TOPICAL REVIEW • OPEN ACCESS

The stochastic digital human is now enrolling for *in silico* imaging trials—methods and tools for generating digital cohorts

To cite this article: A Badano *et al* 2023 *Prog. Biomed. Eng.* **5** 042002

View the article online for updates and enhancements.

You may also like

Progress in Biomedical Engineering

TOPICAL REVIEW

The stochastic digital human is now enrolling for *in silico* imaging trials—methods and tools for generating digital cohorts

A Badano[1,*] , M A Lago[1], E Sizikova[1], J G Delfino[1], S Guan[1], M A Anastasio[2] and B Sahiner[1]

[1] Division of Imaging, Diagnostics, and Software Reliability, Office of Science and Engineering Laboratories, Center for Devices and Radiological Health, US Food and Drug Administration, Silver Spring, MD 20993, United States of America
[2] Department of Bioengineering, The Grainger College of Engineering, University of Illinois, Urbana, IL 61801, United States of America
* Author to whom any correspondence should be addressed.

E-mail: aldo.badano@fda.hhs.gov

Keywords: terminology, representations, population, models, modelings, disease

Abstract

Randomized clinical trials, while often viewed as the highest evidentiary bar by which to judge the quality of a medical intervention, are far from perfect. *In silico* imaging trials are computational studies that seek to ascertain the performance of a medical device by collecting this information entirely via computer simulations. The benefits of *in silico* trials for evaluating new technology include significant resource and time savings, minimization of subject risk, the ability to study devices that are not achievable in the physical world, allow for the rapid and effective investigation of new technologies and ensure representation from all relevant subgroups. To conduct *in silico* trials, digital representations of humans are needed. We review the latest developments in methods and tools for obtaining digital humans for *in silico* imaging studies. First, we introduce terminology and a classification of digital human models. Second, we survey available methodologies for generating digital humans with healthy and diseased status and examine briefly the role of augmentation methods. Finally, we discuss the trade-offs of four approaches for sampling digital cohorts and the associated potential for study bias with selecting specific patient distributions.

1. Introduction

Two decades ago, in the epilogue of their seminal textbook on image science [1], Barrett and Myers pointed out that in the future, sport games might be played with simulated athletes. The advancement of computer graphics and simulation technologies sparked the notion that perhaps the excitement of a real-life sports event could be conducted in the simulation space with digital models of athletes. Since then, continuous advances in computer processing power and modeling techniques have taken place, driven primarily by entertainment applications [2] and quickly becoming a significant component of research and development efforts in a variety of industries[3]. Industries that have widely adopted computational modeling and *in silico* methods throughout the product life-cycle include automotive [3] and manufacturing [4] among others [5]. Medicine lags considerably behind [6] due, in part, to model complexity, challenging validation, associated potential risks for new devices and drugs, and lack of consensus and regulatory standards.

Randomized clinical trials, while often viewed as the highest evidentiary bar by which to judge the quality of a medical intervention, are far from perfect. Common causes of failure include safety issues, difficulties with patient recruitment, enrollment, and retention [7]. In addition, clinical trials can suffer from under-representation of rare subpopulations [8]. These limitations represent a unique opportunity to develop *in silico* trials that are completed as planned, safely, and that include digital cohorts with a representative distribution of subject characteristics and numbers large enough for appropriate statistical

---

[3] To date, Super Bowl games are played with physical-world athletes, in part due to the difficulty of conveying real-life personal struggle, an essential component of the entertainment context for sport players and teams (see, for instance, here).

power. As pointed out in [9], *in silico* data has the potential to address lack of data availability, sharing mechanisms and privacy challenges associated with the use of medical information.

In silico imaging trials are computational studies that seek to ascertain the performance of a medical device for the intended population, collecting this information entirely in the digital world via computer simulations. The benefits of *in silico* imaging trials for evaluating new technology include significant resource and time savings, minimization of subject risk, and ethical considerations [10, 11]. Moreover, *in silico* trials can be used to study devices that do not yet exist or are not practically attainable in the (limited) physical world, allow for the rapid and effective investigation of new technologies [11–13], and facilitate representation from all relevant subpopulations. Each one of these benefits is an essential consideration within the context of the regulatory evaluation of medical technology [11].

The realization that computational models of humans would take center stage in medical imaging system assessment is not new. Full optimization of imaging systems for specific medical tasks requires objects (physical or digital) that represent the variability seen in patients. For many decades, scientists have relied on practical and simpler versions of patients [14]. However, recent advances in computer processing power and simulation methods are now facilitating the development of more detailed and realistic patient models that are based on digital stochastic descriptions of the model components. For instance, a recent report demonstrated the feasibility of an *in silico* trial, the Virtual Imaging Clinical Trial for Regulatory Evaluation (VICTRE), as an alternative approach to establish regulatory evidence in support of medical imaging products [15].

There are numerous parallels between digital- and physical-world trials. Fundamentally, *in silico* trials must include the same essential elements of well-designed physical-world clinical trials. Firstly, the population of subjects for whom the new device or technology is intended must be defined. The study design must contain clear rules for selection and rejection of subjects from a distribution of healthy and diseased subjects. However, *in silico* trials are not subject to effects from covariates in patient selection. For instance, a common problem in evaluating screening tests meant for asymptomatic subjects is that a portion of the enrolled population might be symptomatic [16] with the potential for verification bias [17]. Secondly, when there are two technologies that are being compared, i.e. a new, yet unproven technology and a comparator technology currently in clinical use, both must be unambiguously defined. A good choice for comparator technology should be associated with accurate representations of the device characteristics as supported by validation studies [18]. Thirdly, the study requires a definition of the users of the device's outcome (i.e. images in the case of an imaging device trial). These first three components reflect the physical intended use of the device under investigation, i.e. the intended populations of subjects, the intended device comparison, and the intended image interpreters that will be using the device in the physical world. Finally, whether physical or digital, the trial design must provide a definition of the primary outcome to be evaluated, a protocol and statistical analysis associated with the trial, and an analysis of the risk and benefits introduced by the device under investigation.

Both physical and *in silico* studies require enrollment of representative subjects. In this review, we survey the latest developments in methods and tools for generating the cohorts of digital humans for imaging studies that represent the variability of physical-world subject populations. We refer to the digital cohorts consisting of digital humans (realizations of the digital human models) as 'stochastic humans'. Assessment of new technology and the regulatory evaluation of that technology requires establishing performance levels for intended populations and, therefore, necessitates computational models that allow sampling of the parameter space defining the subject population in the physical world. We propose to name these models digital humans as opposed to digital replicas or twins to avoid confusion.

The review is organized as follows. First, we introduce terminology and representation models regarding the different types of digital humans described throughout the article. Second, we survey available methodologies for generating digital humans with healthy status and for generating diseased cases. Then, we briefly discuss the role of augmentation methods and conclude with an analysis of sampling techniques that may be used to generate the digital cohorts for evaluating the performance of imaging devices.

2. Terminology

A variety of terminologies are being used or proposed for describing digital representations of humans in medicine and other fields. In the literature, some of these are often used without the benefit of a clear definition and, in some instances, wrongly interchangeably.

We propose to use the term stochastic digital human to denote digital representations of humans (or human body parts) generated from multiple random outputs by sampling known distributions for the model characteristics matching the variability observed in human populations. In contrast, non-stochastic representations are deterministic digital versions of a single physical exemplar (e.g. a model of a human body

at a given time) or a group (or family) of physical exemplars which are differentiated by varying physical parameters. Contrary to other terms and concepts currently being discussed including digital families, avatars, chimeras, and digital twins, the concept of a stochastic digital human represents an approach for *in silico* trials and regulatory evaluation that estimates the performance of an imaging device for a population of subjects rather than for an individual patient, thus incorporating the variability observed in the population.

We propose to classify all digital humans as either individual or population models (see figure 1). Individual models are necessarily image-based while population models can be derived either from images or from knowledge of the fundamental characteristics that define the relevant features of a human. Note that we will use the term digital human to refer to the models even if the represented object is a part of the body or the whole body of a subject.

3. Representations

Physical objects (including humans) can be represented using continuous variables. We consider the models of humans as continuous in space (r) and time (t) and described by a coefficient vector affecting a set of model characteristics:

$$\mathbf{f}_m\left(\boldsymbol{r}, t\right) \approx \sum_{n=1}^{N} \theta_n \phi_n\left(\boldsymbol{r}, t\right). \tag{1}$$

Here, N is the dimension of the approximate finite-dimensional representation of the object, and the subscript m indicates the modeling approximation to differentiate from the actual object $\mathbf{f}(\boldsymbol{r}, t)$.

The collection of expansion functions $\{\phi_n(\boldsymbol{r}, t)\}_{n=1}^{N}$ is employed to form $\mathbf{f}_m(\boldsymbol{r}, t)$, and θ_n denotes the nth component of the N-dimensional expansion coefficient vector $\boldsymbol{\theta}$. The quantity $\mathbf{f}_m(\boldsymbol{r}, t)$ constitutes a discrete representation of a digital human that can be readily displayed on a computer or digitally processed. For the case where the expansion functions are defined as indicator functions that describe non-overlapping space-time voxels, $\boldsymbol{\theta}$ can sometimes be interpreted as a digital image whose components θ_n represent the integrated value of the object over the support of the voxel.

