Insight Brief

Digital “Twinning”: How Foundation Models are Transforming Digital Patient Simulation

LUK ARBUCKLE, Global AI Practice Leader and Chief Methodologist, Applied AI Science, AI & Technology Solutions
RAJA SHANKAR, Vice President, Clinical AI Solutions, AI & Technology Solutions
JULIEN HEIDT, Associate Director, Applied AI Science, AI & Technology Solutions
EMILY BRATTON, Senior Director, Applied AI Science, AI & Technology Solutions
NADEA LEAVITT, Senior Principal, Applied AI Science, AI & Technology Solutions
MARTINA PILI, Manager, Clinical AI Solutions, AI & Technology Solutions
JULIA GALLINARO, Senior Consultant, Health Economics Consulting, Commercial Solutions
CHRISTINA MACK, Senior Vice President & Chief Scientific Officer, Applied AI Science, AI & Technology Solutions



Table of contents

Clarifying the core concepts	**1**
External comparators: Borrowing real patients from outside the trial	**3**
Synthetic controls: Modeling outcomes for an external comparator	**4**
Digital twins: Patient-level representations for dynamic inference	**5**
Synthetic twins: Aggregating digital twins into comparator arms	**7**
Synthetic data: Infrastructure, not evidence	**8**
How these approaches fit together	**8**
Modeling digital twins with traditional methods: Strengths and scaling constraints	**10**
Foundation models: Reducing friction, not changing causal requirements	**11**
Conclusion	**13**
References	**14**

Clarifying the core concepts

Clinical development increasingly relies on simulation, prediction, and reuse of existing data to inform decisions before and alongside prospective evidence generation. Advances in large-scale modeling approaches that integrate diverse data sources now make it possible to create richer representations of patient trajectories, diseases and treatments capturing deeper context-based relationships between these concepts. Collectively, these advances support a range of activities across the development lifecycle, from early trial planning through post-approval evidence generation, even though they are often discussed using overlapping terms and labels.

These advances, including large-scale foundation models (i.e., models pretrained on diverse, multimodal datasets that can be adapted across a range of clinical tasks), are contributing to renewed interest in concepts such as digital twins and synthetic comparators, raising a practical question: how do these approaches differ in terms of data, modeling assumptions, and evidentiary use, and where do newer modeling capabilities meaningfully change what can be done versus simply changing how it is implemented? Framing these distinctions clearly is important for how these approaches are evaluated, validated, and ultimately trusted in clinical development.

Collectively, these methods are applied well beyond comparative inference. They are used to inform trial feasibility and design, evaluate endpoint sensitivity, support prognostic modeling and covariate adjustment, enable adaptive and Bayesian frameworks, and facilitate privacy-preserving analyses. In some settings, these methods may also contribute to contextualizing or supplementing clinical evidence, including through external comparator cohorts or model-based comparators, but this represents just one of many ways they support expanded evidence generation. Their value often lies in reducing uncertainty and de-risking decisions upstream, rather than replacing conventional



