Ref. Ares(2023)1186091 - 17/02/2023

Grant agreement no. 823712

CompBioMed2

Research and Innovation Action
H2020-INFRAEDI-2018-1
Topic: Centres of Excellence in computing applications

D2.3 First Report on VVUQ Strategies and Community Application Codes

Work Package:	2	
Due date of deliverable:	Month 24	
Actual submission date:	DD Month 2021	
Start date of project:	01 October 2019	Duration: 48 months

Lead beneficiary for this deliverable: *BSC*
Contributors: *UNIBO, UCL, UvA, UOXF, USFD*

Disclaimer
This document's contents are not intended to replace consultation of any applicable legal sources or the necessary advice of a legal expert, where appropriate. All information in this document is provided "as is" and no guarantee or warranty is given that the information is fit for any particular purpose. The user, therefore, uses the information at its sole risk and liability. For the avoidance of all doubts, the European Commission has no liability in respect of this document, which is merely representing the authors' view.

Project co-funded by the European Commission within the H2020 Programme (2014-2020)		
Dissemination Level		
PU	Public	YES
CO	Confidential, only for members of the consortium (including the Commission Services)	
CI	Classified, as referred to in Commission Decision 2001/844/EC	

*This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement No **823712**"*

Table of Contents

*This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement No **823712**"*

List of Tables and Figures

*This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement No **823712**"*

Version Log

Version	Date	Released by	Nature of Change
V0.1	29/07/2021	BSC	First draft
V0.2	30/07/2021	UNIBO	Add sections on strategy and CoU
V1.0	08/09/2021	BSC	First draft, submitted for Internal Review
V2.0	20/09/2021	BSC	Final version
V2.1	17/02/2021	BSC	Updated version following M36 review

1 Contributors

Name	Institution	Role
Jazmin Aguado-Sierra	BSC	Principal Author
Alfonso Santiago	BSC	Co-Author
Mariano Vazquez	BSC	Co-Author
Marco Viceconti	UNIBO	Contributor
Gabor Zavodszky	UvA	Contributor
Jon McCullogh	UCL	Contributor
Jenny Wang	UOXF	Contributor
Xinshan Li	USFD	Contributor
Alberto Marzo	USFD	Contributor
Antonino Amedeo La Mattina	UNIBO	Contributor
Francesco Marson	UNIGE	Contributor
Alex Wade	UCL	Reviewer
Gavin Pringle	EPCC	Reviewer

*This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement No **823712**"*

2 Definition and Acronyms

Acronyms	Definitions
ASME	American Society of Mechanical Engineers
CBM2	CompBioMed 2
CO	Cardiac Output
CoU	Context of Use. It is a is a concise description of the specified use of the methodology in question in development and regulatory qualification of a class of medical products.
CS	Cardiac Simulator
DDT	Drug Development Tools
DTMRI	Diffusion Tensor Magnetic Resonance Imaging
EDV	End Diastolic Volume
EF	Ejection Fraction
ESV	End Systolic Volume
FSI	Fluid-Structure Interaction
HF	Heart Failure
HPC	High Performance Computing
HR	Heart Rate
LHS	Latin Hypercube Sampling
LV	Left Ventricle
LVAD	Left Ventricular Assist Device
LVESV	Left Ventricular End Systolic Volume
MDDT	Medical Device Development Tools
MRI	Magnetic Resonance Imaging
NCV	Numerical Code Verification
QoIs	Quantities of Interest
RV	Right Ventricle
RVEDV	Right Ventricular End Diastolic Volume
RVESV	Right Ventricular End Systolic Volume
SA	Sensitivity Analysis
SaMD	Software as Medical Device

*This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement No **823712**"*

SQA	Software Quality Assurance
VVUQ	Verification, Validation, and Uncertainty Quantification
V&V	Validation and Verification

*This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement No **823712**"*

3 Public Summary

The aim of this task as part of Work Package 2 is to develop Verification, Validation & Uncertainty Quantification (VVUQ) strategies and techniques, in line with the FDA endorsed draft ASME V&V-40 standard, applicable to the use cases arising from CompBioMed 2 (CBM2).

Model credibility can be established through verification and validation (V&V) activities. Although methods for V&V are becoming well established, guidance is lacking on assessing the relevance and adequacy of the V&V activities for computational models used to support medical device development and evaluation. VVUQ is required to augment the software credibility and, with this, ease of industrial application using the simulation tools. The work within this deliverable is intended to move forward VVUQ techniques on the multiple computational codes involved in CBM2.

To this end, a strategy needs to be established in order to address the particular needs of each application at the various stages of development within CBM2. Therefore, we briefly describe the possible trajectories for computational biomedicine codes, and then we provide some information of the VVUQ Strategy that each trajectory entails for each code level. Last, we propose a process through which all codes under development can consolidate their appropriate VVUQ strategy.

*This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement No **823712**"*

4 Introduction

The aim of this task is to develop Verification, Validation & Uncertainty Quantification (VVUQ) strategies and techniques, in line with the FDA endorsed draft ASME V&V-40 standard, applicable to the use cases arising from CompBioMed2. In the introduction we will:

- Give a synthetic introduction to the matter dealt with.
- Situate the deliverable with respect to the general objectives.
- State and make clear the specific objectives of the deliverable.

