Regulatory Barriers to the adoption of in silico trials

Lowering the regulatory barriers that are slowing the adoption of computer modelling and simulation in the development and de-risking of medical products

INSILICO WORLD

Open Access report on the activities of the In Silico World Consortium

1. Executive Summary

"In Silico World" is a project funded by the European Commission through the H2020 framework programme. It aims to identify and lower all the barriers slowing the adoption of modelling and simulation methodologies in the regulatory evaluation of biomedical products (sometimes referred to as **In Silico Trials**). This extensive report summarises and analyses the consortium's work concerning the regulatory barriers and draws some recommendations for EU policymakers, which are summarised here:

The first recommendation is *separation*. The European Medicines Agency (EMA) should consider separating the panels that evaluate the marketing authorisation of new drugs from those that provide qualification advice on new methodologies. While we were writing this document, the EMA announced the new Methodology Working Party[1].

The second recommendation is *interdisciplinarity*. The composition of the new panel providing qualification advice on new methodologies must include experts with a background in physical sciences (biophysicists, bioengineers, computer scientists, etc.). Medicine is becoming an interdisciplinary science, and regulatory agencies must adapt. Here, there is a long way to go: the new Methodology Working Party includes one physicist out of 21 experts, no bioengineers and no computer scientists.

The third recommendation is *unification*. For several reasons, the regulatory pathways for medicinal products and medical devices are separated in most countries. This is particularly true in Europe, as there is a central authority for medicinal products (EMA), while member states oversee the regulatory evaluation process for medical devices through the so-called notified bodies. While this approach may present several advantages, it is unquestionable that it has impaired the ability of the European Union to adopt technological innovations in the regulatory process quickly. By contrast, the United States Food and Drug Administration (FDA), which is responsible for both drugs and devices, has been leading all initiatives related to in silico trials, also because of the "cultural contamination" that the collaboration of experts from the two regulatory domains has produced. To extend our first recommendation on *separation*, we recommend that the European Commission consider extending the scope of the proposed interdisciplinary scientific working party in charge of qualification advice on new methodologies and also to include the qualification of methodologies for the development of any medical products, including drugs, devices, and Advanced Therapy Medicinal Products (ATMPs).

A last recommendation is not for policymakers but all practitioners in academia and industry forming the in silico trials community of practice. The fourth recommendation is *patience*. The in silico trials community of practice needs to build trust in computational methods first and foremost in clinical research and practice. Once computer models are widely used as clinical decision support systems, high-quality validation datasets are available, and predicted biomarkers are common in clinical research, the regulatory world will also slowly make this transition.

In summary, we recommend policymakers to separate within EMA the scientific working party in charge of qualification advice on new methodologies from that authorising the marketing of new drugs; revise the composition of this new advisory body to include experts with a background in physical sciences and extend the scope of this body to provide qualification advice for methodologies used in the development of all kinds of medical products, not only drugs. While this happens, the In silico Trials community of practice should focus on widening the adoption of modelling and simulation in clinical research and practice.

[1] https://www.ema.europa.eu/en/committees/working-parties-other-groups/chmp/methodology-working-party

Table of Contents

2. Introduction

Despite their enormous potential, the adoption of in silico methodologies in the regulatory evaluation of biomedical products (sometimes referred to as *In Silico* Trials) is growing slowly. The European Commission funded the In Silico World (ISW) project, specifically aimed at analysing and lowering the barriers slowing the adoption. During the initial analysis phase, the consortium identified seven barriers; one was related to regulatory aspects.

During the three years of activity, the ISW consortium conducted two qualification advice procedures with the European Medicines Agency (EMA) and engaged with notified bodies, national authorities, and standardisation bodies. In this report, we provided detailed documentation of all these activities so that any other interested party can learn from our experience.

Members of the ISW consortium were also instrumental in developing an open-access book published by Nature-Springer, entitled "Toward Good Simulation Practice: best practices for the use of computational modelling & simulation in the regulatory process of biomedical products". This book results from a long grass-root consensus process involving hundreds of international experts from academia, industry, clinical practice, and regulatory agencies. As we led this important debate and captured it in this book, several concepts that influenced the present report emerged.

As some of our experts were also involved in the Mobilise-D project, which aimed to define regulatory pathways for wearable mobility monitors to be used in drug trials, we closely monitored the two qualification advice procedures that the Mobilise-D consortium pursued with the EMA and one with the United States Food and Drug Administration (FDA), Center for Drug Evaluation and Research (CDER), and the request for feedback (Q-Sub) to the FDA Center for Devices and Radiological Health (CDRH). While it is up to the Mobilise-D consortium to share the documentation associated with these activities, we will summarise some public domain resources.

But this document's ambition is more than a merge report of these activities. In the following pages, we will reflect on how biomedical regulatory science can cope with the technological revolution pervading the medical field. We will report in detail the various exploratory activities we conducted as part of the In Silico World project, and from the lessons learnt from those, we will propose an analysis of the reasons for some resistance, which we believe goes back to the differences in pragmatic epistemologies of the expert scientists involved in the regulatory panels. Lastly, based on this analysis, we will draw some conclusions and make recommendations for policymakers and the in silico trials community.

The partners of the In Silico World consortium wrote this report. A pre-final draft was circulated to several external experts and then discussed with them on March 14th, 2024, in a public workshop the consortium organised in Catania (and online). The **Catania Workshop** gave us the opportunity to debate with a broader audience (around 100 experts; see the detailed programme in Annexe 3). All participants were invited to submit written comments to the draft report; this final version accounts for all comments collected during the workshop and afterwards.

3. Regulatory Science for In Silico Trials: exploratory activities

3.1. Technical standards for credibility assessment in the EU market

This regulatory barrier is described in detail in a 2022 paper (Pappalardo et al., 2022).

While the qualification of in silico methodologies as drug development tools remains challenging, the US FDA Center for Devices and Radiological Health (CDRH) offers a well-

established regulatory pathway for the credibility assessment of these methodologies when used to develop medical devices. This has the in silico alternative viable, and the medical device industry is quickly adopting it; for example, since 2002, at least 21% of 565 original premarket approval (PMA) applications for medical devices included computational modelling efforts provided in the Summary of Safety and Effectiveness Data (SSED). These figures are likely to have grown since introducing the technical standard that is the centrepiece of this regulatory pathway, the ASME VV-40:2018.

While the CE Marking system for medical devices has various advantages, the lack of a central authority like EMA for the drugs makes it difficult to find the right interlocutor. The authority of medical devices is devolved to the member states, which authorise independent certification organisations called notified bodies to evaluate and provide marketing authorisation to new medical devices. Notified bodies are relatively small organisations, frequently for profit and thus focused on maximising their revenues. The recent introduction of the new Medical Device Regulation (MDR, also referred to as Regulation 2017/745), which requires the re-certification of all medical devices already on the market, has saturated the notified bodies, which have little time to discuss innovations. Any attempt we made to approach selected notified bodies produced no effect.

Similar considerations can be made for the European Commission services in the Directorate-General for Health and Food Safety (DG SANTE), which oversees the EU sectorial harmonisation legislation on medical devices. These entities are busy monitoring the introduction of the MDR and the In Vitro Diagnostic Medical Device Regulation (IVDR, or Regulation 2017/746), and any attempt to bring the topic of in silico methodologies to the attention of the Medical Device Coordination Group Working Group failed.

However, despite these failed attempts, nothing in the current EU regulation would prevent companies from producing regulatory evidence supporting the marketing authorisation of a medical device obtained in silico. In fact, some companies already use finite element model results to support their request for the CE marking of new orthopaedic devices; because no harmonised standard is available, they sometimes follow the ASTM F2996-20 "Standard Practice for Finite Element Analysis (FEA) of Non-Modular Metallic Orthopaedic Hip Femoral Stems" to support the credibility of these in silico evidence.

A tangible effect of the work we and others have done in promoting this debate in the regulatory community is the recent publication of the revised technical standard ISO 21535:2023 "Non-active surgical implants. Joint replacement implants. Specific requirements for hip-joint replacement implants". The standard states: "*Theoretical analysis and modelling, including finite element analysis, can also be used to select the most appropriate size(s) of component(s) for testing the worst case(s) (e.g. see ASTM F2996[17]). If used, the credibility of such modelling for its context of use shall be demonstrated (see ASME V&V 40-2018[24])*". This is an important precedent, and not only for orthopaedic companies. As an official ISO standard acknowledges the VV-40 as the best way to assess the credibility of a predictive model, this is not limited to hip replacement; in principle, any company could produce in silico evidence to a notified body for any medical devices supported by a VV-40 credibility assessment referring to this ISO standard as a precedent.

But more is needed. With the support of the Deutsches Institut für Normung (DIN), the prestigious German standardisation body and partner of the In Silico World Consortium, we approached various officers of the International Organization for Standardization (ISO) and of the International Electrotechnical Commission, two international standardisation bodies whose technical standards can be harmonised in the CE marking system. As a result of this initiative of our project, the IEC Technical Committee TC62 "Medical equipment, software, and systems" is promoting, in concert with ISO and in close collaboration with ASME, the development of a new technical standard that includes the risk-based credibility approach of the VV-40 but extends the scope also to data-driven models such as machine learning

predictors. In September 2023, during an official meeting of IEC TC 62 in Korea, the creation of a new workgroup was approved that, in collaboration with ISO/TC 276/WG5 and ASME V&V40 committees, will develop this new standard. Since the rules of IEC prevent a temporary consortium such as the In Silico Word project from joining directly as a Category A liaison, the Avicenna Alliance (a non-profit alliance of the industries active in *in silico* medicine with the Virtual Physiological Human (VPH) institute which represents the academic and non-profit stakeholders active in *in silico* medicine) took up this role. In December 2023, the Alliance officially announced that their application to join IEC TC 62 and ISO/TC 276/WG5 Committees has been approved. We expect this new initiative to have two positive effects: the first, in the medium term, will be the introduction, also in the EU regulatory system, of a harmonised IEC-ISO standard to assess the credibility of predictive models used in healthcare; the second is that until then companies requesting the CE mark for their new medical devices will be able to use modelling evidence supported by model's credibility assessments according to the ASME VV-40, considering that IEC and ISO are also working on a similar standard.

