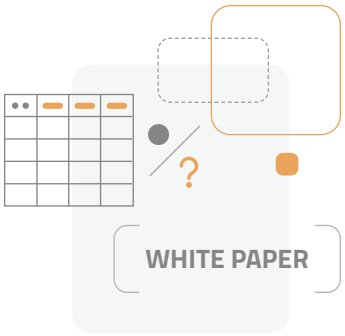
KYTHERA
DECIPHERING HEALTHCARE



Bridging the Data Divide: Solving for the 'Unknown Denominator' in Claims Coverage

Part of Kythera Labs' Series on Coverage, Fidelity, and Trust in Healthcare Data

Executive Summary

Accurate healthcare measurement starts with knowing the total population represented in the data. In claims analytics, that total is often uncertain, a challenge known as the "unknown denominator." When the denominator is incomplete or unclear, every rate, percentage, and market estimate built on it carries an unknown level of uncertainty.

Healthcare business and strategy leaders rely on accurate data to understand their markets, patient populations, and competitive position. Yet even large, seemingly comprehensive datasets can distort reality by double-counting encounters, omitting certain payer segments, or redacting sensitive services. These gaps create an illusion of completeness that can mislead forecasts, misstate market share, and weaken strategic planning.

Open claims datasets, whether publicly available or commercially licensed, have become essential tools for market insight, but their coverage is uneven. Some regions are richly represented while others are sparse, and key categories of care may be excluded for privacy or proprietary reasons. Without a known denominator, organizations are left to infer completeness, introducing uncertainty into every analysis built on that data.