Computational models have been used to support the design of medical devices for many years, without any specific guidance on how to assess their credibility. Model credibility can be established through verification and validation (V&V) activities. Standardization of the VVUQ process has already been addressed by the first two ASME V&V subcommittees, namely the V&V10 and V&V20 [4.1/1, 4.1/2] for solid mechanics and fluid dynamics respectively. Although methods for V&V are becoming well established, guidance is lacking on assessing the relevance and adequacy of the V&V activities for computational models used to support medical device development and evaluation. The V&V40 Subcommittee therefore set out to provide guidance on the application of V&V practices for medical devices. Given the inherent risk of using a computational model as a basis for predicting medical device performance, the ASME V&V 40 Subcommittee has developed a risk-informed credibility assessment framework.

The goal of the risk-informed credibility assessment framework is to empower the medical device industry or the regulatory agencies to determine and justify the appropriate level of credibility for using a computational model to inform a decision. The decision could be internal to an organization or part of a regulatory activity. The risk informed framework (Fig. 4.1.1) initiates with the definition of a question of interest to be answered. Subsequently, a Context of Use (COU) needs to be defined. The COU is the specific role and scope of the computational model used to address a question of interest. The assessment of model risk depends on the model influence and the decision consequences. This leads to the establishment of the multiple credibility goals for the multiple stages involving verification and validation of the computational model. With the desired goal the VVUQ plan is established and executed. If the proposed target goals are finally obtained, the process and results are documented. If the target goals aren't reached, the framework should be iterated to improve the flaws detected.

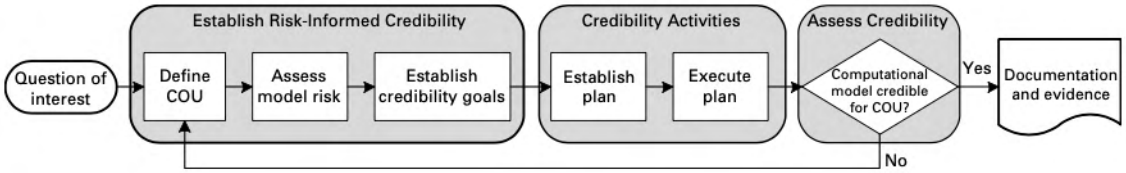


Figure 6.1.1.1.Process Diagram of the Risk-Informed Credibility Assessment Framework. Reprinted from ASME V&V 40-2018, by permission of The American Society of Mechanical Engineers. All rights reserved.

While the ASME V&V40 Standard provides the risk-informed credibility assessment framework, it is application agnostic. That is why, the team in charge to define and execute the VVUQ plan should follow the case-specific guidelines. E.g. a fluid-dynamic based application should follow the ASME V&V 10 Standard [4.1/2] guidelines designed specifically for VVUQ of fluid dynamic models.

*This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement No **823712**"*

This deliverable is intended to move forward VVUQ techniques on the multiple computational codes involved in CompBioMed2. VVUQ is required to augment the software credibility and, with this, ease industrial application of the simulation tools and/or regulatory acceptance. Executing these V&V frameworks for complex simulation codes that reproduce incompletely known physical behaviors, is a challenging task by itself that drives innovation in each research field. Therefore, the support provided by the project to the multiple research partners contributes to the project's Objective 1: "To fully support the Computational Biomedicine community and its diverse set of applications (…)" and Objective 2: "To promote innovation in the field of computational biomedical modelling and simulation(…)". As stated above, credible numerical software is mandatory for industrial adoption of simulation code, which makes VVUQ mandatory to engage with a range of industries across the entire healthcare value chain, from healthcare providers to pharmaceutical and medical device manufacturers, which is part of CompBioMed Objective 6.

While most scientific manuscripts describing a novel numerical model provide some sort of visual comparison of the simulation results with a similar physical phenomenon, that sort of qualitative comparison can't be qualified as validation as per the ASME V&V40 standard. This deliverable is intended to push the biomedicine simulation software boundaries towards more credible numerical results by following the most strict protocols published. Due to the novelty of the ASME V&V40 standard, publications following the risk assessment framework are scarce [4.1/3], therefore the task of applying the ASME standard for a new application involves a considerable step towards more credible simulation results.

The specific objectives of this deliverable are:

- Describe a strategy for computational biomedicine codes.
- Describe the activities carried out to provide BSC's Alya red as a regulatory tool to predict intra-LV flow stagnation after LVAD implantation.
- Describe a sensitivity analysis carried out with Alya for a biventricular electromechanical model of the heart.
- Describe the VVUQ activities to execute with UCL's HemeLB code.
- Describe results obtained after the sensitivity analysis using CovidSim solver.
- Describe how easyVVUQ toolkit is used for a sensitivity analysis for the molecular dynamics solver.
- Describe the Sensitivity analysis carried out with HemoCell.
- Describe the VVUQ plan for CT2S/BoneStrength.
- Describe the VVUQ plan for Palabos, from the University of Geneva.

The document is organized following the structure of the specific objectives above.

*This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement No **823712**"*

5 Developing a VVUQ strategy for computational biomedicine codes

5.1 Rationale

In most other scientific domains, the development of a Verification, Validation, and Uncertainty Quantification (VVUQ) strategy is mostly a technical activity, possibly informed by specific technical standards. When predictive models are to be used in biomedicine, the process is much more complex, and it involves technical and non-technical elements.