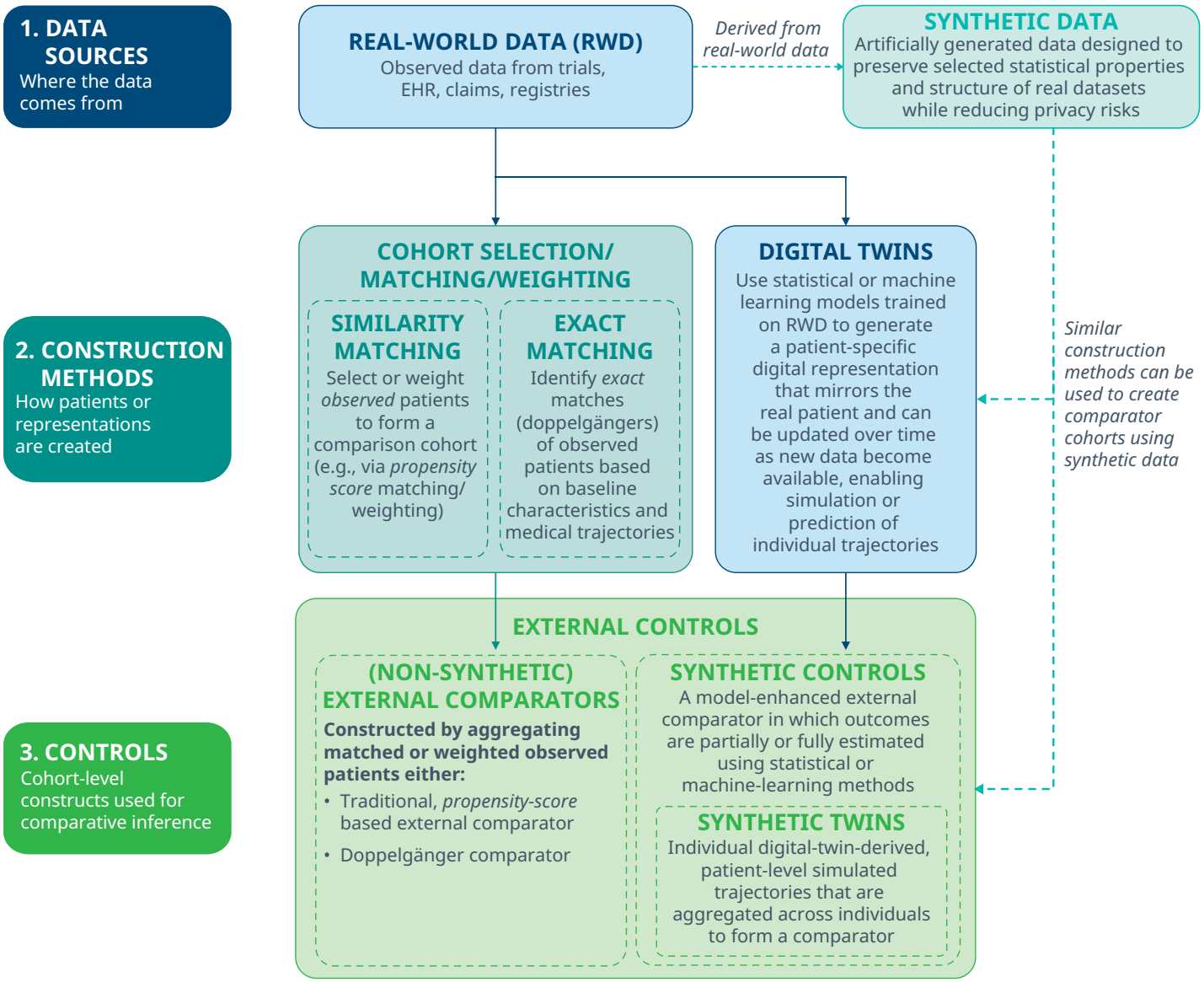
sources of causal evidence, at least in the short term. Over time, as evidence accumulates on the ability of these approaches to predict or represent real-world outcomes, they may, in some settings, reduce reliance on traditional approaches, for example by supporting smaller control arms or augmenting control data, subject to regulatory acceptance.

Randomized placebo or standard-of-care control arms remain the reference standard for causal inference in clinical trials and, where feasible and ethical, continue to anchor regulatory decision making. At the same time, there are well-recognized settings in which traditional internal comparators are constrained, such as small or heterogeneous populations, rapidly evolving standards of care, or operational limitations that affect trial size or informativeness. In these contexts, external comparators, patient-level simulation, and other data-driven approaches can provide critical context and insight when appropriately validated and transparently applied. From a patient perspective, these approaches are often motivated by reducing unnecessary placebo exposure, improving trial feasibility in rare or heterogeneous diseases, and ultimately accelerating clinical development and access to new treatments.

Despite growing adoption, terminology in this space remains inconsistent. Concepts such as *external comparators*, *synthetic controls*, *synthetic data*, and *digital twins* are frequently conflated, obscuring important differences in data provenance, modeling dependence, evidentiary intent, and regulatory implications. This insight brief clarifies these distinctions, explains how the approaches relate to one another, and describes how advances in large-scale modeling are expanding what is possible in digital patient simulation.

These advances enable new types of questions to be explored and new forms of evidence to be generated, while in some settings also reducing operational complexity, without relaxing expectations for rigor, validation, or causal defensibility. Because terminology varies across standards bodies, academia, and industry, we use working definitions and highlight where definitions differ. Our aim is to distinguish the approaches in ways that matter for data provenance, model dependence, and evidentiary use while clarifying tradeoffs and contexts of use (Figure 1).

Figure 1: Conceptual map of comparator approaches and key terms as they relate to data source definitions



Note: "Control" is used in line with clinical trial conventions; "comparator" is more commonly used in Real-World Evidence (RWE) studies

External comparators: Borrowing real patients from outside the trial

When a concurrent randomized control arm is infeasible or insufficient, well-characterized external data can provide context for interpreting treatment effects. An external comparator cohort draws on data from patients who were not enrolled in the trial to serve as a comparator for the treated population. These patients are real individuals, treated in routine clinical practice or in prior studies, with outcomes observed in data collected during the same or an earlier time period. External comparators may be constructed from historical trials or real-world data (RWD) sources and typically require careful cohort definition, time alignment, and endpoint harmonization to support valid comparison.