In parallel, the In Silico World project contributes to the critical revision of the ASME VV-40:2018. On August 21st, 2023, we presented our experience of using the ASME VV-40 in the qualification of BBCT to the ASME VVUQ SC40, the scientific committee in charge of revising the ASME VV-40. That also allowed us to make some recommendations for future revisions. Recently, Dr Cristina Curreli (former UNIBO post-doc, now IOR researcher) joined the VVUQ 40 End-To-End Example working group as a volunteer. The group will soon publish a technical report summarising a comprehensive application of the ASME V&V 40 credibility assessment framework.

The last part of the Catania Workshop was dedicated to a roundtable on the issue of assessing the credibility of data-driven and hybrid models. While the debate was rich and inspiring, all participants agreed that we were far from a consensus. It was suggested that the In Silico World Community of Practice on Slack could be used to develop consensus on some good practices regarding this specific aspect.

The work the In Silico World consortium is doing in this domain has been noticed internationally. In February 2023, a delegation of representatives from various regulatory agencies in Japan visited Bologna specifically to meet us and be updated on our activities. The minutes of the meeting are available in Annexe 2.

Technical standards can be seen as compendia of existing knowledge; as knowledge constantly evolves, technical standards on the model's credibility will also have to continue evolving. In parallel, the research community must publish the application of the VV40:2018 and other technical standards derived from it: exemplar cases ground the abstract process that standards provide into specific cases (Pathmanathan et al., 2019; Luraghi et al., 2021; Lopez Poncelas et al., 2022; Santiago et al., 2022; Galappaththige et al., 2022; Aldieri et al., 2023).

3.2. Uncertainty management

During the Catania workshop, one of the attending experts posed an interesting question: Which approaches should we use for uncertainty management? A trivial answer is the call of the VV40:2018 to conduct more or less extensive uncertainty quantification. However, uncertainty management becomes more complicated when we model living organisms. Michael Weisberg provides a widely accepted taxonomy of the types of idealisation processes in science (Weisberg, 2007). Weisberg proposes three idealisation processes: *Galilean*, *Minimalist*, and *Multi-Model*. Physical sciences and the pragmatic epistemology that has inspired the Verification, Validation, and Uncertainty Quantification (VVUQ) assume that we have enough mechanistic knowledge of the phenomenon being modelled to formulate a theory

that fully explains the phenomenon. We then use Galilean idealisation to "reduce" the order of this model to make it tractable from a mathematical/computational point of view. We rarely use this approach to develop physics-based models of living organisms, which are extremely complex and "entangled" (Viceconti, 2012; Day et al., 2024). We more frequently resort to minimalist idealisation, where we search for the minimum input set that lets us predict a specific output of the modelled living system with the necessary accuracy. We are fully aware that those inputs are necessary but not sufficient to describe the system of interest in full. In the context of Galilean idealisation, uncertainty mainly originates from the errors affecting the quantifications of the input values. Here, uncertainty quantification simply requires quantifying the uncertainty affecting each input of the predictive model and then calculating how much this uncertainty propagates in the predicted outputs. However, for a model built with a minimalist idealisation, there is a much bigger source of idealisation due to all the processes that affect my quantity of interest, which I simply ignored in building my models.

This problem is challenging: one's first reaction could be: how can I account for the uncertainty caused by the cause of the phenomenon I do not know? Here, we use a modified version of the known-unknown matrix used in the so-called *Johari window* psychological technique (Luft and Ingham, 1955). In our case, the two variables are inputs we know/do not know to have an effect on the output and whether we know/do not know how such input affects the output. This creates four options, usually called "known knowns", "known unknowns", etc. But to be more precise, we will call them "explainable knowns", "explainable unknowns", "unexplainable knowns", and "unexplainable unknowns".

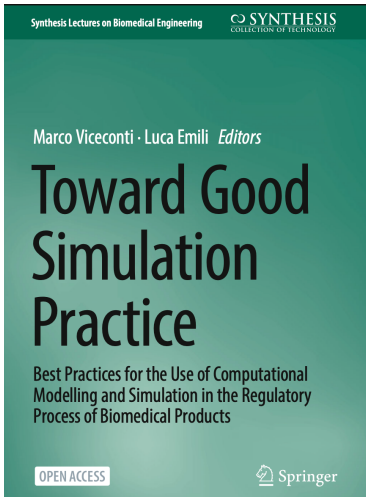
Physics-based models are built with "explainable knowns". On the other hand, we can do nothing to account for "unexplainable unknowns": we are unaware they have an effect, and even if we did, we would not know how to model it. But we argue it is possible to account for the other two. The "explainable unknows" can be modelled by including their effect as a stochastic process. The effects of "unexplainable knowns" can be explored by building another model, including additional knowns to investigate their possible effect on the model output. This *de facto* introduces the third type of idealisation, the Multi-Model Idealisation (MMI).

Explainable knowns Most physics-based models are built using explainable knowns.	**Explainable unknowns** These are the things known to have an effect, but we do not know how to model it. We can represent their effect as uncertainty.
Unexplainable knowns These are the things that we would know how to model, but we do not know if they have effect. We can explore their effect building another model (MMI process).	**Unexplainable unknowns** These are things we are not aware they have an effect, and even if we did but we would not know how to model it. There is nothing we can do about these.

3.3. The Good Simulation Practice

All new medical products, whether medicinal products, medical devices, or biological products, must be developed and tested extensively before being sold to ensure they are safe and effective in treating the intended conditions. Traditionally, all this is done experimentally, studying the new products *in vitro* (which means in the glass, e.g., in cell cultures) or *in vivo* (e.g., in living organisms), through animal experimentation or clinical studies on volunteers. In Silico Trials are computer modelling and simulation technologies that can be used to refine, reduce and even, in some cases, replace these experimental studies. But the development of In Silico Trials, the assessment of their credibility (e.g., how reliably they can be used in place of experimental studies), and their use in the evaluation of safety and efficacy of new medical products is a very new field, for which there is little experience. The Community of Practice led by the VPH Institute, the Avicenna Alliance, and the In Silico World consortium has brought together 138 experts in In Silico Trials working in academia, the medical industry, regulatory bodies, hospitals, and consulting firms. Through a consensus process, these experts produced the first attempt to define some Good Simulation Practices for developing, evaluating, and using In Silico Trials. Good Simulation Practice constitutes an indispensable guide for anyone planning to engage with In Silico Trials at any title.

The book, published by Nature Springer as an open-access eBook[2] in March 2024 (Viceconti and Emili, 2024), provides a theoretical foundation for verification, validation, and uncertainty quantification at the heart of good simulation practice. It discusses model development, especially software life cycle management and quality assurance. The book also covers model credibility and using the standard ASME VV-40:2018. It highlights how the current qualification pathways for in silico methodologies are somewhat inadequate and proposes alternative models. In the last part, the book covers using in silico methodologies for health technology assessment and how the roles of ethical review boards, the sponsor, and the investigator may change when in silico methodologies are adopted.



3.4. Lesson learnt from the Mobilise-D project

Inertial Measurement Units (IMUs) are a class of wearable sensors aimed at quantifying mobility outcomes such as the number of steps walked or the average walking speed. Originally, these devices were introduced as motivational tools (e.g., Fitbit), but they eventually evolved into full-blown measurement tools certified as medical devices. Still, the analytics software of these devices, which converts the raw signals from the sensors into Digital Mobility Outcomes (DMOs), showed limited accuracy when the wearer walked slowly or pathologically. The Mobilise-D project[3], supported by an IMI grant, developed a new class of algorithms that overcomes this limitation and conducted extensive technical and clinical validation studies to demonstrate that with these new algorithms, the DMOs measured during daily living in the real world for a sufficiently long period of time (e.g., a week) could be accurate

[2] https://doi.org/10.1007/978-3-031-48284-7
[3] https://mobilise-d.eu/

and informative enough to be used as secondary biomarkers in efficacy studies of new drugs to treat diseases that affect mobility such as Parkinson's Disease, Multiple Sclerosis, Chronic obstructive pulmonary disease, or the outcome of a fragility hip fracture in elders.

To this purpose, a team of specialists from academia and some of the leading pharma companies, coordinated by Prof Marco Viceconti (UNIBO) and Dr Wim Darte (Novartis), submitted two requests for qualification advice to the EMA, aimed to define the best regulatory pathway for the qualification of DMOs as secondary biomarkers.

This experience is described in detail in two publications (Viceconti et al., 2020, 2022) and two letters of support of EMA[4] to the Mobilise-D consortium. Here, we summarise some of the lessons learned in that process relevant to what is discussed in this report.

The first important point to notice is that the Mobilise-D analytics software, which converts the raw signals recorded by three accelerometers and three gyroscopes, the IMU includes while absolving a metrology function (to measure DMOs), could also be seen as a predictive model. In fact, the software includes machine learning modules and biophysics-informed modules to recognise walking activities from the continuous recording, split such activities into *bouts* of continuous walking, and finally compute for each bout the average step length, the average cadence, and the symmetry index, essential to calculate any other DMO. In fact, partner UNIBO has recently submitted to the FDA CDRH a qualification submission (essentially a request for clarification) on whether such analytics software, when to be certified as software as a medical device, could be assessed as metrology software or as a predictive model using the VV-40. The response of the FDA was negative. They recommended following the metrological pathway; however, their primary motivation is that the current version of the VV-40 only addresses the credibility of biophysical models, and the Mobilise-D software also includes machine learning components. This need to extend the VV-40 also to cover the credibility assessment of data-driven model or to have a separate technical standard for such purpose has been reported by the In Silico World consortium to ASME and IEC-ISO panels.

However, besides this specific issue, there is a general topic of predictive models being handled as measurement tools.

In the qualification advice with the EMA, we decided to consider the software merely as a component of a measuring device to be evaluated together with the hardware from a metrology point of view. This is because, in various preparatory informal meetings with EMA officers, we felt EMA would have resisted the idea that DMOs are predicted rather than measured.