For sake of clarity, and at risk of oversimplifying some situations, we will describe the process used to develop a computational VVUQ strategy in this context with a three-levels representation:

1) Solver level: in some cases, a computational biomedicine solution is developed in a monolithic code, but in most cases a general-purpose solver is used.
2) Solution level: using a solver we implement a computational biomedicine solution that typically predicts certain quantities relevant for biological, physiological, or pathological process.
3) Methodology level: we use the solution to develop a methodology to be used for a specific Context of Use (CoU) in the development of a medical product or a clinical decision support system.

At the present stage some of the codes being developed in the CompBioMed2 project are still defined only at the solver level, some also at the solution level, and very few up to the methodology level. Being this a research project, not all code must necessarily develop to the top layer within the duration of the project. However, it is important to understand the trajectory expected for each code, and align, for each of the three layers the VVUQ practices to the requirements that trajectory will impose. At this time, the results presented in this deliverable, did not require the application support from Task 2.4, since most of the partners have had abundant experience running their applications efficiently using HPC resources.

First of all, we briefly describe the possible trajectories for computational biomedicine codes, and then we provide some information of the VVUQ Strategy that each trajectory entails for each code level. Last, we propose a process through which all codes under development can consolidate their appropriate VVUQ strategy.

5.2 Exploitation trajectories for computational medicine codes

Computational biomedicine codes can be used as ***Digital Patients*** tools or as ***In Silico Trials*** tools, which defines their regulatory pathway and therefore VVUQ activities they should undergo (Fig 5.1.1.1).

Digital Patient models (aka Digital Twin) are used as stand-alone clinical decision support systems, or as embedded systems in medical devices. From a regulatory point of view, since they contribute to the treatment of individual patients, they are considered as medical devices, in the category called Software as Medical Device (SaMD) and must undergo *Certification* before clinical application. Depending on their Context of Use, they belong to different risk classes, which involve more or less extensive levels of scrutiny.

*This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement No **823712**"*

In Silico Trials models are used as Drug Development Tools (DDT) or Medical Device Development Tools (MDDT). They usually aim to refine, reduce, or replace an experimental study, done *in vitro/ ex vivo*, on animals, or on humans. From a regulatory point of view, they must undergo *Qualification*, which typically requires a technical validation and a clinical validation. Their regulatory pathway is not entirely defined yet, especially in Europe, however there is consensus that the VV-40 standard may be accepted for technical validation. The level of scrutiny is usually defined starting from a risk analysis, where refinement is less risky than reduction, which in turn is less risky than replacement; and reduction of *in vitro* test is less risky than animal reduction, which is less risky than human reduction (see Fig. 5.1.1.2).

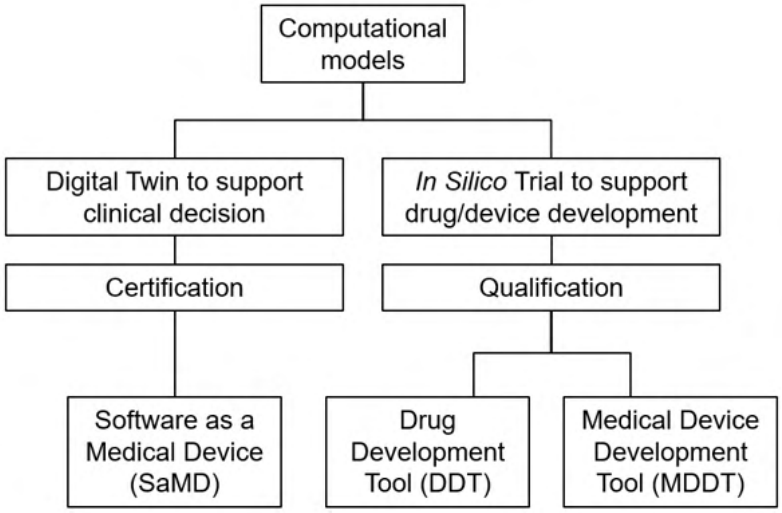


Figure 6.1.1.1 Exploitation trajectories for computational medicine models.

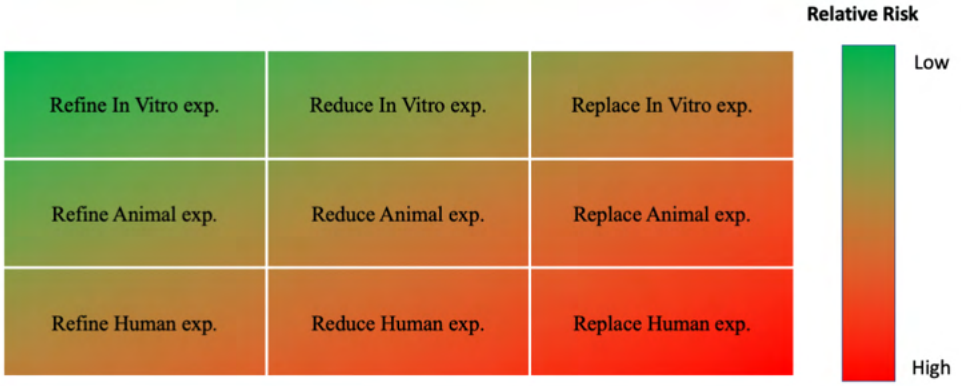


Figure 6.1.1.2 Relative risk of *In Silico* Trials solutions as a function or the experiment they target, and the type of substitution.

Codes that are still being developed at the solver level do not need to specify any of these categorisations. Codes at the solution level normally should be positioned as Digital Patient or *In Silico* Trials tools. Methodologies need to target a specific context of use, from which the risk level is defined, and a proper VVUQ strategy can be identified.

