The goal of precision medicine is to provide tailored therapy to each patient with considering therapeutic benefits and risks. Strides toward this goal have been made by harnessing the benefits of big data, the development of mathematical and data-based computational models, and the use of artificial intelligence (AI) and machine learning (ML) algorithms (1-3). The currently available mathematical models lack fidelity, are unable to represent changes in real-time, and are far from being the optimal tools to provide robust predictive enrichment that can be of use in clinical medicine. In their article, Liu et al. (4) performed an exhaustive bibliometric analysis of the currently available applications, specifically in the arena of individualized diagnosis and treatment of the critically ill patients. There has been an increasing number of articles published over the past decade, however, the clinical relevance and real-world application remains debatable. To circumvent these issues, the construction of digital twin models based on research data and physiological properties has been proposed.

**What is a “digital twin” and what is its utility in healthcare?**

A digital twin is a scientifically sound concept that creates an “in-silico” model, a computerized replica, of a patient and their physiology that can be used in either a clinical or research setting. In other words, a digital twin is an exact replica of a patient at a baseline clinical state or an exact match at baseline created for each subject in clinical trials. These virtual models, or digital twins, are designed and validated using real-life patient data and a conceptual understanding of physiology to simulate a continuum of scenarios or interventions without putting real patients at risk (5-7). Another attempt at defining a digital twin would be “a living model of physical asset or system, which continually adapts to operational changes based on the collected online data and information, and can forecast the future of the corresponding physical counterpart” (8).

Digital twins have been widely utilized in healthcare scenarios ranging from studies of ventricular electrophysiology to modeling disease progression in stroke and multiple sclerosis patients (9-11). Outside of the world of medicine and healthcare, the concept of the digital twin has been increasingly pervasive and has been extensively used all the way from modeling of smart cities, anomaly detection in the automobile industry to the design of electrical systems (12-14). It is quintessential to distinguish the concept of “digital twin” from simulation. While both the processes utilize digital platforms to replicate a process or a system, the “digital twin” platforms create a virtual environment
with a modifiable degree of fidelity for a richer study. In addition to this, whereas a simulation conventionally studies one process, a digital twin platform can run multiple intertwined processes. The level of complexity does not end there, while the simulations do not benefit from having real-time access to data and feedback loop, the digital twin platforms have the advantage of a “two-way” flow of information (from digital twin to end-user and from end-user to the digital twin to improve on the next iteration of modeling). In-silico modeling pairs individual patient’s physical artifacts with digital models, reflecting patient status in real-time. These models provide a unique opportunity to better calculate and assess risk without endangering actual patients. We would not have to perform a real-life evaluation of the risk of death or a major traumatic injury in participants jumping from an aircraft with and without a parachute (15).

While AI holds much promise in structure data (imaging, electrocardiography, robotics-assisted surgery), there will be different challenges for healthcare organizations, providers, and patients in the adoption and integration of AI in healthcare (16-18). The degree of trust between clinicians and the AI system is a vital but difficult to quantify determinant of how AI will be used and adopted in the clinic. Trust in AI systems can be improved by increasing transparency, ensuring robustness, and encouraging fairness in the design of the AI system, but a healthy amount of skepticism should not be discouraged (19). AI can help address disparities by expanding care to underserved populations, but AI may also act to reinforce disparities in disproportionately disadvantaged and underrepresented groups by training the algorithms on a pre-selected group of patients which may not be representative of a diverse real world patient population (20). Applications of ML have progressed rapidly as highlighted in the article by Liu et al., but many clinicians still lack experience in properly understanding and interacting with these models (4,21). Proper regulation of software as medical devices (SaMDs) will be essential to ensure the safe, reliable and fair integration of AI across clinical settings that clinicians and patients can trust.

Over the past decades, multiple digital tools have been developed using AI to aid in clinical decision making and in research. In research, AI has been used to generate large scale synthetic data to train other ML algorithms (22). Patient generated health data, including patient reported outcomes, treatment histories, and biometric data are also increasingly being integrated into the fields of research and clinical care (23).

The Introduction of Internet of things (IoT) devices such as sensors, softwares, and AI/ML algorithms are not without risk and there are regulatory issues regarding safety, efficacy, and quality. Unexpected and unpredictable results from the AI algorithms in healthcare rattle clinicians’ confidence in these new technologies. The failure of IBM's Watson is a living example of when a computer algorithm was unable to provide expected reproducible results and outcomes (24). Major healthcare center partnered with IBM to develop an advisory tool to be used in clinical practice. While the tool performed well in the training environment, its performance in the clinical setting lacked promise (25). The US Food and Drug Administration (FDA) in their position paper in 2019 proposed a regulatory framework for modifications to the AI or ML based software (26). This was proposed with a vision to appropriately regulate the tailored total produce lifecycle-based oversight on these softwares by labeling them as “medical devices”.