Another essential element to highlight is that in the preparatory meeting, we stressed how none of the EMA Scientific Advisory Working Party experts have the background required to evaluate some complex technology from a technical point of view. EMA agreed and invited for the first qualification advice two ad hoc members, who were selected because they were specialists in these technologies. As far as we know, this is unusual for EMA and constitutes an important precedent.

The technical validation plan, designed as an accurate metrology study on volunteers and patients of the various disease groups, received little attention from most experts in the panel except the two technical specialists. As they were fully satisfied, the technical validation plan was approved and not discussed further. So, the rest of the two qualification advice procedures focused on clinical validation.

[4] https://www.ema.europa.eu/en/human-regulatory/research-development/scientific-advice-protocol-assistance/novel-methodologies-biomarkers/opinions-letters-support-qualification-novel-methodologies-medicine-development

To this purpose, DMOs were treated like any other measured biomarker and evaluated in terms of construct validity, predictive capacity, and ability to detect change in clinically relevant variables. Each DMO should be evaluated for each specific context of use and disease. The idea that one could demonstrate that the methodology accurately measures the DMOs for all diseases was not an option. This is because the regulator focuses on how this additional information may improve (or not) the regulatory decision the CoU defines. To some extent, the fact that the measurement tool is accurate is a given; the point of debate is how useful that measurement is in the decision process.

While preparing for the Catania workshop, we had some interesting exchanges with colleagues working at the Chemical Safety and Alternative Methods Unit at the European Commission on the subtle distinction between measured and predicted quantities. One specific example they reported is organ-on-chip (OoC) methodologies used as possible non-animal alternatives. While OoC are experimental methods, they are sometimes combined with computational models such as physiology-based pharmacokinetics (PBPK), for example, to extrapolate between *in vitro* and *in vivo* exposure. In some regulatory contexts, these are still seen as predominantly experimental methods, and their credibility is tested as such. As we mentioned before, this poses a general problem of demarcation. An ambitious solution to such a problem would be the development of a general framework for assessing the *credibility of information*, which can be measured, inferred, or predicted. Such a general theoretical framing would make it possible to assess the credibility of information produced with hybrid methodologies that combine information measured, inferred, and predicted.

3.5. ISW qualification advice with the EMA

3.5.1. BBCT qualification advice

Osteoporosis is a common disease affecting millions of people worldwide (Clynes et al., 2020). It is a silent disease until an impact, even so low energy that should have no effect on a healthy subject, causes a bone fracture. Current pharmacological treatments reduce the incidence of low-energy impact fractures by 50-60% at most (Odén et al., 2013; Wainwright et al., 2005), leaving room for improvement. But despite this, there are currently no new molecules targeting osteoporosis in phase I clinical studies. This is due to a number of reasons, but an important one is that historically, regulators have required fractures as the primary endpoint to demonstrate efficacy. As hip fractures occur with an incidence of less than 1% per year, even in high-risk populations, a phase III clinical study must enrol a very large number of patients and then follow them up for years to observe just a few hundred fractures. The pharma industry has been asking for surrogate endpoints (for example, Areal Bone Mineral Density, aBMD), but none of these biomarkers has the necessary stratification accuracy in separating fractured and non-fractured cases.

The Bologna Biomechanical Computed Tomography (BBCT) at the hip is a digital twin technology able to estimate the risk of fracture at the proximal femur for an individual. This is done by orchestrating 1) a stochastic mathematical model which, starting from the subject's height and weight, calculates possible forces acting at the hip due to a fall on the side and 2) a CT-based finite element model of the subject's femur which predicts the femur strength in all plausible femur poses. By simulating one million falls and comparing the impact force due to a fall and the force required to fracture the femur for each simulated fall, the absolute risk of fracture at the time of the CT, named ARF0, can be extracted. In a retrospective clinical study (Bhattacharya et al., 2019; Aldieri et al., 2022), BBCT was demonstrated to be more accurate than the areal bone mineral density (aBMD) in the prediction of hip fractures. Therefore, the ISW consortium submitted to EMA a request for qualification on the possibility of using the absolute risk of fracture predicted by the BBCT-Hip digital twin as a new response variable to be used in clinical trials for antiresorptive treatments efficacy assessment instead of aBMD.

Because some phase III clinical studies where aBMD was considered the primary endpoint could be found in CHMP assessment reports related to treatments for which EMA granted market authorisation, the initial Context of Use (CoU) we envisioned for BBCT was within Phase III clinical studies.

On 04/12/2021, the Applicant Mimesis S.r.l. requested qualification advice for BBCT-Hip. The submission contained the main briefing book document and two annexes. The first Annex contained a detailed description of BBCT technology, while the second contained BBCT technical validation carried out according to ASME V&V40-2018 technical standards. The main Briefing Book document was articulated into two main sections: the first one contained a general summary with background information motivating the rationale and the need for the proposed methodology and describing it. The second section contained four questions for EMA and the respective positions of the Applicant. The four questions posed were:

- Question 1: *Does EMA agree that the definition of Biomarker as "A biological molecule found in blood, other body fluids, or tissues that can be used to follow body processes and diseases in humans and animals" can broadened so as to include an in silico-based prediction and therefore that term Biomarker applies also to BBCT-hip ARFO?*
- Question 2: *Does EMA agree that the Context of Use clearly describes how BBCT will be used to provide a new surrogate of the fracture endpoint in Phase II clinical Trials?*
- Question 3: *Does the EMA agree that the proposed technical validation strategy is acceptable to assess the precision and accuracy of the BBCT-hip methodology in predicting the absolute risk of proximal femur fracture?*
- Question 4: *Does the EMA agree that the provided clinical validation plan for BBCT-hip is sufficient in supporting the qualification of the use of BBCT as a surrogate of the fracture risk in clinical trials?*

The clinical validation plan proposed in the fourth question in particular involved the conventional design used for any drug development biomarker, i.e., the concepts of construct validity, predictive capacity, and ability to detect change. The cohort on which such clinical validation was proposed was made up of around 100 patients selected according to specific inclusion and exclusion criteria within the HipOp collection available at partner Rizzoli Orthopaedic Institute including CT scans of the hip region originally collected for computer-aided preoperative planning of total hip replacement. Four of those had experienced a proximal femur fracture.

On February 10th, 2022, we had a preliminary informal meeting with the appointed Qualification team, who strongly suggested lowering the CoU to Phase II studies and framing it in greater detail. In particular, the Qualification Team stressed that Phase III clinical trials for antiresorptive treatments look at fracture incidence as the primary outcome, and that would not change. Hence, in light of the comments received, the CoU was thoroughly modified, positioning BBCT as a response variable for use in place of the aBMD in Phase II dose-response clinical studies.

The Briefing Book was subsequently resubmitted in its modified version, and the procedure started on the 3rd of April 2022.

On the 13th of May 2022, EMA sent the applicants a list of issues, including 11 issues to be discussed with the Qualification Team during the Discussion meeting and 7 to be addressed only in written form. Although most of the issues involved methodological and/or technical details to be clarified and better described, the most critical issues dealt with the clinical validation plan originally proposed, which was judged far from sufficient despite the reduced CoU proposed. After submitting the response to the issues document, a discussion meeting with the EMA Qualification Team took place on September 1st, 2022, where the issues and the

proposed answers were discussed. The main points of debate were related to the clinical validation plan: a full-scale clinical trial was mandatory, according to EMA, for such an in silico methodology to be judged reliable enough to be used in the proposed CoU.

On the 15th of September 2022, the EMA Committee for Medicinal Products for Human Use adopted the advice to be given to the Applicant. The complete documentation of the qualification advice is available in Annex 1.

3.5.2. UISS-TB-DR qualification advice

Tuberculosis (TB) is a significant global health issue caused by Mycobacterium tuberculosis. Primarily affecting the lungs, TB can also manifest in other areas of the body (Koch and Mizrahi, 2018; Furin et al., 2019). Many infected individuals exhibit no symptoms, known as latent TB infection (LTBI), with about 10% progressing to active TB, which can be fatal if untreated (Kim and Kim, 2018). Common symptoms of active TB include a chronic cough, night sweats, fever, and weight loss. Treatment typically involves a combination of antibiotics. The first line of therapy for those with low drug-resistance risk includes drugs like isoniazid, rifampicin, pyrazinamide, and ethambutol or streptomycin. For drug-resistant TB, second-line drugs categorised by the World Health Organization are used. Treatment duration varies from 6 months for drug-susceptible TB to 24 months for drug-resistant strains. Monitoring of active TB involves acid-fast bacillus staining of sputum smears or cultures. The efficacy of treatment is measured by the time to sputum culture conversion and the frequency of relapse within 24 months (Zumla et al., 2015). Recent advancements indicate that combining standard chemotherapy with adjuvant treatments like TB vaccines can shorten the time to sputum culture conversion and reduce relapse rates. Therapeutic vaccines like Mycobacterium Vaccae and RUTI are being developed, targeting either the host's immune response or the TB bacilli directly. Clinical trials have shown promising results in the safety and efficacy of these vaccines (Cardona, 2016). Reducing treatment duration is vital, especially in developing countries where treatment adherence is challenging. Shorter treatment times can decrease the incidence of relapses and the development of more drug-resistant TB strains, aligning with WHO's objectives (Seung et al., 2015).