More generally, a digital human model can be established by integrating the continuous representation $\mathbf{f}_m(\boldsymbol{r}, t)$ over a collection of N voxels as

$$f_n = \int_{v_n} \mathbf{f}_m\left(\boldsymbol{r}, t\right) \mathrm{d}^3\boldsymbol{r}\, \mathrm{d}t, \qquad n = 1, \ldots, N, \tag{2}$$

where v_n denotes the support of the nth spatial-temporal voxel and f_n denotes the nth component of a N-dimensional vector $\mathbf{f}$ that represents the digital human.

As discussed below, the choice of the expansion functions and associated expansion coefficients can be specified in different ways, with the general goal of making $\mathbf{f}_m(\boldsymbol{r}, t)$ an accurate approximation of $\mathbf{f}(\boldsymbol{r}, t)$. The expansion functions can depict geometry (e.g. size, morphology), material properties (e.g. x-ray interaction cross-sections, elasticity) or other relevant features (e.g. radioactivity, blood oxygenation levels). For simplicity, we will consider that the stochastic human does not vary with time and proceed only with the spatial dimension $\boldsymbol{r}$. However, the concepts that follow can readily be generalized to model time-varying descriptions [19].

In practice, the coefficient vector $\boldsymbol{\theta}$ can be modeled as a random vector and the expansion functions $\{\phi_n(\boldsymbol{r})\}_{n=1}^{N}$ as random processes. Methodologies for generating large cohorts of digital stochastic models of humans for *in silico* imaging trials, including models for organs and tissues with appropriate variability, can rely on either sampling $\boldsymbol{\theta}$, ϕ_n or both from appropriate distributions representing the intended population. We can denote the cohort of digital stochastic humans as follows,

$$\{\mathbf{f}_s\}_{s=1}^{S} = \sum_{n} \theta_n^s \phi_n\left(\boldsymbol{r}\right), \tag{3}$$

where s denotes a particular state or random realization of a digital human in a cohort of size S.

When ϕ_n are known, analytically or numerically, the stochastic models are referred to as procedural. In this case, the modeler is left with choosing the coefficient vector defining the object ($\boldsymbol{\theta}$). In cases for which the defining characteristics are unknown, θ_n and ϕ_n can be estimated from imaging data.

In the following sections, we review available methods and tools for generating digital human models and digital cohorts. We present a classification of available approaches in figure 1.

Figure 1. Classification of methods to generate digital humans for *in silico* clinical trials.
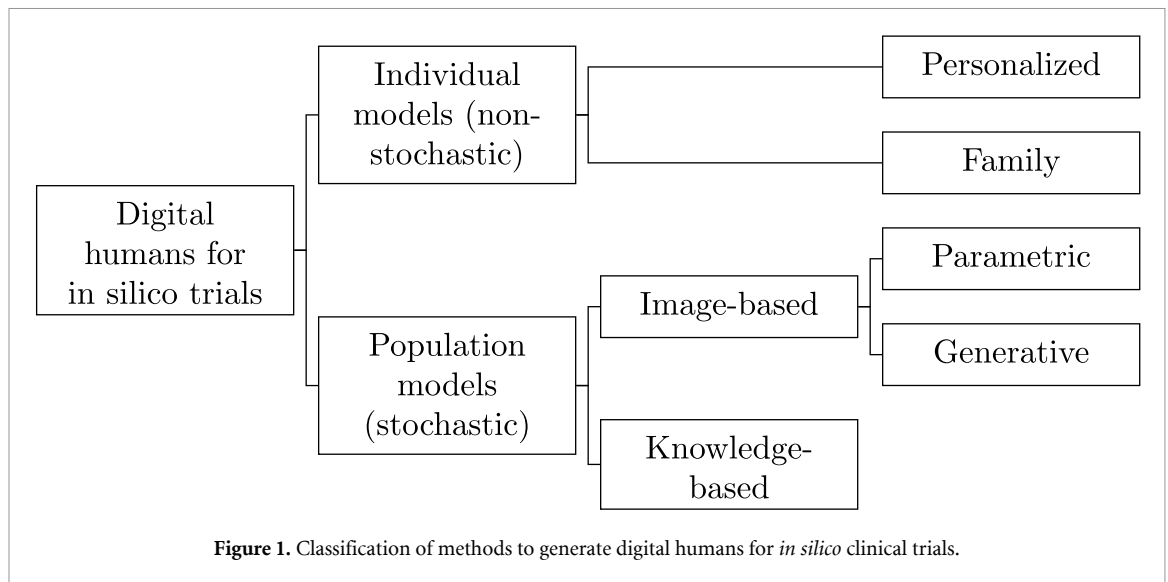
4. Individual models

Individual models attempt to create a digital replica of a specific physical object. Individual models can be categorized as personalized and family models. These models are not stochastic since they are meant to represent individual subjects with as much detail and accuracy as achievable from the image data. In this respect, the representation introduced in section 3 applies only with $S = 1$ resulting in a single coefficient vector (θ_n) defining the individual.

The digital representation in these cases is typically a multidimensional voxelized array that can be segmented into structures such as tissues and organs. Early attempts relied on geometrical volumes represented by analytical expressions altered to generate a wide variety of sizes and shapes. In other words, ϕ_n are described by quadrics and θ_n represent properties of the volumes defined by the surfaces (e.g. x-ray attenuation and scattering properties). These computational models have proved useful in areas of quality control of imaging systems [20, 21] and in radiation dosimetry [22]. Even with more sophisticated geometrical structures [23–25] and more spatial detail, these approaches lack the ability to accurately represent the statistical variability found in humans, organs and tissues. While these simpler models remain practical and useful for some tasks, the lack of realism and variability makes them unsuitable for generating digital humans for *in silico* imaging trials.

4.1. Personalized models or digital twins

Personalized models aim to capture patient-specific information in a digital representation [26]. Medical digital replicas of human subjects are *in silico* representations of an individual in terms of anatomy and physiology. Sometimes referred to as digital twins [27], these replicas are designed to simulate parts or the whole body of a subject for prognostic or predictive assessments.

A critical element of the concept of digital twin is the inclusion of detail found in the individual patient. For example, Jirsa *et al* [28] describe a method to obtain digital virtual brains by mapping the brain network of a subject with epilepsy, using data derived from magnetic resonance imaging (MRI) images of the subject. The digital twin model can be used clinically to estimate the extent, localization and organization of the epileptogenic regions related to seizure. The authors postulate that virtual brains could one day be used as part of the clinical decision-making to improve localization precision for seizure activity and for personalized surgical planning. However, the authors claim that low spatial and temporal resolution and lack of validation currently limit the use of the models in clinical settings.

These models, including digital twins, can be continuously updated from multimodal medical data if the characteristics change over time[4]. Digital twins are of interest in the context of evaluating and selecting optimal medical treatments [27] or imaging procedures [29] within clinical practice, and can also be incorporated into other *in silico* applications [30]. For instance, Wang *et al* [31] suggested three applications in the areas of medical imaging: optimal selection of scanning techniques (so called 'virtual comparative

---

[4] A related concept is an avatar, an artistic and sometimes aspirational digital representation of the human in the digital world for inter-activity purposes.

scanning'), data sharing from *in silico* scanning of the digital replica to the open source community, and improvement of the regulatory process of image reconstruction algorithms. Patient image datasets can also be used to generate models of specific tissues and organs. For instance, the Visible Human project [32] was first made available in 1994 by the National Library of Medicine (NIH) to facilitate anatomy visualization applications and includes a detailed data set of cross-sectional photographs of the human body.

4.2. Family models

Personalized models of a small number of subjects can be assembled into families to generate a collection of a small number of digital humans spanning a common set of parameters, such as subjects' body size and age. These models are based on image acquisitions using different modalities including computed tomography (CT), MRI and chest radiographs (CXRs).

An example of a family model is the Virtual Family [33], released by FDA[5] in 2012. The Virtual Family consists of a set of detailed, anatomically correct whole-body models of an adult male, an adult female, and two children based on high-resolution MRI data of healthy volunteers. Organs and tissues are represented using computer-aided design techniques where each component is a high-resolution, non self-intersecting mesh. In this case, the models are used for electromagnetic, thermal and acoustic simulations in the safety assessment of active and passive medical implants [34]. Safety evaluations do not require full sampling of the intended population and can be performed with a small number of exemplars, provided the exemplars adequately cover the needed parameter space.

Similar approaches are utilized in efforts to provide models of patient anatomy using patient images as the basis for development of cohorts including using MRI and CT images for modeling lungs [35] and torso [14]. More recently, image-derived digital and physical models of the breast have been proposed by Kiarashi *et al* [36] and Bliznakova *et al* [37]. In this approach, a voxelized breast model is derived from patient images through image segmentation for determining the composition of each voxel [38–44]. Patient-derived models are limited to the imaging characteristics of the acquisition system and are also affected by the imperfections of the segmentation methods. The resulting models can also be augmented with physiological features to facilitate imaging studies involving contrast agents [45].

5. Population models

Testing new imaging devices, however, requires the availability of large digital cohorts of stochastic digital humans that can be assembled to properly power a study not only on the aggregate (i.e. for the entire population), but also to analyze for specific subgroups with notable characteristics, including under-represented populations. In this section, we focus our attention on models suited for the generation of large cohorts of digital humans to be enrolled within *in silico* imaging trials.

