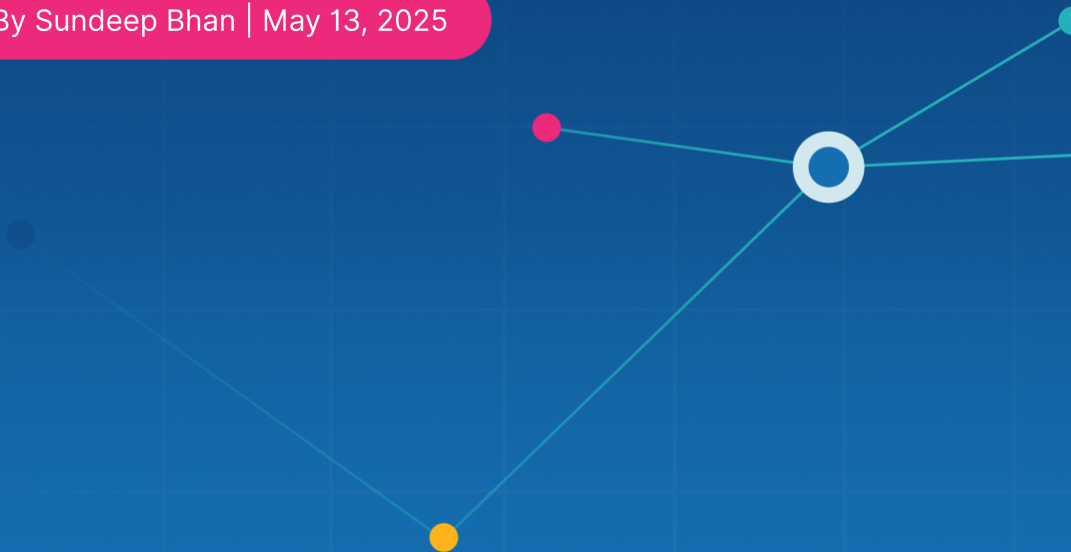


The Challenges and Opportunities with the Fragmentation of Lab Data

By Sundeep Bhan | May 13, 2025



Laboratory data

Is a cornerstone of modern healthcare, influencing nearly every aspect of patient diagnosis, treatment, and management. It provides objective, quantifiable insights that guide clinical decision-making, improve patient outcomes, and support broader public health initiatives.

70%

Of Medical Decisions rely
on lab test results

In fact, approximately **70% of medical decisions rely on lab test results**, which help detect diseases early, often before symptoms appear, reducing the risk of misdiagnosis and ensuring evidence-based clinical assessments.

By offering critical information about a patient's condition, lab data enhances treatment decisions, enabling precision medicine by tailoring therapies to an individual's biomarkers, guiding medication dosing, and preventing adverse drug reactions through genetic insights.

Beyond treatment

Lab data plays a crucial role in preventive and predictive healthcare by identifying risk factors for chronic diseases and guiding targeted therapies. In oncology, for instance, non-small cell lung cancer (NSCLC) treatment decisions increasingly depend on lab data to identify actionable genetic mutations such as EGFR, ALK, and KRAS, which determine whether a patient is eligible for targeted therapies or immunotherapy.

Without access to comprehensive lab data, patients may receive less effective treatments, delaying their chances of remission or survival. Similarly, in the case of rare diseases, lab data is vital for identifying specific genetic mutations that define treatment pathways.

For example Spinal Muscular Atrophy (SMA), a severe neuromuscular disorder, is caused by mutations in the SMN1 gene, which can be detected through genetic testing. Early diagnosis via lab data enables clinical trials optimization, timely intervention with disease-modifying therapies like gene therapy, improving motor function and survival outcomes in affected infants.

