QUANTHEALTH

WHITEPAPER

AI-POWERED CLINICAL TRIAL SIMULATION

The Data Difference: Why Category-Leading Clinical Trial Simulation Must Be Built on a Decision-Ready Data Framework

An AI-driven clinical trial simulation platform – built on real-world data, a biomedical knowledge graph, and foundation models – can predict clinical trial outcomes before a single patient is enrolled.

QUANTHEALTH
www.quanthealth.ai · info@quanthealth.ai

EXECUTIVE SUMMARY

In the new era of AI-powered clinical development, a simple truth has become paramount: the quality of a model's predictions is inextricably linked to the quality of its underlying data. As clinical trial simulation emerges as a category-defining technology, pharmaceutical leaders must examine both the strength and the completeness of the data that enables the foundational models.

This white paper argues that in the field of clinical trial simulation — and especially as more biopharma organizations move toward simulation-first clinical development — the data asset is one of the most important drivers of predictive accuracy. At QuantHealth, our reported 85 to 90 percent predictive accuracy is a direct result of our investment in building a large and highly integrated pharmaco-clinical data asset.

The pages that follow explore the essential attributes of a world-class data foundation, detail the composition of QuantHealth's proprietary data asset, and demonstrate how this data difference powers our simulation capabilities and aligns with an evolving, data-driven regulatory landscape.

85–90%	600+	18,430
PREDICTIVE ACCURACY REPORTED	CLINICAL TRIALS IN VALIDATION SET	COMPOUNDS IN OUR KNOWLEDGE GRAPH

1. The Foundational Principle: Garbage In, Garbage Out

For any artificial intelligence or machine learning (AI/ML) system, the adage "garbage in, garbage out" is not just a cliché — it is a fundamental constraint. An AI model, no matter how sophisticated its architecture, can never outgrow poor data quality. Its understanding of the world is bounded by the breadth, depth, and quality of the information it has been shown.

This principle applies with particular force to clinical trial prediction, where the biology is complex, the patient populations are heterogeneous, and the cost of a wrong answer is measured in years and in hundreds of millions of dollars. A model attempting to predict the performance of a drug in a specific patient population must be trained on data that is not only vast, but that also possesses four critical attributes.

Data Attribute	Why It Matters for Simulation Accuracy
Scale and breadth	The model must learn from a large and diverse patient population, spanning ethnicities and standards of care, so that its predictions generalize rather than reflect the biases of a narrow dataset.
Longitudinal depth	The data must capture the full patient journey over time — diagnoses, comorbidities, procedures, and prescriptions — in order to model disease progression and treatment effects realistically. Snapshots in time are insufficient to understand causality.

Data Attribute	Why It Matters for Simulation Accuracy
Multi-modality	The data must integrate different types of information, including clinical data (EHR and claims), trial data, clinical notes, and biological data, to create a holistic, multi-dimensional view of the drug, the disease, and the patient population.
Biological grounding	The model must represent the underlying biological mechanisms linking a drug to a disease, not merely correlate clinical events and outcomes. A knowledge graph mapping the relationships between drugs, targets, and pathways is one of the most powerful tools for this.

Without a data asset that meets these criteria, a clinical trial simulation framework is simply a black box making predictions from incomplete and potentially biased information. The risk of generating a confident but incorrect prediction is unacceptably high — and a confident wrong answer is more costly to a development program than no answer at all.

2. The QuantHealth Data Asset: A Class of Its Own

Recognizing that a superior simulation engine must be built on a superior data foundation, QuantHealth dedicated years to curating a comprehensive pharmaco-clinical data foundation. This is not an off-the-shelf database — it is a purpose-built, proprietary asset designed specifically for the task of clinical trial simulation, and it integrates four distinct classes of evidence.