Analytic methods are used to address differences in the trial population and the external cohort. These may include propensity score-based matching or weighting, outcome modeling approaches such as g-computation, doubly robust estimators that combine both or others.[1] In some settings, external data are used for aggregate-level benchmarking only with limited or no statistical adjustment; such analyses are descriptive and not sufficient for causal inference. Regardless of implementation, the defining feature of an external comparator is its reliance on observed (real) patient data, with causal contrasts estimated under explicit modeling assumptions.[2,3]

Regulatory agencies have long acknowledged the role of external comparators in constrained settings, particularly in rare diseases or when randomization raises ethical or practical challenges.[4] Acceptance depends heavily on whether the external patients are sufficiently comparable to the trial population. This is the exchangeability problem, and it drives most regulatory scrutiny.

Key questions regulators ask include:

- *Are eligibility criteria aligned?*
- *Are baseline prognostic factors measured and balanced?*
- *Is standard of care comparable over time and geography?*
- *Are endpoints defined and measured consistently?*

Importantly, external comparator describes the source of comparator data, not a specific analytic method. External comparators are best understood as an extension of observational real-world evidence (RWE), applied in a trial context, with familiar strengths and limitations. Notably, many regulatory examples sometimes described as “synthetic controls” are, in reality, externally controlled studies based primarily on observed patient data with statistical adjustment rather than fully simulated outcomes.[3] This distinction is important because the degree of model dependence directly affects regulatory risk and expectations.



Synthetic controls: Modeling outcomes for an external comparator

Synthetic controls are a subset of external comparator approaches in which patient-level data from outside of the trial arm are algorithmically combined or reweighted, and in some cases supplemented with modeled or simulated outcomes to construct a virtual comparator aligned to the treated population (Figure 1).[5] This virtual comparator may consist of either individually matched digital counterparts for each treated patient or a synthetic cohort of digital subjects that preserve population level statistics.[6,7] Rather than relying solely on observed outcomes from external patients, outcomes are partially or fully modeled or imputed using statistical or machine learning (ML) methods.

It is important to note that terminology in this space is not uniformly applied. In particular, some methodological literature reserves the term 'synthetic control' for aggregate-level reweighting approaches used in economics and population-level studies and cautions against its use to describe externally controlled clinical study designs.[3] In this brief, we use the term more broadly to describe model-enhanced external comparator approaches, consistent with its evolving use in clinical research.

External datasets may be incomplete, have limited follow-up, or be misaligned with specific trial endpoints. In some settings, synthetic control methods can be used alongside these data to help contextualize trial outcomes by estimating expected trajectories based on observed baseline characteristics and longitudinal patterns.

From a regulatory perspective, this distinction matters. A synthetic control is still external, but it is no longer purely observational. Its credibility depends not only on data provenance and exchangeability, but also on the validity of the modeling assumptions used to generate outcomes.

This is why agencies consistently emphasize[4] that synthetic controls must be:

- Prospectively planned
- Fully specified in the statistical analysis plan
- Accompanied by sensitivity analyses that test alternative assumptions

A USEFUL WAY TO FRAME THE DIFFERENCE IS THIS:

- External comparators ask: "Can we find similar patients outside the trial?"
- Synthetic controls ask: "Given what we know about similar patients, can we reasonably estimate what would have happened?"

Synthetic controls can increase precision and effective sample size, especially when observed outcomes are sparse, but they are not a shortcut around bias. Rather, they move some of that risk from data selection into model specification.

In practice, the boundary between external and synthetic controls is often blurred. Many approaches labeled as synthetic controls are externally controlled comparisons built from real patients, then enhanced through statistical matching, weighting, or modeling. For example, in 2020 the FDA supported Medicenna Therapeutics' use of Medidata's "synthetic control arm" approach in a hybrid design for a Phase 3 recurrent glioblastoma trial, reducing the number of concurrently randomized controls[8]; however, the comparator was constructed from matched patients in historical trial data, not from simulated or imputed individuals. In these cases, the comparator remains grounded in observed outcomes. As a result, many so-called synthetic controls are more accurately viewed as model-enhanced external comparator approaches rather than fully simulated comparators.

Digital twins: Patient-level representations for dynamic inference

In engineering and systems science, a digital twin is defined by an ongoing connection between a physical system and its digital representation, with information flowing between the two to maintain synchronization over time. In this brief, that linkage is treated as a defining feature. Some digital twins may be defined at a single point in time, while others are maintained through ongoing or periodic alignment with the patient's evolving clinical data[9]; what matters is that a defined linkage between the patient and the digital representation can be re-established as needed. Patient-level simulation alone does not make an approach a digital twin unless the digital representation can remain aligned to the patient's evolving clinical state. In the medical evidence generation context, rather than constructing a comparator at the cohort level, digital twins use predictive modeling to simulate the disease trajectory of an individual patient under different scenarios while maintaining a connection to the patient's evolving clinical data.[10] The reliability of these simulated trajectories depends on how well disease progression and treatment-response relationships are captured in the underlying data, with greater uncertainty in settings where these relationships are less well characterized.