*This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement No **823712**"*

5.3 VVUQ Strategies for each code level

According to the VV-40, a full model credibility assessment is carried out by following specific activities, as reported in Table 5.3.1.

Activity	Credibility factor
Verification	
Code	Software quality assurance Numerical code verification
Calculation	Discretization error Numerical solver error Use error
Validation	
Computational model	Model form Model inputs
Comparator	Test samples Test conditions
Assessment	Equivalency of input parameters Output comparison
Applicability	Relevance of the quantities of interest Relevance of the validation activities to the COU

Table 5.3.1 Verification, validation and applicability activities reported in VV-40 standard.

Based on the three levels we identified in section 5.1, we have proposed their own appropriate requirements in terms of VVUQ.

Codes developed at the ***solver level should undergo code verification***. This involves *Software Quality Assurance* (SQA) and, in some cases, *Numerical Code Verification* (NCV). SQA aims to demonstrate that the software is functioning correctly and can be implemented using a number of different strategies, more or less demanding. However, for a code to be used as a medical device, compliance with the standard EN 62304:2006 "Medical device software. Software life-cycle processes" is required. While the certification of the software according to such standard might be beyond the scope of a research project, it is recommended that the software development practices are aligned with such standard, which will make easier a future certification. NCV aims to demonstrate that a specific algorithm is correctly implemented in terms of numerical solution, with respect to exact analytical solutions. Numerical code verification does not apply to data-driven models, but only to mechanistic models. The most common methods for code verification of continuum mechanics solvers are analytical solution benchmarks, the Method of Manufactured Solutions, and Numerical Solutions Benchmarks. Specific verification strategies are required for molecular dynamics solvers (e.g. https://doi.org/10.1371/journal.pone.0202764), and Agent-Based solvers (e.g. https://doi.org/10.1002/cnm.3470).

Codes developed at the ***solution level should undergo full technical validation***. The technical standard of reference for technical validation of Digital Patient solutions is the ASME VV-40:2018. However, it has been recently suggested that the same standard is also suitable for the technical validation of *In Silico* Trials solutions targeting medical devices (https://doi.org/10.1097/mat.0000000000000996), and also targeting medicinal products

*This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement No **823712**"*

(https://doi.org/10.1002/psp4.12479). The VV-40 standard provides only a general framework for a risk-based definition of the VVUQ plan, and not the details of how to perform the VVUQ, which is solution specific. In general terms, all technical validation plans should include a solution verification, a validation against controlled experiments, and an uncertainty quantification. Specific standards like ASME VV-10 or VV-20 exist for certain classes of models, whereas for others, one has to rely on the best practices that emerge from the scientific literature. Again, while the certification of the software according to such standards might be beyond the scope of a research project, it is recommended that VVUQ activities in the project are aligned with such standards or best practices, which will make easier a future certification.

Codes developed at the ***methodology level should prepare to pursue a regulatory certification/ qualification***. As reported previously, solutions targeting a specific Digital Patient CoU are *certified* as medical products. Solutions targeting a specific *In Silico* Trials CoU are *qualified* by regulatory authorities as drug or medical device development tools (DDT/MDDT). In addition to the technical validation VVUQ, such certification/qualification usually also involves a clinical validation. Depending on the specific CoU the requirements in terms of clinical validation can be very different; however, partner UNIBO can provide support to other partners that are willing to prepare for such activities.

5.4 VVUQ strategy development

As a first step we asked each partner developing a code in the project to report their planned VVUQ activities. The answers are provided in section 7 of this deliverable.

Starting from these replies, partner UNIBO will engage with each of these partners and guide them to a redefinition of their VVUQ plans that is consistent with what was exposed in the previous sections, so to better align these codes to their future exploitation. The revised VVUQ plans will be provided in deliverable D2.5 “Intermediate Report on VVUQ, Exemplar Research Integration and Community Applications”, due at M36.

The full results of such revised VVUQ plans will be provided in full detail in deliverable D2.7 “Final Report on VVUQ, Exemplar Research Integration and Community Application Codes”, due at M47.

6 Planned activities or Activities Carried out

6.1 Alya as a regulatory tool (Barcelona Supercomputing Center)

Background: Over 5 million people suffer heart failure (HF) in the U.S. alone, with ~1 million new cases annually [6.1/1]. From these patients, about 10% is in Stage D [6.1/2] condition, being heart transplant the gold standard treatment. The limited organ availability is making left ventricular assist devices (LVADs) a leading treatment option, with a ~90% 1-year survival rate [6.1/3]. LVADs are centrifugal or axial pumps apically implanted that help support the heart to reach the required cardiac output (CO) to sustain life.
There is evidence [6.1/4] that inflammation is associated with LVAD use, which combined with the endothelial lesion and the abnormal flow patterns [6.1/5] become the three composing

