IMPERIAL

Evidence generation for digital health technologies: Clinical simulation and its regulation

INSTITUTE OF GLOBAL HEALTH INNOVATION

Contents

Authors

Fiona O'Driscoll,
Niki O'Brien,
Gianluca Fontana,
Matthew Prime,
Chaohui Guo,
Ara Darzi,
Saira Ghafur

Suggested citation:

O'Driscoll F, O'Brien N, Fontana G, Prime M, Guo C, Darzi A, Ghafur S.
Evidence generation for digital health technologies: Clinical simulation and its regulation.
Imperial College London (2024).

Acknowledgements

We would like to thank the following people who contributed to this document through attendance at a round table discussion and have agreed to be acknowledged.

Brad Spring, Roche Diagnostics

Stephen Gilbert, Technische Universität Dresden

Xenophon Papademetris, Yale School of Medicine

This report was funded by Roche Information Solutions (RIS). The views expressed in this report are based on interviews and the round table discussion. The authors take final responsibility for the contents of this report.

Foreword

Following the COVID-19 pandemic, digital health technologies (DHTs) have rapidly increased in prevalence and have significantly changed the landscape of e-healthcare delivery. These technologies offer unparalleled benefits to patients and clinicians; however, these same attributes that differentiate DHTs, such as the use of artificial intelligence (AI), also contribute to their challenges. This can result in difficulties in generating the evidence that is required to provide regulators, and ultimately patients, clinicians, and the public, the reassurance that these technologies are safe and effective for their intended use. Given these challenges, there is a significant lack of knowledge about how these rapidly evolving DHTs can be assessed appropriately.

Much of our recently published work has focused on this core issue – how we can generate evidence for DHTs that meet regulatory needs. It is a growing research area, that requires regulators, industry, innovators, and researchers to work together and understand the unique complexities of DHTs. This paper intends to add to the evidence base and show our commitment to working at the forefront of this topic alongside international cross-sectoral experts.

We hope that this paper acts as a comprehensive summary of our recent research on the use of clinical simulation to generate evidence for DHTs, more specifically using Software as a Medical Device (SaMD) as a use case. It aims to consolidate and share the key findings from our recently published eDelphi study, which produced the Simulation for Regulation of SaMD (SIROS) framework. This framework intends to provide regulatory structure to innovators using clinical simulation as an evidence generation method. This is complemented with an in-depth discussion about clinical simulation and DHTs, based on the framework, background literature and a round table discussion held with international experts on the topic. Therefore, this paper should provide readers with a comprehensive overview of the clinical simulation landscape and its current and future use to evaluate DHTs. We hope that this will ensure that DHTs are assessed using the most appropriate and cost-effective methods, providing the essential evidence for regulators, and resulting in transformative healthcare for patients and clinicians.

Dr Saira Ghafur
Lead for Digital Health at the Institute of Global Health Innovation, Imperial College London, and Honorary Consultant at the Imperial College Healthcare NHS Trust

Lord Ara Darzi
Co-Director of the Institute of Global Health Innovation, Imperial College London

Executive summary

Clinical simulation offers a novel, cost-effective method of generating evidence for digital health technologies (DHTs). This evidence is required by regulators to ensure that DHTs are safe and effective before they are approved for patient or clinician use. However, the nature of clinical simulation and its differences to traditional research studies, means that we need to change the way that we assess and regulate its use for DHTs.

Understanding this background, this paper aims to summarise the challenges of DHTs, introduce clinical simulation and then propose and discuss the Simulation for Regulation of SaMD (SIROS) framework, developed to assess the use of clinical simulation in evaluating DHTs.[1] Through this framework, we aim to further the evidence gap in DHT regulation and increase the use of clinical simulation methods in regulatory approvals.

Software as a Medical Device (SaMD) is a category of DHT that has transformed the way healthcare is delivered, through applications such as remote monitoring or clinical decision support tools. However, these new technologies also require a new approach to regulation, since traditional academic research cannot adequately assess SaMD due to the technology's agile nature, along with being unfeasible for most organisations developing DHTs to undertake. To overcome these challenges, industry and regulators are working together to develop new regulatory tools and pathways that allow the appropriate development and regulation of DHTs.

Clinical simulation involves creating realistic clinical scenarios with real end users. It has traditionally been used in medical education to train clinicians, however, now is increasingly applied to assess DHT scenarios. By putting clinicians and other end-users in realistic clinical settings, with synthetic patient cases, they can test how they would use a DHT in practice without requiring a real healthcare setting or putting patients at risk. Clinical simulation is uniquely placed to evaluate DHTs due to its speed, low cost and ability to replicate clinical scenarios with intended end-users of the DHT, therefore assessing how it would be used in real-life.

The Simulation for Regulation of SaMD (SIROS) framework

Working with Roche Information Solutions (RIS), the Institute of Global Health Innovation (IGHI) at Imperial College London conducted a research study to identify the standards and criteria for using clinical simulation as a research method to evaluate SaMD from a regulatory perspective. This resulted in the SIROS framework, which outlines the key criteria that regulators should assess against and includes seven domains: background & context, overall study design, study population, delivery of the simulation, fidelity, software & AI, and study analysis.

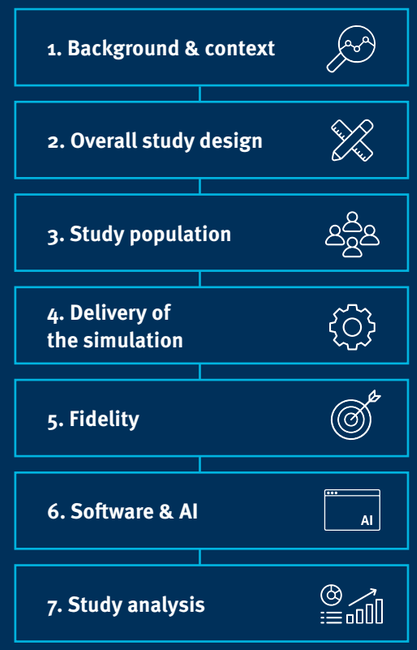


Figure 1: The seven domains of the Simulation for Regulation of SaMD (SIROS) framework

We recommend that the SIROS framework is used by industry and regulators to assess clinical simulation used for SaMD. It may be iterated to incorporate feedback and ensure its continued relevance and applicability in the evolving DHT landscape. The global network of industry and digital health experts, researchers, and regulators (including the FDA and NICE) should continue to collaborate on this topic and further explore emerging areas such as AI and machine and learning, which will continue to change and require attention over the coming years.

Section 1
Introduction: Highlighting the challenges of regulating digital health technologies

The potential for digital health technology (DHT) to revolutionise health and healthcare has been well documented. Communication software enabling the delivery of remote care services can reach patients previously underserved, offer patients greater flexibility, and improve provider efficiency [2], while clinical decision support tools have been found to reduce medical errors, increase adherence to clinical guidelines, and can lead to efficiencies with implications for health system cost savings.[3] In both cases, the technology is regulated as software as a medical device (SaMD), defined as "software intended to be used for one or more medical purposes that perform these purposes without being part of a hardware medical device".[4] Despite increasing regulation of SaMD in many settings globally, limited progress has been made in developing innovative methods for evaluating DHT that consider the fast-paced and adaptive nature of its development.