Accordingly, digital twins are often framed as generating individual-level trajectories, with uncertainty, that can be revised as new data become available.[9] These trajectories are interpreted in the context of that underlying linkage, which distinguishes digital twins from standalone simulation models. On their own, digital twins are not comparator arms or evidence. They function as an enabling layer, supporting a range of applications, including:

- Simulating counterfactual outcomes for individual trial participants.
- Informing trial design decisions before enrollment begins, including dose selection and enrichment strategies based on predicted treatment response.
- Supporting prospectively specified adjustment using baseline predictions as covariates to increase precision, without updating those predictions during the trial.
- Enabling extrapolation of efficacy and safety to populations not directly studied in clinical trials.
- Supporting individualized clinical decisions, including treatment selection and dosing, by estimating patient-specific benefit–risk profiles.

When digital twin outputs are aggregated across patients, they can be used to construct a comparator arm (Figure 1). This is where digital twins can intersect with synthetic controls. The overlap does not mean the terms are interchangeable. Approaches that generate patient-level predictions without an ongoing link to evolving patient data are better described as patient-level synthesis or prediction, not necessarily as a digital twin. Stated differently, digital twins are fundamentally an individualized simulation framework, whereas synthetic controls are fundamentally a comparative inference framework. The two can overlap, but they are not synonymous.

The modeling approaches used to implement digital twins vary. Some are based on mechanistic or systems pharmacology frameworks that explicitly represent biological processes, while others rely on statistical or ML methods trained on observed data. Between these poles, structured knowledge representations such as ontologies or knowledge graphs can be used to constrain predictions and improve clinical plausibility without requiring a fully mechanistic model. Hybrid approaches that combine mechanistic structure with data-driven components are also increasingly explored.[11] In this brief, we focus primarily on data-driven implementations as they relate to patient-level simulation using real-world and clinical data, while noting that the broader concept of digital twins is not tied to a single modeling paradigm.

Regulatory experience with digital twins remains limited, and the examples discussed in this space are heterogeneous. Some methods described publicly alongside digital twins are more accurately characterized as prospectively specified prognostic adjustment or model-informed augmentation, rather than continuously updating digital twins. Key considerations for evidentiary use include clearly defining the context of use, validating model performance, and demonstrating that simulated trajectories behave plausibly across relevant patient subgroups.



Synthetic twins: Aggregating digital twins into comparator arms

A synthetic twin refers to a digital-twin-derived, patient-level simulated comparator that is aggregated across individuals to form a comparator (Figure 1). Conceptually, synthetic twins sit at the intersection of digital twins and synthetic controls. The simulation occurs at the individual level, using patient-specific models that generate counterfactual trajectories, and those simulated outcomes are then combined to produce cohort-level estimates. This distinction is important because not all digital twins become synthetic controls. A digital twin used solely for prognosis, enrichment, or clinical decision support remains a patient-level simulation tool. It becomes a synthetic comparator only when aggregated and used to estimate treatment effects relative to observed outcomes.

From a regulatory perspective, synthetic twins attract heightened scrutiny because they combine two sources of uncertainty: reliance on external data and dependence on model-generated outcomes. In effect, they inherit the exchangeability concerns of external comparators and the assumption sensitivity of synthetic control methods, while adding the complexity of individualized simulation. As a result, their use is generally limited to well-defined contexts with prospective specification, transparent model documentation, and extensive sensitivity analyses.

The FDA's draft guidance on Bayesian methodology in clinical trials[12] provides the formal statistical framework for incorporating externally derived information into trial analyses through informative priors, including requirements for quantifying the degree of borrowing, managing prior-data conflict through dynamic discounting, and evaluating robustness through sensitivity analyses. These are considerations that apply directly when synthetic twin outputs are used to augment concurrent controls.



Synthetic data: Infrastructure, not evidence

Synthetic data are often mentioned alongside digital twins, but they serve a different role and should not be conflated with patient simulation.

Synthetic data are artificial data that aim to preserve certain underlying statistical properties and structure of original datasets and may support de-identification to reduce re-identification risk, depending on the generation method and disclosure controls used.[13,14] Their primary purpose is to enable data sharing and analytic development with privacy protection as an enabling constraint.

In practice, synthetic data are useful for:

- Early feasibility assessments
- Method development and testing
- Collaboration across institutions where patient-level data cannot be shared

Synthetic datasets can be designed for infrastructure and development rather than patient-faithful simulation. Even when synthetic data are generated at the patient level, they still need a fit-for-purpose validation before being used in settings that depend on preserving causal structure for regulatory inference. They should not be positioned as substitutes for external comparators, synthetic controls, or digital twins in submissions intended to support efficacy or safety claims.