The Universal Immune System Simulator – Tuberculosis model (UISS-TB) is an In Silico Trial technology that owns a sophisticated, bi-dimensional agent-based simulator encompassing multi-scale and multi-organ perspectives. This model intricately simulates the proliferation of a specific Mycobacterium tuberculosis (MTB) strain within the pulmonary compartment, intricately considering the immune system's response and the impact of potential treatments aimed at curbing bacterial growth or enhancing the immune reaction. In particular, UISS-TB-DR is a verticalization of the UISS-TB solution specifically tailored to predict the dose-response relationship of therapeutic vaccines. This model stands as a novel paradigm in "in silico trials," where instead of direct biomarker measurement, responses are computationally forecasted. When furnished with 22 input features, experimentally derived from individual patients, UISS-TB-DR can project the trajectory of bacterial load within a specific anatomical compartment, factoring in therapeutic interventions. The lung compartment is represented as a two-dimensional domain. Within this domain, each biological agent operates autonomously, capable of movement, state changes, molecular interactions, and other activities, all governed by established rules of interaction. The system's evolution mirrors principles of chemistry, physics, and biology, adhering to mass and energy conservation laws and entity interaction rules. Key interactions are modelled based on factors like proximity, entity states, local chemical concentrations, and molecular presentation patterns. The state of the system at each time step is defined by the count, type, state, position, and presentation pattern of entities, along with the local concentration of chemical species. UISS-TB-DR simulates MTB infection in the lung, incorporating genetic diversity in the MTB pathogen and various immune cells such as macrophages, neutrophils, dendritic cells, regulatory T cells, B

cells, and T helper cells, alongside relevant molecular entities like cytokines and chemokines. Patient-specific data informs initial infection characteristics and immune system status. Treatment modelling, including vaccines, considers known action mechanisms, administration plans, and individual drug susceptibilities. Vaccine simulations require additional input, such as formulation, dosage, biochemical properties, and interactions with the simulated immune system, derived from preclinical studies. This approach is similarly applicable to the modelling of antibiotic treatments.

On 25/09/2022, the Applicant Mimesis S.r.l. requested scientific advice for their product Universal Immune System Simulator – Tuberculosis Disease Model (UISS-TB) under Article 57(1)(n) of Regulation (EC) 726/2004 of the European Parliament and of the Council. The submission contained the main briefing book document and four annexes. The first Annex contained a detailed description of UISS-TB-DR technology; the second contained UISS-TB-DR the credibility assessment plan carried out according to ASME V&V40-2018 technical standards; the third annexe dealt with the procedure for the generation of digital twins; the fourth annexe contained the relevant sources of information. Appendix A also describes the procedure for Levene's Test for Equality of Variances. The main Briefing Book document was articulated into two main sections: the first one contained a general summary with background information motivating the rationale and the need for the proposed methodology and describing it. The second section contained four questions for EMA and the respective positions of the Applicant. The four questions posed were:

- Question 1: *Does EMA agree that the Context of Use clearly describes how UISS-TB-DR will be used to inform vaccine dose decision-making in phase IIa clinical trials dealing with pulmonary TB?*
- Question 2: *Does EMA agree with the overall strategy in general, and the risk assessment in particular, that we propose to evaluate the validity of the UISS-TB-DR as an in silico methodology for the optimisation of dose-response phase IIa clinical trials?*
- Question 3: *Does the EMA agree that the proposed credibility assessment plan is adequate to support the request for qualification of the UISS-TB-DR in silico methodology for the proposed context of use?*
- Question 4: *Does the EMA agree that a qualitative research based on clinical meaningfulness could be proposed in the credibility assessment validation plan to provide a more in-depth understanding of the predictive capability of the computational model demonstrating the ability of the model to reproduce general pathophysiological behaviours in TB?). Does also the EMA agree that the same qualitative approach can be used to reinforce the credibility assessment of CoU-related evidence i.e., support the escalation dose of a in silico phase IIa trial?*

The clinical validation plan proposed in the third question in particular, compared the circulating IFN-γ levels in TB patients treated with RUTI in an escalating dose trial, measured at various time points, against the dose-response curve predicted for a similar cohort of virtual patients. The primary metric for comparison was the predictive capacity of the model regarding IFN-γ Spot Forming Units. Given the model's focus on the same biomarker already utilised in such studies, the necessity for demonstrating construct validity or the evaluation of the Minimal Important Difference (MID) was deemed redundant. The clinical validation process involved data from a phase IIa, randomised, escalating dose clinical trial. This trial evaluated the safety, tolerability, and immunogenicity of three doses of the RUTI vaccine following a one-month treatment with isoniazid. Participants in the trial were randomised to receive either a placebo or one of the RUTI doses (5, 25, or 50 μg), administered subcutaneously in the deltoid muscle area of alternate arms, with a 28-day gap between

doses. Monitoring extended up to one month following the second RUTI inoculation. The validation process included a comprehensive examination of the test sample and conditions, focusing on the number and characteristics of the test samples used, along with precise measurements of the test conditions. The final stage of the validation involved an assessment of the model by analysing the Equivalency of Input Parameters and Output Comparison.

On November 22nd, 2022, we had a preliminary informal meeting with the appointed Qualification team, who strongly suggested narrowing the CoU. In particular, the current context of use was very broad and did not specify: 1) the stage of the disease (active, latent), 2) the type of vaccine platform (protein fusion, live attenuated, vector, whole-cell), 3) population (age, HIV infection etc), 4) resistance. Moreover, a request to substantiate the role of IFN-gamma as an established biomarker for therapeutic vaccine dose/regimen selection was made.

The Briefing Book was subsequently resubmitted in its modified version, and the procedure started on the 11th of April, 2023.

On the 31st of July 2023, a list of issues was received by EMA, including 6 issues to be discussed with the Qualification Team during the Discussion meeting and 11 to be addressed only in written form. Although most of the arisen issues involved methodological and/or technical details to be clarified and better described, the most critical issues were:

- Assumption of Dose Levels: The CHMP notes that the assumption that the three-dose levels standard for prophylactic vaccines will be suitable for therapeutic vaccines is unsubstantiated. They advise labelling the products as 'immunotherapy medicinal products' to avoid confusion with traditional prophylactic vaccines.
- Dose Selection Complexity: The process of dose selection is highlighted as being more complex than just evaluating an intermediate dose, a minimum effective dose (MED), and a maximum tolerated dose (MTD) in phase 2 trials. The CHMP suggests that a more elaborate approach for dose selection and regimen might be more appropriate.
- Phase 2a Study and Dose Selection: There's an assumption that selecting a dose in the Phase 2a study is a crucial milestone. However, the CHMP views this phase as exploratory, primarily focusing on safety, immunogenicity, and characterising the dose/regimen response. They believe this assumption needs more explanation.
- IFN-γ as a Biomarker for Dose Selection: The CHMP questions the assumption that changes in circulating interferon-gamma (IFN-γ) over time can inform clinical dose selection for an immunotherapy in latent tuberculosis (TB). They point out that the connection between the time course of IFN-γ and clinically relevant endpoints, such as the incidence of active TB, is not yet substantiated and needs further evidence.

After the submission of the response to the issues document, a discussion meeting with the EMA Qualification Team took place on September 25th, 2023, where the issues and the proposed answers were discussed. The main points of debate were related to the contest of use for which it was assumed that circulating interferon-gamma (IFN-γ) changes over time are established as the basis for clinical dose regimen selection. For EMA CHMP, the link between the time course of IFN-γ and the prevention of active TB disease still needs to be substantiated.

On the 21st of December 2023, the EMA Committee for Medicinal Products for Human Use adopted the advice to be given to the Applicant. The complete documentation of the qualification advice is available in Annex 2.

3.5.3. Reflection on the outcomes of the two qualification advice procedures

Besides the specifics of the two qualification procedures, which can be found in full documentation in the appendix, some general considerations can be drawn.

A first general comment is that the communication between the ISW consortium and the EMA SAWP was challenging. The two groups of experts did not always share the same terminology, and the ISW consortium was not always familiar with the EMA qualification advice process. The best evidence is the back and forth around the definition of the Context of Use (CoU) in both cases, which was considered too broad or too narrow. At the beginning of the procedure, the ISW consortium was informally advised by EMA officers to formulate narrow CoUs; but in one case, what the consortium considered a fairly narrow CoU was judged way too broad, while in another, the consortium CoU was criticised for being too narrow.

Regulatory procedures are not designed to be educational; in general, the proponent makes a statement, and the regulator says if they agree and, if not, why. However, the regulator is not expected to advise the proponent on how something should be stated to be acceptable. In this regard, the proponents' limited experience and communication problems did not help.

Despite these challenges, the ISW consortium identified an opportunity for greater understanding throughout the process. Throughout both procedures, every scientific discussion, whether in writing or during meetings, highlighted the need for improved communication between both parties. It was evident that the ISW consortium and the EMA SAWP experts approached problems with different perspectives rooted in their unique scientific philosophies. This realisation opened the door for dialogue to bridge these differences, fostering a collaborative environment where mutual decision-making in scientific research could thrive.

For example, in the BBCT-Hip submission, the ISW consortium asked whether EMA would consider the biomarker definition to include measured values and values predicted by sufficiently credible models. EMA SAWP replied: "This question is not judged very important [...]. A biomarker is a characteristic that is objectively measured [...]". The fact that the SAWP experts not only rejected the notion that a biomarker might be predicted but even considered the question unimportant suggests the differences are bigger than terminology or procedure.

There is a fundamental challenge around interdisciplinary decision-making that the work we have done in the ISW project has made evident and that we try to explain (to the best of our current understanding) in Chapter 4.

On the other hand, for the UISS-TB-DR submission, the ISW consortium obtained a Letter of Support from EMA, endorsing the model-informed drug development (MIDD) approach. However, EMA SAWP highlighted the need for further clinical data collection to ascertain IFN-γ's value as a biomarker and to address platform validation limitations, suggesting an expansion of the platform's application to various therapeutic vaccines and doses.

3.6. EMA reflection paper on AI

In July 2023, the EMA published a "Draft reflection paper on the use of Artificial Intelligence (AI) in the medicinal product lifecycle" (EMA/CHMP/CVMP/83833/2023)[5]. The paper, open for consultation, saw the contribution among the others of one of the researchers involved with the In Silico World project, Elisabetta Biasin (KU Leuven, and external collaborating expert at the EMA), for the ethical and legal aspects.