The COVID-19 pandemic highlighted the potential for DHT to address health challenges on a global scale, as technology was utilised for disease surveillance, including testing and tracking, maintaining socially distanced health services, and developing vaccines. The pandemic also highlighted the importance of appropriate regulation in enabling innovation, as regulatory updates on data sharing allowed large-scale clinical trials to take place, with tangible impacts in addressing COVID-related health challenges.[5] Such an innovation-driven approach to regulation can also be utilised in the context of evaluating different types of DHT, including SaMD.

As the use of DHTs is an increasingly essential aspect of health service delivery, it is critical that robust and novel approaches to evaluating technologies are developed and implemented. This will ensure that innovations can be adopted more rapidly at scale, and ensuring the safety, efficacy, and cost-effectiveness at a population level. However, the growing complexity of DHTs, including artificial intelligence and machine learning, presents a regulatory challenge.

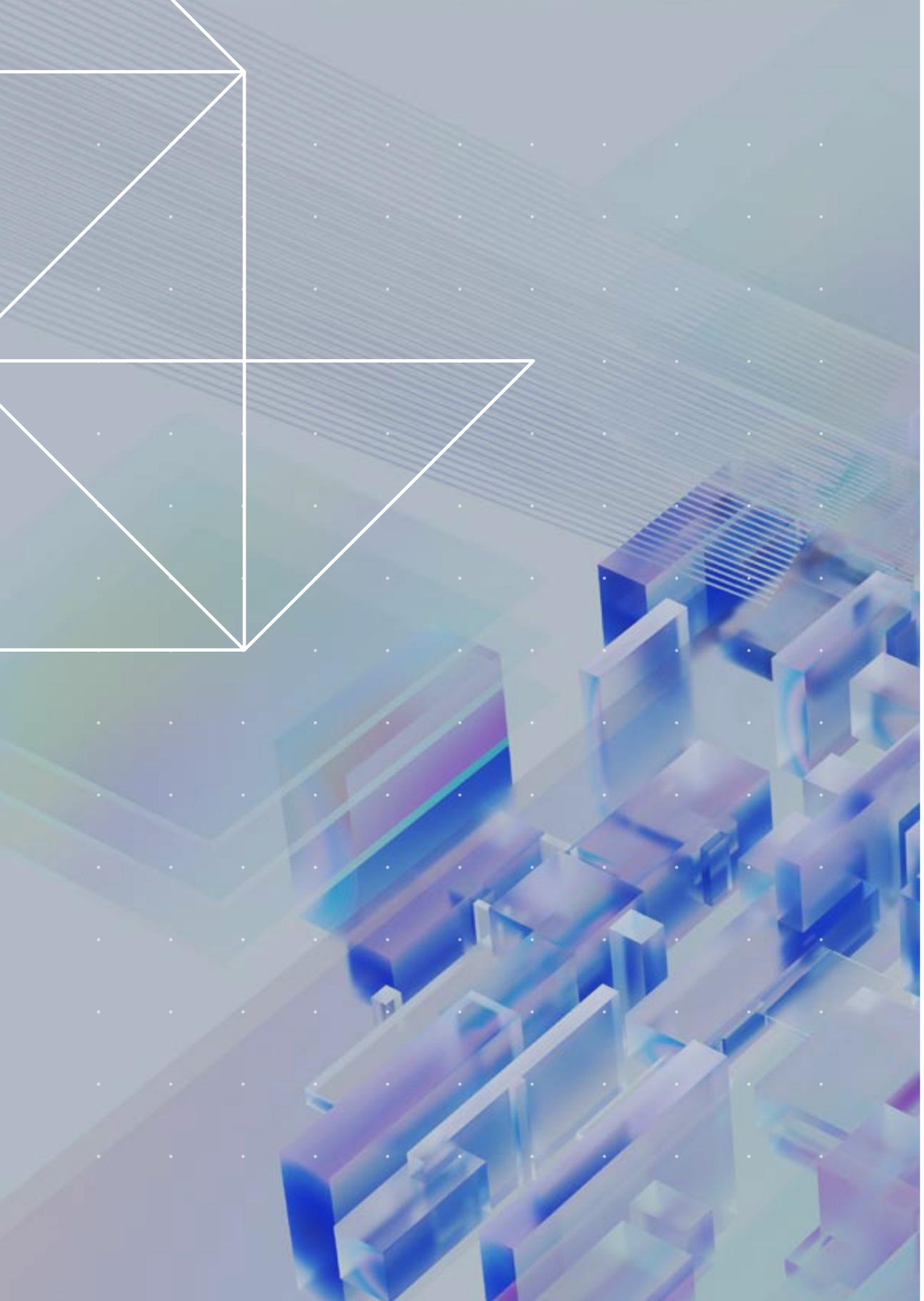
Regulating DHTs

Maximizing the potential of DHTs requires regulators to develop regulatory pathways that support innovation and industry to generate evidence in new ways that balance the importance of quality and safety with the fast-moving pace of DHT innovation. In a 2022 study, nearly half of the start-ups in the study cohort (44%) were found to have a clinical robustness score of 0, having made no FDA filings or clinical trials on ClinicalTrials.gov.[6] While only a snapshot of the larger technology industry, such findings indicate the challenges faced by technology developers in generating an evidence base on the effectiveness of their products.

The crux of the problem is the difficulty of utilizing traditional randomised control trials (RCTs), still considered to be the "gold standard" of academic research [7], to generate evidence to submit to regulators. Historically health technologies, including novel drugs and medical devices, have been tested and validated through RCTs, but increasingly, DHTs have fast cycles of iteration and require constant updates and improvements not feasible in the RCT setting, which typically requires 5.5 years from enrolment to publication and methodology that mandates no changes mid-trial.[8] Though estimates on costs vary [9], RCTs are also expensive which is incompatible with the spectrum of companies, including startups and small and medium-sized enterprises (SMEs), developing DHTs.

Innovation-enabling regulatory pathways

The challenges faced by innovators in generating evidence for DHTs are increasingly clear to regulators who are in turn working with relevant stakeholders to simplify pathways. In the United States, the Digital Medicine Society brought together stakeholders including the US Food and Drug Administration (FDA), industry, startups, and other stakeholders which resulted in the creation of a regulatory compass tool (the RegPath Decision Tool) enabling innovators to understand whether their DHTs falls within FDA regulation, and if so, which regulatory pathway is relevant.[10] In the United Kingdom, the Innovative Devices Access Pathway (IDAP) pilot was launched in 2023 to facilitate the development of innovative technologies, by providing innovators with a multi-partner support service to bring new products to patients sooner.[11] While in Europe, Germany was the first country to launch a "fast track" pathway for digital health applications (DiGA) to be reimbursed by statutory health insurance (e.g., apps on prescription),[12, 13] although this pathway is limited to certain solutions, where applications such as clinical decision support are not within its scope. Following this approach, France has adopted a similar route through the Prise en Charge Anticipée (PEC-AN) programme which promotes early reimbursement for innovative DHTs.[14]



These recent developments have begun to pave the way for the generation of different types of necessary evidence. Real-world evidence (RWE), defined by the FDA as data relating to patient health status and/or the delivery of health care routinely collected from a variety of sources (e.g., electronic health records, medical claims data, data from digital technologies)[15], is increasingly being generated in health facilities and used as evidence in regulatory approvals.[16, 17] Germany's DiGA pathway has advanced the use of real-world evidence for DHT regulatory approvals and actively supports its inclusion in submission packages. However, this has been seldom used to date, thought to be due to the lack of prior use and uncertainty associated with this approach.[18]

As DHTs are an increasingly essential aspect of health service delivery it is critical that these robust novel approaches to evaluating technologies are developed and implemented to ensure innovations can be adopted more rapidly at scale, and ensuring the safety, efficacy, and cost-effectiveness at a population level.