A helpful way to think about synthetic data is scaffolding. They support the development and validation of analytic approaches, but they are not themselves evidence.

How these approaches fit together

Regulatory positioning across these four approaches varies considerably by maturity and intended use. Table 1 below provides an evidence-class mapping.

Synthetic data supports feasibility and method development, not efficacy or safety claims



Artificial data
- Preserve statistical properties
- Not typically patient level

In practice
- Early feasibility assessments
- Method development, testing

Generative
- From imputation to AI
- With or without privacy protection

Not evidence
- Overlapping AI/ML methods
- Do not support efficacy or safety

Example: Test a cohort definition, analysis code, and power calculation on a synthetic claims dataset before requesting access to the real patient-level data

Table 1: Evidentiary mapping of digital comparator approaches by regulatory evidence class, maturity, and primary agency focus

APPROACH	REGULATORY EVIDENCE CLASS	MATURITY	PRIMARY REGULATORY FOCUS
External comparators	Observational RWE	High	Exchangeability, bias control, a priori SAP specification
Synthetic controls	Observational RWE (constructed)	Moderate–High	Robustness of modeling assumptions, sensitivity diagnostics
Digital twins (patient-level models)	Model-informed evidence	Emerging	Validation framework, context of use, interpretability
Synthetic twins	External comparator + modeling	Emerging/high scrutiny	Transparency, causal defensibility, audit trails

These approaches address different challenges in clinical development and may be used together when traditional comparators are constrained. They differ primarily in where comparator data comes from and how much modeling is involved. In practice, they support a range of recurring use cases across development.

De-risking trial design upstream

Digital twins can be used before trials begin. By simulating individual patient trajectories under different assumptions, they inform eligibility criteria, endpoint selection, follow-up duration, and expected event rates. This application operates entirely at the design stage and does not contribute to the evidentiary dataset used for regulatory evaluation.

Augmenting and contextualizing limited or single-arm evidence

When enrollment is limited or randomization is infeasible, external comparators can supplement small, randomized controls or provide benchmarks for single-arm studies using well-matched real-world or historical patients. The goal is to strengthen interpretation of treatment effects, not to eliminate empirical comparison.

Synthetic control methods can improve precision or interpretability when follow-up is incomplete, endpoints are sparse, or patient heterogeneity is high. This includes approaches that incorporate modeled or simulated outcomes, such as those derived from patient-level simulation frameworks (e.g., digital twins) when aggregated to form a comparator. External comparators derived from fit-for-purpose data and appropriate study designs may support causal comparisons, while more model-dependent approaches can be used to contextualize findings, assess plausibility, or augment other sources of comparative evidence.

Supporting adaptive and hybrid designs

In contrast to upstream design applications, model-based predictions can also be incorporated prospectively into trial execution and analysis without altering the randomized comparison. In this setting, the prognostic score is fixed before treatment effects are observed and is used as a covariate rather than as an updating estimate of the patient's post-randomization journey. Prognostic adjustment, implemented through covariate adjustment using baseline risk predictions, reduces variance and improves precision of treatment effect estimates without compromising randomization. Predicted outcomes may also inform predefined interim analyses by providing expectations for event accrual, response heterogeneity, or subgroup performance. In these settings, model-based predictions guide interpretation and decision-making but do not replace observed outcomes or serve as primary evidence.

As these applications move from borrowing observed external data toward simulating counterfactual outcomes at the individual patient level, the analytical demands increase. Digital twins, in particular, require models that can represent disease trajectories over time and explore how those trajectories might differ under alternative treatments. Over time, this need has pushed development teams toward increasingly sophisticated modeling approaches to support individualized prediction and simulation.

Modeling digital twins with traditional methods: Strengths and scaling constraints

ML methods have been used for disease-trajectory modeling for decades. Approaches such as joint longitudinal-survival models, multi-state models, PK/PD frameworks, and Bayesian hierarchical methods demonstrate that individual-level outcome prediction is feasible and can be statistically rigorous.

In practice, however, applying these approaches to high-dimensional, heterogeneous data settings can require substantial effort.[15] As data sources expand and diversify, model development often involves non-trivial data curation steps, including coding reconciliation, unit normalization, and missing-data handling. Many traditional implementations have primarily relied on structured clinical or electronic health record (EHR) data, and integrating modalities such as imaging, genomics, or irregular longitudinal follow-up may introduce additional complexity.

More fundamentally, many established supervised learning approaches are designed to optimize prediction of observed outcomes under observed treatments. Extending these models to generate counterfactual predictions, which are central to digital twin applications, may require additional assumptions and modeling choices to support causal interpretation. In this context, the distinction between predictive accuracy and causal validity remains an active area of methodological development.