*This project has received funding from the European Union’s Horizon 2020 research and innovation programme under the Grant Agreement No **823712**“*

parts of the Virchow's triad [6.1/6] for thrombus formation. The local flow conditions influence the type of thrombus created. White thrombus are formed in regions with high velocities and high shear stresses that lead to platelet activation [6.1/7] and fibrin aggregation. On the contrary, red thrombus are created by stagnant and slow recirculating flows with low shear stresses that lead to an aggregation of all blood components [6.1/8, 6.1/9]. While the latest LVADs generation have a reduced white thrombus formation due to the novel magnetic and hydrodynamic rotors, the patients still suffer thromboembolic events. The reason for this is that the abnormal LV flow patterns combined with the low shear stresses suggest the LV as a relevant site for red thrombus formation.
While there is an extensive number of publications dealing with multiple LVAD factors like ventricular size [6.1/10], cannula implantation position [6.1/11], implantation depth [6.1/12, 6.1/13, 6.1/14] or angulation [6.1/15], none of them provide credibility evidence as suggested in the recent ASME V&V40 [6.1/16], neither guided by the historical V&V20 [6.1/17] specifically designed for computational fluid dynamics (CFD) more than 10 years ago. The reason for this is, most probably, that such a validation requires a thorough comparison of the simulation results against experiments and hundreds of executions of the numerical model, which involves a large computational cost. This work follows the V&V40 pipeline for a computational model of a benchtop LV-LVAD system to quantify intraventricular flow patterns.

Methods: The bench experiments were performed with the San Diego State University (SDSU) cardiac simulator (CS). This CS is a mock circulation loop of the heart and the circulatory system with an apically implanted LVAD that has been reported previously in [6.1/18, 6.1/19]. It involves a silicone LV based on an idealised geometry, immersed in a water-filled tank and connected to an external circulatory loop mimicking the systemic circulation. The tank is fully watertight, so when the piston pump generates negative pressure, the LV expands to the end diastolic volume (EDV). Two beating modes and three pump speeds are used for six validation points. The condition 22[%] @ 68.42[bpm] has ejection fraction (EF) = 22[%], and heart rate (HR) = 68.42[bpm], with end systolic volume (ESV)=180[cm3] and end diastolic volume (EDV) =230[cm3]. The condition 17[%] @ 61.18[bpm] has EF = 17[%] and HR = 61.18[bpm] with ESV=180[cm3] and EDV=216.86[cm3]. The pump speeds used for the validation points are 0[rpm], 8k[rpm] and 11k[rpm].
The computational domain is created from the exact same computer geometry used to manufacture the silicone ventricle. To obtain a computationally inexpensive and accurate way of deforming the ventricle, a unidirectional fluid-structure interaction (FSI) approach is used to deform the LV (similarly as [6.1/13]). A pressure is imposed in the external solid domain which afterwards deforms the CFD domain between the ESV and the EDV. Once the simulation pipeline is completed, the input files are modified to work as a template. This template is used by Dakota server (DARE) for the sensitivity analysis (SA) and the uncertainty quantification (UQ). DARE is an automating tool that works coupled with Sandia's Dakota and allows automatically encoding, submitting and retrieving jobs to any high-performance computing (HPC) infrastructure. Dakota allows to characterise and sample model inputs for multiple analysis types like SA, UQ or optimisation. The Dakota+DARE pair runs in a computer external to the HPC machine for as long as the analysis under execution may last (up to several weeks in this work). DARE receives Dakota's chosen inputs, processes the simulation templates with an encoder, submits the job to the supercomputer queue, waits for the jobs to be finished and processes the final results to feed Dakota with the obtained outputs. A combination of Dakota's restart capabilities with DARE failure capture capabilities makes this a robust framework for the required analysis.

*This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement No **823712**"*

VVUQ plan: The V&V 40 [6.1/16] standard provides a framework for assessing the relevance and adequacy of the completed VVUQ activities.

- Question of interest: For an apically implanted LVAD, does the numerical model that includes as inputs: (a) the pump H-Q (head-flow) performance curve, (b) the heart rate (HR), and (c) the pre-LVAD implantation Ejection Fraction (EF); produce flows and velocity fields that agree with the bench experiment?
- Context of Use (CoU): The heart-LVAD computational model may be used to assist in the preclinical design and development of LVAD, by characterising the intraventricular flows for a given pump performance curve. The presented credibility evidence consist of: code and numerical verification by computing the observed rate of convergence in a manufactured solution and a mesh convergence study; UQ with mixed aleatory-epistemic inputs using validation against a bench experiment with six operating conditions. The heart-LVAD computational model will then be used to characterise ventricular flows and derived quantities of interest (QoIs), but by no means replacing animal experiments or clinical trials.
- Model influence: Although the numerical test will augment the evidence provided by the bench test, animal testing and clinical trials are still required. Therefore, the model influence can be categorised as low as it only supports the evidence, and it doesn't solely rely on this computational evidence.
- Decision consequence: The model is only intended to augment the bench test experiment information related with intra-LV flow fields and not to extract any clinical-related conclusion. Despite this, this method can model the device design in operation points in between the operation points. Therefore, the decision consequence is categorised as medium.
- Model risk assessment: As the model influence has been categorised as "low" and the decision consequence as "medium", the LV-LVAD model is categorised with a risk of 2 on the 1-5 scale, therefore requiring a mid to low level goals in the VVUQ plan.

These goals are defined in Table 1. The steps to achieve them are:

1. Provide verification evidence: software quality assurance (SQA) practices should be followed to ensure reproducibility. A strict numerical code verification is mandatory to ensure correctness in the coding of the models. Numerical calculation verification is mandatory to ensure a correct spatial discretisation of the problem
2. Execute a sensitivity analysis in the operation range: A non-linear Sensitivity Analysis (SA) on the operation range of the cases should be executed to (a) understand the impact of each input in the QoIs, (b) Safely reduce the variables for the UQ.
3. Proceed to the uncertainty quantification in multiple validation points: The extreme cases and a middle point should be investigated to ensure a credible solution. A comparison of the QoIs distribution is required including multiple metrics that allow comparing the results with other similar works and future projects.
4. Provide an overall evaluation of the UQ results: a final analysis in which range the model is credible is required to a safe use of the model for predictions.