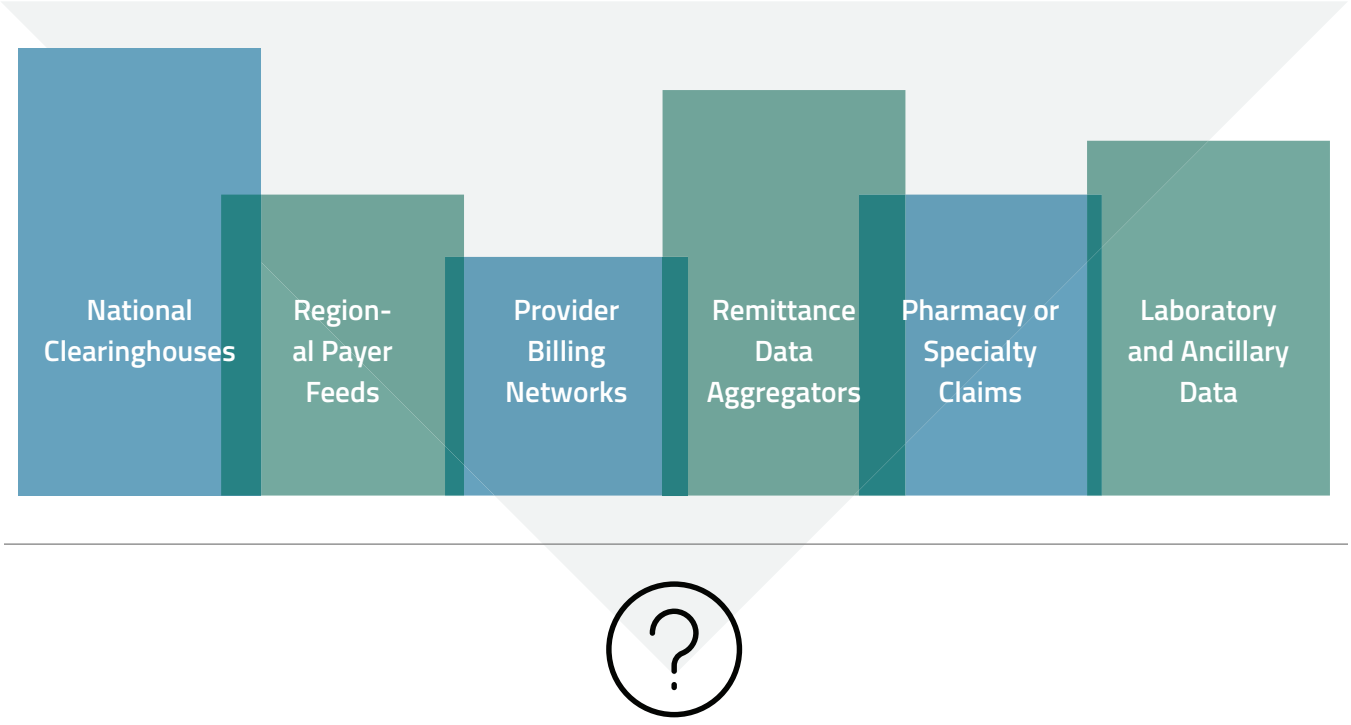
KYTHERA

Kythera Labs addresses this challenge by quantifying, normalizing, and reconciling open claims coverage to establish a measurable baseline of fidelity. By reconstructing the denominator, Kythera helps organizations distinguish what is visible in the data from what is real in the market. This provides a verified foundation for planning, forecasting, and strategic decision-making that reflects true healthcare activity rather than assumed completeness.

This paper explores the implications of the unknown denominator in open claims analytics and introduces Kythera's approach to restoring measurable fidelity so that confidence in coverage is proven, not assumed.

The Core Challenge: The Data Divide in Open Claims

Open claims datasets provide valuable visibility into healthcare activity beyond any single payer or provider network, but they are not designed for accuracy. Each dataset reflects the rules, incentives, and reporting behavior of its contributors. After all, open claims only exist in the first place so that providers can get paid. The opportunity to mine such data for market intelligence is only incidental to its original purpose. In addition, the data collection points are distributed all across the country with data originating in as many locations as there are provider organizations needing to submit for payment. No single aggregator captures the full market, and participation varies by clearinghouse, payer, and geography.



Open claims data are compiled from many distributed sources. Partial overlaps and missing contributionscreate uneven visibility and an incomplete denominator.

The result is a data divide, a structural disconnect between what open claims appear to show and what actually occurs in the market. While these datasets capture more encounters than any single payer file (a common alternative to open claims for certain use cases), they lack a stable denominator. There is no definitive count of the patients, encounters or payers missing from view, which makes it impossible to determine true completeness.

For healthcare leaders, this creates a coverage paradox. Two datasets may each claim 80 percent market coverage, yet they may represent entirely different populations or payer mixes. The apparent similarity hides significant differences in what is actually measured. This uncertainty undermines confidence in metrics such as market share, leakage and service-line demand, all of which depend on knowing the total population represented in the data. Without that known denominator, even precise analytics can rest on an uncertain foundation.

The Unknown Denominator Problem

Even when a dataset reports millions of lives or encounters, the true denominator, which is the total number of patients and events in the market, is unknown. Without that baseline, percentages, market share and growth rates cannot be accurately calculated or compared.

Uneven Coverage: Data Gaps Within Open Claims Networks

Open claims data provides valuable breadth, but it is not a complete view of the healthcare landscape. Certain payers and delivery systems restrict or limit how their claims appear in open datasets, either for contractual, competitive, or regulatory reasons.

Large integrated organizations that act as both payer and provider, such as Kaiser Permanente, often maintain data within closed networks that are not routinely shared externally. Similarly, some national insurers have been selective about which claims are made available through third-party data aggregators or open claims feeds. These variations can create blind spots in markets where those entities have significant presence.

Even within open networks, sensitive categories of care such as behavioral health, reproductive health, or HIV treatment may be partially redacted or excluded under HIPAA expert determination privacy standards. The result is uneven visibility across populations and service lines, with certain encounters overrepresented and others underreported.

Network Type	Typical Data Accessibility
National, multi-payer networks	High data availability across many states
Regional Blue Cross Blue Shield plans	Partial availability; varies by region
Closed integrated systems (e.g., Kaiser Permanente)	Limited or no external data sharing
Government programs (Medicare, Medicaid)	Selective sharing under regulatory agreements
Behavioral and reproductive health networks	Often redacted for privacy reasons

Variability in network participation and data accessibility contributes to gaps in coverage and reinforces the challenge of an unknown denominator.