At the same time, these considerations are evolving rapidly. Advances in automated ML, multimodal modeling, and agentic frameworks are beginning to streamline aspects of model selection, training, and deployment, with the potential to reduce some of the practical barriers described above. As such capabilities mature, the balance between bespoke model development and scalable implementation is likely to continue shifting.

As a result, building a credible ML-based digital twin for a given indication may still involve meaningful upfront investment, particularly in aligning data, modeling assumptions, and intended use.



Foundation models: Reducing friction, not changing causal requirements

Foundation models are discussed in connection with digital twins because they can reduce some of the practical barriers to building and maintaining patient-level simulation models at scale. These models are first trained on large and diverse healthcare datasets, such as EHR, clinical notes and medical images, which enables them to learn general patterns about how patient data are structured and how clinical events unfold over time. They can then be adapted to specific tasks, allowing the same underlying model to be reused across different applications.[16,17,18]

The relevance to digital twins and synthetic controls lies in the possibility of improving scalability and reuse. Repeatedly rebuilding data pipelines, reconciling coding systems, and retraining bespoke models constrain the practical use of digital twins beyond narrow, one-off applications. Foundation models may support greater reuse across indications and data sources, although the extent of efficiency gains may be context dependent.

Traditional ML and statistical methods remain fit-for-purpose in many contexts (e.g., single-task predictions with well-curated, domain-specific training data). Foundation models are best understood as complementary capabilities: particularly valuable in data-rich, multimodal, or longitudinal settings where reuse and scalability are important. In practice, foundation models may be combined with traditional statistical, causal, or mechanistic methods, rather than replacing them.

Reduced data harmonization burden

Training models on large, diverse clinical datasets allows them to better handle variations in coding, timing, and gaps in records. This can reduce the amount of repetitive, study-specific data curation typically required, while preserving the need for rigorous validation and data quality assessment.

Consistency of endpoint definitions

Pretrained models can help align different data sources when identifying clinical endpoints, by learning shared structure across differences in coding, timing, and follow-up. This can improve robustness under irregular data. However, they do not fix poorly defined outcomes: ambiguity in endpoint definition or derivation, including imputation choices, can materially affect results and may be amplified by models. Clear endpoint definitions, transparent derivation logic, and sensitivity analyses therefore remain essential.

Multimodal integration

Foundation models can combine information from multiple sources, such as laboratory values, clinical notes, imaging features, genomics, etc., into a single view of the patient. This enables richer patient representations that reflect how clinically relevant information is distributed across data types.

Generalization across indications

Traditional ML models remain effective within specific, well-defined contexts. However, adapting a pre-trained foundation model to a new disease area can require less data and development effort than building a new bespoke model. This improves efficiency for organizations managing multiple development programs without displacing existing modeling approaches.

At the same time, the ability to generalize across indications depends on how well disease processes and treatment-response relationships are captured in the underlying data. In well-characterized settings with relatively stable and understood patterns of progression, model transfer and adaptation may be more reliable. In contrast, rare or less well-understood diseases may introduce greater uncertainty, particularly where relevant heterogeneity is incompletely observed. Foundation models trained on large and diverse datasets may help mitigate some of these limitations by improving representation of heterogeneous patient trajectories, although this remains an active area of evaluation.