Section 2
Clinical simulation: Providing an opportunity to evaluate digital health technologies

Clinical simulation refers to placing actual end-users, such as clinicians, in simulated scenarios to carry out realistic tasks that are part of their usual work practice. Various inputs may be required to carry out the simulation, for example, a realistic simulated environment that mimics healthcare facilities, or clinical scenarios that are representative of what clinicians would commonly face in their day-to-day practice. This ensures that regardless of the purpose of the clinical simulation, it is placing the right user in the right environment to carry out the right task.

To date, clinical simulation has most commonly been used in education, as a training methodology for clinicians. This is due to the ability to simulate complex activities in a safe environment, such as teaching surgical skills. However, more recently it has been used to assess DHTs where it would not otherwise be feasible to do so in a real-life setting, given its relatively low-cost, agility and speed. It is this relatively nascent use of clinical simulation as an investigative methodology that is the focus of our work.

Clinical simulation as a method to assess DHTs

Clinical simulation can be used to assess DHTs in two primary ways:

- Using scripted clinical scenarios performed by actors to replicate real-life patient and clinician interactions.
- Presenting synthetic patient cases to clinicians for them to assess a DHT.

The focus of our work is in the latter, where we work to create realistic synthetic patient cases and present them to clinicians for them to use and assess the DHT in its intended purpose.[19] A helpful case study that explains in more detail what this involves is provided below.

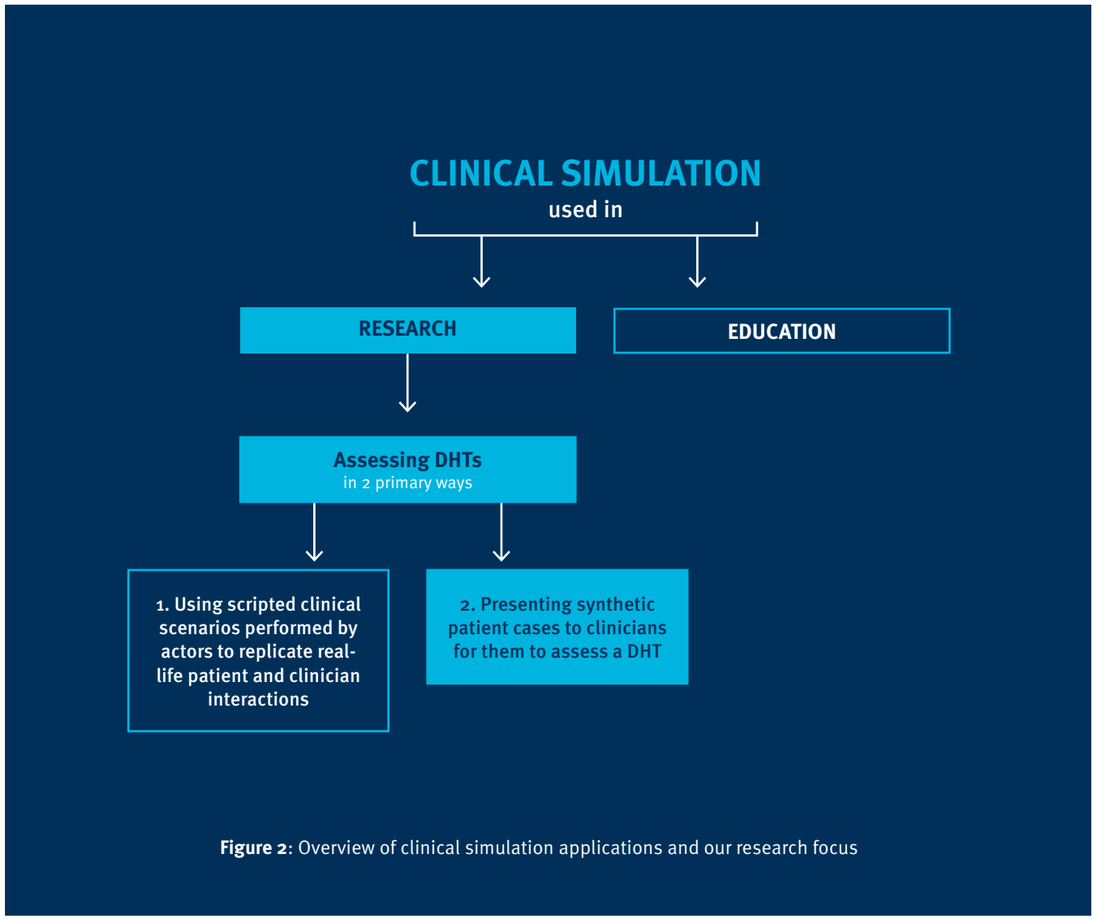


Figure 2: Overview of clinical simulation applications and our research focus

Case study: Assessing a clinical trial match solution using clinical simulation[20]

Background

Clinical trials in oncology are essential for the scientific advancement of treatments and offer an opportunity for patients with cancer to try alternative new treatments if they are eligible. These trials require the enrolment of suitable study participants to assess the clinical efficacy and safety of novel treatments. The opportunities for patients to enrol in clinical trials have increased significantly, as evidenced by the number of active studies listed on ClinicalTrials.gov growing from 2,119 in 2001 to over 400,000 in 2023.[21] It can be challenging for clinicians, such as oncologists or research nurses, to match patients to these cancer clinical trials. These trials often have complicated and lengthy eligibility criteria, resulting in a labour-intensive task and highly manual task completed in demanding clinical environments.

DHT being assessed

The NAVIFY Clinical Trial Match (CTM) application was developed by Roche Diagnostic Information Solutions to reduce the amount of time it takes to match patients to the right cancer clinical trials. This aims to improve the quality of matching and reduce cognitive burden for the decision-maker. NAVIFY CTM analyses data on a patient's condition, genomic alterations, and the hospital's location to automatically find relevant clinical trials.

Clinical simulation approach

Synthetic oncology patient cases were developed with a multidisciplinary clinical team, and these were shared with a group of 10 clinicians. The clinicians were asked to assign five patients to a clinical trial using NAVIFY CTM and five patients to a trial as per usual practice, which is carried out by manually searching online trial databases. The clinicians were asked to use the same rigour as they would in the usual practice. The quality, efficiency, and cognitive burden of the matching process was assessed for each participant.