5.1. Image-based models

Image-based models estimate and sample model components from relevant characteristics within the acquired medical images. Image-based models estimate model components ϕ_n and θ_n in equation (3) from within the acquired medical images. Whether parametric or generative, all image-based models are limited by the quality of the source data (i.e. medical images), including noise, artifacts, and contrast constraints, and do not provide an unequivocal mapping to the underlying tissues. In practice, the use of image-based models should also acknowledge the limitation arising from the existence of a null space of the imaging system [46]. The null space, which typically arises from the mapping of a continuous object to discrete data with an imperfect image acquisition system, results in an unavoidable loss of information regarding the object. Given that the imaging system operator is only partially known for most imaging systems and cannot represent information obscured by the null space of the imaging transformation, image-based models are limited even when imaging system models include noise measurement.

5.1.1. Image-based parametric models

In image-based parametric models, the generation of cohorts is achieved by creating models based on available sets of patient imaging data and model modification techniques including parametric deformation, morphing, and registration. Parametric models (also known as stylized phantoms [47]) capture a population cohort by a set of mathematical equations representing a series of surfaces (e.g. splines) defining organs that are later voxelized into a volumetric model. The popular 4D extended cardiac-torso (XCAT) phantom [48] is

---

[5] www.fda.gov/about-fda/cdrh-offices/virtual-family.

an example of an image-based parametric model, and a survey of other representations can be found in Kainz *et al* [49].

One limitation of this approach is that model development is typically performed on a small number of available patient images. For instance, Erickson *et al* [39] presented a methodology to create a database of anatomically variable 3D digital breast models from dedicated breast CT images using a tissue classification and segmentation algorithm and a fuzzy C-means segmentation algorithm. The study provided a population of 224 breast phantoms incorporating a range of breast types, volumes, densities, and parenchymal patterns. However, using hundreds of images might be insufficient to properly characterize a population for statistically powered *in silico* imaging trials across patient variability.

Some recently released image datasets include a larger number of cases. For example, the Medical Information Mart for Intensive Care CXR dataset [50] contains 227 835 imaging studies from 65 379 patients presenting to the Beth Israel Deaconess Medical Center Emergency Department between 2011 and 2016. Similarly, the Medical Imaging and Data Resource Center effort [51] is undertaking a large, multi-year, systematic effort to collect high-quality COVID data, and over 100 000 imaging studies have been made public after 2 years of work and with significant funding from the NIH.

Cohorts containing multiple realizations of digital humans can be obtained by extending image-derived models to create populations in a statistical manner. For instance, Sturgeon *et al* [52] developed synthetic breast models using principal component analysis (PCA) to describe a small training set of patient images. In this approach, each existing patient breast CT volume was compactly represented by the mean image plus a weighted sum of eigenbreasts. The distribution of weights was sampled to create synthesized breast phantoms that matched fibroglandular density and noise power law exponent distributions in real images. Hence, the distribution of the synthetic model is determined by that of the training data, and, therefore, might suffer from a lack of appropriate representations of cases at the tails of the distribution (e.g. very large or very small, very dense or very glandular breasts). A related concept from the computer vision and graphics community is the statistical human body model, in which a vertex-based model of the body surface is learned, typically via PCA, from subjects' input. The techniques rely on linear blend skinning to constrain the surface vertex deformation with respect to a template bone skeleton [53]. Created for non-medical purposes, these parametric models are typically learned from training examples of lower resolution than what is common in medical imaging.

One alternative approach is to add deformation morphing using an anatomic template [26]. Lee *et al* [47] introduce a hybrid, non-uniform rational B-spline surface based phantom of an infant by combining the expressiveness of a voxel phantom with the flexibility of geometric manipulation and organ positioning in a parametric phantom. Another example is the XCAT Warp [54], where artificial intelligence (AI)-assisted unsupervised registration is used to warp XCAT to patient CT images to capture a more broad set of variations, compared to the existing organ and model scaling offered by XCAT. These methods are suitable for investigating digital-twin approaches where individual models reflecting the characteristics of a single individual are needed.

Limitations of image-based parametric models. Patient data sets collected from well-defined areas are likely still insufficient to capture the total variability in patient images and the large number of subgroups one may find interesting to study[6]. This limitation precludes the use of image-based parametric models for accurately creating digital cohorts for large scale *in silico* trials.

5.1.2. Image-based generative models

Image-based generative models attempt to synthesize a population of stochastic digital humans from information contained in medical images. Ideally this population captures the variability in the anatomy and tissue properties within a specified cohort of to-be-imaged subjects. Consider a collection of N-dimensional digital humans $\{\mathbf{f}_s\}_{s=1}^{S}$ that represents the cohort of interest as described by equation (3). Here, the index s specifies a digital human within the cohort and the variable S denotes the size of the cohort. Although objects are inherently infinite dimensional, we assume that each realization of the stochastic human can be accurately described by a N-dimensional representation as specified in equation (1). Thus, each object is represented by a N-dimensional vector $\mathbf{f}_s$ that resides in a Euclidean vector space.

This setting corresponds to a practical situation in which an *in silico* study employs a fully discrete representation of an imaging system in which a finite-dimensional approximation of an object is mapped to discrete image data. As mentioned in section 3, each digital human $\mathbf{f}_s$ can be interpreted as a realization of a random vector $\mathbf{f}$ that is characterized by an unknown probability density function pr($\mathbf{f}$). The ability to sample

---

[6] 'I cannot breed them. So help me, I have tried. We need more … than can ever be assembled. Millions, so we can be trillions more,' Niander Wallace in Blade Runner 2049 (see www.imdb.com/title/tt1856101/characters/nm0001467).

from pr(**f**) to generate large ensembles of objects for use in *in silico* imaging trials is, at least conceptually, the ultimate objective of a stochastic digital human model. Emerging generative methods that utilize neural networks are being actively developed for this purpose [55]. We refer to these methods as generative models.

Types of image-based generative models. Instead of explicitly modeling pr(**f**), which is difficult due to the high dimensionality of **f**, generative models typically seek to define a stochastic process for drawing samples. In other words, they are implicit. Specifically, they map an analytically tractable, low-dimensional distribution pr(**z**) to the desired samples for the high-dimensional distribution pr(**f**). Various strategies to learn the mappings $f \mapsto z$ and $z \mapsto f$ have been proposed. For example, a *variational autoencoder (VAE)* model [56] constrains pr(**z**) to be an independent and identically distributed standard normal distribution so that samples of the random vector **z** can be readily generated, and applies a reconstruction loss term to the output. To improve output realism, a *generative adversarial network (GAN)* model [57] employs an additional discriminator network, trained simultaneously with the generator, to discriminate between the real and generated examples. The GAN training process is adversarial and approximately solves a min–max optimization problem [57]. GAN models have been extremely popular for various medical image generation tasks [58]. A *flow-based model* [59] learns pr(**f**) by constructing a sequence of invertible transformations, and therefore, calculates the exact log likelihood of the observed sample. This class of models addresses the instabilities of the training process in GANs and VAEs, but typically requires a more specialized architecture, larger number of parameters and higher computational costs. Finally, a *diffusion model* [60, 61] constructs a Markov chain, where noise is gradually added during the forward process, and then removed during the backward process, from a sample. Compared to other models, in diffusion models, **z** has a higher dimensionality. They can significantly outperform GANs in output image quality [62]. To date, almost all studies of deep generative models have focused on synthesizing images rather than object representations. These models have been applied to a variety of medical imaging applications [58], but focused on synthesizing images, rather than object representations.

Limitations of image-based generative models. There are several significant challenges to employing deep generative models to establish stochastic human models. A fundamental and potentially limiting issue is the fact that a collection of objects $\{\mathbf{f}_s\}_{s=1}^S$ describing the population is generally not available. Medical images are degraded by the presence of measurement noise and/or reconstruction artifacts which are a limitation of the imaging system and not representative of the true underlying objects. As such, conventional generative models that are directly trained on degraded images will not learn how to sample from the true distribution of objects. In essence, there is a 'chicken and egg problem' when seeking to establish stochastic human models via deep generative models. There are two possible ways to circumvent this limitation. First, one can utilize high-quality medical images as surrogates of the objects. For example, in certain tomographic imaging modalities and under specific conditions, images of object properties can be reconstructed and accurately approximate the true object properties. In this case, generative models are trained in the conventional manner, with images representing the training data. If these images are representative of the desired subject cohort, the generative model has the opportunity to accurately capture object variability. Second, one can modify the generative model training process to incorporate the image degradation process in training. This approach, referred to as an ambient GAN (AmGAN) [63], utilizes a generator network that is augmented with a measurement operator. Objects produced by the generator are mapped to degraded image data, which are then compared with experimental images by the discriminator network. This permits establishment of an implicit generative model that describes object randomness to be learned from indirect and noisy measurements of the objects themselves. In a preliminary study, AmGAN was explored for establishing stochastic object models from imaging measurements for use in optimizing imaging systems [63].