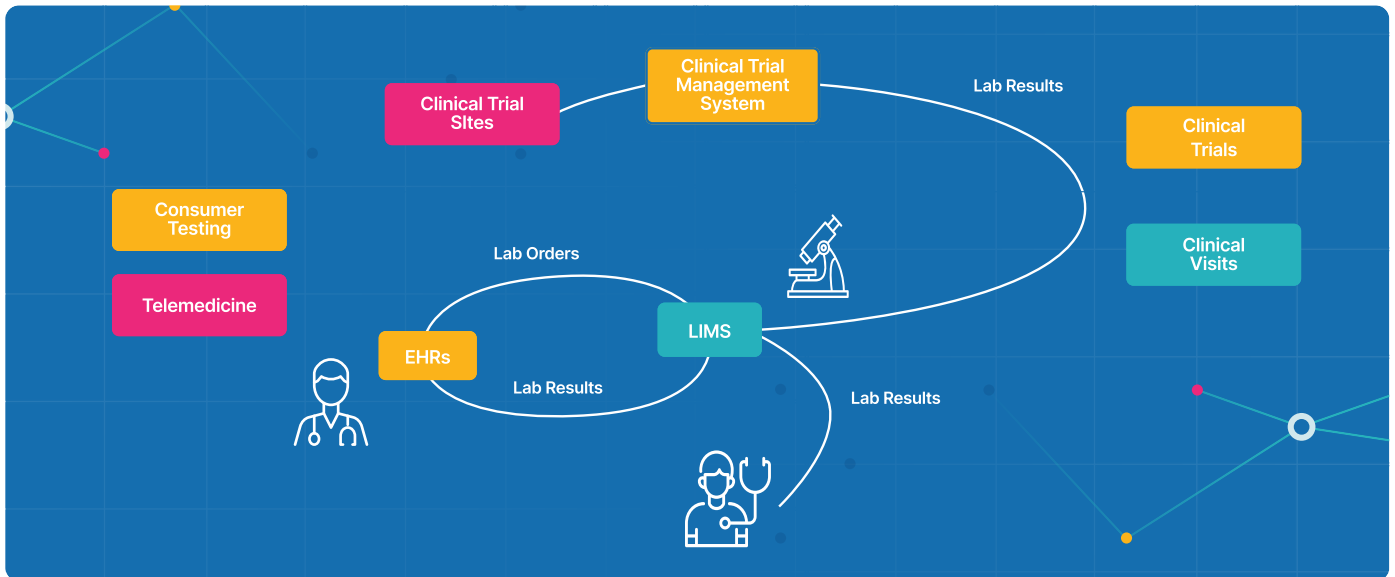


Figure 1 - Lab Testing Workflow

While lab data is critical for clinical decision-making and research, it remains highly fragmented, making it difficult to build a comprehensive view of a patient's diagnostic journey. This fragmentation stems from several structural and technical challenges.

- 1 **First**, approximately half of diagnostic testing in the U.S. is performed by hospital-based laboratories, where data is often stored in disparate, siloed systems that are difficult to aggregate.
- 2 **Second**, the rise of specialized and startup diagnostic companies, particularly in areas like genomics, liquid biopsy, and early cancer detection, has created new silos, as these companies often operate on proprietary platforms with limited data interoperability.
- 3 **Third**, the lack of interoperability and standardization among laboratory information systems (LIS), many of which are custom-built or legacy systems makes it challenging to exchange or harmonize data across labs, providers, and platforms.

Compounding these issues is the fact that lab data can flow through multiple external channels, such as billing systems, clinical connectivity companies, or digital health service providers. While this creates more opportunities for data to be accessed and commercialized, it also introduces additional risk. These pathways often require data rights to be extended through third-party service contracts, which can be revoked, restricted, or expire, creating instability and uncertainty around long-term access.

