Federal Register Notice: 89 FR 51554, Federal Register :: Networking and Information Technology Research and Development Request for Information on Digital Twins Research and Development, June 18, 2024.

Request for Information on the National Digital Twins R&D Strategic Plan

Gary An

Regarding foundational challenges facing the development Biomedical Digital Twins

Gary An, MD, FACS
Department of Surgery
University of Vermont Larner College of Medicine

The primary challenge facing biomedical research

The most impactful and currently intractable challenge for translational biomedical research is the ability to reliably predict the effect of a potential pharmacological intervention (a drug or vaccine with a specific presumed mechanism of action) in the whole-person context for an individual patient. This gap is why the drug development pipeline is so inefficient, prone to failure at the most expensive (both in terms of money and human cost) phase, that of clinical trials, and, consequently, impacting the eventual cost of approved drugs[1]. Addressing this challenge requires:

1. Knowing how the target system works
2. Knowing the origins (mechanistically) of variability between individual instances of the target system and
3. Recognizing that disease and health are inherently dynamic processes that can change the responsiveness of the system to interventions over time.

I assert that addressing these three requirements can only be achieved through the development of biomedical digital twins that are compliant with the NASEM definition referred to in this Request for Information.

The case for compliance with the NASEM Definition specifically regarding Biomedical Digital Twins

Biomedical applications of digital twin technology face a contradiction between the messaging of the promise of digital twins and the reality of the current methodological readiness to deliver on that promise. The impact of this contradiction is accentuated by:

1. The fact that the term “twin” has its intuitive appeal due to the inherently biological origin of the term; upon hearing the term one immediately thinks of and pictures identical human twins (notably, not fraternal ones…).
2. The fact that biological systems are unable to meet the basic preconditions present for successful applications of digital twin technology, notably the lack of first-principles-based trustworthy computational specifications for the digital object and inability to characterize the ground truth of the behavior of the real world twin, namely in of describing the heterogeneity of the real-world population in terms of a “true” probability distribution; this latter fact due to perpetual and intractable data sparsity relative to the number of potential configurations of the real world system.

Because of the intuitive and reflexive lay interpretation to the term “twin”, there is a pervasive danger that projects not in compliance with the NASEM definition that are nonetheless portrayed as “medical digital twins” will invariably fall far short of public expectations by not delivering on what actual compliant digital twins could. This situation has three negative consequences:

1. Early public dissatisfaction with the concept of “medical digital twins” due to the implementations falling far short of the expected (and often promised) benefits. For example, “medical digital twins” are often portrayed as being able to personalize therapies for an individual, optimizing treatments by predicting their effects through execution via the “medical digital twin.” The relative dearth of effective therapeutics for many diseases functionally converts this idea of “having personalized treatments for me” to “this is the best we can do with the set of therapies we currently have.” This is not to say that there isn’t a benefit from being able to do the best we can at the moment (as a clinician I can appreciate this need), but it is a significant downgrade from the expectation a patient might have regarding a promised benefit inherent in the term.
2. Diversion of resources away from investigators actually interested in providing the expected capabilities present in the NASEM compliant digital twin paradigm by diluting such resources to support projects that by their very nature may be more readily implemented (but with an impact far short of what could be achieved with a “true” medical digital twin) at the exclusion of addressing the very real fundamental challenges that face the development, deployment and evaluation of NASEM compliant biomedical digital twins.
3. Limits the cross-disciplinary lessons learned from digital twin research in other domains that are compliant with the NASEM Definition. More explicitly, if the biomedical community assents to a diluted definition of a “medical digital twin” then the transferability of developments from fields that have true digital twins will be severely compromised in terms of a shared vocabulary, necessary preconditions of the system being twinned, and the nature of the challenges associated with digital twin development. Current examples of this disconnect are: 1) the lack of trustworthy computational specifications, 2) the lack of physical assets to provide the ongoing data and 3) inadequate identification of the “ground truth” present in the real world that allows for rigorous Validation and Uncertainty Quantification of the underlying computational specification. All three of these areas are open research and development questions regarding biomedical digital twins.