A systematic review of this document is beyond the scope of this report. This is because the document is largely irrelevant in the context of In Silico Trials. On page 3 of the EMA paper,

[5] https://www.ema.europa.eu/documents/scientific-guideline/draft-reflection-paper-use-artificial-intelligence-ai-medicinal-product-lifecycle_en.pdf

Artificial Intelligence (AI) and Machine Learning (ML) are described as follows: "AI and ML tools can, if used correctly, effectively support the acquisition, transformation, analysis, and interpretation of data within the medicinal product lifecycle". There is no mention of using ML as a predictor to reduce, refine, or replace experimental studies. In other words, AI is seen as a data processing tool that supplements more conventional statistical tools aimed at inference, not prediction. The concept "predict" appears only once on page 10, where they speak of explainability: "However, it is acknowledged that several of the most effective modelling architectures allow only limited insight into the translation from feature space through latent space to prediction or classification". Thus, while EMA is opening at the general idea that some ML methodologies can expand classic statistical modelling, this remains limited to inference and not to prediction in its proper sense of (even partial) replacement of an experiment.

4. A reflection on interdisciplinary decision-making

4.1. Introduction

If we compare the attitude of regulatory agencies toward in silico methodologies, we observe a marked difference between medical device regulators and drug regulators. At first observation, this difference is hard to explain; in both cases, the mission is the same: to make sure that only medical products that are safe and effective are placed on the market, whereas regulatory affairs (administrative aspects of regulation) and regulatory law (legal aspects of regulation) may differ, the foundations of regulatory science are, in both cases, the scientific method, the empirical method for acquiring knowledge humanity has successfully used since the 17th century.

But regulatory science is not, strictly speaking, a scientific domain; it is an interdisciplinary community of practice. Because of this, decision-making in regulatory science is not, and cannot be, based only on the scientific method.

In the following, we will argue that this complex decision-making process heavily depends on the different pragmatic epistemologies that the various disciplines rely on. We believe that this is the main reason behind this marked difference in their perception of in silico methodologies between medical device regulators and drug regulators.

4.2. Interdisciplinary decision-making: the truth equation

The Scientific Method is "the typical modality through which science pursues a knowledge of reality that is objective, reliable, verifiable, and shareable" (Viceconti 2011). It is composed of four distinct steps:

- Observation, description, and formulation of the research question;
- Formulation of theories based on observations and a priori knowledge;
- Falsification of such theories through repeated controlled experiments;
- Confirmation of the surviving theory by predicting a well-known phenomenon, independent but correlated.

However, an implicit assumption of this approach is that all people involved are *peers*, scientists with roughly the same level of expertise in that specific scientific domain. We also assume their opinions may be individually biased but not collectively. Last, we assume communication among peers occurs with perfect efficiency (no information loss, no misunderstanding).

Unfortunately, none of these assumptions holds for any interdisciplinary knowledge domain. Interdisciplinary communities of practice are, by definition, formed by people with different levels of expertise on different subjects; their members cannot be considered peers in the classical meaning of the word. Communication always involves some level of inaccuracy, but within an interdisciplinary community, much more so. Lastly, if we deal with an applied research field like the one considered here, large groups of experts may be similarly biased due to financial and/or career benefits.

So, interdisciplinary science cannot rely only on the scientific method for decision-making. We need to broaden our analysis to understand how these groups make decisions.

Decision theory is a complex philosophical domain[6]. It articulates in:

- Normative decision theory: Concerned with the identification of optimal decisions, where optimality is often determined by considering an ideal decision-maker who is able to calculate with perfect accuracy and is, in some sense, fully rational.
- Prescriptive decision theory: Concerned with describing observed behaviours through the use of conceptual models, under the assumption that those making the decisions behave under some consistent rules.
- Descriptive decision theory: Analyses how individuals actually make the decisions that they do.

A pretty popular approach to normative decision theory is Savage's normative theory, also called "theory of choice under uncertainty":

$f, g \in F$ are acts;

$s_i \in S$ are the possible states of the world;

$f(s_i)=o_i \in O$ is the state s_i is actual, the act f will produce the outcome o_i, among the possible outcomes O.

The utility U of a decision is the summation of the utility of an act for all possible states, weighted for the probability that state occurs:

$$U(f) = \sum_i u[f(s_i)] \cdot P(s_i)$$

It is easy to see how challenging it would be to formulate this utility function $u[\,]$ in the context of our interest.

If we analyse how people make their decisions in daily life, we can observe several factors. We believe something is true because:

- We receive evidence.
- Someone we trust says so.
- Most people we know agree.
- Most people we know disagree (complot theory).
- The statement is aesthetically pleasant.
- The statement confirms or at least does not contradict our worldview.

Following the scientific method, we should rely only on evidence to decide. Evidence can be separated into *inductive* (i.e., empirical, which derives from observation) and *deductive* (which derives from prior knowledge). An interdisciplinary context imposes the need for some

[6] https://plato.stanford.edu/entries/decision-theory/

principle of authority because other people's opinions may carry more weight due to their specialism. Then, there are two key social determinants: conformism and anti-conformism. Different people assign different weights to this, but for sure, our opinions are influenced by others' opinions, one way or another. Last, the two more psychological factors are aesthetic pleasure and personal beliefs.

If we rewrite Savage formula in the light of these considerations, the trust we assign to a statement can be expressed as:

$$T = k_{Ob}Ei + k_{Re}Ed + k_{Au}Au + k_{Co}Co + k_{Ac}Ac + k_{Ae}Ae + k_{Wv}Wv$$

Where:

Ei = Inductive evidence (empirical)

Ed = Deductive evidence

Au = Authoritative opinion

Co = Conformism

Ac = Anti-conformism

Ae = Aesthetics

Wv = World View

This "truth equation" only highlights the various determinants that commonly concur with a decision. It will be helpful to compare how different groups of experts make decisions. But before this, we must introduce another element: the epistemological differences.

4.3. Epistemological differences

Looking closely, we see that each domain of specialism in science has its pragmatic epistemology. Epistemology is the branch of philosophy concerned with knowledge; in science, each specific scientific domain has its approach to deciding what should be considered a scientific truth, which is shared among all practitioners. The epistemological differences between any two scientific domains can be many and sometimes subtle. However, one aspect allows us to cluster them into two groups: *statistical inference*.

Statistical inference is the process of analysing data sampled from a larger population to infer the properties of such a population. We use statistical inference to estimate the type and properties of the statistical distribution of a value over a population from a finite number of samples, as well as if different values observed in a sample appear to be correlated. There are two types of statistical inference:

- In Frequentist Inference, you assign the probability to the data and then use them to accept or reject a hypothesis.
- In Bayesian inference, you assign probability directly to the hypotheses.

Frequentist statistics is entirely based on the frequencies with which things occur. Let us use a classic problem: decide if a treatment T has an effect on the property D of a system, which can be measured experimentally before and after the treatment, that is statistically significant if compared to when no treatment NT is given. Observing how the repeated measurements distribute, we assume they follow a Gaussian probability distribution, defined by a mean *m* and a variance *s*. Because the number of measurements N is finite, we can obtain estimates $(\hat{\mu}, \hat{\sigma})$. We use the repeated measurements to estimate $(\hat{\mu}_T, \hat{\sigma}_T)$ and $(\hat{\mu}_{NT}, \hat{\sigma}_{NT})$, and an appropriate test to calculate the probability p that the two distributions have difference average. This is strictly true only for $N \to \infty$. Frequentist inference uses different approaches to decision

theory. The *epidemiological approach*, based on the Neyman–Pearson decision theory, is commonly used in life science. Set up two statistical hypotheses, H_0 and its opposite H_1, and choose the parameters of the statistical test (α, β) according to best practices and the sample size according to subjective cost-benefit considerations. Choose H_0 so that a small sample size works against you. If the data falls into the rejection region of H_0, accept H_1; otherwise, accept H_0. Accepting a hypothesis does not mean that you believe in it, but only that you act as if it were true.

Bayesian statistics is entirely based on the probability associated with beliefs. To decide if a treatment T affects the property D of a system, which can be measured experimentally before and after the treatment, we express the posterior probability that *T* has an effect given the disease *D* as:

$$P(T|D) = \frac{P(D|T) \cdot P(T)}{P(D)}$$

The *posterior probability* that the treatment affects D is equal to $P(D|T)$, the *likelihood* that we experimentally observe that D changes given T, times $P(T)$, the *prior probability* that T has an effect, divided by $P(D)$, the probability that D is observed.

The difference between these two approaches to inference is profound.

In frequentist statistics, truth is hidden but exists. Data are (statistically) true, while prior knowledge is scarcely informative and is more likely to be a source of bias. The decision is binary: to choose the default action hypothesis in all cases but the one where the data falsifies it, in which case you choose its opposite hypothesis. The default action is what you choose if you know nothing; only observational data may change your mind. This logical framework fits well in legal reasoning. There is a default action (for example, everyone is innocent until proven guilty), and the choice is binary: the null hypothesis H_0 or the alternative (opposite) hypothesis H_1.

There is no truth in Bayesian statistics, only more or less probable beliefs. We have prior beliefs on each statement, which probability can be decreased or increased through experimental observations. Each possible action has a probability associated with it; you choose the one with the highest chance of being true. This fits well with Popper's decision-making framework of the scientific method (Popper, 1990) if we complement it with the concept of Platt's strong inference (Platt, 1964).

This difference is reflected deeply in how scientists with different training form their decisions and the role that a model's prediction may have in forming such decisions.

Scientists providing scientific advisory to regulatory agencies on medicinal products are predominantly biologists, pharmacists, epidemiologists, and psychologists, who tend to rely on the epistemologies of social and natural sciences. In these domains, the inference is mostly Frequentist. Information is rarely quantitative, and when it is, significant systematic errors may affect it. There is an expectation that prior knowledge is scarcely informative due to the complexity of the phenomena under investigation and the potential presence of biases.

For these scientists:

K_{Ei} ↑↑- the accumulation of empirical evidence will help.

K_{Ed} ↓↓- Deductive evidence will not help.

K_{Au} ↓- Lack of interdisciplinarity makes it challenging to define authority.

K_{Co} ↑↑- A large consensus among practitioners is considered evidence, especially in medicine.

$K_{Ac} \rightarrow 0$ - No room for irrational thinking.

K_{Ae} ↑- Discursive narrative, made with the specialist vocabulary we are familiar with, is preferred.

K_{Wv} ↑- Narratives that do not require contradicting my worldview are more welcome.