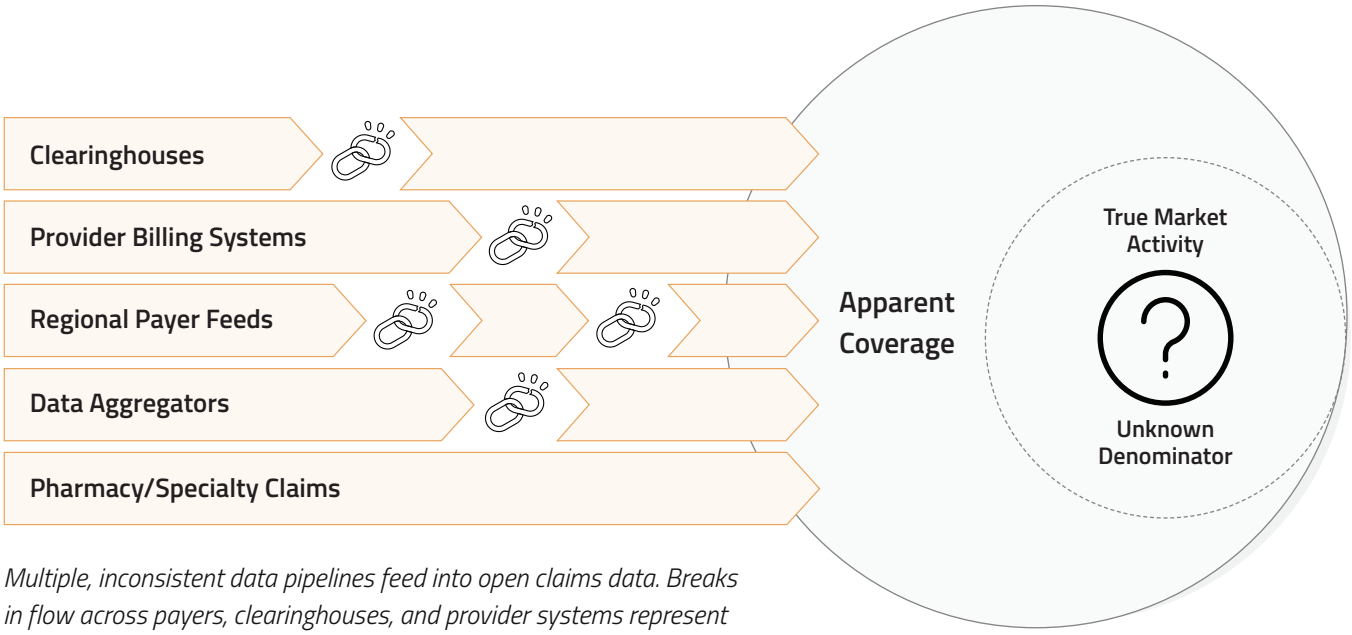
Understanding these structural limitations is essential to interpreting open claims data responsibly. Fidelity begins with transparency, with a clear understanding of which segments of the market are well represented, which are not, and where additional context or linkage is needed to reflect true activity.

Structural Fragmentation: How Inconsistency Distorts Analytics

The structure of claims data reflects the complexity of the healthcare system itself. Each payer, clearinghouse, and data supplier follows its own standards for coding, submission, and validation. Even within a single market, two data sources may record the same patient encounter in different ways, or one may exclude it altogether.

These inconsistencies make it difficult to determine what is truly represented in the data. When the same healthcare event appears multiple times or not at all, the denominator becomes unstable. A dataset may appear complete because it contains a large number of claims, yet without a consistent structure or standard for reconciliation, it is impossible to know how many of those claims represent distinct, verified encounters.

For business and strategy leaders, this inconsistency introduces silent risk into every analysis. Market share, referral patterns, and service-line growth all depend on accurate measures of patient activity. When structural differences distort those measures, confidence in performance metrics erodes, even when the underlying data seems precise.



Multiple, inconsistent data pipelines feed into open claims data. Breaks in flow across payers, clearinghouses, and provider systems represent the structural fragmentation that prevents a complete count of real healthcare activity — the unknown denominator.

Fragmentation across payers and sources is more than a technical inconvenience. It is the primary reason the denominator remains unknown. Without structural consistency, datasets cannot be compared, and trends cannot be validated. The lack of alignment across identifiers, taxonomies, and data models leaves each source isolated within its own logic, unable to support a unified view of healthcare activity.