Latent patient representations

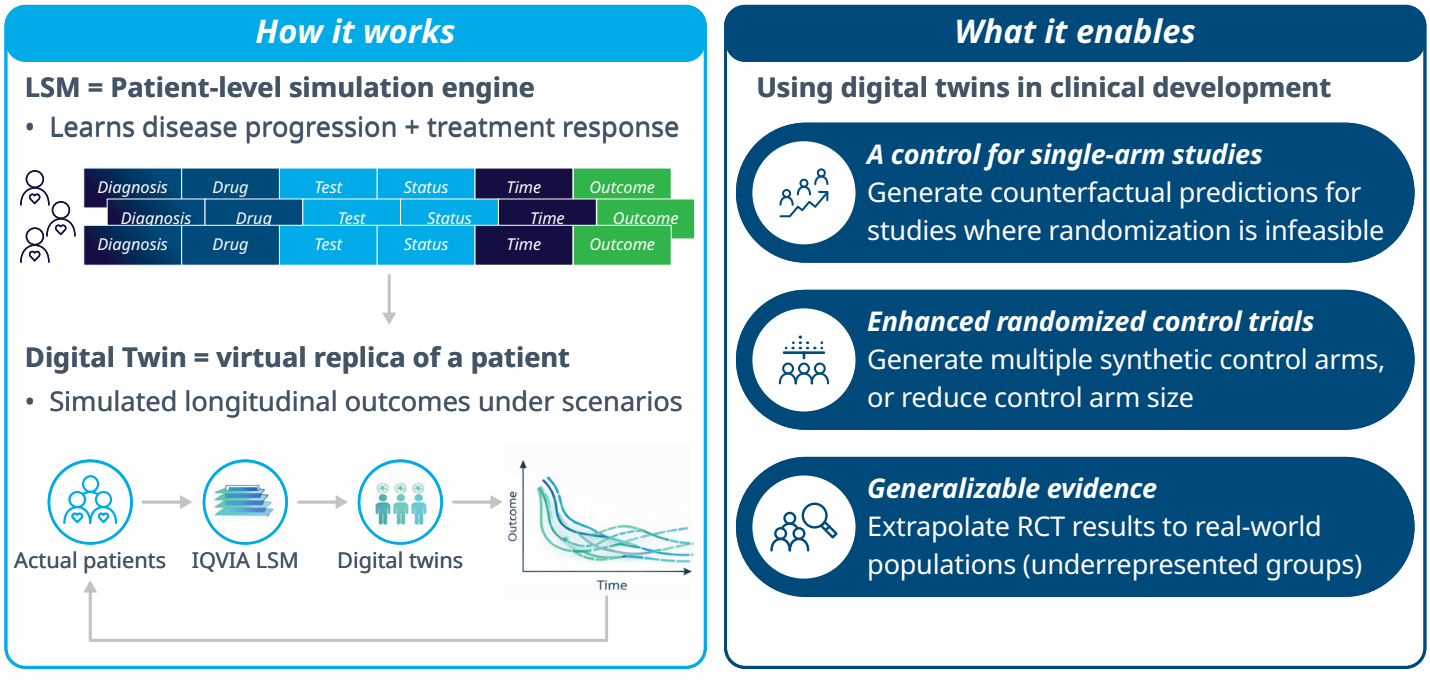
Foundation models allow representing each patient as a compact numerical profile — often called an "embedding" — that summarizes many aspects of their history (diagnoses, treatments, timing of events, and information from multiple data sources). These profiles may make it easier to identify similar patients by comparing overall patterns rather than matching on a limited set of predefined variables. This can improve patient matching for applications such as synthetic controls, while also helping preserve relationships across different data types. These approaches are best seen as complementing, rather than replacing, traditional covariate-based methods.

Improved counterfactual pathways

Emerging work on causally structured foundation models aims to better simulate how a patient's trajectory might change under different treatment scenarios by incorporating causal principles into model design. While this is a promising direction, it does not remove the need to make and clearly state causal assumptions. Rather, they shift where those assumptions are encoded and how transparently they must be documented.

Taken together, these features suggest that foundation models may serve as an enabling component for digital twins, particularly where multimodal data, reuse, or scale are important. By reducing efforts required to integrate heterogeneous data, maintain models across indications and generate individualized trajectories, they could make it more feasible to deploy digital twin frameworks beyond narrow, one-off applications. In practice, this can support broader exploration of disease mechanisms, more systematic evaluation of trial design choices, and more consistent application of simulation approaches across programs. However, this increased scalability does not change the underlying evidentiary requirements: the validity of digital twin outputs remains dependent on data quality, appropriate model specification, and explicit causal assumptions.

Twins enable controls, leaner RCTs, generalizable evidence





Conclusion

External comparators, synthetic controls, digital twins, synthetic twins, and synthetic data are distinct approaches with different purposes, assumptions, and regulatory considerations. In practice, their value lies in strengthening and contextualizing conventional control strategies within a broader evidence framework. As the methodological landscape continues to expand, maintaining clarity on these distinctions becomes increasingly important.

The emergence of foundation models expands what is possible for individualized simulation and trial design and introduces new approaches to patient-level prediction, synthetic cohort generation, and control arm construction that are still being defined. Real progress will depend on clearly distinguishing these approaches and understanding how they can be used together to improve evidence generation.

Progress will also depend on how well models are constrained. Purely data-driven approaches can produce predictions that are statistically valid but clinically implausible when key relationships are missing or underrepresented in the data. Knowledge graphs and ontologies provide structured representations of biological, clinical, and mechanistic relationships, which can be used to impose consistency constraints, guide feature relationships, and restrict outputs to biologically and clinically plausible states. In practice, this helps reduce unrealistic predictions and improves the credibility and interpretability of digital twin outputs.

At the same time, effective, real-world implementation is rarely driven by a single model type. Emerging agentic architectures show how multiple models and data sources can be orchestrated to deliver practical solutions. In many cases, effective digital twin approaches may rely on coordinated ensembles of models, with foundation models as one component alongside other ML approaches.

Ultimately, advances in tooling expand what is possible but do not reduce the need for rigor. Validation, plausibility, and fit-for-purpose regulatory strategy remain essential. IQVIA is actively advancing this space by bringing expertise in healthcare data, ontologies, privacy-preserving methods, and applied AI to develop scalable, credible approaches that align with regulatory expectations. This systems-level perspective will be central to translating digital twins from methodological innovation into reliable components of evidence generation.