*This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement No **823712**"*

Aspect				Evaluation		
				Goal	Obt.	Max.
1. Verification	1.1. Code	1.1.1. Software Quality Assurance		C	C	C
		1.1.2. Numerical code verification		D	D	D
	1.2. Calculation	1.2.1. Discretisation error		C	C	C
		1.2.2. Numerical solver error		B	B	C
		1.2.3. User error		C	C	D
2. Validation	2.1. Computational model	2.1.1. Model form		B	B	C
		2.1.2 Model input	2.1.2.1. Quantification of sensitivities	C	C	C
			2.1.2.2. Quantification of uncertainties	D	D	D
	2.2. Comparator	2.2.1. Test samples	2.2.1.1. Quantity of test samples	A	A	C
			2.2.1.2. Range of characteristics of test samples	B	B	D
			2.2.1.3. Measurements of test samples	C	C	C
			2.2.1.4. Uncertainty of test samples measurements	A	A	C
		2.2.2. Test conditions	2.2.2.1. Quantity of test conditions	B	B	B
			2.2.2.2. Range of test conditions	D	D	D
			2.2.2.3. Measurements of test conditions	B	B	C
			2.2.2.4. Uncertainty of test conditions measurements	A	A	C
	2.3. Assessment	2.3.1. Equivalence		C	C	C
		2.3.2. Output Comparison	2.3.2.1. Quantity	B	B	B
			2.3.2.2. Equivalency of output parameters	C	C	C
			2.3.2.3. Rigour of output comparison	C	C	C
			2.3.2.4. Agreement of output comparison	B	B	C
3. Applicability	3.1. Relevance of the Quantity of interest			B	B	C
	3.1. Relevance of the validation activities to the CoU			D	D	D

Table 6.1.1 Goals for the VVUQ steps after the risk assessment.

Results: The SA is intended to highlight the input parameters with a considerable impact in the QoIs. To proceed with the latin hypercube sampling (LHS) a uniform distribution is considered for the SA. 500 samples are obtained with LHS and shown in the scatter plot at Fig. 7.1.1 together with the Person's correlation coefficient ρ. Pearson's ρ is a measure of the strength of a linear association between the two variables in each bivariate plot. From a visual analysis of the scatter plot, it appears that the data is nonlinear, heteroskedastically distributed, and contains multivariate outliers, failing 3 of the 7 assumptions required for Pearson's analysis. To overcome this issue Sobol indices are calculated by constructing a polynomial chaos expansion (PCE) surrogate model using the Wiener-Askey scheme using an expansion order of 5 and 1000 samples in the emulator was used to compute the total Sobol indices. Sobol indices provide

*This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement No **823712**"*

information of the importance of each input taking into account complex factors like nonlinearities, input interactions, and sample dispersion. The Total Sobol index of each input with respect to each QoI are shown at the tornado plot in Fig. 6.1.1. The larger the index, the more important that input is for the QoI.