Finally, image-based generative models can misrepresent details from the object space. By definition, stochastic digital human models should be independent of the imaging system, measurement noise and any reconstruction method employed in the imaging process. In other words, they should provide an *in silico* representation of the ensemble of subjects to-be-imaged and not estimates of them that would be indirectly measured or computed by imaging systems. While promising, the use of generative models for *in silico* clinical trials is nascent and there remain important topics for future investigation. The objective assessment of these technologies is largely lacking, and there is no consensus regarding what statistical information can be reliably learned. Additionally, current models have largely been applied on 2D images and their extension to three-dimensions is an ongoing topic of research. Finally, as with any data-driven method for establishing stochastic human models, the presence of an imaging system null space will fundamentally limit the ability of GANs to describe certain components of the to-be-imaged objects. The extent to which the null space can be mitigated also remains a topic of ongoing research [63].
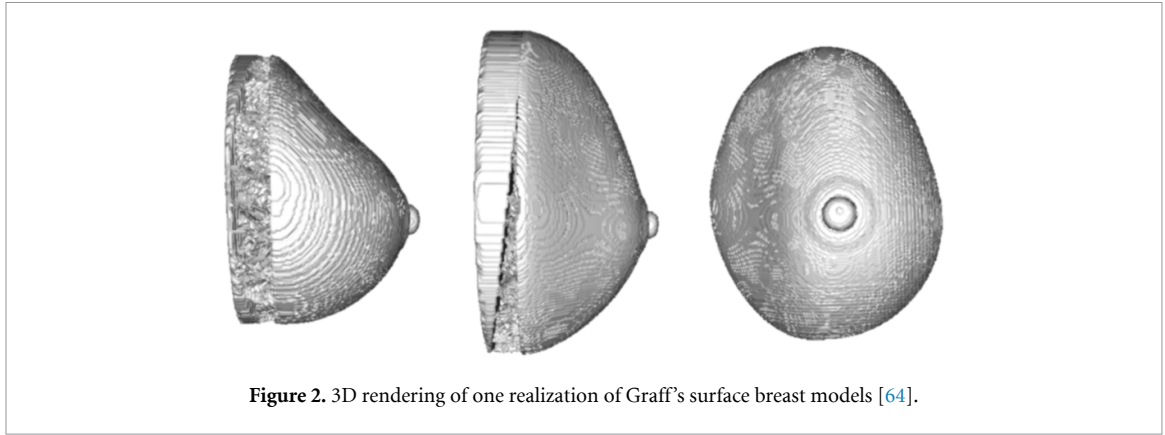
Figure 2. 3D rendering of one realization of Graff's surface breast models [64].

5.2. Knowledge-based models

Knowledge-based (also known as procedural) models are constructed by sampling a set of ϕ_n and θ_n in equation (3) from distributions representing the relevant characteristics of the models. The characteristics of the distributions are often derived from physical or biological measurements. Procedural models allow for an unlimited number of random realizations of the object, leading to the possibility of creating large cohorts of digital humans including the representation of rare cases, and at varying spatial resolution which can properly account for small structures that might be relevant for the specific imaging task being studied. However, they are usually computationally intensive and require a large number of parameters to be defined and estimated based on prior knowledge. Their accuracy and realism depend on the parameter combinations and they can sometimes generate completely unrealistic outputs.

Knowledge-based, procedural models are common in modeling breast anatomy for imaging studies. Graff [64] proposed a detailed model that begins with defining an outside surface using a quadratic hemisphere shell with a skin layer and nipple area overlaid. The shape of the shell is then adjusted for the overall breast volume and surface curvature. Using a Voronoi segmentation approach, the interior is randomly divided into regions of fat or glandular components, with each glandular component containing a ductal network with terminal duct lobular units. The volume is then filled with Cooper's ligaments, chest muscle, and blood vessels. For the VICTRE trial [15], the breast model was sampled with a 50 μm voxel size. The implementation is initiated with a set of random seeds and creates random voxelized breast anatomy objects segmented into nine different tissue types. Several different modeling techniques are employed including a non-isotropic Voronoi segmentation, recursive tree branching algorithms to generate a ductal tree and vascular network, and Perlin-noise perturbed random spheroids to create fat lobules.

$$S(\theta_s, \phi_s) = \begin{cases} (a_{1b} \cos^{\epsilon_1} \theta_s \cos^{\epsilon_2} \phi_s, a_{2l} \cos^{\epsilon_1} \theta_s \sin^{\epsilon_2} \phi_s, a_3 \sin^{\epsilon_1} \theta_s), & 0 \leqslant \theta_s < \pi/2, \pi/2 \leqslant \phi_s < \pi \\ (a_{1t} \cos^{\epsilon_1} \theta_s \cos^{\epsilon_2} \phi_s, a_{2l} \cos^{\epsilon_1} \theta_s \sin^{\epsilon_2} \phi_s, a_3 \sin^{\epsilon_1} \theta_s), & 0 \leqslant \theta_s < \pi/2, 0 \leqslant \phi_s < \pi/2 \\ (a_{1t} \cos^{\epsilon_1} \theta_s \cos^{\epsilon_2} \phi_s, a_{2r} \cos^{\epsilon_1} \theta_s \sin^{\epsilon_2} \phi_s, a_3 \sin^{\epsilon_1} \theta_s), & 0 \leqslant \theta_s < \pi/2, -\pi/2 \leqslant \phi_s < 0 \\ (a_{1b} \cos^{\epsilon_1} \theta_s \cos^{\epsilon_2} \phi_s, a_{2r} \cos^{\epsilon_1} \theta_s \sin^{\epsilon_2} \phi_s, a_3 \sin^{\epsilon_1} \theta_s), & 0 \leqslant \theta_s < \pi/2, -\pi \leqslant \phi_s < -\pi/2. \end{cases} \quad (4)$$

In a knowledge-based digital human model such as the one introduced by Graff [64, 65], components of the model are described by parameterized surface expressions. For instance, the breast surface is modeled as superquadric surfaces parameterized via polar angles ϕ_s and θ_s as shown in equation (4). Parameter set a refers to the volume of the breast for bottom, top, right and left hemispheres and length of the breast, while ϵ adjusts the shape. An example of one random realization for the breast surface obtained from Graff's model is shown in figure 2. In this case, the model is constructed by randomly sampling a set of parameters a and ϵ (equivalent to the basis functions in equation (3)) to obtain a distribution of shapes with tunable variability.

A similar approach by Bliznakova *et al* [37] describes a 3D breast software model for x-ray breast imaging simulations based on a breast external shape, ductal lobular system, Cooper's ligaments and pectoralis muscle. In this approach, a mammographic background texture is added to the tissue regions. Blood vessels, nerves and lymphatics were not modeled explicitly. A similar, more simplistic approach, was developed by Bakic *et al* [66] based on two ellipsoidal regions of large scale tissue elements: predominantly adipose tissue and predominantly fibro-glandular tissue. Internal tissue structures within these regions are approximated by a distribution of elements including shells, blobs, and a ductal tree. Similar approaches have been reported for full-body models [47].

6. Modeling disease

Disease states can be incorporated into digital cohorts using image-based methods or object-space models of the condition. The analogy between digital human models and disease models can be established if we consider lesions as continuous variables in space (r) and time (t), described by a coefficient vector affecting a set of lesion model characteristics. For simplicity, we will consider the disease independent (of the underlying anatomy where the disease is located) and additive. This assumption allows us to represent the disease cases as a sum of the stochastic human model and the disease component, an addition that is typically performed in the voxelized object model or directly within the model images. We recognize this approach is a known simplification, as disease processes often have significant impact on underlying tissues.

Analogously to the description provided by equation (3), we can generate a set of disease models $\{\mathbf{d}_s\}$ defined by:

$$\{\mathbf{d}_s\}_{s=1}^S = \sum_n \lambda_n^s \psi_n\left(\boldsymbol{r}, \boldsymbol{t}\right), \tag{5}$$

where λ_n^s is a disease characteristics coefficient vector described by the function ψ_n over N parameters. Characteristics that define lesions can include geometric functions (e.g. size, morphology), material properties (e.g. x-ray interaction cross-sections, elasticity) or other relevant features (e.g. radioactivity, blood oxygenation levels).

Methodologies for generating and incorporating disease into cohorts of digital stochastic models rely on sampling λ_n and ψ_n from appropriate distributions representing the intended population. In some cases, disease models are specific to a given anatomical location or physiology corresponding to a digital human exemplar. In other cases, disease models are independent of the digital healthy human and are simply added or inserted multiple times into models of healthy anatomy. In both cases, diseased subjects are denoted by a cohort of digital stochastic humans with added disease components:

$$\{\mathbf{f}_s\}_{s=1}^S = \sum_n \theta_n^s \phi_n\left(\boldsymbol{r}\right) + \sum_n \lambda_n^s \psi_n\left(\boldsymbol{r}\right), \tag{6}$$

where $\{\mathbf{f}_s\}_{s=1}^S$ is a cohort of diseased digital humans (for simplicity, and similarly as in the previous section, we choose to omit the time dimension). Similarly to normal models, when ψ_n are unknown, models of disease can be obtained relying on imaging. Alternatively, when ψ_n are known, analytically or numerically, the stochastic disease models are referred to as knowledge-based (also known as procedural).

6.1. Image-based models of disease
Similar to image-based models of the human body, image-based models of disease rely on imaging data for extracting lesion information. Various techniques for capturing disease characteristics, particularly for breast lesions, have recently been explored [67, 68]. Image-based neural network models for disease modeling have also been explored. For instance, Kadia *et al* [69] proposed a method to generate synthetic, infection-like patterns in the lung to create large collections of 2D and 3D training examples for deep segmentation models. While image-based models contain features from actual patient data and thus may look more realistic at first glance, they suffer from limited resolution of the tumor model, largely determined by the imaging acquisition characteristics and limited number of available lesion morphologies, shapes, and sizes. In addition, image-based methods require an institutional review board approval for obtaining and utilizing the diseased case data for research and development, which could delay or disadvantage some analysis efforts.

