](http://www.clarifyhealth.com).