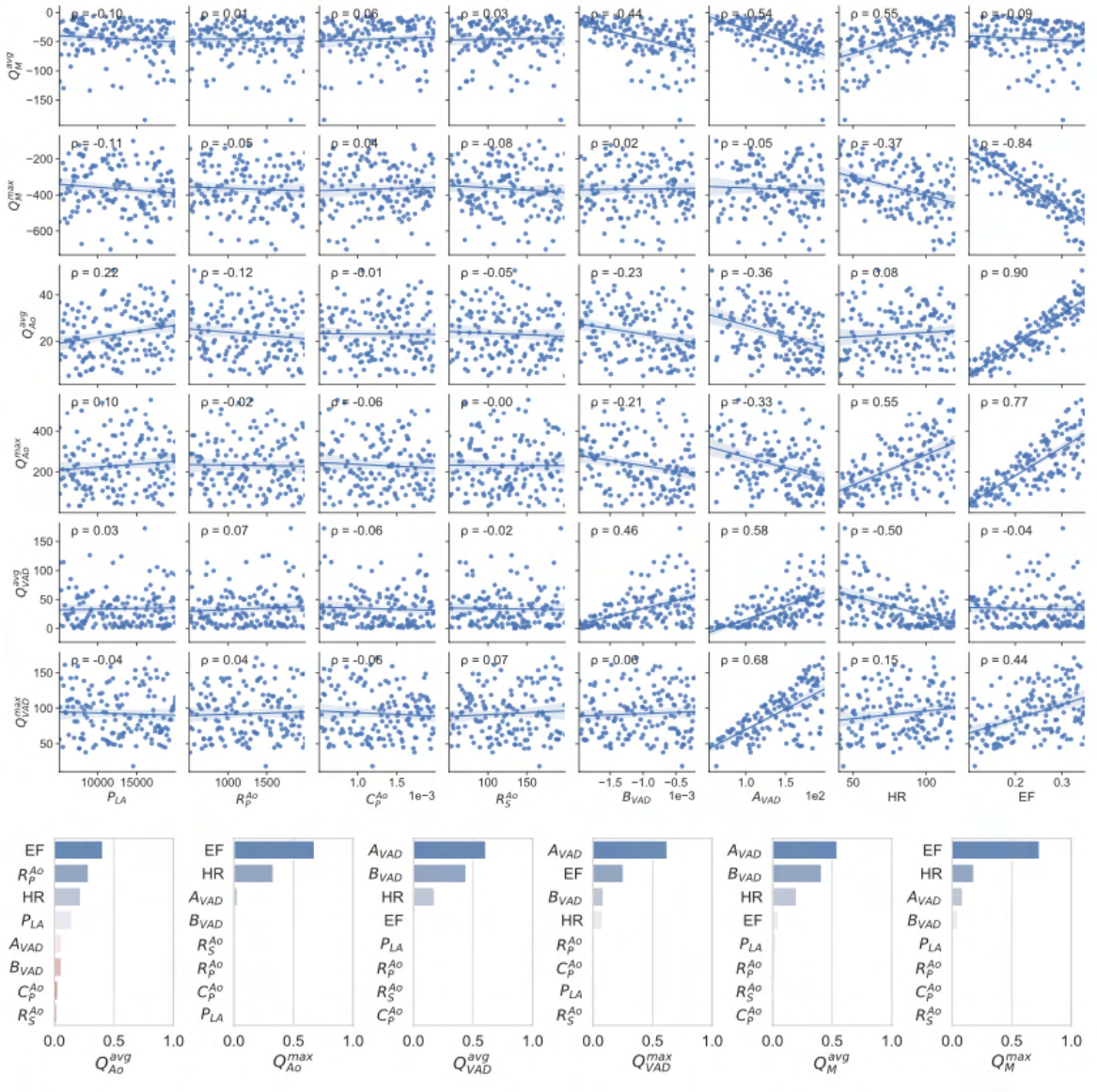
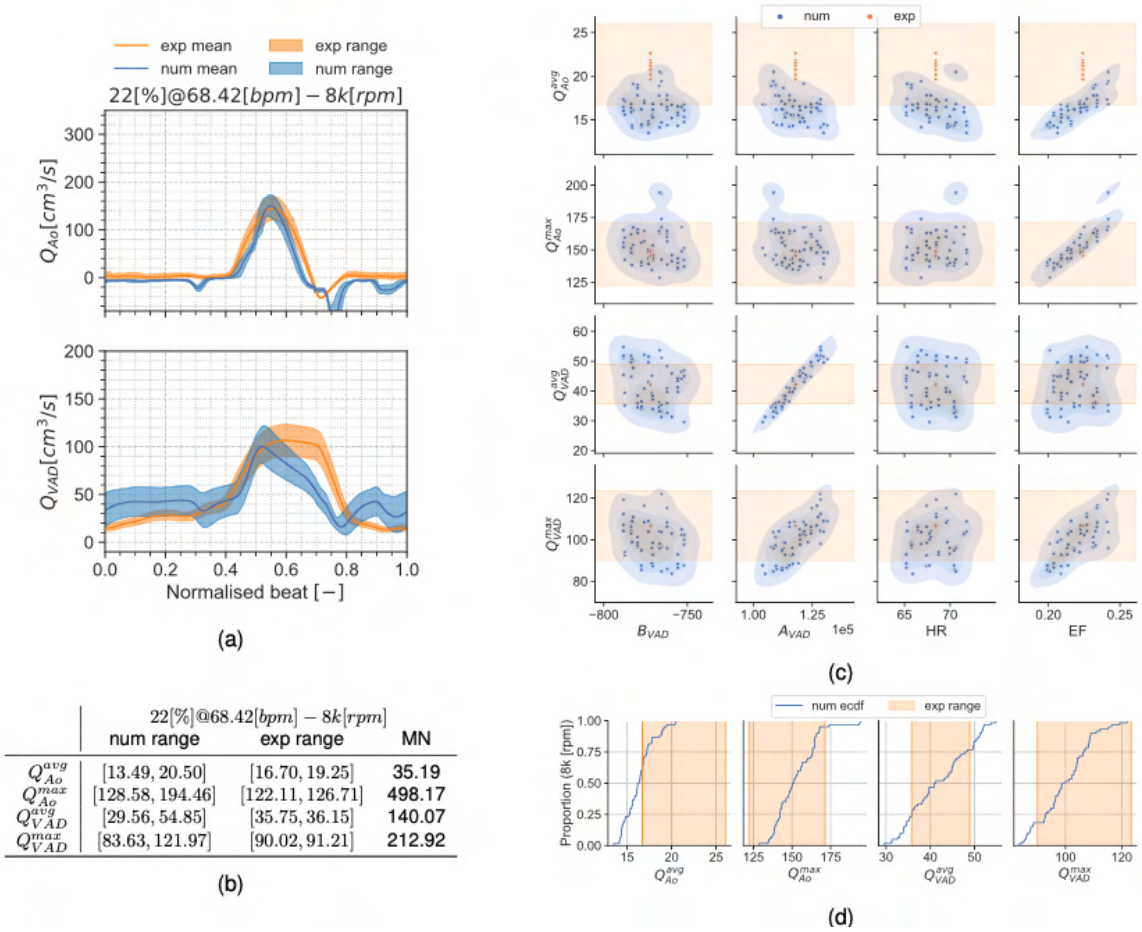


Figure 6.1.1.1 Scatter plots for the inputs and outputs and Sobol indices for the inputs.