Clinical simulation results

Using the NAVIFY CTM tool, participants were able to complete trial matches faster, find more relevant trials and required lower mental effort compared to online trial databases, such as ClinicalTrials.gov.

Expanding the clinical simulation evidence base

While we have published other work surrounding clinical simulation and DHT, this paper focuses on our recent study where we chose to examine SaMD,[1] more specifically SaMD where the intended end-users are clinicians, such as in the case study provided above. As mentioned earlier, given the increasing use of SaMD in clinical practice and the regulatory challenges that it faces in terms of evidence generation, it is an important DHT to consider. Focusing specifically on SaMD allows us to generate more information about a specific area of DHT, while also drawing important insights about the application of clinical simulation in other areas.

Section 3
An eDelphi study designed to develop a clinical simulation regulatory framework

Given the lack of evidence related to clinical simulation being used to develop regulation for digital health technologies (DHTs), the Imperial College London team, working with Roche Diagnostics colleagues, used Software as a Medical Device (SaMD) as a use case to explore this further.[1] We sought to answer the following research question:

> **"What are the standards and criteria for using clinical simulation as a research method to evaluate software as a medical device (SaMD) from a regulatory standpoint?"**

Standards and criteria are used to assess DHTs and their study methodologies before they are approved in a market.[22, 23] They ensure that DHTs meet minimum pre-defined criteria across safety, quality, and efficacy. Using rigorous academic methods, the Delphi study approach was taken, using an online format (eDelphi) to allow the research team to gain consensus from global experts in various sectors such as digital health, medical device regulation, health policy and health systems.

Study preparation and recruitment

Prior to commencing the study, a background literature review was completed to gain a greater understanding of the SaMD and clinical simulation landscape. This also led to the development of the initial 19 potential items for inclusion in the standards and criteria that were proposed to the study participants in the Delphi scoping round.

Recruitment of suitable participants for the study took several months to ensure an adequate sample size was achieved and efforts were made to ensure participation from across geographies and sectors. The recruitment was conducted using purposive and snowball sampling from the research team's existing networks through digital health work.

Study design

Delphi studies are commonly used for developing consensus on various topics where no clear guidance or evidence may exist. This is carried out using a structured communication and deliberation process where participants are asked to consider a set of questions in each Delphi round. After completing each round, the results are compiled and any questions that do not meet the set of pre-defined criteria are removed. The questions are presented again in subsequent rounds to participants until a pre-determined endpoint has been reached.

Four key features of our study that were typical of a Delphi consensus approach[24] include:

1. A group of experts with diverse background irrespective of geographical location were recruited as panellists
2. Anonymity was preserved throughout the questionnaire process to ensure no "bandwagon effect"
3. Iterative rounds were conducted until the pre-defined criteria of consensus was reached
4. Design of subsequent rounds was informed by summarising group response in previous rounds

Participants initially answered questions that related to what they thought the standards and criteria for using clinical simulation as a research method to evaluate SaMD from a regulatory standpoint should be, in a Delphi scoping round, and following this, in two formal Delphi rounds. The pre-determined criteria for including an item for each round was as follows:

- **› 60%** of participants rating an item as very important or important, or
- **› 10%** of participants rating an item as not important at all or not important

Study results

45 participants were initially recruited to take part in the study, with 33 participants progressing to the end of the Delphi. These participants came from a range of geographies and sectors, as represented in the figures below. While there was representation from three continental regions – Europe, Middle East, Africa (EMEA); North America (NA); and Asia, Pacific (APAC) – most of all recruited participants worked in the United Kingdom (53%) and the United States (13%). See figure 3

Participants were recruited from various sectors, with the majority being from academia (36%) or the pharmaceutical industry (33%), reflecting the recruitment strategy used. Of note, some individuals had cross-sectoral experience, however, their primary sector of employment at the time of the study was recorded. See figure 4

In the scoping round, participants were asked if the 19 pre-defined criteria developed by the research team from the literature review were relevant, irrelevant or if they were not sure. They were also

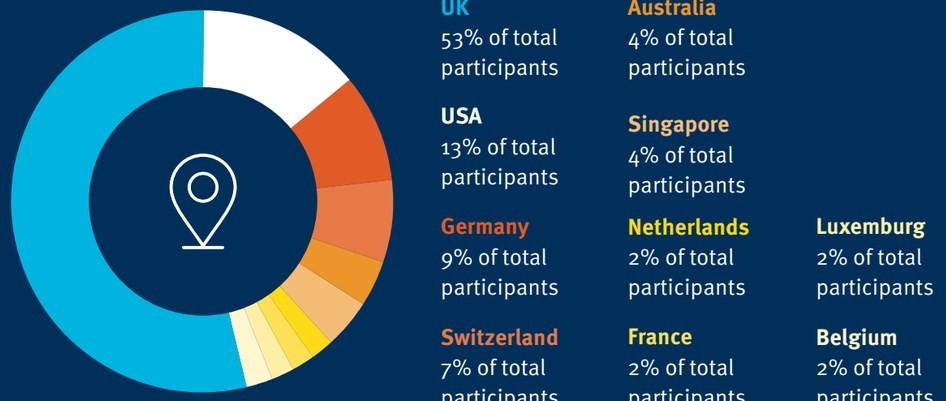


Figure 3: Composition of the participants initially recruited by country

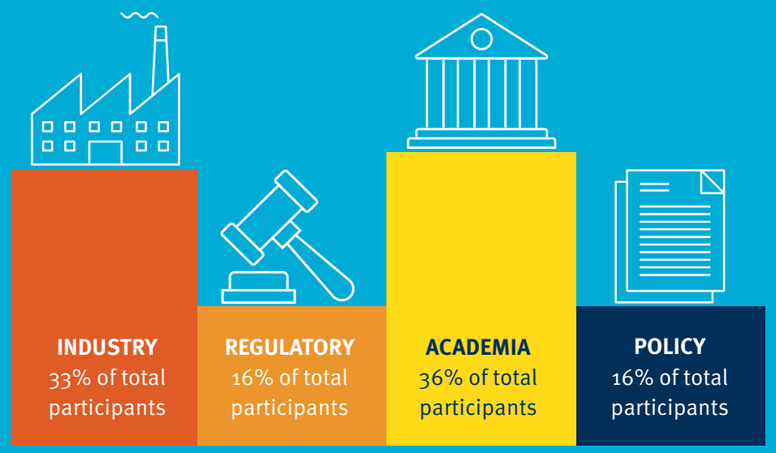


Figure 4: Composition of the participants by sector

asked to add any other items that they thought should be included also. Over 50% of participants rated each criterion as relevant and so they were included in round 1 of the Delphi. An additional 36 items were added to round 1 also based on the additional suggestions provided by participants. The final 55 items generated were grouped into seven domains based on the comments provided and the research team's expertise.

For round 1, participants were asked to rate each of the 55 items brought forward from the scoping round on a scale from 1 (not important at all) to 5 (very important), along with being asked to provide any comment on their thoughts about them. 43 of the 55 items included met the criteria to progress to the next Delphi round.