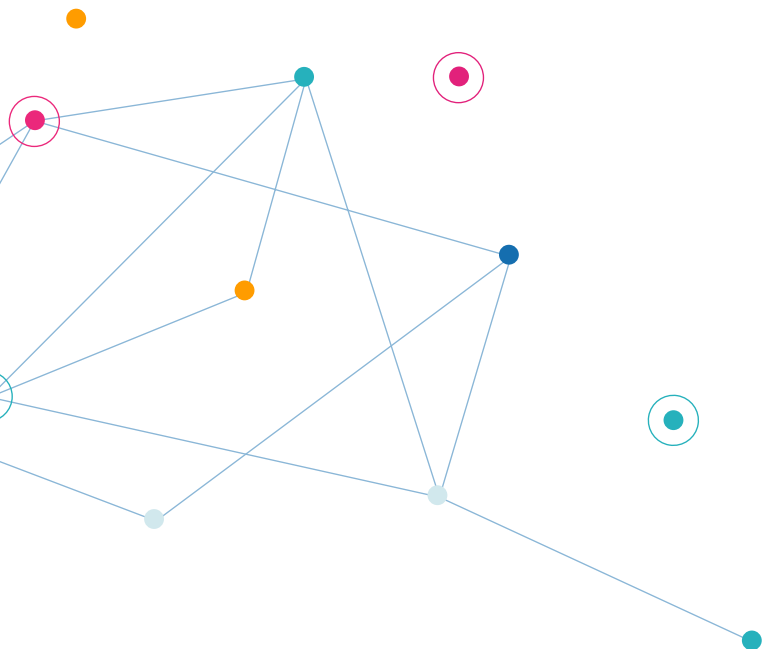
Moreover, as lab data travels through multiple intermediaries, it often leads to data duplication, where the same patient record or test result exists in multiple places without clear provenance, an issue that tokenization alone cannot fully solve.

As a result, companies that rely solely on indirect access to lab data may face limitations in data completeness, fidelity, and continuity. This underscores the strategic advantage and reliability of working with dedicated data marketplaces, where access rights are more secure, data quality is higher, and insights can be delivered with greater precision and consistency.

At a broader scale, lab data is indispensable for public health and medical research, tracking epidemic outbreaks such as COVID-19 and influenza, supporting real-world evidence (RWE) for drug approvals, and guiding healthcare policy decisions by identifying disease trends and disparities.

Additionally, it helps lower overall healthcare costs by preventing unnecessary procedures, avoiding hospital readmissions through post-treatment monitoring, and supporting value-based care models that emphasize data-driven patient management. Ultimately, lab data ensures that healthcare decisions are precise, timely, and evidence-based, driving better patient outcomes, improving system-wide efficiency, and saving lives.

The lab data ecosystem is comprised of three primary categories:



Originators of Data

These are entities that generate laboratory data through diagnostic tests, imaging, and analyses.



Intermediaries

Organizations that aggregate, process, and facilitate the exchange of lab data between originators and end-users.



Users and Customers of Lab Data

Entities that utilize lab data for various purposes, including research, clinical decision-making, and policy development.



Figure 2 - US Lab Data Ecosystem - Prognos Health 2025

1 Originators of Data (Labs & Diagnostic Companies)

These companies produce lab data through various diagnostic services:

National & Global Diagnostic Labs

National reference labs like Labcorp and Quest Diagnostics hold a significant share of the U.S. diagnostic testing market, collectively performing 20–25% of all lab tests nationwide. These labs specialize in high-volume, routine testing such as blood panels, metabolic and lipid profiles, infectious disease screening, and pathology services and serve a broad network of physicians, hospitals, and employers. They also offer specialized testing in areas like genetics, oncology, and women's health, but their core strength lies in efficient, centralized processing of large-scale diagnostic workflows, often through expansive logistics networks and automation infrastructure.

Key Labs - Quest Diagnostics, Labcorp, Sonic Healthcare, Eurofins Scientific



Specialty & Genetic Testing Labs

Specialty diagnostic companies like GeneDx, Guardant Health, Invitae, Exact Sciences, Natera, Foundation Medicine, Myriad Genetics, Ambry Genetics, Caris Life Sciences, Tempus, Prevention Genetics, Veracyte, Neogenomics, and Helix focus on high-complexity testing in oncology, rare diseases, and genetic medicine. These organizations specialize in next-generation sequencing (NGS), molecular diagnostics, and anatomic pathology, playing a critical role in precision diagnostics by offering advanced services such as comprehensive genomic profiling, liquid biopsy, carrier screening, hereditary cancer panels, and tissue-based analysis. Their testing helps identify actionable mutations, guide targeted therapies, and enable earlier detection of disease, and is often essential for treatment selection in conditions like non-small cell lung cancer (NSCLC) or hereditary breast and ovarian cancer, as well as for diagnosing rare diseases linked to specific genetic mutations. Complementing this are key anatomic pathology groups like PathGroup, Pathology Regional Services (Path Regional), Clinical Pathology Laboratories (CPL), and Pathline, which provide critical insights through biopsy and surgical specimen analysis. Together, these specialty and pathology labs are indispensable for advancing personalized medicine, clinical research, and precision oncology.