The UQ consists of six validation points. As we count with a single execution of the bench experiment, the results are treated with an epistemic error range that is intended to account for the user error. On the contrary, the multiple executions of the numerical experiment let us use the statistical data for the output. Figure 6.1.2 shows an example of one validation point (22[%]@68.42[bpm] and 8k[rpm]). Results were analyzed through scatter plots, empirical cumulative distribution functions (ECDFs), and multiple validation metrics. The validation metric

*This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement No **823712**"*

is computed as in [6.1/18], the minimum Minkowski L1 norm (MN) is chosen between the experimental range and the numerical distribution.



(b)

	$22[\%]@68.42[bpm]-8k[rpm]$		
	num range	exp range	MN
Q_{Ao}^{avg}	$[13.49, 20.50]$	$[16.70, 19.25]$	35.19
Q_{Ao}^{max}	$[128.58, 194.46]$	$[122.11, 126.71]$	498.17
Q_{VAD}^{avg}	$[29.56, 54.85]$	$[35.75, 36.15]$	140.07
Q_{VAD}^{max}	$[83.63, 121.97]$	$[90.02, 91.21]$	212.92

Figure 6.1.1.2 Summary for the condition 22[%] @ 68.42[bpm] and 8k[rpm]. 6.1.2/2a: aortic valve and mitral flows. 6.1.2/2b: validation metrics. 2c: scatter plot showing the numerical and bench experiment data. 6.1.2/2d: cumulative distribution for the numerical experiment and epistemic experimental ranges.

6.1.1 VVUQ at the Solution level of Alya (Collaboration between Barcelona Supercomputing Center and University of Oxford)

At the University of Oxford, we are using versatile Alya, developed by Barcelona Supercomputer Center, to unravel mechanisms underlying diseases in the human heart by modelling and simulations. As shown in Figure 6.1.3, the mechanisms in the evolution of human heart diseases are multiscale (subcellular ionic dynamics, myocyte fibre orientation, organ geometry, etc.) and multi-physics (electrophysiology, excitation–contraction coupling, elastic deformation, scar growth and remodelling, blood flow dynamics, etc.), leading to a large number of parameters in computer models. Each parameter plays different roles on the accuracy, efficiency, stability and usability of the model. It is essential to find the optimal parameter set via the VVUQ, before

*This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement No **823712**"*

using our computer models to help doctors make decision in their clinical practices [6.1/20] and to adjust our models to fit other applications, e.g. design of a coronary artery stent or a 3D printer heart simulator. Thanks to the close collaboration with Barcelona Supercomputer Center -- weekly meeting, gitlab code co-developing and ongoing basecamp discussion -- we have verified and validated the state-of-the-art electromechanical coupling models of the human heart for cellular excitation–contraction coupling [6.1/21], primary left ventricle [6.1/22] and biventricular anatomy [6.1/23], and quantified the uncertainty of model to the variability of model parameters [6.1/22, 6.1/23], fibre orientations [6.1/24], geometries and positions [6.1/25]. Although these findings via VVUQ can fulfil the preliminary requirement of our research, an automated and versatile toolkit is really needed for a comprehensive methodology level VVUQ of our models for clinical and industrial applications. More details on this progress and following plan for developing a versatile and user-friendly VVUQ toolkit are reported as follows.

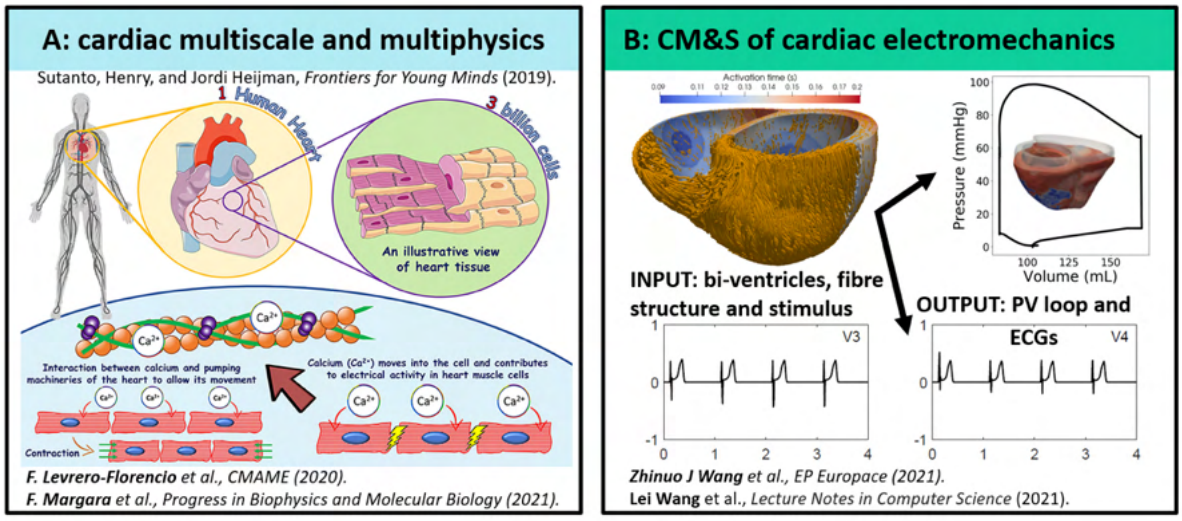


Figure 6.1.1.1 Mechanisms in human heart diseases are multiscale and multi-physics. We are developing computer modelling and simulations (CM&S) to unravel these mechanisms.

Verification and Validation at the Solution Level

The verification of Alya with new functions has already been done by many tests on simple geometries, e.g. see [6.1/21]. For validation, we applied Alya in simulations of 3D human hearts. As shown in [6.1/23], