https://doi.org/10.1038/s41587-025-02847-x

Applications of digital twins in medicine

Research efforts to develop digital twins in medicine are rapidly increasing, with promising emerging applications in oncology, diabetes management and cardiovascular medicine. While medical digital twins hold great promise for personalized healthcare, their implementation is no easy feat. The field faces diverse challenges including collecting data, choosing computational model designs, ensuring safety and efficacy, and preventing biases. What can be learned from the successes achieved? Which are the most promising upcoming applications? We asked experts in the field for their thoughts.

Reinhard Laubenbacher

Mikael Benson at the Karolinska Institute in Sweden is the coordinator of Swedish Digital Twin Consortium; *Tina Hernandez-Boussard* is associate dean of research at Stanford University; *Reinhard Laubenbacher* is a professor at the University of Florida; *Mihaela van der Schaar* is John Humphrey Plummer Professor of Machine Learning, AI and Medicine and director of the Cambridge Center for AI in Medicine at the University of Cambridge, UK.

What is a digital twin, and how does it differ from other computational models?

Reinhard Laubenbacher: The digital twin (DT) concept was pioneered by NASA decades ago in an engineering context and is widespread in industry and other application areas, such as urban planning. As the term implies, there is a pair of objects involved. There is a physical object, such as an engine, a corn field, a city or a patient. Typically, there is an objective associated with this physical object, such as determining preventive maintenance tasks, implementing optimal watering schedules at different times of the year, or determining the optimal amount of insulin to administer to a patient with type I diabetes during different activities. These objectives are to be accomplished with the help of the DT. This DT is based on a computational model that captures features of the physical object relevant to the objective at hand. This model is calibrated with data that characterize the physical object and capture its state as it evolves over time. For instance, a mathematical model of glucose metabolism could be calibrated to an individual patient by including, on the one hand, physiological and lifestyle data, such as age, body composition and activity schedules and, on the other hand, continuous glucose measurements from a subcutaneous sensor. The DT consists of the model, calibrated dynamically to the patient, and the two evolve together over time. This illustrates the main difference between a DT and a computational model. The computational model is generic and the DT is specific. The concept has been described authoritatively in a 2023 report by the National Academies of Engineering, Science, and Medicine (NASEM)[1].

Mihaela van der Schaar: In medicine, DTs are designed to simulate the evolving state of individual patients, not by predicting a single outcome, but by generating a range of plausible future scenarios. This enables clinicians (and, potentially, patients themselves) to explore how a patient's condition could unfold under different clinical decisions, treatments or interventions. DTs provide dynamic, individualized foresight that supports proactive and informed decision-making. This generative capability enables DTs to support understanding of complex patient journeys given different interventions (that is, treatments, life-style changes etc.); for instance, simulating how a patient's cardiac function might respond to different levels of physical activity and therapy adjustments. Furthermore, DTs provide a safe and rapid environment to test clinical hypotheses, such as assessing the potential risks of discontinuing a treatment for a specific patient given their diseases. They also enable foresight in planning interventions, by simulating and comparing outcomes of clinical strategies, such as determining the optimal timing for surgery in patients with frailty, considering multiple possible trajectories of recovery or deterioration.

The generative capacity of DT models is grounded in a fusion of data-driven learning and explicit domain knowledge, which ensures that their simulations are not only plausible but also (clinically or scientifically) meaningful and explainable. Crucially, they must remain 'alive', continuously updating as variables, treatments and clinical understanding evolve. Traditional mechanistic models and machine learning approaches struggle to meet this demand, but new methodologies for building digital twins are capable to deal with these challenges.

Mikael Benson: While many definitions of DTs have been proposed, none is generally accepted. Reasons include the wide variety of potential applications and emerging technologies. Thus, it is likely that definitions will be adapted to different contexts and change continually.

The Swedish Digital Twin Consortium uses a broad definition of medical DTs: virtual representations of healthy or sick processes across life cycles that can be understood, learned and reasoned with using real-time data or simulation models to predict, prevent, or treat diseases. Those DTs can span from whole populations to individuals. DTs should ideally be dynamic representations of health or disease trajectories that encompass whole lifetimes. Such trajectories can be defined in whole populations on the basis of electronic medical records from hundreds of millions of patients. On a population-scale, DTs of disease trajectories may be used to identify and prevent novel disease-causing environmental factors. On the scale of an individual, population-based trajectories in

combination with detailed molecular data may be used to find out if an individual follows a specific trajectory. The next step would be to prevent that trajectory as early as possible, perhaps even in childhood. That prevention could entail AI-based monitoring of multi-scale data for early diagnosis, treatment or life style changes.

The difference compared to other computational methods is context-dependent integration of any method that is optimal for early diagnosis, prevention or treatment of a DT. Such promiscuous integration is similar to how clinicians work: we use any combination of methods to reach diagnostic and therapeutic decisions depending on what is optimal for a given patient.

What is often misunderstood about digital twins?

MvdS: The belief that DTs are fixed models — constructed once and then applied repeatedly without the need for further refinement. A core requirement of an effective DT is its ability to evolve continuously alongside the real-world system it represents. Traditional modeling approaches are often ill-equipped for this task. They are typically designed with fixed input structures and require extensive re-engineering or retraining when new variables or knowledge need to be incorporated. This rigidity hampers their relevance in medicine. Without adaptive capacity, a model risks becoming obsolete the moment the clinical context shifts.

Another common misunderstanding is the tendency to equate DTs with high-fidelity simulations — detailed, static replicas of physical systems that exist merely to mirror reality. While high fidelity is valuable, a DT must go beyond passive simulation. The defining characteristic of a DT is its generative capability — its ability to simulate multiple plausible future trajectories of a dynamical system responding to different interventions in real time. This dynamic, foresight-driven and intervention-aware capacity is what sets DTs apart from conventional, static simulations.

It is also frequently and incorrectly assumed that DTs function as synthetic data generators — tools designed to produce artificial datasets for training machine learning models. DTs are not about fabricating data samples in bulk.

DTs are also fundamentally distinct from 'world models' used in reinforcement learning or robotics, which generalize over abstract environments or fixed tasks. A DT is instance-specific: it models a particular patient, organ, biological system, clinical trial or healthcare process.

Mihaela van der Schaar.

RL: The most common confusion is about the difference between a generic model and a personalized model. Cancer research frequently uses computational models of signaling pathways in cells that allow the simulation of changes introduced through mutations associated with malignant changes. For instance, epidermal growth factor (EGF) binds to its receptor, triggering downstream pathways, such as the MAPK–ERK pathway, which promotes controlled cell division. Dysregulation of EGF function can lead to uncontrolled cell proliferation, one of the hallmarks of cancer. Understanding such general mechanisms can help search for drugs that target such dysregulation. There is no need to personalize the model to an individual patient to understand this relationship. What might be necessary is to determine whether a given patient has a particular mutation to know whether a given drug might be effective. On the other hand, to decide whether a chemotherapy regimen is effective for a given patient, it might well be important to make predictions based on a DT of the patient, calibrated with imaging and other data taken from the patient. One instance of this is the cancer DT program in the Yankeelov laboratory at UT Austin.

What are the most successful examples of digital twins in medicine to date? What are the most promising medical applications of digital twins in the future, and would they outperform AI and other computational models?

RL: One example that comes closest to the general definition laid out in the NASEM report[1] is the artificial pancreas, a US Food and Drug Administration (FDA)-approved device that largely automates insulin injection for patients with type I diabetes. It is based on a mathematical model of human glucose metabolism, calibrated to the characteristics of a particular patient. A subcutaneous sensor provides near-streaming data about glucose levels, and a closed-loop controller integrated with the mathematical model determines the level of insulin required at any given time, which is then injected by an attached insulin pump[2]. This is an example of a DT based on a mechanistic model of a relevant part of human biology.

An example for which the DT is based on data-driven machine learning models is the work in the Trayanova laboratory at Johns Hopkins University, which builds DTs of heart function related to arrythmias and other conditions[3]. There is a wide range of promising applications of DT technology leveraging a good understanding of the relevant human biology, such as in biomechanics, or in data-rich areas, such as neuroscience.

Tina Hernandez-Boussard: The Barcelona Supercomputing Center (BSC) is pushing the boundaries of DTs in cardiovascular medicine and in other areas. They have the computing power to run the necessary simulations, which are at the core of the DT. The BSC has 'Alya Red', which is a high-resolution DT of the human heart comprising around 100 million simulated cells, each governed by ~50 mathematical equations. This detailed model, running on the MareNostrum facility, captures mechanical and electrical heart dynamics. BSC researchers have used Alya Red to simulate scenarios like pacemaker placement and arrhythmia treatment, enabling virtual trials of device interventions.