We believe that the NASEM definition of a digital twin describes unique capabilities that distinguish potential biomedical digital twins from other biomedical computational approaches, such as personalized predictive models, virtual cohorts and *in silico* clinical trials. These other methods, while demonstratively useful, do not incorporate the ongoing updating of the digital object with data from the real-world twin, and therefore are limited in achieving the goal of “true precision medicine.” The goal of medicine is to provide the right intervention(drug) at the right time for the right patient, and the goal of biomedical research is to provide this capability for every potential patient. We have

previously presented Axioms of True Precision Medicine that explicitly note a fundamental description of these desired features[2]:

- Axiom 1: Patient A is not the same as Patient B (Personalization)
- Axiom 2: Patient A at Time X is not the same as Patient A at Time Y (Precision)
- Axiom 3: The goal of medicine is to treat, prognosis is not enough (Treatment)
- Axiom 4: Precision medicine should find effective therapies for every patient and not only identify groups of patients that respond to a particular regimen (Inclusiveness)

The NASEM Report definition of a digital twin with a fit for purpose that includes control discovery matches directly the goals represented in the Axioms of True Precision Medicine listed above.

The case for cellular-molecular mechanism-based NASEM-compliant biomedical digital twins, and challenges that need to be addressed

I assert the following to be true:

- Assertion 1: The primary means by which we seek to treat disease is through drugs.
- Assertion 2: Drugs function by molecular mechanisms that affect the behavior of cells, that result in changes in tissue/organ function that manifest systemically as disease and health.
- *Conclusion 1: Rational design of drugs requires knowledge and representation of relevant cellular and molecular pathways.*
- Assertion 3: Human beings have essentially the same functional cellular and molecular structure.
- Assertion 4: Human beings differ in the exact functional responsiveness of their shared cellular and molecular structure.
- *Conclusion 2: Human beings share a common specification that can be potentially rendered computationally, but individuals represent specific instantiations (e.g. parameterizations) of that common specification.*
- Assertion 5: It is currently (and likely perpetually) impossible to comprehensively characterize the entirety of the cellular and molecular features that make up the human species.
- *Conclusion 3: Any implemented computational specification of a human in perpetually epistemically incomplete.*
- Assertion 6: Even given this limitation of knowledge, the number of possible configurations of the known cellular and molecular features cannot be characterized by sampling the population (Curse of Dimensionality).
- *Conclusion 4: The ground truth of the real-world population distribution in terms of cellular/molecular features is perpetually uncertain.*

Conclusions 1 and 2 point to the need to develop cellular-molecular-based biomedical digital twins in order to meet the promise of “a treatment tailored to you” inherent to the label “medical digital twin.” Conclusions 3 and 4 point to fundamental challenges inherent to the development of such biomedical digital twins, and the difference

between the biomedical area and other areas in which digital twins are being used/developed, namely that these latter areas have trustworthy computational specifications that can be subjected to “classical” verification, validation and uncertainty quantification. Conversely, biology requires a reconceptualizing of “validation” and “uncertainty quantification”. For example, regarding uncertainty quantification, physics-based models may be concerned with numerical error propagation as the equation these models solve is directly reflective of the underlying reality (e.g., a molecular dynamics simulation based on electronic configurations of the simulated molecules, while the uncertainty in knowledge-based cell-level mechanistic models comes primarily from both the epistemic uncertainty as to the causal and hierarchical mechanisms that govern the system behavior and the unquantifiable natural stochasticity and variation inherent to biological processes. We assert that models that represent cellular and molecular biology to generate tissue/organism/individual output have unique properties that call for a readjustment of what validation and uncertainty quantification mean, and the support of research on novel methods of characterizing such systems. The challenges that need to be addressed include, but are not limited to:

1. The transition of existing statistical methods that establish validity and uncertainty quantification at the population level to novel methods that characterize and evaluate individuals.
2. Generally, transitioning population-level statistical methods to novel methods that define individual trajectories.
3. Methods that move beyond traditional conformal prediction to account for model incompleteness/uncertainty and uncharacterized stochasticity.
4. Methods for evaluating potential control modalities for such systems.
5. The need to engage developers of physical assets such that the bidirectional data links can be established at the appropriate level of granularity given the representation level of the computational models.

I acknowledge that this list of challenges is not novel or unique; I make this comment to add emphasis to the importance of these issues to a specific class of potential biomedical digital twins that can both take full advantage of the capabilities inherent to the NASEM definition and meet the public expectation from the lay interpretation of the term “digital twin.”

References:

1. An, G. Specialty Grand Challenge: What it Will Take to Cross the Valley of Death: Translational Systems Biology, “True” Precision Medicine, Medical Digital Twins, Artificial Intelligence and In Silico Clinical Trials. *Front. Syst. Biol.* 2022 April 28 2:901159. https://doi.org/10.3389/fsysb.2022.901159
2. An G, Cockrell C, Day J. Therapeutics as control: Model-based control discovery for sepsis. Complex Systems and Computational Biology Approaches to Acute Inflammation: A Framework for Model-based Precision Medicine. 2021:71-96.