In round 2, participants were asked to rate the 43 items from the previous round on the same scale from 1 – 5 and to provide comments on their decision. This resulted in all 43 items meeting the same pre-determined criteria and the Delphi process ended. The final agreed criteria were classified based on their level of consensus – low (›60%), high (›70%) and very high (›80%).

Research limitations

There were some limitations in the research conducted, such as the geographical spread of the participants. While significant effort was made to include participants from a range of geographical locations, most participants were from high-income countries in Europe and North America (91%). This may have introduced bias in the results as the situation in low- and middle-income countries have not been included.

Development of the Simulation for Regulation of SaMD (SIROS) framework

The outputs of this study are the final criteria that reached a consensus for inclusion to assess clinical simulation as a research method to evaluate SaMD from a regulatory standpoint. This is outlined in detail in the following section.

Section 4
Simulation for Regulation of SaMD (SIROS) framework

As outlined in the previous section, the Simulation for Regulation of SaMD (SIROS) framework was developed as a result of the Delphi process by the Imperial College London research team. The framework includes seven domains, which each include criteria agreed during the Delphi process. These seven domains, and their comprising criteria, have therefore been chosen by the panel of international experts as important to consider when assessing clinical simulation as a research method to evaluate SaMD from a regulatory standpoint.

'Background and context' is the first domain, given its overall importance in setting the scene for the use and purpose of the SaMD, as this will justify the use of clinical simulation as an evaluation methodology. It also acknowledges the importance of providing any pre-existing evidence to support the use of the SaMD while to a lesser degree it acknowledges the importance of disclosing funding and conflicts of interest.

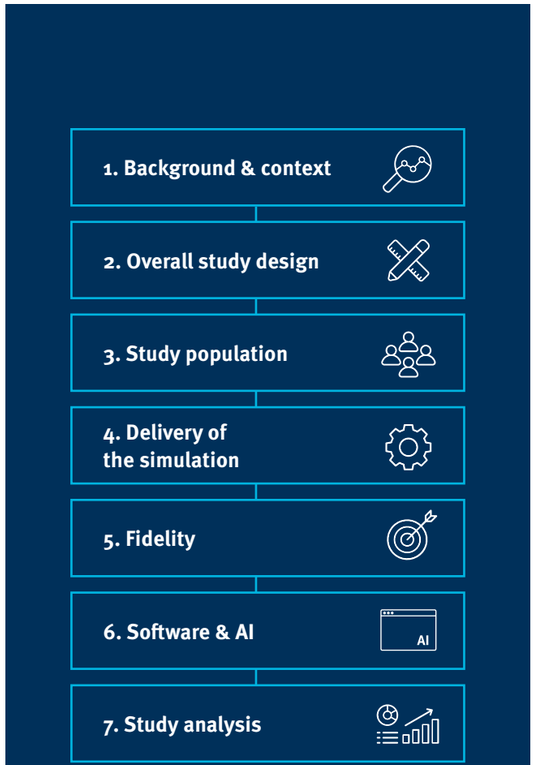


Figure 4: The seven domains of the Simulation for Regulation of SaMD (SIROS) framework

1. BACKGROUND & CONTEXT **Background and context describe the general background and contextual information that is needed by regulators to understand any information that follows, along with justifying the regulatory submission.**	**LEVEL OF CONSENSUS**
Clear description of the SaMD being evaluated, including its purpose and intended end users	Very high
Description and justification of the clinical simulation performed, alongside any other research being conducted to evaluate the SaMD	Very high
Overview of the existing evidence to support the SaMD is provided	High
Sources of funding and other conflicts of interest are declared appropriately	Low

2.OVERALL STUDY DESIGN **Overall study design describes the aspects of the clinical simulation study design to be considered by the regulator.**	**LEVEL OF CONSENSUS**
Potential limitations of the study design are discussed	Very high
Potential biases associated with the study design are discussed	Very high
Strategies to minimise potential study biases are described	Very high
Issues on equity have been considered in the overall study design e.g., high risk patient profiles, racial disparities	Very high
Digital literacy is considered in the study design e.g., digital literacy of clinicians taking part in the clinical simulation or the digital literacy of the intended end users	Low
Risk management in the study is described e.g., impact assessments	High

The second domain of **'Overall study design'** considers the importance of various aspects of the clinical simulation study design. Several items reached very high consensus, that focused on the clinical simulation team discussing the study design limitations, any potential biases and how they were minimised, along with equity being considered across the study. Issues on equity in clinical simulation approaches and SaMD applications are essential for regulators to clearly examine and avoid the introduction of any bias in the study process.

'Study population' focuses on describing the participants of the clinical simulation study which has been performed to evaluate the SaMD. This study focused on clinicians as study participants to assess the SaMD in each clinical scenario. Study participants should be representative of the intended end users of the SaMD, and therefore the eligibility criteria used for recruitment should be tailored to the SaMD end users e.g., their level of experience, practice area, qualification type.

3. STUDY POPULATION **Study population describes the way in which clinical simulation study participants, in this case clinicians, are recruited and took part in the study.**	**LEVEL OF CONSENSUS**
The eligibility criteria for clinicians who took part in the clinical simulation is representative of the intended end users e.g., staff level, qualification, experience	Very high
The sampling and recruitment methods used to recruit clinicians who took part in the clinical simulation is clearly described	Low
The number of clinicians who took part in the clinical simulation is provided	Very high
Issues on equity were considered in the sampling and recruitment process to ensure representativeness	High

4. DELIVERY OF THE SIMULATION **Delivery of the simulation describes the multiple aspects of the clinical simulation and how it took place practically. This allows regulators to understand how the study was performed.**	**LEVEL OF CONSENSUS**
The environment in which the clinical simulation took place is described e.g., physical or virtual location, type of healthcare facility	Very high
The timing of the clinical simulation is described e.g., time of day, length of time taken	Low
The equipment used for the clinical simulation is described	Very high
The facilitator (the individual who facilitated the clinical simulation for the clinicians), if any, is described e.g., what role they took, how many there were, what input they had	Low
The initial orientation and any training provided to the clinicians before taking part in the clinical simulation is described	Very high
When the clinical simulation was being performed, the SaMD was described in sufficient detail to the clinicians taking part in the clinical simulation to allow them to evaluate it	Very high

The fourth domain **'Delivery of the simulation'** refers to clear descriptions of the various aspects specifically related to the way in which the clinical simulation was performed. The regulator should seek information about where, when, and how the clinical simulation took place, for example, a description of what information and training was provided to the study participants before they took part in the clinical simulation. This is important to be able to replicate any studies and understand what may have influenced the results.

Fidelity refers to the degree of exactness with which something is reproduced, and therefore may occur on a spectrum ranging from a feature that is replicated to a high or low degree of fidelity. The fidelity should also relate to the intended end use of the SaMD. For example, high healthcare facilities fidelity is where the clinical simulation took place in an oncology outpatient clinic if that is where the SaMD is intended to be used by the clinician. There was a very high level of consensus for high fidelity being required for the concept in which the clinical simulation is being used, the synthetic patient cases, and the clinical scenarios used in the simulation. Whereas less people agreed on the importance of high fidelity for healthcare facilities in the clinical scenario.

5. FIDELITY **Fidelity describes the multiple aspects of the clinical simulation that are designed to produce a study environment that is as close to real-life as possible or as deemed appropriate by the researchers. Fidelity is not a single one-off consideration, but there are many components to it, such as physical, conceptual and clinical fidelity.**	**LEVEL OF CONSENSUS**
There is a clear analysis, considering the risk and impact, of the different levels of fidelity, e.g., high, medium and low, required for various aspects of the clinical simulation	High

A lack of fidelity in any aspect of the clinical simulation is explained and justified e.g., fidelity in one aspect of the scenario may not be required for the SaMD being assessed	High
The clinical simulation has high conceptual fidelity, that meets the intended use of the SaMD	Very high
The clinical simulation uses high fidelity synthetic patient cases, that meet the intended use of the SaMD	Very high
The clinical simulation has high clinical scenario fidelity, that meets the intended use of the SaMD	Very high
The clinical simulation has high healthcare facilities fidelity, that meets the intended use of the SaMD	High
The methodology and rationale for developing the synthetic patient cases is described	Very high
The overall representativeness of the synthetic patient cases is described	Very high
Potential limitations of the synthetic patient cases are discussed	Very high
Potential data bias in development of the synthetic patient cases is discussed	Very high
Strategies to minimise potential data bias associated with synthetic patient cases are discussed	High

6. SOFTWARE & AI **Software & AI describes the various aspects of machine learning and AI that are a part of the SaMD being evaluated and that should be considered by the regulator.**	**LEVEL OF CONSENSUS**
Any continuous machine learning algorithms embedded in the SaMD are described	Very high
The design and development of any continuous machine learning algorithms embedded in the SaMD are described	Very high
Any continuous machine learning algorithms are reviewed at regular intervals to monitor their changes from the initial set-up	Very high
Any software updates to the SaMD made since the clinical simulation study are described and justified	Very high

Software and AI related to SaMD are important areas for regulators to be aware and informed of, as shown by the results of this study, with all four criteria within this domain reaching very high levels of consensus. These criteria relate to clear descriptions provided of any continuous machine learning algorithms, how they were developed and how they are being monitored over time, along with any software changes to the SaMD.

The final domain is the **study analysis**, which includes various aspects of how the clinical simulation study was analysed that should be considered by the regulator. This includes the outcome measures that were chosen for the study, which should meet the end use and indication of the SaMD. It should also clearly describe how the analysis was performed along with any sensitivity analysis to examine any assumptions and assess the robustness of the findings.

7. STUDY ANALYSIS **Study analysis describes the process that was taken to analyse the results of the study. This includes initially setting out what will be measured at the start of the study, along with data analysis and any other study outcomes.**	**LEVEL OF CONSENSUS**
The primary and secondary outcome measures are clearly defined, including how and when they were assessed	Very high
Rationale and justification for the chosen primary and secondary outcome measures is provided	Very high
The usability of the SaMD is assessed as part of the clinical simulation	High
The feasibility of the SaMD is assessed as part of the clinical simulation	High
The impacts of any unintended consequences e.g., harm/clinical risk from the study are described	Very high
The data analysis performed is clearly described e.g., statistical methods and the unit of analysis used (e.g., individual, team, group)	Very high
The generalisability of the study findings is discussed, e.g., to other populations or clinical scenarios	Very high
Sensitivity analysis is performed to assess the robustness of the clinical simulation findings	Very high

Section 5
Discussing the framework

Following the Delphi study and development of the SIROS framework, a round table discussion was held that included a small group of study participants as representatives from academia, healthcare, industry, and regulatory bodies. The purpose of the round table event, held virtually to facilitate international participation, was to present and discuss the framework, generating in-depth conversation about clinical simulation, DHTs and the regulatory process. Round table discussions are commonly conducted to gain consensus through an informal, facilitated discussion and provide space for discussing areas of contention. The discussion was therefore not meant to be exhaustive and cover all framework domains. Instead, attendees were presented with unusual findings or questions arising from the Delphi study and this is reflected in the summary below, with support provided by background literature and team knowledge.

Using clinical simulation to enhance evidence generation across the DHT development lifecycle

Clinical simulation is a research methodology that can be used at various stages of the development of DHTs to provide evidence supporting the technology, depending on the knowledge required at that stage.

Clinical simulation offers a unique capability to reduce the burden of evidence and costly clinical trials for innovators, while still maintaining the level of evidence required to demonstrate safety and effectiveness. The application of clinical simulation may be more relevant at various stages of the DHT development, depending on the product and evidence requirements at any given time, which is supported by the first section of the SIROS framework where justification of the clinical simulation must be provided along with any existing evidence. Based on the product phase, different evidence generation methods will be required and as a result, there are likely 'different tiers of good application for clinical simulations,' with clinical simulation being available as a low-cost, agile option for specific products or at certain stages.

> *"A lot of it comes down to the product type, the use statement and therefore the classification of the product. Perhaps then we can be more specific about those types of products where simulation could be applicable."*

Therefore, regardless of its use, clinical simulation should not be seen as a replacement for methods such as randomised controlled trials (RCTs), which may be the essential requirement to provide the relevant data in certain cases. Instead, it should be considered amongst the toolkit of other research methods for appropriate DHTs and their use cases. For example, clinical simulation may be a strategic choice at earlier stages of evidence generation, allowing the company to iterate the tool before entering a costly trial or used at different time points in a product development.[20]Post-launch, clinical simulation offers a unique opportunity to assess updated algorithms in pre-determined change control plans.[25] and identify safety issues.

Innovators and regulators should continue to consider the 'applicability and appropriateness' of using clinical simulation to generate evidence for SaMD along with identifying which DHTs are most suitable to assess with clinical simulation based on their classification,[22] thus increasing its use and acceptability for regulatory evidence.

Simulating real-life healthcare environments to assess DHTs appropriately

Clinical simulation allows innovators to create realistic healthcare environments to test DHTs within, by putting the right end user in the right place to test the right product.

Due to logistics or technological constraints, it is often not possible to test a SaMD, or other DHT, within a real-world setting. However clinical simulation offers the unique opportunity to evaluate DHTs within an environment that is as close to reality as possible, based on various elements such as the healthcare facilities, staffing, or clinical scenario used.

> *"Clinical trials rarely actually simulate the real world themselves. You can introduce a lot of situations into simulations that you would not see even if you ran a 10-year clinical trial."*

Along with high conceptual fidelity, the SIROS framework outlines that to evaluate a SaMD appropriately for its intended end use using clinical simulation methods, high fidelity synthetic patient

cases, clinical scenarios and healthcare facilities are all required. Using these carefully designed and data-driven inputs will result in a highly accurate testing environment in which to safely evaluate a DHT in an agile, scalable and low-cost way, as evidenced by our previous work.[19] However, given the innate complexity and unpredictability of healthcare scenarios, our round table experts also discussed the importance of evaluating broader effects, such as the significant cognitive burden placed on clinicians as they are faced with complex treatment decisions in high-pressure, fast-paced environments. This can be carried out through measures such as assessing cortisol levels,[26] or cognitive burden when carrying out tasks.[27] Designing specific stresses into the clinical simulation scenarios could also simulate realistic healthcare environments of device alerts, interruptions and interacting digital systems, further differentiating clinical simulation from the highly controlled RCT environments.

The ability to develop synthetic patient cases for specific high-risk or vulnerable patient groups also provides the opportunity to test DHTs for populations that may be challenging to recruit or not possible to include in RCTs. Therefore, the ability to replicate multiple clinical scenarios in a simulated healthcare environment or include different patient groups ensures that clinical simulation is a practical and insightful method of assessing DHTs to provide evidence that would not otherwise be possible to obtain.

Avoiding bias within the delivery of the clinical simulation

As with all research, it is important that bias is minimised in clinical simulation. However, particular care should be given to how the synthetic patient cases used in the simulation are developed.

Avoiding bias in research is a fundamental principle of well-designed studies to ensure validated research findings, however, it is particularly relevant for clinical simulation. Bias is where systematic error has been introduced to the study process, which affects the research outcomes.[28] Given the nature of clinical simulation, with numerous inputs such as clinical scenarios, study facilities and synthetic patient cases, strategies to identify and minimise potential biases within the study design, and more specifically the development of patient cases, are essential.

> *From a regulatory perspective, we need to see that the population that something has been based on is relevant to the population of the product that it is being applied to.*

This is also important given the use case of SaMD in the SIROS framework, where algorithmic bias in DHTs is a well-recognised issue and can introduce or exacerbate health inequities.[29] This was a key focus area of the round table discussion, which highlighted the need to mitigate against bias in the development of synthetic patient cases. The cases used in clinical simulation studies are developed by the research team to represent various patient groups for the specific clinical scenario, such as synthetic oncology patient cases used to evaluate a clinical decision support tool that matches patients to oncology clinical trials.[27] Historically, clinical case vignettes used to inform online symptom checkers have been developed by clinicians based on their clinical knowledge and experience.[30] However, the inherent bias due to different clinicians' experiences or descriptions and lack of direct patient involvement, has resulted in recommendations for a standard way to develop these vignettes,[30] an approach which could be replicated for synthetic patient cases.

Taking a data-driven approach to developing patient cases also requires close examination, as healthcare data may have systemic collection and recording issues which mean that it is not representative of the population.[29] However, once these issues are accounted for, clinical simulation can be used to counter-act the existing structural bias within research, by designing patient cases directly with and for underrepresented patient groups and public members. This would ensure that an equitable and unbiased approach is adopted, with clinical simulation championing the novel use of synthetic patient cases in this way.

Ensuring transparency in the development of software and AI

International regulation around AI-enabled DHTs is changing to ensure that regulators understand how algorithms change over time and what impact they may have on patient care.

Artificial intelligence is increasingly being used to modify and enhance DHTs, with the field transforming rapidly and offering unprecedented benefits for patients and healthcare workers.[31] However, some

round table panellists noted the lack of transparency around AI algorithm development as an issue for its regulation, in particular 'black box' AI where the machine learning process cannot be explained.

> *How do we handle these products in post-market world, especially when you have a machine learning AI algorithm that could continuously change?*

Transparency in AI development is required to monitor and mitigate potential risks to the public, such as inequitable health outcomes or patient safety concerns. Regulators must be aware of how algorithms have been developed and how they change over time, with regulatory oversight now seen as an ongoing process post-launch due to their ever-changing nature.

Regulatory bodies are working to address this, with the FDA proposing new guidance that supports continuous AI improvements in software through the submission of Predetermined Change Control Plans (PCCP), ultimately aiming to improve patient access and safety to AI-enabled technologies.[25] There are clear information requirements to submit as part of the PCCP to promote transparency, including the planned changes and how they will be implemented and assessed. While several international regulatory bodies, including the FDA, the Medicines and Healthcare products Regulatory Agency (MHRA) in the UK and Health Canada worked together to publish Good Machine Learning Practice for Medical Device Development: Guiding Principles,[32] the European Union (EU) is currently preparing the world's first comprehensive AI law that aims to regulate the use of AI throughout the EU.[33] The AI Act takes a risk-based approach, with medical devices classified as high-risk as they already are regulated by EU product safety legislation, and therefore are subject to prohibited practices. Whilst this regulation is a much-needed structure to ensure better healthcare products and outcomes, our experts also highlighted that the nature of AI requires a sufficiently flexible framework to adapt to ongoing innovation, while maintaining appropriate risk management. This adaptability will promote DHT innovation within a safe framework while ensuring that transparent AI practices benefit patients and clinicians.

Conclusion and next steps

Clinical simulation provides a cost-effective and agile approach to evaluating SaMD, given its ability to replicate real-life clinical scenarios and test products with intended end users. Novel methods such as clinical simulation are required for innovators and regulators alike, to adapt to the ever-changing DHT and regulatory landscape. DHT features such as continuous machine learning and rapid innovation pathways have resulted in expanded patient and staff benefits, however, they cause regulatory challenges due to the inability of existing regulatory processes to adapt to their assessment. The current lack of clear regulatory frameworks for SaMD, and other DHTs, is a major cause for concern and one that we aim to address through the Simulation for Regulation of SaMD (SIROS) framework. The development of the SIROS framework has provided an initial regulatory structure for the application of clinical simulation to generate evidence for SaMD. The framework's seven domains will provide innovators with the information required to generate sufficient evidence demonstrating safety and effectiveness to regulators, and future SaMD users. However, additional work is needed to expand the use of clinical simulation in evaluating DHTs and provide further regulatory pathways to adopt similar novel methods of evidence generation.

To build upon the SIROS framework, we advise that it be used in practice by innovators and regulators to evaluate SaMD. To successfully apply the framework, we recommend that it is operationalised into a checklist that provides a user-friendly interface and support structure. Case studies that showcase its use could be developed by innovators to encourage adoption and expand the use of clinical simulation to assess DHTs across other areas. The framework should also be iterated based on real-world feedback over time, creating a comprehensive refined version to support adoption and real-world applicability. Furthermore, we believe that the SIROS framework should be modified and adopted to other DHTs like SaMD, given that they also may face the same evidence generation issues and require new regulatory approaches.

Novel evidence generation techniques, such as clinical simulation, offer benefits to all those involved in the development of DHTs and require a usable, regulatory structure to support their approval. The SIROS framework aims to bridge this gap between innovators and regulators and support the continued rapid development of new DHTs. We look forward to seeing how this area continues to evolve and work with others in this space to further expand the use of clinical simulation for DHTs.

References

1. O'Driscoll F, O'Brien N, Guo C, Prime M, Darzi A, Ghafur S. Clinical Simulation in the Regulation of Software as a Medical Device: An eDelphi Study. JMIR Form Res 2024;8:e56241

2. OECD. The COVID-19 Pandemic and the Future of Telemedicine. Paris: OECD Publishing; 2023.

3. Sutton RT, Pincock D, Baumgart DC, Sadowski DC, Fedorak RN, Kroeker KI. An overview of clinical decision support systems: benefits, risks, and strategies for success. NPJ Digit Med. 2020;3(1):17.

4. International Medical Device Regulators Forum SaMD Working Group. Software as a Medical Device (SaMD): Key Definitions. International Medical Device Regulators Forum; 2013 Dec 18.

5. Ghafur S, O'Brien N, Howitt P, Painter A, O'Shaughnessy J, Darzi A. NHS data: Maximising its impact for all. Imperial College London; 2023.

6. Day S, Shah V, Kaganoff S, Powelson S, Mathews SC. Assessing the Clinical Robustness of Digital Health Startups: Cross-sectional Observational Analysis. J Med Internet Res. 2022;24(6):e37677.

7. Jones DS, Podolsky SH. The history and fate of the gold standard. Lancet. 2015;385(9977):1502-3.

8. Pham Q, Wiljer D, Cafazzo JA. Beyond the Randomized Controlled Trial: A Review of Alternatives in mHealth Clinical Trial Methods. JMIR Mhealth Uhealth. 2016;4(3):e107.

9. Speich B, von Niederhäusern B, Schur N, Hemkens LG, Fürst T, Bhatnagar N, et al. Systematic review on costs and resource use of randomized clinical trials shows a lack of transparent and comprehensive data. J Clin Epidemiol. 2018;96:1-11.

10. Croxen J, Kaganoff S. A rising tide lifts all boats: Navigating digital health regulation: Rock Health; 2023 [Available from: https://rockhealth.com/insights/a-rising-tide-lifts-all-boats-navigating-digital-health-regulation/.

11. Medicines and Healthcare products Regulatory Agency. The Innovative Devices Access Pathway (IDAP): GOV.UK; 2023 [Available from: https://www.gov.uk/government/publications/the-innovative-devices-access-pathway-idap]

12. Sauermann S, Herzberg J, Burkert S, Habetha S. DiGA - A Chance for the German Healthcare System. J Eur CME. 2022;11(1):2014047.

13. Schliess F, Affini Dicenzo T, Gaus N, Bourez J-M, Stegbauer C, Szecsenyi J, et al. The German Fast Track Toward Reimbursement of Digital Health Applications (DiGA): Opportunities and Challenges for Manufacturers, Healthcare Providers, and People With Diabetes. Journal of Diabetes Science and Technology [Internet]. 2022. Available from: https://journals.sagepub.com/doi/abs/10.1177/19322968221121660.

14. Haute Autorité de Santé. Digital medical devices: early support: Haute Autorité de Santé; 2023 [Available from: https://www.has-sante.fr/jcms/p_3376633/fr/dispositifs-medicaux-numeriques-la-prise-en-charge-anticipee#:~:text=La%20prise%20en%20charge%20anticip%C3%A9e%20concerne%20les%20DMN%20pr%C3%A9sum%C3%A9s%20innovants,le%20type%20de%20DMN%20concern%C3%A9.

15. U.S. Food & Drug Administration. Real-World Evidence: fda.gov; 2023 [Available from: https://www.fda.gov/science-research/science-and-research-special-topics/real-world-evidence#:~:text=Real%2Dworld%20evidence%20is%20the,derived%20from%20analysis%20of%20RWD.

16. Baumfeld Andre E, Reynolds R, Caubel P, Azoulay L, Dreyer NA. Trial designs using real-world data: The changing landscape of the regulatory approval process. Pharmacoepidemiol Drug Saf. 2020;29(10):1201-12.

17. U.S. Food & Drug Administration. FDA Approval Demonstrates the Role of Real-World Evidence in Regulatory Decision-Making on Drug Effectiveness: fda.gov; 2021 [Available from: https://www.fda.gov/drugs/news-events-human-drugs/fda-approval-demonstrates-role-real-world-evidence-regulatory-decision-making-drug-effectiveness.

18. Stern AD, Brönneke J, Debatin JF, Hagen J, Matthies H, Patel S, et al. Advancing digital health applications: priorities for innovation in real-world evidence generation. Lancet Digit Health. 2022;4(3):e200-e6.

19. Lau K, Halligan J, Fontana G, Guo C, O'Driscoll FK, Prime M, et al. Evolution of the clinical simulation approach to assess digital health technologies. Future Healthcare Journal. 2023;10(2):173-5.

20. Prime MS, Guo C, Fontana G, Ghafur S, Halligan J, Gardner C. An assessment of the NAVIFY clinical trial match application using clinical simulation-based research methods. Journal of Clinical Oncology. 2021;39(15_suppl):e13552.

21. ClinicalTrials.gov. Trends, Charts, and Maps: ClinicalTrials.gov; 2023 [Available from: https://classic.clinicaltrials.gov/ct2/resources/trends#RegisteredStudiesOverTime.

22. National Institute for Health and Care Excellence. Evidence standards framework for digital health technologies. 2018.

23. ORCHA. How do Digital Health Standards Assess Evidence: Orchahealth.com; 2021 [Available from: https://orchahealth.com/how-do-digital-health-standards-assess-evidence/.

24. Junger S, Payne SA, Brine J, Radbruch L, Brearley SG. Guidance on Conducting and REporting DElphi Studies (CREDES) in palliative care: Recommendations based on a methodological systematic review. Palliat Med. 2017;31(8):684-706.

25. U.S. Food & Drug Administration. Marketing Submission Recommendations for a Predetermined Change Control Plan for Artificial Intelligence/Machine Learning (AI/ML)-Enabled Device Software Functions. 2023.

26. LeBlanc V, Regehr C, Tavares W, Scott A, MacDonald R, King K. The Impact of Stress on Paramedic Performance During Simulated Critical Events. Prehospital and Disaster Medicine. 2012;27(4):369-74.

27. Gardner C, Halligan J, Fontana G, Fernandez Crespo R, Prime M, Guo C, et al. Evaluation of a clinical decision support tool for matching cancer patients to clinical trials using simulation-based research. Health Informatics J. 2022;28(2).

28. Pannucci C, Wilkins E. Identifying and avoiding bias in research. Plast Reconstr Surg. 2010;126(2):619-25.

29. O'Brien N, Van Dael J, Clarke J, Gardener C, O'Shaughnessy J, Darzi A, et al. Addressing Racial and Ethnic Inequities in Datadriven Health Technologies. Imperial College London; 2022.

30. Painter A, Hayhoe B, Riboli-Sasco E, El-Osta A. Online Symptom Checkers: Recommendations for a Vignette-Based Clinical Evaluation Standard. J Med Internet Res. 2022;24(10):e37408.

31. World Economic Forum. Global Health and Healthcare Strategic Outlook: Shaping the Future of Health and Healthcare. 2023.

32. Good Machine Learning Practice for Medical Device Development: Guiding Principles. U.S. Food & Drug Administration, Health Canada, Medicines and Healthcare products Regulatory Agency; 2021.

33. European Parliamentary Research Service. Artificial Intelligence Act: European Parliament; 2023 [Available from: https://www.europarl.europa.eu/RegData/etudes/BRIE/2021/698792/EPRS_BRI(2021)698792_EN.pdf.

IMPERIAL

INSTITUTE OF
GLOBAL HEALTH
INNOVATION