Understanding this fragmentation is essential because it defines the point where standardization must begin. Only when data is normalized to a common structure can the denominator be reconstructed and fidelity measured with confidence. This is the point where Kythera's approach begins, restoring structure, reconciling records, and creating a consistent foundation for measurement.

Kythera's Approach: Normalizing the Divide

Kythera Labs addresses structural inconsistencies in claims data through its proprietary Common Data Model (CDM), a unifying framework that standardizes, validates, and harmonizes disparate data sources. The CDM applies consistent identifiers, controlled vocabularies, and quality checks across every data element to ensure that each record represents a verified, singular healthcare event. Proprietary normalization and deduplication processes reconcile overlapping patient lives and restore coherence to the data.

The CDM operates as a measurable framework for fidelity. By aligning data to a common structure, Kythera redefines completeness in practical terms, quantifying how much of the market is actually represented in the data and where visibility remains limited. This approach transforms uncertainty about coverage into a verifiable measure of confidence.

Key elements of Kythera's normalization process include:

- **Remastering and Alignment:** Unifying payer, provider, and patient identifiers across sources.
- **Deduplication and Validation:** Eliminating redundant claims and verifying that each remaining record reflects a real encounter.
- **Completeness Scoring:** Measuring and documenting confidence levels for each record and data slice.
- **Structural Mapping:** Linking open and closed claims, where available, to reconstruct longitudinal journeys and payer transitions.

Building a Common Language for Coverage

Kythera's Common Data Model aligns every record to a unified structure, creating a reproducible foundation for measuring healthcare activity across markets, payers, and providers.

This process does more than clean data. It reconstructs the denominator, filling in the missing context needed to interpret market activity accurately. By restoring structure and consistency, Kythera's CDM turns fragmented data into a coherent, measurable representation of healthcare activity that can be trusted for analysis and decision-making.

Establishing a Standardized Baseline for Healthcare Analytics

When fragmentation is resolved and the denominator is defined, open claims data becomes not just accessible but trustworthy. Kythera's Common Data Model (CDM) enables healthcare organizations to evaluate markets using consistent definitions, metrics, and denominators, creating a standardized baseline for analytics.

Establishing a standardized baseline allows organizations to measure healthcare activity with consistency across time, regions, and data sources. Kythera's CDM enforces a uniform structure that normalizes key metrics such as prevalence, utilization, and cost. These measures become reproducible and comparable, giving every stakeholder a common reference point for performance and planning.

Metric	Traditional Open Claims	Kythera CDM Normalized
National Coverage	Variable	Enhanced, consistently measured
Denominator Accuracy	Imprecise	Verified and documented
Bias in Prevalence Estimates	High	Quantified and reduced
Update Frequency	Inconsistent	Regular and standardized

The CDM establishes a consistent analytical baseline where completeness, accuracy, and frequency can be measured and reproduced.

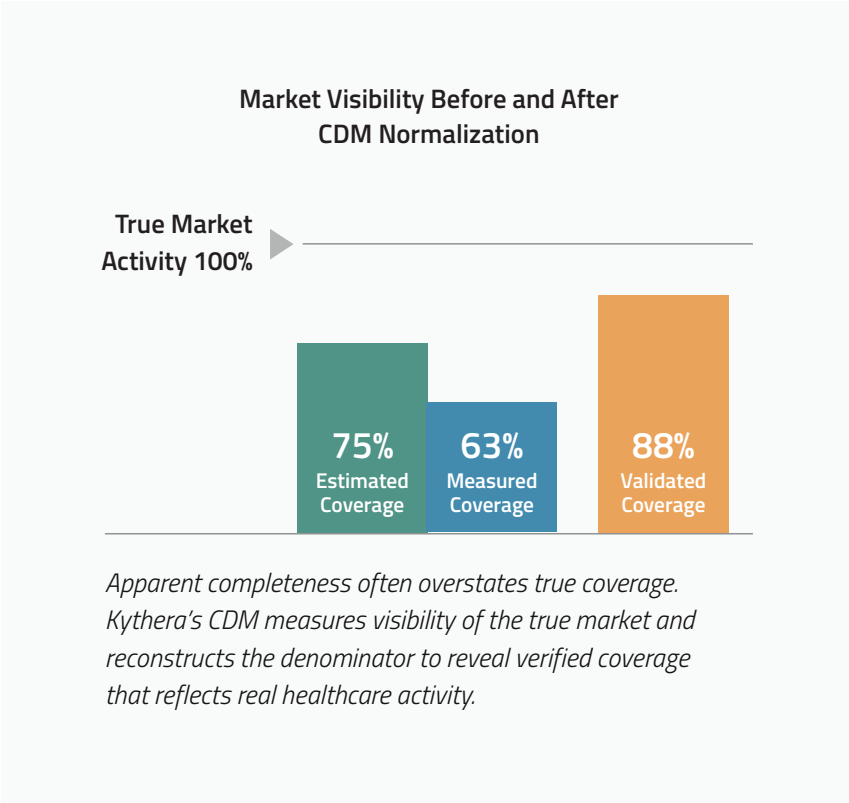
A standardized baseline is the foundation of reliable analytics. It ensures that every measure, from prevalence to utilization, can be traced to a known denominator and reproduced with confidence. This foundation benefits every stakeholder. Healthcare organizations can evaluate performance with accuracy, researchers can harmonize studies across datasets, and innovators can compare outcomes using consistent, reproducible metrics.