This effort has led to gaining consensus from the major stakeholders about the good ML practice (GMLP). In addition to the harmonization of GMLP, including data management, feature extraction, training, evaluation, and documentation, the oversight for these activities was also emphasized. In 2011, the FDA identified eight priority areas where “new or enhanced engagement” at the regulatory science front was found to be quintessential. Subsequently, in 2013, a ninth priority area was also added. The regulatory issues surrounding the development and quality control of SaMDs (digital twins and AI/ML algorithms) could overlap with more than one priority areas identified by the FDA such as priority area 4, “Ensure FDA readiness to evaluate innovative emerging technologies” and priority area 3, “support new approaches to improve product manufacturing and quality and potentially others”.

Since this field of SaMDs and AI/ML algorithms in healthcare is a nascent and evolving, the regulatory guidance is not as well defined as for conventional drug, device and/or biologics (development and approval). The Global Harmonization Task Force (established in 1993) and the International Medical Device Regulators Forum (IMDRF) currently guide the regulation of SaMDs for the FDA in the United States and in Europe. To ensure the safety, efficacy and performance of SaMDs; the IMDRF outlined principles for quality management (27). The FDA also published a working model for the software precertification program in 2019, which is a voluntary pathway for manufacturers of SaMDs with unwavering commitment to quality, excellence
Digital technologies and AI also have the potential to contribute to the realms of research and medical education. In clinical research, digital technologies can help enhance clinical trials including recruitment through online engagement, health data collection through smart devices, and analytics utilizing AI/ML technologies (29). Utilizing digital technologies in clinical trials can help reduce costs, improve data fidelity, and allow studies to reach more diverse populations (30) (Figure 1).

In education, AI is anticipated to change not only how information is taught, but also what topics will be focused on. As AI becomes more pervasive in the clinical setting, future providers will need to adapt to be proactive in the design, development of AI diagnostic systems, and how to optimize these AI systems for the most optimal patient care (31). Since the 1920s, AI has been incorporated into education as personal tutors and typing instructors. In the 1980s, a dialogue based tutoring system (DBTS) known as CIRCSIM was introduced to medical students, allowing students to better understand and retain what they had learned, but since then there has been a paucity of literature on the application of AI to the education of nursing, allied health, and medical education (11,32). AI may potentially be so integrated into medical education one day as to be able to recommend modules, identify areas of weakness, and even augment learning during an operation throughout a resident’s day (33).

**Current gaps and potential areas for future development**

There is an evident lack of a mechanism for head-to-head comparison of AI algorithms used in healthcare to assess their efficacy and superiority (over the conventional regression models and amongst themselves). An example could be cited from the arena of intensive care medicine where AI/ML algorithms have been used for prognostic and predictive enrichment in the clinical setting. Some of these algorithms and softwares are purely data driven (34,35). Other algorithms are grounded in the pathophysiological basis of the disease (36). Currently, FDA regulations and guidance fall short of comparing two similar SaMDs for efficacy, safety, and quality. This may diminish the motivation for the developers and the clinical scientists to strive for continual improvement or even quality control.

There is also insufficient characterization and oversight for development and performance testing. The current SaMDs development process is affected by similar biases as in clinical trials. The FDA and National Institute of Health (NIH) now require that the patient population recruited and targeted during a drug development should have a good
representation of historically under-represented patient populations such as women and minorities. However, during the development of AI/ML algorithms, the datasets used for training and validation could have an inherent bias if the dataset is primarily composed of a skewed population.

**Current oversight overview, standards, and recommendations for product/technology evaluation**

The United States government has been advocating to emphasize the use of AI application for the public good while maintaining the aspects of fairness, safety, and governance (37). In the year 2020, the White House published an executive order for the regulation for AI applications highlighting the ten main principles of public trust in AI, public participation, scientific integrity and information quality, risk assessment and management, benefits and cost, flexibility, fairness and nondiscrimination, disclosure and transparency, safety and security and interagency coordination (38).