6.2. Knowledge-based models of disease
Knowledge-based models of disease are constructed by sampling a set of known (or assumed known) ψ_n and λ_n in equation (5) from distributions representing the relevant characteristics of the disease, where distributions are often derived from physical or biological measurements. In contrast to image-based models, knowledge-based models enable the generation of unlimited numbers of lesion shapes with variable resolution. Examples of knowledge-based models include de Sisternes [70] spiculated breast cancer mass model and Sengupta [71] growing breast mass models. In [71], a breast lesion growth method based on biological and physiological phenomena accounting for the stiffness of surrounding anatomical structures was introduced. Breast ligaments were considered as rigid structures with elastic moduli in the range of 8×10^4–4×10^5 kPa, while fat (elastic modulus varying from 0.5 to 25 kPa) and glandular tissues (elastic modulus varying from 7.5 to 66 kPa) constituting the more elastic regions of the breast. In this approach, tumor cells are less likely to grow through stiffer structures and instead, preferentially proliferate through the more elastic regions of the breast. Depending on the breast local anatomical structures, a range of unique lesion morphologies can be realized, allowing lesions to blend naturally into the anatomical regions.

A common simplifying assumption is to define the disease model independent from other human model components. For example, in VICTRE [15] and in Sengupta *et al* [72], breast cancer mass lesions are added to the normal breast models by replacing voxels in the breast with voxels of the lesion model, without modification to adjacent voxels. This approach, while practical, does not account for the significant effect of the growing tumors on its surrounding tissues, typically visible in x-ray images as architectural distortions suggestive of abnormalities. To consider these effects, equation (6) needs to be modified to account for the interaction between normal and disease models.

7. Role of augmentation methods

Augmentation methods start with an already-defined object, image or a set of defined objects, and generate new examples based on properties of inputs, as well as pre-defined or data-driven transformations (in contrast, digital human models start with only an object description, such as that given in equation (1)). GAN-based models (see section 5.1.2) are similar to augmentation methods in that they employ complex transformations derived with the help of training data sets. Augmentation methods typically employ analytically-defined or stochastic operators that do not require the use of neural networks, and can be applied both in the object domain and in the acquired image domain. Techniques in the latter group generate examples that could be obtained through an imaging system applied to an object with an accompanying degradation (e.g. smoothing, noise, reconstruction artifacts).

Geometric transformations, intensity operations, and spatial filtering are among the most basic types of augmentation methods. Geometric transformations redefine the spatial relationships among voxels or geometrical locations in an object, and include affine (scaling, rotation, translation, reflection and shearing), as well as non-affine transformations, such as non-linear warping and morphing [73]. Intensity operations modify intensity values in a grayscale image or channel values (e.g. RGB or CMYK) in a color image. Examples include operations such as a family of gamma corrections, linear contrast adjustments, and remapping voxel values using a pre-defined or pseudo-random remapping curve [74, 75]. Spatial filtering (using a filter mask) is another possibility for generating a new image or object based on an existing one. Spatial filtering can be linear (in which case it can be implemented by a convolution operation) or non-linear (e.g. median filtering), and can be implemented to smooth or sharpen to emphasize certain features. Finally, all three types of augmentations can be combined using a continuous mapping from the parameter space of transformations to the image or object space [76].

Noise injection is an image augmentation method that enhances robustness of machine learning models and belongs to the family of domain randomization methods [77]. Although noise injection after data acquisition does not generate a new member of a patient population, it can generate a different representation of an object in the image domain, and can be useful for augmenting patient cohorts obtained with *in silico* modeling. Some earlier and non-medical applications of noise injection in machine learning sought to augment the image data sets without regard to the physics of image acquisition [78, 79]. Other works used physics-based techniques for noise modeling and addition, improving realism of the noise appearance in the augmented images [80, 81]. The main benefit of noise injection in the image domain for *in silico* trials is that it may allow for the rapid generation of different representations of the same object at different noise levels, leading to comparisons that may require less computational power compared to a full implementation of image acquisition physics applied to a digital stochastic object model. Addition of texture to a model in the object domain has similarities to noise injection in the image domain in that both techniques aim at producing noise-like properties (e.g. using a noise power spectrum in modeling), but are different in that addition of texture in the object domain does not attempt to model the noise from data acquisition [82].

Combination of objects or images is another popular augmentation technique. In the object domain, combination of an object model for a normal (non-diseased) patient with a lesion model (as described in section 6) can be thought of as an example of this type of augmentation. Generating new members of a patient population based on an eigenspace analysis of existing patient objects, as was done in [52] and described in section 5.1.1 is another example of augmentation in the object domain. In the image domain, researchers investigated tools for the extraction of image parts from one clinical image and then their insertion into a new location on the same or different image. Pezeshk *et al* [83] used an image blending technique based on Poisson image editing to insert pulmonary nodules extracted from one chest CT exam into another. Augmenting a training data set for a machine learning model using this technique can improve the model performance on independent, real test datasets [84]. Likewise, Ghanian *et al* [85] used a similar technique to insert microcalcification clusters extracted from one mammogram into another mammogram, and showed that experienced observers cannot reliably distinguish between computationally inserted and native clusters. Besides the ability to convince experts, desirable properties for such combination techniques

include acceptable noise properties in the combined image, plausible lesion-background combinations (that might require the intervention of an operator during the augmentation process), and a sufficient range of variation in the combined images that can be generated, which are often difficult to satisfy simultaneously.

The main advantage of data augmentation methods is their practicality. For example, existing images or models both for normal and diseased patients can be manipulated (with relative ease) with geometric transformations leading to expanded patient representations. When implemented in the image domain, augmentation methods are fast, bypassing the stage where a model for the imaging system is applied to the object to yield an image. However, important shortcomings accompany these advantages. Unless deliberate attention is paid, augmentation methods may yield objects or images that are biologically or physically implausible. An extreme example may be an intensity transformation that results in bones with lower Hounsfield units than soft tissue. Although this can be avoided easily by using an intensity transformation that is monotonically increasing, most augmentation methods and transformations need careful planning to avoid such inconsistencies, and it may not be possible to avoid all inconsistencies. The consequences of such implausible images or objects on the results of an *in silico* imaging trial should be carefully considered. In addition, many augmentation techniques do not result in an independent, new representation from the population, but rather in representations that are highly dependent on the original objects or images used as inputs to the augmentation method. For example, lesion insertion methods described in the previous paragraph do not increase the number of lesions in the augmented data set, but only the lesion-background combinations that are generated. Again, the consequences of this limitation in the range of variation of generated images should be an important consideration in an *in silico* imaging trial that uses augmentation.

8. Considerations for sampling digital cohorts

In silico studies require careful study planning and good clinical trial design. Even if and when methodologies for developing digital stochastic models of humans for imaging studies become widely available, enrolling digital cohorts needs an understanding of the trade-offs and potential for bias associated with selecting a specific distribution of study subjects. At the start of the design of an *in silico* imaging trial is the challenging task of scoping the population of the digital humans to be included in the study. For instance, a number of previous computational studies in breast imaging using procedural models used a uniform sampling with a desired average of 50% adipose and 50% fibroglandular voxels [86] with an uncompressed breast size of 14 cm. Another example of enrollment strategy can be found in the OpenVCT platform, where a range of size and glandularity is specified and then uniformly and randomly sampled [87]. A more recent in silico imaging study used sampling from a multi-class distribution identifying four different breast densities resulting in the characteristics of the intended population [15].

The importance of sampling strategy for digital cohorts was highlighted by Romero *et al* [88]. Romero described a study comparing bootstrapping, GAN and Gaussian sampling methods for the generation of digital cohorts of aortic aneurysm geometry. The sampling based on a GAN approach achieved the highest efficiency (i.e. ratio of generated cases deemed as acceptable and belonging to the target population), but was sensitive to sample size and susceptible to losing statistical properties of the sample. On the other hand, sampling based on bootstrapping and Gaussian distribution sampling were less efficient, but better captured statistical properties, with Gaussian distribution more suited for sampling underrepresented cases (i.e. distribution tails). Their study emphasized the need to capture the variability in the data to a degree that depends on the goals of the *in silico* study. In particular, it is important to determine if the goal of the study is to replicate the variability seen in a clinical trial population or, rather, to investigate the performance of a device across all population subgroups.

Through *in silico* enrollment, digital cohorts $\{\mathbf{f}_s\}_{s=1}^{S}$ are generated. We denote the distribution of the population of digital humans as $\mathbf{f}_d$, where d represents the digital world, and the distribution of subjects in the intended population as $\mathbf{f}_i$. In this context, the goal of the *in silico* enrollment is to minimize the difference $\Delta \mathbf{f} = |\mathbf{f}_d - \mathbf{f}_i|$ between the digital (d) and physical-world intended distributions, where $|.|$ denotes a statistical distance measure. Clinical trial enrollment programs in the physical world require strategies to ensure a reasonable $\Delta \mathbf{f}$ given available sampling resources. The goal of *in silico* enrollment is to capture the intended distribution to a greater extent than possible in the corresponding clinical trial patient enrollment.