References

1. Hernán, M. A., & Robins, J. M. (n.d.). Causal inference: What if. https://miguelhernan.org/whatifbook

2. Burcu, M., Dreyer, N. A., Franklin, J. M., et al. (2020). Real-world evidence to support regulatory decision-making for medicines: Considerations for external control arms. Pharmacoepidemiology and Drug Safety, 29(10), 1228–1235. https://doi.org/10.1002/pds.4975

3. Rippin, G., Largent, J., Hoogendoorn, W. E., Sanz, H., Bosco, J., & Mack, C. (2024). External comparator cohort studies: Clarification of terminology. Frontiers in Drug Safety and Regulation

4. U.S. Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research, Center for Biologics Evaluation and Research, & Oncology Center of Excellence. (2023, February). Considerations for the design and conduct of externally controlled trials for drug and biological products (Draft guidance). https://www.fda.gov/media/164960/download

5. Medidata. (2021). Synthetic control arms white paper. https://www.medidata.com/wp-content/uploads/2021/09/SCA-Whitepaper.pdf

6. Björnsson, B., Borrebaeck, C., Elander, N., et al. (2020). Digital twins to personalize medicine. Genome Medicine, 12(1), 4. https://doi.org/10.1186/s13073-019-0701-3

7. Eadie, A.L., Lynch, H.F., Scheinerman, N. et al. The arrival of digital twins and in silico trials in drug development. Nat Med 32, 1967–1971 (2026). https://doi.org/10.1038/s41591-026-04344-3

8. Medidata. (n.d.). Medicenna gains precedent-setting FDA approval to design a hybrid synthetic control arm for a phase 3 registrational trial. https://www.medidata.com/en/medicenna-gains-precedent-setting-fda-approval-todesign-a-hybrid-synthetic-control-arm-for-a-phase-3registrational-trial/

9. Katsoulakis, E., Wang, Q., Wu, H., et al. (2024). Digital twins for health: A scoping review. npj Digital Medicine. https://www.nature.com/articles/s41746-024-01073-0

10. Akbarialiabad, H., Pasdar, A., Murrell, D. F., Mostafavi, M., Shakil, F., Safaee, E., Leachman, S. A., Haghighi, A., Tarbox, M., Bunick, C. G., & Grada, A. (2025). Enhancing randomized clinical trials with digital twins. npj Systems Biology and Applications, 11(1), Article 110. https://doi.org/10.1038/s41540-025-00592-0

11. Chen, Y., Yang, O., Sampat, C., Bhalode, P., Ramachandran, R., & Ierapetritou, M. (2020). Digital twins in pharmaceutical and biopharmaceutical manufacturing: A literature review. Processes, 8(9), Article 1088. https://doi.org/10.3390/pr8091088

12. U.S. Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research, & Center for Biologics Evaluation and Research. (2026, January). Use of Bayesian methodology in clinical trials of drug and biological products (Draft guidance). https://www.fda.gov/media/190505/download

13. El Emam, K., Mosquera, L., & Bass, J. (2020). Evaluating identity disclosure risk in fully synthetic health data: Model development and validation. Journal of Medical Internet Research, 22(11), e23139. https://www.jmir.org/2020/11/e23139/

14. Duke University. (2025). Synthetic data generation using generative AI: Policy brief. https://healthpolicy.duke.edu/sites/default/files/2025-06/Synthetic%20Data%20Generation%20Using%20Generative%20AI_Policy%20Briefv2_0.pdf

15. Wiegrebe, S., Piller, J., Gorski, M., Behr, M., Küchenhoff, H., Heid, I. M., & Bender, A. (2025, September 24). Multi-state models for modeling disease histories based on longitudinal data [Preprint]. arXiv. https://doi.org/10.48550/arXiv.2509.19956

16. Moor, M., Banerjee, O., Abad, Z. S. H., et al. (2023). Foundation models for generalist medical artificial intelligence. Nature, 616(7956), 259–265. https://doi.org/10.1038/s41586-023-05881-4

17. Singhal, K., Azizi, S., Tu, T., et al. (2023). Large language models encode clinical knowledge. Nature, 620(7972), 172–180. https://doi.org/10.1038/s41586-023-06291-2

18. Khan, W., Leem, S., See, K. B., Wong, J. K., Zhang, S., & Fang, R. (2024, June 15). A comprehensive survey of foundation models in medicine [Preprint]. arXiv. https://doi.org/10.48550/arXiv.2406.10729

CONTACT US

iqvia.com

© 2026. All rights reserved. IQVIA® is a registered trademark of IQVIA Inc. in the United States, the European Union, and various other countries.
07.2026.ATS. 26-CRS-2106-06JUN