Hospital & Academic Lab Networks

Hospital and academic medical center laboratories perform an estimated 50% of all diagnostic testing in the U.S. and collectively represent the largest category of lab data originators, with over 7,000 hospital-based labs nationwide. These labs support both inpatient and outpatient care, delivering a wide spectrum of routine and specialized tests that are deeply embedded in clinical workflows. In addition to standard panels like CBCs and metabolic profiles, they frequently conduct high-complexity and esoteric testing, including infectious disease diagnostics, pathology, and transplant-related testing. Academic medical centers also lead in diagnostic innovation, offering custom assays, genomic research, and clinical trial support. While this data is rich, timely, and highly contextualized within patient care, it is often siloed in proprietary or homegrown lab information systems (LIS), which limits interoperability and makes cross-institutional aggregation and standardization a major challenge.

For example: Mayo Clinic Laboratories, Cleveland Clinic Laboratories, UCSF Health Lab Services



Retail & Direct-to-Consumer (DTC) Labs:

Direct-to-consumer (DTC) testing companies like Everlywell, LetsGetChecked, 23andMe, AncestryDNA, and Color Health have rapidly expanded access to lab testing by enabling individuals to request and complete tests from home, bypassing traditional healthcare settings. These companies focus on wellness, preventive health, genetic insights, and early risk detection, offering services such as hormone testing, food sensitivity panels, STI screening, ancestry reports, and polygenic risk scores for conditions like cancer and heart disease.

While not all results are intended for diagnostic use, many are CLIA-certified and physician-reviewed, with some companies increasingly partnering with employers and health systems for broader population health initiatives. Though these labs represent a smaller share of total testing volume, they generate large volumes of consumer-authorized, structured data—particularly in genomics—that can be valuable for research, product development, and long-term health monitoring. However, data access and use are tightly governed by user consent and privacy policies, creating variability in how and where the data can be shared or applied.



23andMe

everlywell

ancestryDNA

color

2 Intermediaries (Data Aggregators, Tokenization, Exchange and Marketplace Platforms) Role in Transforming Lab Data into Actionable Insights

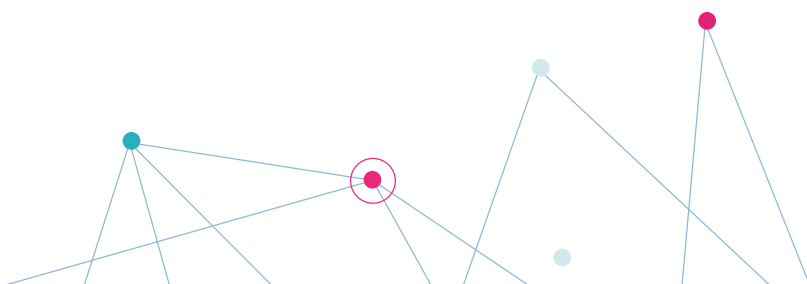
These companies facilitate data exchange between originators and end-users, ensuring data is aggregated, de-identified, tokenized, and analyzed for commercial or research applications.

The fragmentation of lab data presents significant challenges for healthcare organizations, pharmaceutical companies, and researchers who rely on timely, accurate, and comprehensive insights to drive better patient outcomes. Raw lab data, often scattered across multiple sources in disparate formats, is difficult to utilize without proper ingestion, harmonization, and standardization. This is where intermediaries play a crucial role by collecting lab data from various sources, structuring and normalizing it, and ensuring it is both **accurate and accessible** for downstream users.

Effective intermediaries must build robust **data ingestion pipelines** that can integrate structured and unstructured lab data from national reference labs, hospital networks, and specialty diagnostic providers. They must also employ **harmonization techniques**, such as standardizing lab test names, normalizing result values, and mapping different reporting formats to a unified ontology. Without this process, lab data remains fragmented, inconsistent, and difficult to integrate into real-world applications like clinical decision-making, population health studies, and precision medicine.

Beyond harmonization, intermediaries must ensure accessibility by delivering lab data in formats that seamlessly integrate into customer workflows. This requires building scalable, privacy-compliant platforms that allow real-time access to lab data while maintaining strict data security and patient privacy regulations. Additionally, intermediaries must enable linkability—leveraging tokenization and de-identification technologies to connect lab data with other healthcare datasets such as claims, EHRs, and genomics.

The true value of lab data is realized when it becomes interoperable and actionable, allowing pharmaceutical manufacturers to optimize their targeting strategies, enabling clinical researchers to identify biomarker-driven patient cohorts, and helping healthcare providers make data-driven decisions at the point of care.



Healthcare Data Marketplaces and Exchanges

While platforms like **Snowflake Marketplace, HealthVerity, AWS Marketplace, Google Cloud Marketplace, and Databricks Marketplace** offer access to a wide range of healthcare datasets, **Prognos Health** stands apart as the **only dedicated marketplace solely focused on lab data**.

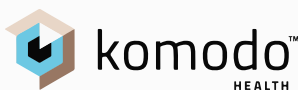
With the most **comprehensive coverage and depth** in the industry, Prognos aggregates lab results via direct agreements with **national, specialty and genetic testing companies**, and **hospital and academic labs**.



Data Aggregators and Analytics Companies with access to lab data

Data aggregators and analytics companies like IQVIA, Komodo Health, Veradigm, Truveta, MMIT, Diaceutics, and Optum play an important but varied role in the lab data ecosystem. While most focus primarily on claims, EHR, or prescribing data, their access to lab data is typically limited, indirect, or acquired through specific partnerships. Some, like Truveta, access lab data through affiliated health systems, while MMIT may obtain reference lab data on a deal-by-deal basis.

Diaceutics stands out by focusing on smaller, fragmented labs with access to data from 500 of the 7000 community and hospital-based labs but lacking access to national or specialty and genetic testing companies. Overall, lab data is often a secondary or supplemental asset for these firms, with only Diaceutics and MMIT, making strategic use of it for diagnostics or market access analytics.



Tokenization & Privacy Solutions

Tokenization and privacy solutions companies like **Datavant**, **LexisNexis Risk Solutions**, **Management Science Associates (MSA)**, and **HealthVerity** play a foundational role in enabling the secure and compliant use of lab data across the healthcare ecosystem. These companies specialize in **de-identifying patient data and generating privacy-preserving tokens** that allow disparate datasets such as lab results, claims, EHR, and genomic data to be linked at the patient level without exposing identifiable information.

Among them, **Datavant is the market leader**, offering an **industry-leading tokenization solution** and the **largest ecosystem of connected healthcare organizations**, including life sciences companies, providers, payers, labs, and research institutions. While these companies do not aggregate lab data themselves, their infrastructure enables lab data to be **interoperable and linkable** with other datasets, unlocking its full value for real-world evidence generation, clinical research, and commercial applications while maintaining strict compliance with HIPAA and data privacy standards.

datavant

 LexisNexis®

 **health**verity

 **msa** Management
Science
Associates, Inc.

Third-Party Services and Diagnostics Companies:

Third-party services and diagnostics companies such as Hologic, Siemens Healthineers, Abbott Diagnostics, Thermo Fisher Scientific, Beckman Coulter, Sysmex, Roche Diagnostics, Becton-Dickinson, Danaher Corporation, Illumina, PerkinElmer, and Qiagen play a critical role in powering the infrastructure of lab testing itself. These companies develop and manufacture the analyzers, reagents, instruments, and molecular testing platforms used by laboratories around the world. While they do not typically own or license lab data, they influence the structure, format, and quality of data that originates from diagnostic tests.

Complementing these are third-party clinical connectivity companies such as HC1, Moxe Health, and Ellkay, along with laboratory information system (LIS) vendors, which help labs manage, store, and exchange their test data. Several companies, like CornerstoneAI, Atria, and ZS, provide lab data clean up and standardization. These solutions often operate behind the scenes, providing connectivity, data normalization, and integration with EHRs or analytics platforms. Though not data marketplaces, these organizations are essential enablers of the lab data ecosystem, shaping how lab data is generated, formatted, and routed through clinical and commercial workflows.

3 Users and Customers of Lab Data:

Entities that utilize lab data for various purposes, including research, clinical decision-making, and policy development. Lab data serves a wide range of customers across healthcare, life sciences, research, and business sectors. Below is a comprehensive list of all major customer types categorized by their primary use cases.

Pharma & Biotech

Commercial teams, R&D, CROs, CDMOs, Gene therapy, and rare disease companies

Healthcare Providers

Hospitals, ACOs, IDNs, Academic medical centers, Urgent care

Payers & Insurance

Health insurers, Medicare/Medicaid, PBMs, Self-insured employers

Life Sciences & Research

Public health agencies, Genomics companies, AI & ML companies, Non-profits

Digital Health & Wellness

Telemedicine, DTx, At-home testing, Pharmacy clinics

Legal & Regulatory

FDA, CDC, Legal compliance firms, Forensic labs

Finance & Investment

Private equity, VCs, Hedge funds, Actuarial firms

Government & Public Health

NIH, WHO, State health departments, HIEs

1 Pharmaceutical & Biotech Companies

Use Case: Drug development, commercialization, clinical trial optimization, and regulatory submissions.

- **Pharmaceutical Commercial Teams** – Use lab data for sales, marketing, and patient/HCP segmentation.
- **Pharmaceutical R&D Departments** – Leverage lab data for biomarker discovery, drug efficacy studies, and precision medicine.
- **Biotech Startups** – Utilize lab data for novel drug and diagnostic test development.
- **Gene Therapy & Cell Therapy Companies** – Require lab data to track biomarker efficacy and patient response.
- **Contract Research Organizations (CROs)** – Conduct clinical trials and research using lab-derived real-world evidence.
- **Contract Development & Manufacturing Organizations (CDMOs)** – Use lab data for quality control in drug manufacturing.

2 Healthcare Providers & Health Systems

Use Case: Clinical decision-making, patient monitoring, population health management, and public health surveillance

- **Hospitals & Health Systems** – Integrate lab data into EHRs for diagnosis and treatment optimization.
- **Academic Medical Centers** – Use lab data for research and advanced clinical trials.
- **Independent Physician Associations (IPAs)** – Utilize lab results for coordinated patient care.
- **Accountable Care Organizations (ACOs)** – Leverage lab data to track patient outcomes and cost efficiency.
- **Integrated Delivery Networks (IDNs)** – Use lab data for value-based care models and cost savings.
- **Urgent Care Centers** – Require quick lab insights for immediate diagnosis and treatment.



3 Payers & Insurance Companies

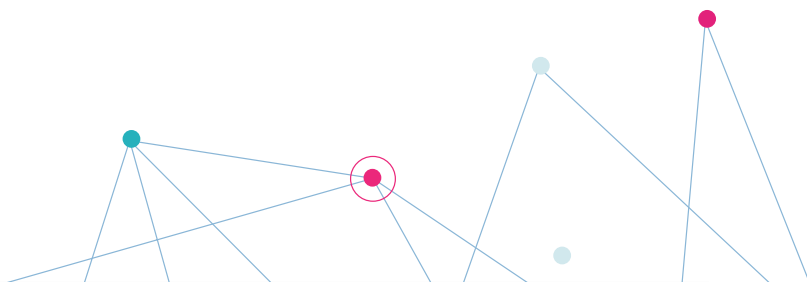
Use Case: Risk assessment, fraud detection, and healthcare cost management.

- **Health Insurance Companies** – Analyze lab data for underwriting and risk stratification.
- **Medicare & Medicaid Payers** – Use lab data for claims validation and population health analytics.
- **Pharmacy Benefit Managers (PBMs)** – Optimize drug formularies based on lab test trends.
- **Self-Insured Employers** – Track employee health trends and predictive analytics for wellness programs.