MvdS: Some of the most impactful applications of DTs in medicine have emerged in cardiology. Cardiac DTs simulate the electrophysiological behavior of individual hearts, allowing clinicians to model arrhythmias and virtually test interventions such as ablation or pacemaker implantation. These simulations provide a level of personalization that substantially improves procedural planning and patient outcomes.

In oncology, DTs have been used to model tumor growth and treatment response, supporting oncologists in simulating how a tumor might evolve under different chemotherapy, radiotherapy or immunotherapy regimens. Our HDTwin project[4] has demonstrated how

these models can be used to personalize therapy plans.

However, most existing computational methods for building DTs remain rigid, often requiring extensive manual re-engineering to incorporate new data, biomarkers or treatments. Their architectures are typically fixed at design time, which limits their adaptability and scalability in a rapidly evolving clinical environment. The next generation of DTs will overcome these limitations by being adaptive. Our CALM-DT[5] (Context-Adaptive Language Model-based Digital Twin) exemplifies this shift. CALM-DT can incorporate new data, variables and clinical knowledge on the fly, without retraining or architectural redesigns. This adaptability will enable DTs to evolve in real time with their physical counterparts, making them truly 'living models'.

Mikael Benson.

What types of patient data are needed to train digital twins? What are the main challenges in obtaining these data?

THB: DTs require multimodal, longitudinal patient data to enable accurate, personalized simulations. Key data types include clinical data (electronic health records, lab results, treatments), omics data (genomics, proteomics), imaging and biosensor data (radiology, wearables), behavioral and social determinants (lifestyle, socioeconomic factors, barriers to care) and environmental exposures (geographic location, pollution). Critically, these data must represent diverse populations — not just 'convenient' cohorts — to avoid creating a flawed 'average virtual patient' that fails to reflect real-world variability.

RL: The availability of data in sufficient quantity and quality is of crucial importance and is a challenge for practically all DT applications. In addition, many things cannot be measured in patients or cannot be used for training models on a large scale. The specific requirements depend on the intended application. Treating sepsis in the intensive care unit, providing decision support during heart surgery, optimizing chemotherapy or keeping a healthy person healthy all happen on very different timescales. Most applications work best when dynamically calibrated with longitudinal data, such as periodically repeated measurements of cytokine levels in the blood of a patient with COVID to monitor the strength of the immune response.

The biggest challenge is the lack of noninvasive sensing technology and imaging modalities. While information about a patient's genome is helpful in some applications, like cancer therapy, it must be supplemented by other patient information. The number of patients for whom genomic information is available is growing rapidly, but so is our knowledge about all the different factors that influence a patient's response to a condition or treatment. This makes clustering or correlation models less effective since the size of the clusters is decreasing rapidly.

MB: Because most diseases involve thousands of genes across multiple cell types and organs, multi-omics data at the level of spatial single cells are needed. Moreover, as disease-associated changes differ between stages of the same disease, longitudinal data are needed. This is almost impossible for most diseases because they can develop over years or decades before symptoms and diagnosis. An alternative solution can be to infer molecular disease trajectories based on analyses of different stages of the same disease in the same organ, as we recently showed in a study of prostate cancer[6]. Population-wide studies of longitudinal electronic medical records needed to characterize such disease-trajectories are available in the United States, Scandinavia and an increasing number of other countries.

MvdS: While structured data such as electronic health records, lab results, medication histories, imaging studies, and physiological signals from wearables form the backbone of DT models, unstructured data — clinical notes, radiology reports, pathology findings and patient-reported outcomes — are equally important as they often contain nuanced contextual information. DTs also rely on external knowledge sources such as clinical guidelines, biomedical literature, expert-defined mechanistic models and real-world evidence datasets. These sources provide foundational knowledge about disease mechanisms, treatment pathways and causal relationships that may not be directly observable in patient data.

However, assembling and integrating this diverse data landscape poses considerable challenges. Many patients, especially those with rare diseases or complex multimorbidity, have fragmented and incomplete data histories. Data may be inconsistently recorded across different care settings, with varying levels of granularity and quality. Even for common conditions, longitudinal datasets often contain gaps and irregular sampling intervals and are missing key variables, which limit the robustness of traditional modeling approaches.