Within this pragmatic epistemology, predictive models can be used at most to design experiments because only empirical evidence (observational data) can be trusted, while deductive evidence cannot. The inherent interdisciplinarity required to evaluate an in silico methodology is currently not reflected in the scientific advisory panels of EMA and FDA-CDER. Hence, the evaluation tends to focus only on those aspects where the experts in the panel feel qualified, downplaying the importance of the other aspects. The innovative nature of in silico methodologies produces a lack of consensus among practitioners on their validity. The way in silico methodologies are presented, frequently through mathematical language, does not help. Last but not least, because the idea that something can be predicted rather than tested experimentally would require a considerable change in the worldview of these experts, there is an inherent resistance.

The scientists providing scientific advisory to regulatory agencies on medical devices are, at least in part, physicists, chemists, engineers, and computer scientists by training who rely on a cultural background that is based on physical sciences. In these domains of knowledge, there is an expectation that most observations are quantitative and should not be free of systematic errors (unbiased); prior knowledge, in the form of laws of physics that have resisted extensive falsification attempts, is usually considered informative. Their inference is primarily Bayesian, with an expectation that prior and posterior are similar, which is the basis for the concept of models' validation.

For these scientists:

K_{Ei} ↑ ↑ - the accumulation of empirical evidence will help.

K_{Ed} ↑ ↑ - deductive evidence will help.

K_{Au} ↑ - Interdisciplinarity makes it easier to recognise authority.

K_{Co} ↑ - Conformism affects individuals in all domains of science (Kuhn and Hacking, 2012)

$K_{Ac} \rightarrow 0$ - No room for irrational thinking.

KAe ↑ - Narratives made with math are more explicit and less ambiguous.

K_{Wv} ↑ – the use of models does not contradict their worldview.

In this context, it is customary to use predictive models built with prior mechanistic knowledge and/or prior data to refine, reduce, and sometimes even replace experimental studies. For these scientists, inductive evidence is not inherently superior to deductive evidence. They always work in interdisciplinary panels with biomedical scientists and clinical experts, making it easier to respect the respective authority during the discussions. While a certain amount of conformism is weighed in every regulatory decision, these experts recognise the importance of radical innovations. When in silico methodologies are presented with a mathematical language, this is accepted as a sign of clarity and precision. The footing in physical sciences makes the idea of using predictive models not necessarily in contradiction with their worldview.

In this context, predictive models are accepted, at least in principle, to refine, reduce, and even replace experiments. Any engineer trusts the second laws of dynamics more than their experimental results, comforted by the fact that Newton's laws have resisted any falsification attempt for centuries. The strong interdisciplinarity in the medical device panels allows them to tackle complex and radically new methodologies with the confidence that someone in the panel has the necessary expertise to assess each aspect involved. In medical device panels, there is the same caution we observe in drug panels towards methodologies for which there

is scarce experience. However, their greater familiarity with the mathematical language and the fact that modelling aligns more with their worldview makes medical device panels more open to this type of innovation.

5. Discussion

Two important books surely influence the reflections in this report. The first is 1962 Thomas Kuhn's "The Structure of Scientific Revolutions" (Kuhn and Hacking, 2012). In this book, Kuhn theorises that scientific progress in each sub-domain of science is not linear. At any point in time, every sub-domain tends to conform to a specific paradigm and builds its pragmatic epistemology around it. This resistance persists until anti-conformist scientists highlight "anomalies" that eventually drive revolutions that impose new paradigms and an adapted pragmatic epistemology. The second book is 1966 Michel Foucault's "Les mots et les choses: Une archéologie des sciences humaines" (The Order of Things: An Archaeology of the Human Sciences) (Foucault, 1995). In it, Foucault proposes that every historical period has underlying epistemic assumptions and ways of thinking, which determine what truth is and what an acceptable discourse about a subject is.

However, an important point frequently overlooked in the debate around these two books is that these paradigms and their epistemic assumptions result from the scientific discoveries in that sub-domain until then. We call it *pragmatic epistemology* to stress that it is developed within a scientific sub-community in each historical period as a pragmatic choice, a good practice informed by the collective experience of that community. We are trying to say that psychologists have good reasons to adhere to a frequentist view, as physicists have good reasons to adhere to the Bayesian view. The history of their sub-domain has guided them to agree on that particular pragmatic epistemology.

The problem is that the introduction of computer modelling and simulation and, more recently, artificial intelligence (AI) has forced each scientific domain to change its paradigm. Biology and medicine have resisted more than other domains because such paradigm change requires a fairly significant change in their pragmatic epistemology and imposes a re-thinking of education in these domains. However, as Kuhn has convincingly argued by looking at the history of science, paradigm changes may be delayed, but eventually, they are inevitable.

We believe that biological and medical sciences will be forced to undergo a dramatic paradigm shift in the next 10-20 years, which might be painful for a fraction of the involved scientists, and will produce conflicts. Researchers from fields already engaged in interdisciplinary research applications (such as tissue engineering) have started this process earlier, due to the strong presence of scientist from physical and engineering sciences.

It should be noted that a similar argument could be made for AI technologies. AI plays a role in *in silico* medicine as an alternative method to develop predictors when only implicit knowledge of the data is available. But it also has a separate role as a method to automate repetitive human decision-making. These two roles are sometimes confused. The first fits entirely in what has been discussed here. The only difference is that the inductive approach to the development of a predictor, entirely based on data without prior knowledge, makes frequentists a bit more comfortable. But in the end, in the frequentist world, the goal is inference; prediction is seen as an excess of *hubris*. The discussion on the second role is complex because the concept of credibility in the sense of the VV-40 is replaced by the concept of "Would another human believe you are not a machine?". A perfect example is the Turing test used to decide if an AI is distinguishable from human intelligence[7]. We see a potential danger in mixing these two planes. The credibility of a predictive model aimed to

[7] https://en.wikipedia.org/wiki/Turing_test

replace an experiment can be framed in metrological terms, whereas that of an AI mimicking human behaviour falls in the realm of experimental psychology. However, the separation in some cases, such as radiomics, is not so easy.

And meanwhile? We believe the only option is to acknowledge that the drug regulators' resistance to using in silico methodologies is just the tip of the iceberg and that the cultural roots of such resistance go deep. This goes in both directions: physical sciences experts need to develop a pragmatic epistemology that account for the challenges that *in vivo* experimentation face. On the other hand, biomedical sciences experts need to accept that much of their research will see a growing role of technologies and computational methods in particular. Because they have not been trained in these subjects, biomedicine is becoming an interdisciplinary domain where biologists and medical doctors need to work with engineers and with computer scientists. In some areas of medicine, this is already obvious to everyone, and it has been happening already for some time (i.e., radiology); in other domains, such as pharmacology, this has just started, and some resistance is unavoidable. As we asked various colleagues to revise this document, we found it enlightening that, commenting on the section entitled "A reflection on interdisciplinary decision-making", a colleague with a psychology background told us "all the equations were unnecessary", whereas another with a physics background praised the use of equations to clarify the concepts. This is exactly the cultural tension we must mediate by forming truly interdisciplinary scientific advisory panels.

Starting from this consideration, the In Silico World Consortium has compiled some recommendations for policymakers and our own community of practice.

6. Recommendations

6.1. Recommendations for EU Policymakers

The EMA Scientific Advisory Working Party (SAWP) is currently providing scientific advice on the requests for marketing authorisation for new drugs and on the qualification of the methodologies used to develop them. This organisational model, which has made sense for a long time, now needs to be questioned. The qualification requests will see a progressive growth of candidate methodologies based on radically innovative technologies. These evaluations cannot be selective, focusing on some aspects and neglecting others. But this implies a composition of expertise in the panel quite different from that required by the scientific advice on new drugs. The second is seen as one of the most critical functions of the EMA; thus, its needs always take precedence. Thus, our first recommendation to EU policymakers is **separation**. The European Medicine Agency (EMA) should consider separating the panels that evaluate the marketing authorisation of new drugs from those that provide qualification advice on new methodologies. This would allow each panel to have an optimal composition of expertise. As we were finalising this report, we were delighted to hear that this step has already been undertaken. The EMA has recently established the *Methodology Working Party*[8]: "The Methodology Working Party (MWP) was established by the Committee for Medicinal Products for Human Use (CHMP) in order to pool and use expertise in key areas such as biostatistics, modelling and simulation, pharmacokinetics, pharmacogenomics, and real-world evidence". As far as we understand, the responsibility of producing qualification opinions on new methodologies will continue to rest on the SAWP of the CHMP, and the MWP will have more of a consulting role. However, we invite the EMA to consider in the future the possibility of formally separating the methodological qualification from the marketing authorisation decision.

[8] https://www.ema.europa.eu/en/committees/working-parties-other-groups/chmp/methodology-working-party

Separating the panel that provides qualification advice and opinions will make it easier to reform its composition. This panel must see the presence of bioengineers, medical physicists, computer scientists, and any other technological expertise required alongside those competencies traditionally represented in the SAWP. In this regard, we were not impressed by the composition of the first MWP panel: a scan of the CV of the 21 experts the EMA indicated for this panel only one has a background in physics. We acknowledge that a more interdisciplinary composition for the MWP would pose some challenges, as these experts will need to learn to work together, and their different epistemologies will inevitably clash. However, the European Commission has significant experience in moderating interdisciplinary panels in the evaluation of research proposals; such experience could be capitalised on here to establish appropriate procedural rules on how such interdisciplinary panels form their decisions. Thus, the second recommendation is **interdisciplinarity**. The composition of the new panel providing qualification advice on new methodologies must also include experts with a background in physical sciences (biophysicists, bioengineers, computer scientists, etc.). Experts in biomathematics should be chosen to include specialists in Bayesian inference. Medicine is becoming an interdisciplinary science, and regulatory agencies should adapt.