4 Life Sciences & Medical Research Organizations

Use Case: Real-world evidence, population studies, and biomarker research.

- **Epidemiologists & Public Health Agencies** – Analyze lab data for underwriting and risk stratification.
- **Medical Device Companies** – Leverage lab data for device validation and performance monitoring.
- **Genomics & Precision Medicine Companies** – Require lab and biomarker data for personalized treatment plans.
- **Academic & Government Research Institutions** – Utilize lab data in clinical studies and scientific research.
- **AI & Machine Learning Companies** – Use lab data to develop predictive models for diagnostics and treatment.
- **Non-Profit & Advocacy Groups** – Analyze lab data for disease awareness and research funding.



5 Digital Health & Consumer Wellness Companies

Use Case: Personalized healthcare, remote monitoring, and patient engagement.

- **Digital Therapeutics (DTx) Companies** – Use lab data for behavioral and disease management programs.
- **Telemedicine & Virtual Care Providers** – Require lab data for remote patient diagnosis and monitoring.
- **Wearable & Remote Monitoring Companies** – Combine lab data with sensor data for health tracking.
- **At-Home Testing & Direct-to-Consumer (DTC) Labs** – Provide lab data directly to patients for self-health management.
- **Pharmacy & Retail Health Clinics** – Leverage lab data for customer health screening programs.

6 Legal, Compliance, and Regulatory Bodies

Use Case: Compliance monitoring, forensic investigations, and regulatory reporting.

- **Regulatory Agencies (FDA, EMA, CDC, WHO)** – Use lab data for drug approvals and epidemiological surveillance.
- **Legal & Compliance Firms** – Require lab data for malpractice cases, insurance disputes, and compliance audits.
- **Occupational Health & Workplace Safety** – Monitor employee health risks through lab data.
- **Drug Testing & Forensic Labs** – Utilize lab data for toxicology, criminal investigations, and workplace testing.

7 Financial & Investment Sectors

Use Case: Market intelligence, investment due diligence, and healthcare forecasting.

- **Healthcare Private Equity Firms** – Analyze lab data to evaluate investment opportunities.
- **Venture Capitalists in HealthTech** – Use lab trends to identify innovative startups.
- **Hedge Funds & Market Analysts** – Track lab data for predicting pharma stock performance.
- **Reinsurance & Actuarial Firms** – Utilize lab insights for pricing life and health insurance policies.

8 Government, Policy, and Public Health Organizations

Use Case: Population health surveillance, disease control, and healthcare policy decisions.

- **Centers for Disease Control & Prevention (CDC)** – Use lab data to track infectious disease outbreaks.
- **World Health Organization (WHO)** – Analyzes global lab data for pandemic monitoring and response.
- **National Institutes of Health (NIH)** – Funds and conducts research using aggregated lab data.
- **Local & State Health Departments** – Monitor regional health trends through lab test analytics.
- **Health Information Exchanges (HIEs)** – Use lab data for interoperability and patient health records.

About Prognos Health

Prognos Health is the leading marketplace dedicated to unlocking the power of lab data to improve health outcomes. With direct partnerships across national, specialty, and hospital-based labs, Prognos delivers the most precise, comprehensive, and timely lab data in the industry. By extracting granular insights from structured and unstructured lab sources, Prognos enriches patient profiles, boosting data completeness from 60% to 100% in critical areas like oncology and rare diseases. With market-leading coverage in key conditions such as non-small cell lung cancer (NSCLC), Prognos empowers pharmaceutical and biotech companies to accelerate precision medicine strategies and real-world evidence initiatives. Designed for speed and impact, Prognos can deliver data within hours of ingestion and provide initial datasets within 24 hours of contract signing, transforming how lab data fuels innovation across the healthcare ecosystem.

Special thanks to

Su Huang, Susan Williams, Nate George, Merwin Lau, Arnaub Chatterjee, Travis May, Jason Bhan, Patrick Aysseh, and Bill Paquin for their feedback and input on this article