Another critical challenge is the integration of multimodal data sources into a coherent, unified DT framework. Structured and unstructured data, clinical guidelines and evolving scientific knowledge need to be harmonized in a way that preserves their complex interdependencies. This becomes increasingly difficult when trying to model not just a single organ or disease, but the interactions across systems and comorbidities that define real-world patient trajectories.

Data privacy and governance constraints complicate the large-scale collection and sharing of patient data. Regulations rightly protect patient confidentiality but create barriers to aggregating data across institutions and populations, often limiting the diversity and richness of training datasets.

Fortunately, recent advances in data-centric AI are making these challenges more tractable. Techniques for harmonizing heterogeneous data, imputing missing values, aligning multimodal streams and extracting structured insights from unstructured text are increasingly robust. These advances enable the development of DTs even when data are sparse, fragmented or incomplete, making it feasible to build useful models in real-world clinical settings where ideal datasets are rarely available.

What are the main computational challenges of building a digital twin for medical applications?

RL: This depends on how the DT is to be used and what type of computational model it depends on. A DT for drug development, for

instance, will generally not require real-time computational results, as opposed to decision support for stabilizing the blood pressure of a patient with sepsis, which must be updated on a time scale of minutes. DTs that are based on AI-models will often require more easily computable surrogate models for deployment in a clinical setting. Likewise, complex mechanistic models that require repeated numerical solution of high-dimensional equations – for instance, to simulate blood flow in the heart – depend on appropriate surrogate models.

One challenge facing medical DTs is that many aspects of human biology – for instance, the immune system – cannot easily be described by physics-based models. This requires different modeling paradigms, ranging from AI models to Gaussian processes, various Bayesian learning algorithms, agent-based models, all the way to systems of ordinary or partial differential equations. These all have differing data requirements and computational complexity.

MvdS: One fundamental challenge lies in managing evolving modeling environments. For a DT to remain relevant and useful, it must be able to seamlessly integrate new variables, data and knowledge, a task that remains one of the core computational hurdles.

Another critical challenge is the need to handle flexible dimensionality in state–action representations. Patients differ in their biological makeup and in the types and amounts of data available about them. A DT must be capable of simulating trajectories across a wide range of input and output dimensionalities, accommodating situations where new variables are added or certain data points are missing.

A third computational challenge revolves around the integration of domain knowledge with data-driven insights. Medical decision-making is informed by a combination of mechanistic understanding (for example, physiological laws, pharmacokinetics) and empirical data patterns derived from clinical observations. DTs must bridge these two worlds, embedding explicit expert knowledge while leveraging machine learning to capture complex, data-driven dynamics. This integration is not trivial, as domain knowledge is often unstructured or incomplete and blending it effectively with data-driven components requires hybrid modeling approaches.

This challenge is compounded by the fact that biomedical systems are more complex and less well understood than engineered systems, where DTs have traditionally thrived. In engineering domains, the physical laws

Tina Hernandez-Boussard.

and system behaviors are often precisely defined and measurable; in medicine, many biological mechanisms remain only partially understood, variable across individuals or completely unknown. This makes it significantly harder to model patient trajectories accurately and increases the burden on computational methods to accommodate uncertainty, incomplete knowledge and biological heterogeneity.

Computational scalability is another concern, particularly for DT architectures that rely on large language models. While large language models offer a powerful mechanism for flexible, context-adaptive simulations, they are constrained by finite context windows during inference. Managing how information is selected and supplied to the model, especially in complex simulations with vast state-action spaces, is a non-trivial computational task.

Finally, DTs designed for clinical decision support must be equipped to handle uncertainty in the scenarios they generate. A DT must help clinicians understand which scenarios are more likely, which are edge cases, and where significant uncertainty remains. Developing robust mechanisms to generate, prioritize and interpret these possible futures remains a major computational frontier.

What are the advantages and disadvantages of the different computational approaches under development (for example, data-driven versus mechanistic)?

RL: In industry, almost all DTs are based on physics-based mechanistic models. If we want to use DTs for control, having mechanisms available to identify optimal control inputs is crucial. In medicine, too, most problems we would like to solve are control problems, identifying optimal interventions to maintain or alter a patient's health trajectory. So, basing DTs on mechanistic models is desirable. The artificial pancreas is an example[2].