But this change will pose another problem. Many experts in the SAWP come from national drug authorities, primarily pharmacologists and epidemiologists. Where can EMA find these technical and technological experts? Another issue for the EU market is that the separation of the regulatory authorities for drugs and medical devices makes it more difficult to cope with radical innovations such as in silico methodologies. In particular, while the US FDA can provide qualification advice on new methodologies to develop drugs or medical devices, in Europe, the EMA can provide such advice only for drug development methodologies. No one is currently organised to provide advice on methodologies to develop medical devices. However, these two problems could become a solution. This newly established panel for the qualification of new methodologies could be given a broader scope, which also includes methodologies for developing medical devices. This will necessarily expand the experts involved to those working on medical devices, who are more likely to come from those technical and technological backgrounds. Thus, the third recommendation is **unification**. We recommend that the European Commission consider extending the scope of the EMA Methodology Working Party to provide qualification advice and qualification opinions on new methodologies for the development of every medical product, including drugs, medical devices, and Advanced Therapy Medicinal Products (ATMPs). This would not change the responsibilities of the marketing authorisation of medical devices, which would remain on the national authorities through the notified bodies. However, such notified bodies could trust that evidence obtained with new methodologies that received a positive qualification opinion from the EMA MWP could be reliably used in the regulatory assessment of medical devices.

It should be noted that this proposed change would not imply a change of remit for EMA or for the regulatory framework defined by the MDR/IVDR regulations. Drugs would still be authorised by EMA, and medical devices CE marked by notified bodies. However, the EC would recognise that only the EMA has the organisational capabilities to provide scientific advice on new methodologies for all kinds of products. EMA already has a similar role in the MDR on high-risk medical devices, where "EMA supports the medical device expert panels that provide opinions and views to notified bodies on the scientific assessment of certain high-risk medical devices and in vitro diagnostics". This is called the Clinical Evaluation Consultation Procedure (CECP).

It should be clear that these changes are non-trivial and may require some time to be implemented and some progressivity in their introduction. It should also be clear that such changes will not magically disappear all the issues this report has highlighted. Some of the epistemological differences we reported run deep, and it will take a long time to perform that

paradigm shift we mentioned above. But at least these different views will be confronted within interdisciplinary panels, from which balanced compromises will surely emerge.

In summary, we welcome the creation of the EMA Methodology Working Party; to accelerate the adoption of in silico methodologies, we recommend that its composition be made much more markedly interdisciplinary, it be given full responsibility for the qualification of new methodologies, and expand such role for all biomedical products, including medical devices and ATMPs.

6.2. Recommendations for the IST community of practice

A last recommendation is not for policymakers but all practitioners in academia and industry forming the in silico trials community of practice. The fourth recommendation is **patience**. The in silico trials community of practice needs to build trust in computational methods first and foremost in clinical research and practice. Once computer models are widely used as clinical decision support systems and predicted biomarkers are common in clinical research, the regulatory world will also slowly make this transition. We understand that because of the enormous costs associated with bringing a new drug to the market, this is a very appealing market segment for in silico trials. But it is also the one with the tallest regulatory barriers right now, rooted in profound epistemological differences that will take time to mediate.

Meanwhile, we recommend that academia and industry focus on developing clinical decision support systems and in silico methodologies to be used in pre-regulatory stages, for example, as alternatives to animal experimentation or as tools for designing and optimising clinical trials. Practitioners should also focus on contexts of use where clinical trials are impossible (virtual placebo, rare diseases, etc.). The recent FDA announcement[9] of a Pilot Program Model-Integrated Evidence for Generic Drugs on using modelling techniques instead of conducting *in vivo* bioequivalence (BE) studies to facilitate generic drug development goes exactly in this direction of small steps.

Also, an important recommendation for practitioners is to work on high-quality data registries and make them available to the scientific community to validate computational models. The EMA itself mentioned the possible use of registry data for the clinical validation of in silico methodologies in the Qualification Advice List of Issues received in May 2022 relative to the BBCT solution. To build high-quality data and, at the same time, trust in in silico methodologies, there could be a "transitional phase" where these new methodologies are first “gradually integrated" - where possible – into prospective clinical studies. In this case, the model predictions would simply provide additional data related to that clinical trial, with a very low impact on the final decision. There are two main advantages to this "gradual" approach. On the one hand, clinical data would be systematically extracted, focusing on quantities of interest useful for rigorous comparison with predictions. Furthermore, if the prospective comparison between clinical study results and predictions confirms that the computational model provided reliable estimates, trust in the computational model would be built step by step. The generated evidence will then support the use of the model in a “relatively” different context (for example, to predict the effectiveness of a new drug) with increasing impact (for example, to select doses to be tested in a new dose-selection study).

The publication of the book on Good Simulation Practice is an important step: but the in silico medicine Community of Practice should not stop investing time and money into capturing best practices on modelling & simulation in healthcare through consensus processes. We invite the industry associations such as COCIR, EFPIA, Vaccines Europe, EuropaBio and MedTech Europe, together with the specialistic organisations (Avicenna Alliance and VPH Institute), to

[9] https://www.fda.gov/drugs/news-events-human-drugs/deep-dive-fdas-model-integrated-evidence-mie-industry-meeting-pilot-program-generic-drugs-01182024

explore how the public-private partnership Innovative Health Initiative (IHI) could support the continuation of such collective effort.

7. References

Aldieri, A., Curreli, C., Szyszko, J.A., La Mattina, A.A., Viceconti, M., 2023. Credibility assessment of computational models according to ASME V&V40: Application to the Bologna Biomechanical Computed Tomography solution. Comput. Methods Programs Biomed. 240, 107727. https://doi.org/10.1016/j.cmpb.2023.107727

Aldieri, A., Terzini, M., Audenino, A.L., Bignardi, C., Paggiosi, M., Eastell, R., Viceconti, M., Bhattacharya, P., 2022. Personalised 3D Assessment of Trochanteric Soft Tissues Improves HIP Fracture Classification Accuracy. Ann. Biomed. Eng. 50, 303–313. https://doi.org/10.1007/s10439-022-02924-1

Bhattacharya, P., Altai, Z., Qasim, M., Viceconti, M., 2019. A multiscale model to predict current absolute risk of femoral fracture in a postmenopausal population. Biomech. Model. Mechanobiol. 18, 301–318. https://doi.org/10.1007/s10237-018-1081-0

Cardona, P.-J., 2016. The Progress of Therapeutic Vaccination with Regard to Tuberculosis. Front. Microbiol. 7, 1536. https://doi.org/10.3389/fmicb.2016.01536

Clynes, M.A., Harvey, N.C., Curtis, E.M., Fuggle, N.R., Dennison, E.M., Cooper, C., 2020. The epidemiology of osteoporosis. Br. Med. Bull. 133, 105–117. https://doi.org/10.1093/bmb/ldaa005

Day, T.C., Zamani-Dahaj, S.A., Bozdag, G.O., Burnetti, A.J., Bingham, E.P., Conlin, P.L., Ratcliff, W.C., Yunker, P.J., 2024. Morphological Entanglement in Living Systems. Phys. Rev. X 14, 011008. https://doi.org/10.1103/PhysRevX.14.011008

Foucault, M., 1995. Les mots et les choses: une archéologie des sciences humaines, Collection Tel. Gallimard, Paris.

Furin, J., Cox, H., Pai, M., 2019. Tuberculosis. Lancet Lond. Engl. 393, 1642–1656. https://doi.org/10.1016/S0140-6736(19)30308-3

Galappaththige, S., Gray, R.A., Costa, C.M., Niederer, S., Pathmanathan, P., 2022. Credibility assessment of patient-specific computational modeling using patient-specific cardiac modeling as an exemplar. PLoS Comput. Biol. 18, e1010541. https://doi.org/10.1371/journal.pcbi.1010541

Kim, H.W., Kim, J.S., 2018. Treatment of Latent Tuberculosis Infection and Its Clinical Efficacy. Tuberc. Respir. Dis. 81, 6–12. https://doi.org/10.4046/trd.2017.0052

Koch, A., Mizrahi, V., 2018. Mycobacterium tuberculosis. Trends Microbiol. 26, 555–556. https://doi.org/10.1016/j.tim.2018.02.012

Kuhn, T.S., Hacking, I., 2012. The Structure of Scientific Revolutions: 50th Anniversary Edition, Fourth edition. ed. University of Chicago Press, Chicago ; London.

Lopez Poncelas, M., La Barbera, L., Rawlinson, J.J., Crandall, D., Aubin, C.E., 2022. Credibility assessment of patient-specific biomechanical models to investigate proximal junctional failure in clinical cases with adult spine deformity using ASME V&V40 standard. Comput. Methods Biomech. Biomed. Engin. 25, 543–553. https://doi.org/10.1080/10255842.2021.1968380

Luft, J., Ingham, H., 1955. The Johari window, a graphic model of interpersonal awareness. Proc. West. Train. Lab. Group Dev. 246.

Luraghi, G., Bridio, S., Miller, C., Hoekstra, A., Rodriguez Matas, J.F., Migliavacca, F., 2021. Applicability analysis to evaluate credibility of an in silico thrombectomy procedure. J. Biomech. 126, 110631. https://doi.org/10.1016/j.jbiomech.2021.110631

Odén, A., McCloskey, E.V., Johansson, H., Kanis, J.A., 2013. Assessing the impact of osteoporosis on the burden of hip fractures. Calcif. Tissue Int. 92, 42–49. https://doi.org/10.1007/s00223-012-9666-6

Pappalardo, F., Wilkinson, J., Busquet, F., Bril, A., Palmer, M., Walker, B., Curreli, C., Russo, G., Marchal, T., Toschi, E., Alessandrello, R., Costignola, V., Klingmann, I., Contin, M., Staumont, B., Woiczinski, M., Kaddick, C., Salvatore, V.D., Aldieri, A., Geris, L., Viceconti, M., 2022. Toward A Regulatory Pathway for the Use of in Silico Trials in the CE Marking of Medical Devices. IEEE J. Biomed. Health Inform. 26, 5282–5286. https://doi.org/10.1109/JBHI.2022.3198145

Pathmanathan, P., Cordeiro, J.M., Gray, R.A., 2019. Comprehensive Uncertainty Quantification and Sensitivity Analysis for Cardiac Action Potential Models. Front. Physiol. 10, 721. https://doi.org/10.3389/fphys.2019.00721

Platt, J.R., 1964. Strong Inference. Science 146, 347–353. https://doi.org/10.1126/science.146.3642.347

Popper, K.R. (Karl R., 1990. The logic of scientific discovery. London : Unwin Hyman.