- **Longitudinal electronic health record (EHR) and claims data:** These provide the deep, real-world context of the patient journey, capturing how patients present, progress, and respond outside the controlled conditions of a trial. This information is triangulated across datasets and curated by a team of medical informatics experts into dense, signal-rich timelines.
- **Millions of clinical notes and rich biomarker sets:** Leveraging the most modern approaches to natural language processing and clinical abstraction to uncover rich patient descriptions, disease states, and biomarker sets.
- **Data from tens of thousands of completed clinical trials:** These provide direct evidence of how drugs have performed in controlled research settings, including the design choices that shaped those results.
- **A proprietary biomedical knowledge graph:** This component maps the relationships between drugs, their pharmacological features, molecular targets, biological pathways, and clinical effects — allowing our models to represent the why behind a treatment effect, not only the what.

Together, these components enable clinical trial simulations with or without sponsor-provided data. This gives biopharma companies a clear, competitive — yet practical — advantage: they can assess likely outcomes in weeks, with high confidence, at a fraction of the cost of running the trial itself.

This multi-modal asset is the training ground for our Large Real-World Drug Model (LRDM), the transformer-based foundational layer through which our models learn biology and its links to clinical manifestations —

capturing the complex, non-linear relationship connecting a trial's design to its probable outcome. The scale and richness of the underlying data are what allow the LRDM to generalize, including to novel drugs, rare diseases, and complex trial designs where conventional approaches have little to work from.

Proprietary Knowledge Graph: Simulation Without Sponsor Data

The QuantHealth Knowledge Graph (QHKG) is a proprietary biomedical knowledge graph designed to support AI-driven prediction of clinical outcomes. It provides a structured representation of compounds, diseases, biological processes, pathways, targets, cells, anatomy, and clinical context across multiple therapeutic modalities.

The graph is organized around a compound-centric view of human biology, enabling detailed modeling of how therapeutic interventions interact with biological systems. By integrating diverse biomedical sources, structural features, and mechanism-specific attributes, it creates a foundation for interpretable and scalable prediction workflows.

To distinguish compounds with similar mechanisms of action (MoA), the knowledge graph captures fine-grained MoA differences, distinct structural and pharmacological features, and detailed chemical and protein structure information.

The resulting graph contains **166,261 nodes** across **27 node types** and **8,106,906 edges** across **129 relation types**, integrated from biomedical data sources. In total, QHKG represents **18,430 approved and experimental compounds** across multiple therapeutic modalities.

3. The Result: How the Data Foundation Translates to Predictive Accuracy

The data difference is not a theoretical advantage. It is the direct, empirical reason for our platform's reported 85 to 90 percent accuracy in predicting trial outcomes — a level of performance observed across a validation set of more than 600 trials. Five capabilities follow directly from the data foundation.

- **Reasoning about novel mechanisms:** The biomedical knowledge graph allows the LRDM to relate a novel mechanism of action to known pathways and targets, supporting credible predictions for first-in-class assets that have no direct precedent.
- **Robust generalizability:** Training on diverse patient populations reduces the biases inherent in some datasets, supporting more reliable predictions across different trial settings and geographies.
- **Indication-specific insight:** Through parameter-efficient fine-tuning, our models gradually learn indication-specific patterns while retaining the general clinical understanding inherited from the foundational LRDM model.
- **Predictive competitive positioning:** By integrating clinical and commercial intelligence into R&D, organizations can make better-informed decisions across portfolio management, trial design, evidence generation, and go-to-market planning — while assets are still in early clinical development.

- **Accuracy in rare disease:** The scale of the data asset means we can often construct robust virtual cohorts even for rare diseases — a task that is difficult or impossible with smaller, less diverse datasets.

Case in Point: AI Simulation Anticipates a Cardiovascular Outcomes Trial

Cardiovascular drug development has among the lowest success rates in the industry: only an estimated 20% of Phase 1 cardiovascular assets are approved by the Food and Drug Administration.[1] Outcomes studies compound the difficulty because they require large populations, long follow-up, and substantial investment — leaving critical decisions under-informed for years at a time.

The QuantHealth data framework fueled an AI-based simulation that predicted the direction, magnitude, subgroup effects, and temporal dynamics of the treatment effects later observed in the VESALIUS-CV trial. The model's trial-level performance was validated against 22 randomized cardiovascular outcomes trials, and its predictions aligned closely with the VESALIUS-CV results subsequently presented at the American Heart Association Scientific Sessions in November 2025 and published in the *New England Journal of Medicine*.[2]

Findings of this kind position AI-driven simulation as a primary tool for probability-of-success assessment, protocol optimization, and portfolio-level decision-making — applied before a single patient is enrolled.

4. A Future-Ready Foundation: Alignment with the Regulatory Landscape

QuantHealth's data framework is increasingly aligned with the direction of regulatory science — a direction that is still actively evolving rather than settled. Since the start of 2025, the FDA has taken a series of concrete steps that expand the role of AI models and real-world evidence in drug development.

- **January 2025:** The FDA issued its first draft guidance on the use of AI to support regulatory decision-making for drugs and biologics, proposing a risk-based credibility assessment framework organized around a model's defined context of use, spanning the nonclinical, clinical, and post-marketing phases.[3]
- **February 2026:** FDA leadership established a single adequate and well-controlled trial plus confirmatory evidence as the default basis for approval — placing greater weight on the quality of supportive evidence, including mechanistic and real-world data.[4]
- **April 2026:** The FDA opened a Request for Information on a pilot to optimize early-phase clinical trials with AI, targeting dose selection, safety monitoring, and early go/no-go decisions.[5]

Taken together, these moves point toward a regulatory environment in which credible, well-characterized evidence from large-scale real-world data and validated models can play a more central role in development and approval decisions. QuantHealth's platform — built on large-scale real-world data, a biomedical knowledge graph, and foundation models — is positioned to generate precisely this class of evidence, provided it meets the fit-for-purpose and credibility standards the FDA is now formalizing.

5. Conclusion: The Unspoken Differentiator

As AI reshapes drug development, the conversation is maturing beyond data toward measurable outcomes: speed to clinical and commercial success, or the ability to fail faster and redirect capital to better assets. However, it remains a critical factor in delivering the accurate predictions. In the emerging category of clinical trial simulation, the platforms that deliver the most reliable predictions will become the trusted standard.

At QuantHealth, our category leadership is a direct result of our commitment to high-quality, robust data curation. It is, however, just one of the reasons we are able to predict trial outcomes with the accuracy we report.

> Learn more about the QuantHealth simulation platform at **www.QuantHealth.ai** or contact us at info@quanthealth.ai.

References

1. Biotechnology Innovation Organization, Informa Pharma Intelligence, QLS Advisors. Clinical Development Success Rates and Contributing Factors 2011–2020. Biotechnology Innovation Organization; 2021. Accessed July 30, 2026. https://go.bio.org/rs/490-EHZ-999/images/ClinicalDevelopmentSuccessRates2011_2020.pdf
2. Bohula EA, Marston NA, Bhatia AK, et al; VESALIUS-CV Investigators. Evolocumab in patients without a previous myocardial infarction or stroke. N Engl J Med. 2026;394(2):117-127. doi:10.1056/NEJMoa2514428
3. US Food and Drug Administration. Considerations for the Use of Artificial Intelligence to Support Regulatory Decision-Making for Drug and Biological Products. Draft guidance. Center for Drug Evaluation and Research; January 2025. Docket No. FDA-2024-D-4689. Accessed July 30, 2026. https://www.fda.gov/regulatory-information/search-fda-guidance-documents/considerations-use-artificial-intelligence-support-regulatory-decision-making-drug-and-biological
4. Prasad V, Makary MA. One pivotal trial, the new default option for FDA approval — ending the two-trial dogma. N Engl J Med. 2026;394(8):815-817. doi:10.1056/NEJMp2601234
5. US Food and Drug Administration. AI-enabled optimization of early-phase clinical trials pilot program; request for information. Fed Regist. 2026;91(82):23100-23104. Published April 29, 2026. Docket No. FDA-2026-N-4390. Accessed July 30, 2026. https://www.federalregister.gov/documents/2026/04/29/2026-08281/ai-enabled-optimization-of-early-phase-clinical-trials-pilot-program-request-for-information