Analyzing $\Delta \mathbf{f}$ corresponding to an *in silico* enrollment strategy may be needed to understand how the difference across study subject characteristics could affect the outcome of the trial. Here, we discuss a simplified test case scenario (see figure 3) that compares different enrollment strategies for an *in silico* trial comparing digital mammography and digital breast tomosynthesis (DBT) derived from the VICTRE [15] project. In this section, we utilize metrics and analysis analogous to those used in the VICTRE study. Here, we assume the populations (digital and physical) consist of normal and diseased subjects with a prevalence of 0.5. These two classes of patients are therefore sampled with equal probability. We calculate the difference of
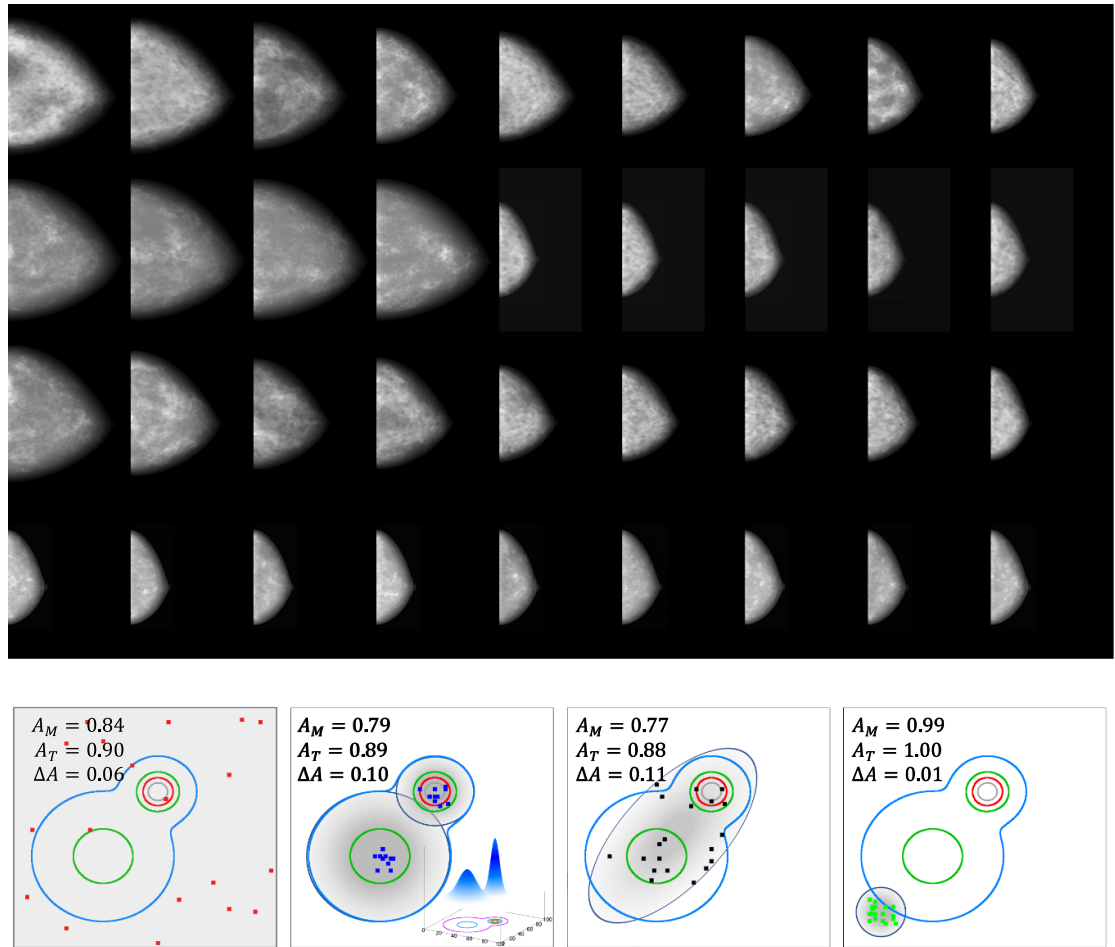
Figure 3. Effect of sampling strategies on performance assessment. Top: sample mammographic images from the four different distributions considered in this work, from top to bottom: uniform, matched, simpler, and narrow. Bottom: sampling is from a bimodal distribution of subjects (seen in 3D insert in the second panel from the left) described by two random parameters: (from left to right) uniform, matched, simpler, and narrow. Only 20 samples are shown here for ease of visualization. The gray shading depicts the distribution from which samples are taken in each of the four cases. The contour lines indicate 0.1, 0.01, and 0.001 levels with respect to the maximum in the distributions. Dots represent data points illustrating the coordinates of sample cases. A_M, A_T, and ΔA refer to the lesion detection average AUC for mammography, average AUC for digital breast tomosynthesis, and the average AUC difference, respectively.

performance (measured using the area under the receiver operating characteristic curve, or AUC, in the task of differentiating between normal and disease subjects) between mammography and DBT. We consider the following four sampling approaches. In the first approach (uniform), $\mathbf{f}_i$ is unknown and subjects are sampled uniformly within a range of interest, from all possible combinations of the input parameters that define $\mathbf{f}$. In the second approach (matched), $\mathbf{f}_i$ is known and subjects are sampled from the true underlying distribution. In the third approach (simpler), $\mathbf{f}_i$ is unknown, but can be approximated by another, simpler distribution from which samples are obtained. Finally, in the fourth approach (narrow), $\mathbf{f}_i$ is known to be a narrow, well-defined subset of the general population of subjects of particular interest (e.g. rare diseases or very obese subjects).

For this simplified example, let $\mathbf{f}_i$ be a bimodal distribution defined by two parameters (e.g. breast size and glandularity). Using equation (3), we can express the model through two expansion functions $\phi_{1,2}$, each associated with one of the two random variables affected by a random parameter set given by $\theta_{1,2}$. As seen in figure 3, one of the modes of the distribution has twice the amplitude and half the variance of the other. The four density plots illustrate a top view of the distribution contour plot with the individual samples drawn using the four different sampling strategies. The results demonstrate that the choice of sampling strategy can have a significant effect on the difference in AUC, which for this example case, ranges from a difference of 0.01 (almost zero) to 0.11 in terms of device performance.

A graphical representation of the populations obtained for each sampling approach is depicted in figure 3) where the model variability from each sampling approach are observed. The first row shows samples from the uniform approach with variability across size and density. The second row shows the

matched approach with samples that are drawn primarily from the two components of the bimodal distribution $\mathbf{f}_i$. The third row depicts more gradual variability as the sampling draws examples from wider distributions around the peaks, while the fourth row clearly depicts samples from a narrow distribution with well-defined size and density. It is notable that a relatively small change in the population of models obtained from using different sampling strategies yields measurable and significant change in the figure-of-merit metric used to compare imaging technologies.

The effect of patient distributions on the results of a trial has been well documented [89, 90]. Depending on the objective of the trial or the claims of the device, the sampling strategy needs to be considered. Full statistical representation of all possible cases (or completeness) of the population participating in a clinical trial is an onerous target. However, digital human cohorts built stochastically from *in silico* models offer the potential to reduce bias [91], either by complementing the patient population in a clinical trial [92] or by providing a diverse all-*in-silico* population [15, 93]. The test case described in this section is an example of how validated and detailed computational tools can be leveraged to simulate and analyze rare cases or underrepresented populations. Specifically, we show that by independently sampling breast density and breast size one can create cases that would fall far from the average, rendering them hard to obtain in a real-world clinical trial.

9. Summary and conclusions

In silico trials are an emerging area of regulatory research that offer the ability to capture highly diverse patient distributions at a significant time and cost savings, compared to traditional physical clinical trials. To conduct *in silico* trials, realistic digital representations of humans are needed. In this paper, we reviewed and discussed existing techniques for generating digital humans, including disease models, for *in silico* imaging trials. Digital humans can be created using image-based or knowledge-based techniques. In summary, we favor techniques with object-based representations (rather than images of objects) in order to decouple the characteristics of the image acquisition system from the characteristics of the object (true representation of the physical-world human). In generating digital humans for *in silico* trials, one should consider the quality and quantity of the source data or knowledge used, and whether the models represent a single patient, a small cohort, or a sizable population with realistic patient variability.

It remains a crucial next step to evaluate the quality of the digital human models and the images that can be generated with them. In particular, it is essential to carefully identify the patient distribution that the particular digital human model can and cannot capture, in order to prevent misuse and ensure patient safety. We need to study to what extent model-derived data contributes to our understanding of performance levels for populations with rare diseases or for populations underrepresented in traditional clinical trials. Future work should examine the ethical and safety considerations of relying on digital humans for clinical trials. Overall, the use of *in silico* imaging trials and *in silico* trials in medicine is a rapidly developing field and has the potential to address many of the emerging challenges in the regulatory evaluation of medical devices.

Data availability statement

All data that support the findings of this study are included within the article (and any supplementary files).