Santiago, A., Butakoff, C., Eguzkitza, B., Gray, R.A., May-Newman, K., Pathmanathan, P., Vu, V., Vázquez, M., 2022. Design and execution of a verification, validation, and uncertainty quantification plan for a numerical model of left ventricular flow after LVAD implantation. PLoS Comput. Biol. 18, e1010141. https://doi.org/10.1371/journal.pcbi.1010141

Seung, K.J., Keshavjee, S., Rich, M.L., 2015. Multidrug-Resistant Tuberculosis and Extensively Drug-Resistant Tuberculosis. Cold Spring Harb. Perspect. Med. 5, a017863. https://doi.org/10.1101/cshperspect.a017863

Viceconti, M., 2012. Multiscale Modeling of the Skeletal System. Cambridge University Press.

Viceconti, M., Emili, L. (Eds.), 2024. Toward Good Simulation Practice. Best Practices for the Use of Computational Modelling and Simulation in the Regulatory Process of Biomedical Products, Synthesis Lectures on Biomedical Engineering. Springer Cham.

Viceconti, M., Hernandez Penna, S., Dartee, W., Mazzà, C., Caulfield, B., Becker, C., Maetzler, W., Garcia-Aymerich, J., Davico, G., Rochester, L., 2020. Toward a Regulatory Qualification of Real-World Mobility Performance Biomarkers in Parkinson's Patients Using Digital Mobility Outcomes. Sensors 20, 5920. https://doi.org/10.3390/s20205920

Viceconti, M., Tome, M., Dartee, W., Knezevic, I., Hernandez Penna, S., Mazzà, C., Caulfield, B., Garcia-Aymerich, J., Becker, C., Maetzler, W., Troosters, T., Sharrack, B., Davico, G., Corriol-Rohou, S., Rochester, L., Mobilise-D Consortium, 2022. On the use of wearable sensors as mobility biomarkers in the marketing authorization of new drugs: A regulatory perspective. Front. Med. 9, 996903. https://doi.org/10.3389/fmed.2022.996903

Wainwright, S.A., Marshall, L.M., Ensrud, K.E., Cauley, J.A., Black, D.M., Hillier, T.A., Hochberg, M.C., Vogt, M.T., Orwoll, E.S., Study of Osteoporotic Fractures Research Group, 2005. Hip fracture in women without osteoporosis. J. Clin. Endocrinol. Metab. 90, 2787–2793. https://doi.org/10.1210/jc.2004-1568

Weisberg, M., 2007. Three Kinds of Idealization. J. Philos. 104, 639–659.

Zumla, A., Chakaya, J., Centis, R., D'Ambrosio, L., Mwaba, P., Bates, M., Kapata, N., Nyirenda, T., Chanda, D., Mfinanga, S., Hoelscher, M., Maeurer, M., Migliori, G.B., 2015. Tuberculosis treatment and management--an update on treatment regimens, trials, new drugs, and adjunct therapies. Lancet Respir. Med. 3, 220–234. https://doi.org/10.1016/S2213-2600(15)00063-6

8. Annexe 1: full documentation of the qualification advice procedures

The complete documentation of the two qualification advice procedures with the EMA is available in Open Access on Zenodo.

The complete documentation on the Qualification Advice with EMA for the BBCT-Hip solutions is here: https://zenodo.org/uploads/10938447.

The complete documentation on the Qualification Advice with EMA for the UISS-TB solution is here: https://zenodo.org/uploads/10939450.

9. Annexe 2: minutes of the meeting with Japanese regulators

EU-Japan bilateral meeting on standards, regulatory pathways, and regulatory science for in silico trials

Bologna, February 3rd, 2023

Organised by the In Silico World project, with the collaboration of the VPH Institute and the Avicenna Alliance

Participants:

- **Masanori Kikuchi**: Japan head delegate of TC150 (Implant for surgery), Professor in Polymer and Biomaterials Field, National Institute for Materials Science, Japan.
- **Kiyoyuki Chinzei**: Secretary of TC150/WG14 (Models of tissue for mechanical testing of implants), Professor of Health and Medical Research Institute, AIST, Japan.
- **Keiko Koyanagi**: Japan Fine Ceramics Associations.
- **Kohei Murase**: Convener of TC150/JWG1 (Additive manufacturing in surgical implant applications), Professor of Osaka University, Japan.
- **Marco Viceconti**: Coordinator of the In Silico World project, full professor at the University of Bologna.
- **Cristina Curreli**: Post-doctoral researcher at the University of Bologna (IT), specialist in the verification validation and uncertainty quantification of predictive models for in silico trials.
- **Vincenzo Carbone**: Senior Product Manager MedTech at In Silico Trials Technologies, leader of the commercial exploitation work-package in the In Silico World project.
- **Goran Stanic**: Communication officer In Silico World project, VPH Institute.
- **Francesco Pappalardo**: Full professor at the University of Catania (IT), expert in in silico trials for transmissible diseases treatments, leader of the regulatory work-package in the In Silico World project (attend online).
- **Liesbet Geris**: Full professor at the University of Liege (BE), Director of the VPH Institute, the non-profit organisation representing the academic component of the international community of practice formed around in silico medicine and in silico trials (attend online).
- **Heike Moser**: Expert for standards in the field of medical informatics at Deutsches Institut für Normung e. V. (DIN) Medical Standards Committee (NAMed) (attend online).
- **Thierry Marchal**: Industry Director for Healthcare Solutions and Ansys, Director of the Avicenna Alliance, the non-profit organisation representing the industrial component of the international community of practice formed around in silico medicine and in silico trials (attend online, to be confirmed).
- **Marc Horner**: Distinguished Engineer, Ansys Corp. and vice-chair of the ASME VV-40 Sub-Committee.

Meeting minutes

Prof Viceconti welcomes all delegates.

Prof Viceconti provides a summary of the European agenda around in silico medicine, particularly on the current situation regarding regulatory pathways for predictive models as a medical device and as medical product development tools in Europe and the USA. He stresses the vast difference between device and drug regulators, which he believes are largely explained by their different backgrounds.

Prof Chinzei confirms that the situation in Japan is very similar, and the backgrounds of the two groups of regulators are also vastly different.

Prof Murase presents specific use cases where modelling was used to refine or partially replace pre-clinical experiments in certifying new orthopaedic devices.

Prof Chinzei gives an overview of the Japanese regulatory context on the subject.

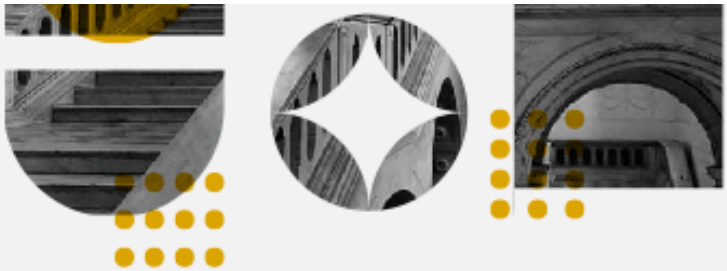
A discussion follows. It is agreed that the situation in Japan is not substantially different from that in the EU or the USA. Thus, it makes sense to explore forms of collaboration and regular exchange of information.

Prof Viceconti provides a summary of the Good Simulation Practice initiative. He asks his Japanese colleagues to pass to all interested colleagues in their country the invitation to join the In Silico World Community of Practice on Slack and, where they want, through this, contribute to the final revision of the GSP Book to be published in a few months.

Dr Curreli provides an in-depth presentation of the ASME VV-40:2018 technical standard. Dr. Horner complements this by providing information on the current activities of the VV-40 subcommittee, which he is co-chair of. He also takes the opportunity to inform the Japanese delegates of various international activities conducted by the Avicenna Alliance.

A discussion follows. It is confirmed that while the VV-40 is not officially recognised in Japan, local regulators may consider it a good basis to demonstrate the credibility of mechanistic (deductive) models. The Japanese delegates are invited to promote their country the membership in the VPH Institute for academic individuals and organisations and to the Avicenna Alliance for industrial organisations to create a concertation table and to work together to pursue common objectives.

10. Annexe 3: detailed programme of the Catania workshop



INSILICO WORLD

14TH MARCH 2024 | h 14:00–18:40

WORKSHOP: THE REGULATORY BARRIERS TO IN SILICO TRIALS

h 14:00–15:50

PRESENT

14:00
Opening remarks
Francesco Pappalardo (Host and WP4 leader -In Silico World consortium)

14:10
Review of current regulatory pathways in Europe and the USA to certify DTHs and MDDT/DDT
Cristina Curreli (In Silico World consortium)

14:30
Principles and regulatory application of credibility assessment with ASME V&V 40-2018
Jeff Bischoff (Zimmer-Biomet; chair of the ASME V&V40 committee)

14:50
Bringing ASME VV-40 to the international level: the new IEC/ISO work group
Regina Geierhofer (Siemens Healthineers; Secretary of the IEC TC62)

15:10
Toward Good Simulation Practice: from position report to future technical standards
Vincenzo Carbone and Klaus Zeier (In Silico World consortium)

15:30
EMA Qualification advice on BoneStrength and on UISS-TB
Alessandra Aldieri and Giulia Russo (In Silico World consortium)

15:50
Coffee break

h 16:00–18:40

FUTURE

16.10
Credibility of Computational Models Program: new guidance from FDA-CDRH
Prasanna Pathmanathan(FDA-CDRH)

16.40
Devices vs. Drugs: statistical inference approaches at the root of an epistemological divide.
Marco Viceconti (In Silico World consortium)
Presentation and open discussion on the In Silico World Report on regulatory barriers to in silico methodologies.

17.40
Presentation and roundtable - Assessing the credibility of data-driven and hybrid models. Is a general credibility framework possible?
Marco Viceconti, Saverio Ranciani, Jeff Bischoff, Regina Geierhofer, Prasanna Pathmanathan

18:40
end of workshop

Monastero dei Benedettini, Catania (Italy) info@insilico.world

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101016503