In many cases, we don't know mechanisms or don't have good models for them. In other cases, stochastic features determine outcomes so statistical models might be more appropriate. Ultimately, humans are complex systems and likely cannot be described by mechanistic models. Missing mechanistic knowledge can be in part ameliorated by correlations derived from large patient cohorts with similar characteristics. Thus, DTs that are based on models that combine available mechanistic information with AI/machine learning models will likely be the most common technology.

THB: Data-driven and mechanistic models each have distinct strengths and limitations. Data-driven methods, particularly deep learning, excel at identifying complex patterns in high-dimensional data (for example, genomics, radiology), enabling rapid insights from large datasets. However, they often lack interpretability, require massive training datasets and struggle with generalizability to under-represented populations. In contrast, mechanistic models incorporate established biological principles, offering transparency and robustness in scenarios with sparse data, but they may oversimplify disease complexity and fail to capture nonlinear interactions evident in real-world patient data.

A critical challenge and opportunity lies in multimodal data fusion, which integrates diverse clinical data (molecular, imaging, clinical, etc.) to overcome the limitations of single-modality analysis. While deep learning has advanced single-modality processing (for example, tumor detection in radiology), progress in integrating modalities (for example, linking genomics to treatment response) remains slow due to technical and methodological hurdles. For example, tumors exhibit heterogeneity that no single data type can fully characterize; effective fusion could align histopathology with spatial transcriptomics to reveal actionable insights. However, key obstacles include data sparsity (missing modalities for certain patients), lack of standardization (varied formats and scales), and interpretability gaps (for example, 'black box' fusion models). Federated learning and attention-based neural networks are promising tools to

address these issues, but clinical validation remains a bottleneck.

The future of DTs hinges on hybrid approaches that combine data-driven and mechanistic modeling, leveraging the scalability of AI with the interpretability of domain knowledge.

Given the uncertainty inherent in digital twin models, what safety concerns should be considered when using them in patient care?

RL: In most cases, DTs can be viewed as medical devices, expert systems providing decision support, or computational tools used in drug development. The FDA and similar agencies around the world already regulate these. The approval standards will need to be updated to encompass this new type of computational model. The FDA has long been a leader in working with the modeling community to update standards. There are new aspects of patient privacy and ethics that need to be considered. If we envision a future in which each patient has a DT designed for them at birth that evolves with the person over time, then an important question is who owns a person's DT, who has access, where does it reside, and how are forecasts of a patient's health to be acted on or not.

THB: The clinical use of DTs requires careful consideration of uncertainty and fairness to ensure patient safety. DTs must quantify uncertainty for each individual patient. This includes accounting for data limitations (for example, missing or sparse patient records), model variability and gaps in biological understanding. Without clear uncertainty metrics — such as prediction confidence intervals or reliability scores — clinicians may over-trust recommendations, potentially leading to harmful decisions. For example, a twin might predict a treatment's success without revealing whether that estimate is based on robust data or speculative assumptions. To mitigate this, models must be transparent, such as having an associated model card.

A critical challenge is ensuring fairness at the individual level. Traditional fairness checks evaluate whether a model performs equally for different populations, but DTs require deeper scrutiny because biases can emerge in highly personalized predictions. For instance, a twin might underestimate risks for a low-income patient with complex comorbidities if its training data primarily reflects healthier, wealthier populations. To address this, models must be stress-tested on diverse edge cases, and fairness should be measured by consistency in outcomes for similar individuals — not group averages.

Finally, operational and ethical frameworks must evolve alongside the technology. Clinicians may develop overreliance on DTs, trusting predictions even when uncertainty is high, and liability questions remain unresolved if a model's recommendation causes harm. Additionally, patients must be informed about the limitations and evolving nature of these tools, requiring dynamic consent processes. To ensure DTs enhance rather than compromise care, regulators and developers must prioritize transparency, ongoing validation and equitable access.

Should digital twins always be fit-for-purpose, or is there potential benefit in building multi-scale, comprehensive models?

MvdS: DTs should be fit-for-purpose — but with an eye toward composability and, where appropriate, integration into broader multiscale frameworks. In clinical settings, the most immediate value comes from DTs that are designed to address specific, well-defined problems. A twin built to evaluate the short-term effects of adjusting medication in heart failure, for example, should focus on relevant cardiovascular variables and the time horizon required for safe and actionable decision-making. Building large, all-encompassing models in such scenarios would not only introduce unnecessary complexity but could also increase the risk of overfitting, reduce interpretability and delay deployment.