ORCID iDs

A Badano ⓘ https://orcid.org/0000-0003-3712-6670
M A Anastasio ⓘ https://orcid.org/0000-0002-3192-4172

References

[1] Barrett H H and Myers K J 2013 *Foundations of Image Science* (Wiley)
[2] Magnenat-Thalmann N and Thalmann D 2005 *Handbook of Virtual Humans* (Wiley)
[3] Dosovitskiy A, Ros G, Codevilla F, Lopez A and Koltun V 2017 CARLA: an open urban driving simulator *Conf. on Robot Learning* (PMLR) pp 1–16
[4] Cimino C, Negri E and Fumagalli L 2019 Review of digital twin applications in manufacturing *Comput. Ind.* **113** 103130
[5] Tao F, Zhang H, Liu A and Nee A Y 2018 Digital twin in industry: state-of-the-art *IEEE Trans. Ind. Inform.* **15** 2405–15
[6] Thelen A, Zhang X, Fink O, Lu Y, Ghosh S, Youn B D, Todd M D, Mahadevan S, Hu C and Hu Z 2022 A comprehensive review of digital twin—part 1: modeling and twinning enabling technologies *Struct. Multidiscip. Optim.* **65** 354
[7] Fan J, Liu X, Li Y, Xia H, Yang R, Li J and Zhang Y 2022 Quality problems of clinical trials in China: evidence from quality related studies *Trials* **23** 1–11
[8] Food U *et al* 2022 Diversity plans to improve enrollment of participants from underrepresented racial and ethnic populations in clinical trials; draft guidance for industry; availability (available at: www.fda.gov/regulatory-information/search-fda-guidance-documents/diversity-plans-improve-enrollment-participants-underrepresented-racial-and-ethnic-populations)

[9] Rajotte J-F, Bergen R, Buckeridge D L, El Emam K, Ng R and Strome E 2022 Synthetic data as an enabler for machine learning applications in medicine *iScience* **25** 11

[10] Abadi E, Segars W P, Tsui B M, Kinahan P E, Bottenus N, Frangi A F, Maidment A, Lo J and Samei E 2020 Virtual clinical trials in medical imaging: a review *J. Med. Imaging* **7** 042805

[11] Badano A 2021 *In silico* imaging clinical trials: cheaper, faster, better, safer and more scalable *Trials* **22** 1–7

[12] Abadi E, Segars W P, Sturgeon G M, Roos J E, Ravin C E and Samei E 2018 Modeling lung architecture in the XCAT series of phantoms: physiologically based airways, arteries and veins *IEEE Trans. Med. Imaging* **37** 693–702

[13] Wedlund L and Kvedar J 2021 Simulated trials: *in silico* approach adds depth and nuance to the RCT gold-standard *npj Dig. Med.* **4** 121

[14] Segars W P and Tsui B M 2009 MCAT to XCAT: the evolution of 4-D computerized phantoms for imaging research *Proc. IEEE* **97** 1954–68

[15] Badano A, Graff C G, Badal A, Sharma D, Zeng R, Samuelson F W, Glick S J and Myers K J 2018 Evaluation of digital breast tomosynthesis as replacement of full-field digital mammography using an *in silico* imaging trial *JAMA Netw. Open* **1** e18547

[16] Pepe M 2005 Evaluating technologies for classification and prediction in medicine *Stat. Med.* **24** 3687–96

[17] Arifin W N and Yusof U K 2022 Correcting for partial verification bias in diagnostic accuracy studies: a tutorial using R *Stat. Med.* **41** 1709–27

[18] Berti F, Antonini L, Poletti G, Fiuza C, Vaughan T J, Migliavacca F, Petrini L and Pennati G 2021 How to validate *in silico* deployment of coronary stents: strategies and limitations in the choice of comparator *Front. Med. Technol.* **3** 37

[19] Barrett H H and Caucci L 2020 Stochastic models for objects and images in oncology and virology: application to PI3K-Akt-mTOR signaling and COVID-19 disease *J. Med. Imaging* **8** S16001

[20] Shepp L A and Logan B F 1974 The Fourier reconstruction of a head section *IEEE Trans. Nucl. Sci.* **21** 21–43

[21] Martin J, Ruthven M, Boubertakh R and Miquel M E 2020 Realistic dynamic numerical phantom for MRI of the upper vocal tract *J. Imaging* **6** 86

[22] Snyder W S, Ford M R, Warner G G and Fisher H Jr 1969 Estimates of absorbed fractions for monoenergetic photon sources uniformly distributed in various organs of a heterogeneous phantom *Technical Report* (Oak Ridge National Lab., Tenn.)

[23] Caon M 2004 Voxel-based computational models of real human anatomy: a review *Radiat. Environ. Biophys.* **42** 229–35

[24] Zu X G 2005 The VIP-man model-a digital human testbed for radiation simulations *SAE Trans.* 779–87 (available at: www.sae.org/publications/technical-papers/content/2005-01-2713/)

[25] George Xu X 2014 Computational phantoms for organ dose calculations in radiation protection and imaging *The Phantoms of Medical and Health Physics* (Springer) pp 225–62

[26] Fu W, Sharma S, Abadi E, Iliopoulos A-S, Wang Q, Lo J Y, Sun X, Segars W P and Samei E 2021 iPhantom: a framework for automated creation of individualized computational phantoms and its application to ct organ dosimetry *IEEE J. Biomed. Health Inform.* **25** 3061–72

[27] Kamel Boulos M N and Zhang P 2021 Digital twins: from personalised medicine to precision public health *J. Pers. Med.* **11** 745

[28] Jirsa V, Wang H, Triebkorn P, Hashemi M, Jha J, Gonzalez-Martinez J, Guye M, Makhalova J and Bartolomei F 2023 Personalised virtual brain models in epilepsy *Lancet Neurol.* **22** 443–54

[29] Pesapane F, Rotili A, Penco S, Nicosia L and Cassano E 2022 Digital twins in radiology *J. Clin. Med.* **11** 6553

[30] Erol T, Mendi A F and Doğan D 2020 The digital twin revolution in healthcare *2020 4th Int. Symp. on Multidisciplinary Studies and Innovative Technologies (ISMSIT)* (IEEE) pp 1–7

[31] Wang G *et al* 2022 Development of metaverse for intelligent healthcare *Nat. Mach. Intell.* **4** 9220929

[32] Spitzer V, Ackerman M J, Scherzinger A L and Whitlock D 1996 The visible human male: a technical report *J. Am. Med. Inform. Assoc.* **3** 118–30

[33] Christ A *et al* 2009 The virtual family, development of surface-based anatomical models of two adults and two children for dosimetric simulations *Phys. Med. Biol.* **55** N23

[34] Fujimoto K, Zaidi T A, Lampman D, Guag J W, Etheridge S, Habara H and Rajan S S 2022 Comparison of SAR distribution of hip and knee implantable devices in 1.5T conventional cylindrical-bore and 1.2T open-bore vertical MRI systems *Magn. Reson. Med.* **87** 1515–28

[35] Duetschler A *et al* 2022 Synthetic 4DCT(MRI) lung phantom generation for 4D radiotherapy and image guidance investigations *Med. Phys.* **49** 2890–903

[36] Kiarashi N, Nolte A, Sturgeon G M, Segars W P, Ghate S V, Nolte L W, Samei E and Lo J Y 2014 Development and application of a suite of 4-D virtual breast phantoms for optimization and evaluation of breast imaging systems *IEEE Trans. Med. Imaging* **33** 1401–9

[37] Bliznakova K, Bliznakov Z, Bravou V, Kolitsi Z and Pallikarakis N 2003 A three-dimensional breast software phantom for mammography simulation *Phys. Med. Biol.* **48** 3699

[38] Li C M, Segars W P, Tourassi G D, Boone J M and Dobbins III J T 2009 Methodology for generating a 3D computerized breast phantom from empirical data *Med. Phys.* **36** 3122–31

[39] Erickson D W, Wells J R, Sturgeon G M, Samei E, Dobbins J T, Segars W P and Lo J Y 2016 Population of 224 realistic human subject-based computational breast phantoms *Med. Phys.* **43** 23–32

[40] Hsu C M, Palmeri M L, Segars W P, Veress A I and Dobbins III J T 2013 Generation of a suite of 3D computer-generated breast phantoms from a limited set of human subject data *Med. Phys.* **40** 043703

[41] Elangovan P, Mackenzie A, Dance D R, Young K C, Cooke V, Wilkinson L, Given-Wilson R M, Wallis M G and Wells K 2017 Design and validation of realistic breast models for use in multiple alternative forced choice virtual clinical trials *Phys. Med. Biol.* **62** 2778

[42] García E, Fedon C, Caballo M, Martí R, Sechopoulos I and Diaz O 2020 Realistic compressed breast phantoms for medical physics applications *Proc. SPIE* **11513** 30–37

[43] Sarno A, Mettivier G, di Franco F, Varallo A, Bliznakova K, Hernandez A M, Boone J M and Russo P 2021 Dataset of patient-derived digital breast phantoms for *in silico* studies in breast computed tomography, digital breast tomosynthesis and digital mammography *Med. Phys.* **48** 2682–93

[44] Caballo M, Rabin C, Fedon C, Rodríguez-Ruiz A, Diaz O, Boone J M, Dance D R and Sechopoulos I 2022 Patient-derived heterogeneous breast phantoms for advanced dosimetry in mammography and tomosynthesis *Med. Phys.* **49** 5423–38

[45] Sauer T J, Abadi E, Segars P and Samei E 2022 Anatomically and physiologically informed computational model of hepatic contrast perfusion for virtual imaging trials *Med. Phys.* **49** 2938–51

[46] Tam L K, Stockmann J P, Galiana G and Constable R T 2012 Null space imaging: nonlinear magnetic encoding fields designed complementary to receiver coil sensitivities for improved acceleration in parallel imaging *Magn. Reson. Med.* **68** 1166–75

[47] Lee C, Lodwick D, Hasenauer D, Williams J L, Lee C and Bolch W E 2007 Hybrid computational phantoms of the male and female newborn patient: NUBRS-based whole-body models *Phys. Med. Biol.* **52** 3309