Use Case Scenario: Applying a Standardized Baseline to Reconstruct Coverage

The following example illustrates how Kythera's Common Data Model (CDM) establishes a measurable foundation for confidence in coverage. The scenario represents a modeled analysis designed to show how normalization and reconciliation within the CDM can clarify apparent completeness and reconstruct the denominator in claims data.

In this example, a regional health system estimates that its data provides visibility into roughly 75 percent of patient encounters within its market. A deeper assessment using Kythera's standardized baseline reveals duplication across clearinghouse feeds and uneven participation among major payers. When these inconsistencies are measured and corrected through the CDM's normalization process, verified visibility drops to approximately 63 percent, reflecting the true denominator within the available data.

After the CDM process aligns identifiers, reconciles duplicates, and applies completeness scoring, validated coverage increases to approximately 88 percent. This improvement demonstrates how fidelity can be quantified and completeness made measurable once the denominator is properly defined.



Apparent completeness often overstates true coverage. Kythera's CDM measures visibility of the true market and reconstructs the denominator to reveal verified coverage that reflects real healthcare activity.

Through this process, the health system achieves a stable analytical baseline. Market share, leakage, and referral analyses now reflect verified patient behavior rather than assumptions about completeness. The same principle applies across other analytical domains, such as service-line demand, payer mix, or competitive dynamics, where confidence depends on knowing the total population represented in the data.

Restoring Confidence Through Fidelity

Kythera's Common Data Model transforms apparent completeness into measurable fidelity, allowing organizations to quantify what they know, what remains uncertain, and how that certainty changes as data improves.

Conclusion

Every measure of healthcare performance depends on understanding the total population represented in the data. When that denominator is unknown, every subsequent calculation, from market share to utilization to growth potential, rests on uncertain ground.This uncertainty has long limited how healthcare organizations interpret open claims data and the confidence they place in the insights derived from it.

Kythera Labs resolves this challenge through its Common Data Model (CDM), which transforms fragmented datasets into a coherent, standardized representation of healthcare activity. By estimating a consistent denominator, the CDM restores measurable fidelity and establishes a verifiable foundation for analytics.

With this foundation in place, healthcare leaders can evaluate markets, monitor performance, and plan strategically with greater confidence in what the data actually represents. Coverage is no longer an assumption; it is a measurable fact that can be documented, compared, and improved.

When the denominator is known, analysis becomes not just more accurate but more meaningful. It allows organizations to make decisions grounded in truth rather than inference, ensuring that the data they rely on reflects the healthcare system as it truly operates.

Connect with Kythera. Kythera is a data technology company that brings unprecedented clarity and structure to complex real-world healthcare data. Kythera's Wayfinder Technology Platform, supported by pre-configured pipelines, processing libraries, analysis tools and remastered datasets, helps Healthcare and Life Sciences organizations work with greater speed, scale and confidence. Learn more at www.kytheralabs.com.

KYTHERA™
DECIPHERING HEALTHCARE

346 Main Street
Franklin, TN 37064
www.kytheralabs.com