Currently, AI applications that have a role in the clinical decision-making process, either in diagnostic realm or treatment, are regulated by the FDA as SaMDs. However, barriers to more specific regulations such as the ambiguous nature of AI/ML, concerns about cybersecurity, and the rapid development of these technologies remain (39). The FDA stratifies potential risk posed by an SaMDs into classes I (low risk) to class IV (highest risk) based on the severity of a healthcare condition (non-serious to critical) and the extent to which the SaMDs is involved in clinical decision making (purely informational to treatment delivery). While current FDA regulations cover fixed rule-based AI, there is little oversight on more advanced, continuously learning AI applications that incorporate ML to adapt to new information (19).

There also remain areas of potential testing that can further improve the development process and regulatory oversight of SaMDs. Some of these areas include: (I) self-awareness of limitations (AI algorithms if not trained appropriately, lack the concept of contextualizing and can often disregard important cues when those lay at or outside the limit of their competencies); (II) transparent logic (40) (moving away from black-box algorithms and providing a transparent interface to build clinical trust and engagement); and (III) auditability or accountability (providing independent means to evaluate the software’s continuing performance). The suggestions made above are by no means an extensive or complete list and there remains a plethora of issues that need close regulatory oversight.

**A word on ethical issues surrounding development of SaMDs and digital twins in healthcare**

The use of AI based SaMDs and digital twins have a huge potential to transform the healthcare delivery for the better, but, as highlighted above, there exist also major ethical concerns around the governance of such algorithms. The five major components of biomedical ethics (patient autonomy, non-malfeasance, distributive justice, utility and beneficence) can be compared to the regulatory issues in AI/ML (informed consent, algorithm fairness and biases, intellectual property law, data privacy, and safety and transparency) (41) (Figure 2).

Depending on the development process of the digital twin (use of AI/ML, IoT, big data etc.), it can be affected by multiple socio-ethical issues. These can include the appropriate representation of under-represented patient population (women and minorities) and training and validation cohorts may have a population selection bias (western vs. developing world patient population) to name a few. Other major areas of improper conduct could involve privacy and property of data, patient autonomy and freedom. Concerns exist that digital twin technology could further worsen the health disparity by facilitating medical treatment or education in the developed world, compared to the other regions where it may lack the same degree of availability. Despite the challenges, the overall sentiment towards the future development of SaMDs and digital twin platforms remains positive, and experts believe in the overall social benefit brought in by this innovation and that digital twins promise to replace “subjective data” with the “objective data” (42,43).

**Summary and conclusion**

SaMDs and digital twin platforms in healthcare have an enormous potential. Considering the pace of development and interest of the stakeholders, appropriate oversight is pivotal. At this point, although there are no clear FDA guidelines specifically for AI/ML algorithms, the guidelines for SaMDs still apply and should be diligently followed. Despite the current checks and balances in place by the FDA and IMDRF there are multiple areas where the safety, efficacy and quality concerns could be improved upon. Some of the suggested ways to address those concerns would

© Annals of Translational Medicine. All rights reserved.  
Ann Transl Med 2022 | https://dx.doi.org/10.21037/atm-22-4203
be bringing in place measures to improve accountability (independent auditing), transparency (transparent logic), and performance testing.

**Acknowledgments**

**Funding:** Supported by the National Center for Advancing Translational Sciences, No. UL1 TR002377 (to AL).

**Footnote**

**Provenance and Peer Review:** This article was commissioned by the editorial office, *Annals of Translational Medicine*. The article did not undergo external peer review.

**Conflicts of Interest:** All authors have completed the ICMJE uniform disclosure form (available at [https://atm.amegroups.com/article/view/10.21037/atm-22-4203/coif](https://atm.amegroups.com/article/view/10.21037/atm-22-4203/coif)). AL serves as an unpaid editorial board member of *Annals of Translational Medicine* from September 2022 to August 2024. The other authors have no conflicts of interest to declare.

**Ethical Statement:** The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

**Open Access Statement:** This is an Open Access article distributed in accordance with the Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International License (CC BY-NC-ND 4.0), which permits the non-commercial replication and distribution of the article with the strict proviso that no changes or edits are made and the original work is properly cited (including links to both the formal publication through the relevant DOI and the license). See: [https://creativecommons.org/licenses/by-nc-nd/4.0/](https://creativecommons.org/licenses/by-nc-nd/4.0/).

**References**

32. Randhawa GK, Jackson M. The role of artificial intelligence in learning and professional development for healthcare professionals. Healthc Manage Forum
Cite this article as: Lal A, Dang J, Nabzdyk C, Gajic O, Herasevich V. Regulatory oversight and ethical concerns surrounding software as medical device (SaMD) and digital twin technology in healthcare. Ann Transl Med 2022. doi: 10.21037/atm-22-4203