[48] Segars W P, Sturgeon G, Mendonca S, Grimes J and Tsui B M 2010 4D XCAT phantom for multimodality imaging research *Med. Phys.* **37** 4902–15

[49] Kainz W *et al* 2018 Advances in computational human phantoms and their applications in biomedical engineering-a topical review *IEEE Trans. Radiat. Plasma Med. Sci.* **3** 1–23

[50] Johnson A E, Pollard T J, Berkowitz S J, Greenbaum N R, Lungren M P, Deng C-y, Mark R G and Horng S 2019 Mimic-CXR, a de-identified publicly available database of chest radiographs with free-text reports *Sci. Data* **6** 1–8

[51] Medical Imaging and Data Resource Center (MIDRC) (available at: www.midrc.org/) (Accessed 10 January 2023)

[52] Sturgeon G M, Park S, Segars W P and Lo J Y 2017 Synthetic breast phantoms from patient based eigenbreasts *Med. Phys.* **44** 6270–9

[53] Lewis J P, Cordner M and Fong N 2000 Pose space deformation: a unified approach to shape interpolation and skeleton-driven deformation *Proc. 27th Annual Conf. on Computer Graphics and Interactive Techniques* pp 165–72

[54] Chen J, Li Y, Du Y and Frey E C 2020 Generating anthropomorphic phantoms using fully unsupervised deformable image registration with convolutional neural networks *Med. Phys.* **47** 6366–80

[55] Galbusera F, Niemeyer F, Seyfried M, Bassani T, Casaroli G, Kienle A and Wilke H-J 2018 Exploring the potential of generative adversarial networks for synthesizing radiological images of the spine to be used in *in silico* trials *Front. Bioeng. Biotechnol.* **6** 53

[56] Kingma D P and Welling M 2013 Auto-encoding variational Bayes (arXiv:1312.6114)

[57] Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, Courville A and Bengio Y 2020 Generative adversarial networks *Commun. ACM* **63** 139–44

[58] Singh N K and Raza K 2021 Medical image generation using generative adversarial networks: a review *Health Informatics: A Computational Perspective in Healthcare* (*Studies in Computational Intelligence* vol 932) ed R Patgiri, A Biswas and P Roy (Springer) pp 77–96

[59] Rezende D and Mohamed S 2015 Variational inference with normalizing flows *Int. Conf. on Machine Learning* (PMLR) pp 1530–8

[60] Ho J, Jain A and Abbeel P 2020 Denoising diffusion probabilistic models *Advances in Neural Information Processing Systems* vol 33 pp 6840–51

[61] Croitoru F-A, Hondru V, Ionescu R T and Shah M 2022 Diffusion models in vision: a survey (arXiv:2209.04747)

[62] Dhariwal P and Nichol A 2021 Diffusion models beat GANs on image synthesis *Advances in Neural Information Processing Systems* vol 34 pp 8780–94

[63] Zhou W, Bhadra S, Brooks F J, Li H and Anastasio M A 2022 Learning stochastic object models from medical imaging measurements by use of advanced ambient generative adversarial networks *J. Med. Imaging* **9** 015503

[64] Graff C G 2016 A new, open-source, multi-modality digital breast phantom *Proc. SPIE* **9783** 72–81

[65] Chen D T, Kakadiaris I A, Miller M J, Loftin R B and Patrick C 2000 Modeling for plastic and reconstructive breast surgery *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2000: 3rd Int. Conf., (Pittsburgh, PA, USA, 11–14 October 2000) (Proc. 3)* (Springer) pp 1040–50

[66] Bakic P R, Myers K J, Glick S J and Maidment A D 2016 *Virtual Tools for the Evaluation of Breast Imaging: State-of-the Science and Future Directions* vol 6 (Springer) pp 518–24

[67] Dukov N, Bliznakova K, Feradov F, Buliev I, Bosmans H, Mettivier G, Russo P, Cockmartin L and Bliznakov Z 2019 Models of breast lesions based on three-dimensional x-ray breast images *Phys. Med.* **57** 80–87

[68] Bliznakova K *et al* 2019 Development of breast lesions models database *Phys. Med.* **64** 293–303

[69] Kadia D, Nguyen T V and Asari V 2022 Lesion synthesis for robust segmentation of infected lung region on small-scale data *SSRN* (available at: http://dx.doi.org/10.2139/ssrn.4029426)

[70] de Sisternes L, Brankov J G, Zysk A M, Schmidt R A, Nishikawa R M and Wernick M N 2015 A computational model to generate simulated three-dimensional breast masses *Med. Phys.* **42** 1098–118

[71] Sengupta A, Sharma D and Badano A 2021 Computational model of tumor growth for *in silico* trials *Proc. SPIE* **11595** 1262–70

[72] Sengupta A, Sharma D and Badano A 2021 Computational model of tumor growth for *in silico* trials *Proc. SPIE* **11595** 115954S

[73] Wolberg G 1988 Geometric transformation techniques for digital images: a survey

[74] Chlap P, Min H, Vandenberg N, Dowling J, Holloway L and Haworth A 2021 A review of medical image data augmentation techniques for deep learning applications *J. Med. Imaging Radiat. Oncol.* **65** 545–63

[75] Hesse L S, Kuling G, Veta M and Martel A L 2021 Intensity augmentation to improve generalizability of breast segmentation across different MRI scan protocols *IEEE Trans. Biomed. Eng.* **68** 759–70

[76] Tian K, Lin C, Lim S N, Ouyang W, Dokania P and Torr P 2021 A Continuous Mapping for Augmentation Design *Advances in Neural Information Processing Systems* vol 34, ed M Ranzato, A Beygelzimer, Y Dauphin, P Liang and J W Vaughan (Curran Associates, Inc) pp 13732–43

[77] Noh H, You T, Mun J and Han B 2017 Regularizing deep neural networks by noise: its interpretation and optimization (arXiv:1710.05179)

[78] Moreno-Barea F J, Strazzera F, Jerez J M, Urda D and Franco L 2018 Forward noise adjustment scheme for data augmentation *2018 IEEE Symp. Series on Computational Intelligence (SSCI)* pp 728–34

[79] Bae H-J, Kim C-W, Kim N, Park B, Kim N, Seo J B and Lee S M 2018 A Perlin noise-based augmentation strategy for deep learning with small data samples of HRCT images *Sci. Rep.* **8** 17687

[80] Omigbodun A O, Noo F, McNitt-Gray M, Hsu W and Hsieh S S 2019 The effects of physics-based data augmentation on the generalizability of deep neural networks: demonstration on nodule false-positive reduction *Med. Phys.* **46** 4563–74

[81] Fabian Z, Heckel R and Soltanolkotabi M 2021 Data augmentation for deep learning based accelerated MRI reconstruction with limited data *Proc. 38th Int. Conf. on Machine Learning* (*Proc. Machine Learning Research* vol 139) ed M Meila and T Zhang (PMLR) pp 3057–67

[82] Abadi E, Segars W P, Sturgeon G M, Harrawood B, Kapadia A and Samei E 2019 Modeling 'textured' bones in virtual human phantoms *IEEE Trans. Radiat. Plasma Med. Sci.* **3** 47–53

[83] Pezeshk A, Sahiner B, Zeng R, Wunderlich A, Chen W and Petrick N 2015 Seamless insertion of pulmonary nodules in chest CT images *IEEE Trans. Biomed. Eng.* **62** 2812–27

[84] Pezeshk A, Petrick N, Chen W and Sahiner B 2017 Seamless lesion insertion for data augmentation in CAD training *IEEE Trans. Med. Imaging* **36** 1005–15

[85] Ghanian Z, Pezeshk A, Petrick N and Sahiner B 2018 Computational insertion of microcalcification clusters on mammograms: reader differentiation from native clusters and computer-aided detection comparison *J. Med. Imaging* **5** 1

[86] Gong X, Glick S J, Liu B, Vedula A A and Thacker S 2006 A computer simulation study comparing lesion detection accuracy with digital mammography, breast tomosynthesis and cone-beam CT breast imaging *Med. Phys.* **33** 1041–52

[87] Barufaldi B, Higginbotham D, Bakic P R and Maidment A D A 2018 OpenVCT: a GPU-accelerated virtual clinical trial pipeline for mammography and digital breast tomosynthesis *Proc. SPIE* **10573** 1057358

[88] Romero P, Lozano M, Martínez-Gil F, Serra D, Sebastián R, Lamata P and García-Fernández I 2021 Clinically-driven virtual patient cohorts generation: an application to aorta *Front. Physiol.* **12** 713118

[89] Norori N, Hu Q, Aellen F M, Faraci F D and Tzovara A 2021 Addressing bias in big data and AI for health care: a call for open science *Patterns* **2** 100347

[90] Celi L A *et al* 2022 Sources of bias in artificial intelligence that perpetuate healthcare disparities-a global review *PLoS Digit. Health* **1** e0000022

[91] Sinisi S, Alimguzhin V, Mancini T, Tronci E and Leeners B 2020 Complete populations of virtual patients for *in silico* clinical trials *Bioinformatics* **36** 5465–72

[92] Pappalardo F, Russo G, Tshinanu F M and Viceconti M 2019 *In silico* clinical trials: concepts and early adoptions *Briefings Bioinform.* **20** 1699–708

[93] Myles P, Ordish J and Tucker A 2023 The potential synergies between synthetic data and *in silico* trials in relation to generating representative virtual population cohorts *Prog. Biomed. Eng.* **5** 013001