That said, there is substantial long-term potential in modular, composable DTs that can be extended or connected to other models to achieve multiscale or system-wide insights. For example, a DT of a patient's immune system may need to interact with a tumor model to simulate immunotherapy outcomes, or a twin of a clinical trial cohort may need to integrate individual patient twins to understand population-level dynamics. Similarly, models of individual organs may need to interoperate within broader physiological systems to simulate cross-organ effects.

What is essential is flexibility. The ability to build fit-for-purpose DTs for immediate clinical use should not preclude the possibility of expanding or linking those twins when a broader perspective is needed. Architectures like CALM-DT make this possible by allowing the twin to operate in environments of varying complexity, ingesting new data or knowledge without architectural redesign.

Moreover, not all scenarios require the same scale, resolution or data depth. The right level of modeling depends on the specific question, the available data and the decisions that need to be supported. In some cases, a focused twin simulating the dynamics of a tumor in response to radiotherapy may be sufficient. In others — such as planning organ transplantation or managing multisystem disease — there is value in integrating across time scales, biological layers or patient cohorts.

RL: The authors of the NASEM report[1] emphasize the requirement to be fit for purpose. This implies that the DT should not capture parts of human biology and spatial and temporal scales not pertinent to the application. For instance, if the application is focused on optimal use of a respirator for a patient, then the immune response in the lungs is likely not relevant.

But there are intriguing possibilities that arise if we envision a 'full' DT of a patient, in all its complexity. For instance, we could then simulate perturbations, such as prolonged exposure to pollution or chronic high stress levels, and could simulate the effects on the patient over extended time spans in all its ramifications, expected and unexpected — a true systems approach to human health. While this may seem far-fetched, there are attempts in this direction. The Virtual Physiological Human Institute (https://www.vph-institute.org/) in Europe has as its vision a comprehensive physiological model of a human. The National Institutes of Health recently held a grant competition for a Whole Person Research and Coordination Center (https://www.nccih.nih.gov/grants/whole-person-research-and-coordination-center). This center will be integral in coordinating the initial and future research programs and networks in whole person health research. Part of the center's activities will be to begin the construction of a comprehensive computational model of a whole person.

Several proof-of-principle studies have demonstrated the feasibility of digital twins to model cardiovascular disease or cancer. What are the main barriers to broad clinical application of such models, and what evidence thresholds must be met for their adoption?

MB: Many examples of DT models of cardiovascular disease are mainly based on imaging methods. While they can model physiological processes, they lack cellular and molecular resolution. At the recent AACR

conference in Chicago, the clinical relevance of imaging-based DTs was questioned for this reason.

The main barriers to clinical implementation of DT models are integration of imaging and high-resolution molecular data into DTs and computational treatment of those DTs to find optimal drugs for the individual patient. Evidence for their adoption will require clinical trials showing their superiority compared to existing clinical methods.

RL: The evidence threshold is clear: a DT must do better than the standard of care for the condition in question. This will need to be established through clinical trials and FDA approval. If thresholds are met, physicians will be eager to use DTs for better patient care, and insurers will hopefully be eager to pay for the additional costs their use might incur. For DTs that are focused on preserving health rather than restoring it, the challenge is more complex. But the broad adoption of a range of commercial devices such as smart watches with health apps suggests that there will be a broad base of people who are willing to serve as patients in trials, whether clinical or commercial.

I see the main barrier to broad clinical application in the state of the technology infrastructure. There is no common scientific, computational or medical infrastructure to support extensive development of the tools needed for academic laboratories and companies to work together on R&D projects without continually reinventing the wheel. So, the challenges are similar to those of the green energy industry 20 years ago. Without stable and sustained funding for startup partnerships between academia and industry, this technology may not get off the ground for some time. The experience of the solar industry in the United States, where relatively little state support was provided, which allowed companies to fail repeatedly before being successful, and China, where extensive support was provided, is very instructive in this regard.

THB: Broader adoption requires addressing four practical and evidence-based hurdles. First, trust must be established through rigorous validation. Clinicians will demand evidence that twins outperform current risk models. Patients, meanwhile, need transparency about how predictions are generated, how uncertainty is communicated and how their data will be used and shared. Second, compute power remains a bottleneck for real-time use; while cloud-based solutions can help, safety-critical applications may require local deployment, raising costs. Third, clinical integration poses technical and cultural challenges — electronic health record interoperability is limited, and clinicians resist disruptive workflow changes. Finally, added burden is a deal-breaker: if DTs require manual data entry or excessive review, adoption will falter.

What industry standards are necessary to validate digital twins in medicine?

MB: The same principles for clinical trials that have already been applied to test drug treatments for tumors based on genome sequencing. In other words, compare DT-based drug predictions with routine treatment protocols. Since DTs can predict response to thousands of drugs, clinical trials may be divided into two steps. First, test if DTs can predict response to standard treatments better than existing methods. Second, if yes, test drug predictions in patients that do not respond to standard drugs, either based on biomarkers derived from the first step or repeated construction of high-resolution DTs to identify non-responders to standard treatments.

RL: An important lesson we can learn from industry is that standards are absolute key to the success of a technology that is to be applied in a wide range of healthcare settings, and that depends on the integration of results from biomedicine, computational modeling, high performance computation, real-time data collection under uncertainty, and humans in the loop in many essential roles. In all these fields, very few standards exist. The systems biology community, for instance, has worked on establishing standards for mechanistic mathematical models for decades and is just now starting to make substantial inroads. For AI models there are virtually no widely accepted standards available. Data collection is not standardized across hospitals and sometimes not even within hospital systems.

Mathematical standards for the assessment of DTs, their specifications, comparison and performance measurements are largely missing. For instance, there is no agreed-upon method to compare two stochastic models of the same biology and assess whether they agree.

Should the field aim to build digital twins that represent patient subgroups within a given disease indication?

RL: In most cases, this will be the only feasible approach for some time to come. Two different patients, with different genomes, ages and body compositions, can respond in ways to a given health perturbation that are indistinguishable to us. And even if we could tell them apart, the treatment modalities we have available might not be able to account for these differences. In many cases, therefore, there would be no benefit to go to a resolution beyond subgroups of patients.

THB: Absolutely, particularly for underserved populations. By focusing on specific subgroups — whether defined by clinical, demographic or socioeconomic factors — the DT can uncover insights that would be lost in population-wide averages. For example, a diabetes twin tailored to Federally Qualified Health Center populations could reveal how food insecurity or transportation barriers impact medication adherence, enabling targeted interventions. Similarly, subgroup twins could help optimize resource allocation by predicting how many bilingual providers are needed in specific communities or how clinic staffing affects chronic disease outcomes.

The key is strategic prioritization: developing subgroup twins where heterogeneity most impacts outcomes and where the insights will drive meaningful action. High-impact applications might include maternal health models for racial and ethnic minority groups or hypertension twins for rural populations with limited specialist access. While resource-intensive, these targeted models can deliver more equitable care than one-size-fits-all approaches. Crucially, they must be developed in partnership with the communities they aim to serve, ensuring the models capture real-world barriers and needs rather than researcher assumptions.

What can the biomedical field learn from the implementation of digital twins in other areas, such as in industry (for example, manufacturing) or earth system science?

RL: One lesson to learn from industrial DT projects is the order of magnitude of funding required to create a medical DT science and technology base that is scalable and can support a medical DT industry. Available programs for medical DT research are several orders of magnitude too small if the goal is a viable technology in the next 20 years.

Another lesson to be learned from engineering, climate science, weather forecasting, city planning, agriculture and a plethora of other applications, is that there are formidable

technical, computational and scientific challenges that must be solved in a truly transdisciplinary approach. We have made only small progress in the last 30 years in bringing together quantitative scientists and life scientists, as most biological research is still not integrated with computational approaches. For DT research and development, we need to add clinicians to the teams. I see the team-building challenge as possibly the most difficult one to meet.

THB: The field can gain valuable insights from DTs in other complex systems, moving beyond traditional industrial examples like aviation to learn from adaptive systems such as urban resilience models and ecological simulations. While manufacturing twins excel in precision engineering (for example, predictive maintenance for jet engines) and smart city twins optimize dynamic systems (for example, traffic flows or disaster response), biomedicine must integrate these strengths with lessons from ecosystems like coral reefs, which model stressor interactions and collapse thresholds. For instance, our proposed work with a county health department will use a community-scale DT to simulate extreme weather responses, such as optimizing cooling center placements or power grid shutoffs during extreme heat events, demonstrating how population health can benefit from the real-time adaptability of urban twins and the network thinking of ecological models.

However, translating these lessons to biomedicine requires navigating unique challenges, including stricter regulatory frameworks, fragmented data ecosystems and higher ethical stakes. Industrial and urban twins often operate with standardized, high-frequency data (for example, Internet of Things sensors in smart cities) while healthcare struggles with siloed electronic health records and privacy constraints. Moreover, unlike a manufacturing twin's fail-safe simulations, medical errors can have irreversible consequences. The path forward lies in selectively adapting cross-disciplinary strategies: embracing modularity for scalable organ-specific models, leveraging ecological principles to study interconnected health determinants, and adopting cities' real-time monitoring capabilities. By curating the best of these approaches, biomedicine can build DTs that are not just technologically advanced, but also socially accountable.

Looking ahead, what do you think is feasible in the next five years and what is unlikely to be realized soon?

MB: Within the next five years DTs will be implemented for individualized drug treatment of serious diseases that require costly drugs and health care. A strategy for this has already been published by us and is ready for clinical trials[7]. General implementation of DTs to predict and prevent whole disease trajectories in individuals will be based on studies during the next five years but implemented after that for suitable trajectories.

RL: With the right resources, a lot could be accomplished in five years. Prototypes of DTs could be built, a major R&D program to support a medical DT industry could be set up, and partnerships between academia, the government and industry could be established to everybody's benefit. There is a lot of excitement in all three of these domains. I am most familiar with the academic sector. Every conference I go to has numerous talks about different aspects of DTs and their challenges. The field is ready and only requires the right incentives. For instance, as one example, the EDITH Virtual Human Twin (https://www.edith-csa.eu/) project has just concluded with a strategic roadmap for the field, with a large cast of stakeholders, funded by the European Commission.

THB: In the near future, disease-specific DTs (for example, for diabetes management, oncology treatment planning) will gain traction in well-resourced health systems. These will focus on discrete clinical decisions — like optimizing chemotherapy regimens or insulin dosing — with measurable outcomes.

Comprehensive multi-organ simulations (for example, a full human physiome) remain scientifically and computationally impractical due to unresolved biological complexity and data integration hurdles.

Despite progress, resource-intensive twins will predominantly benefit wealthy health systems, with low-resource settings relying on pared-down versions or being excluded entirely.

Are there any other issues you would like to raise?

RL: The Scottish physician Archibald Pitcairne, in a sequence of publications beginning in 1688, proposed a mathematical explanation of circulation, a *Principia medicinae*, that mirrored Newton's *Principia Mathematica*, published the year before, with the beating heart playing the role of gravity. This is in its essence surprisingly modern: medicine today makes heavy use of chemistry, biology and physics, which, in turn, rely profoundly on mathematical sciences. Coupled with data derived from technological advances in engineering and information technology, we find ourselves in a better position than Pitcairne to develop such a mathematical 'theory' of health and disease, with the prospect of biomedical 'laws' and practical mathematical tools for the researcher and the clinician alike. This body of mathematical theory forms the underpinning of digital twin technology that will, in many ways, be the culmination of Pitcairne's vision. And this is the right time. A confluence of high-performance computation, new mathematical and computational advances, very promising new sensor technologies and unprecedented advances in imaging can help us find new and better ways to meet the ever-growing healthcare needs of humankind.

Interviewed by Iris Marchal

Published online: 14 October 2025

References

1. Willcox, K., Chung, C., Kinter, J., Qualters, I. & Segundo, B. *Foundational Research Gaps and Future Directions for Digital Twins* (National Academies Press, 2023); https://doi.org/10.17226/26894
2. Cobelli, C. & Kovatchev, B. *J. Diabetes Sci. Technol.* https://doi.org/10.1177/19322968231195081 (2023).
3. Roney, C. H. et al. *Circ. Arrhythm. Electrophysiol.* **15**, e010253 (2022).
4. Holt, S., Liu, T. & van der Schaar, M. *Adv. Neural Inf. Process. Syst.* **37**, 72170–72218 (2024).
5. Amad, H., Astorga, N. & van der Schaar, M. Preprint at *arXiv* https://doi.org/10.48550/arXiv.2506.12091 (2025).
6. Smelik, M. et al. *Cancer Res.* **85**, 2514–2526 (2025).
7. Schäfer, S. et al. *Genome Med.* **16**, 42 (2024).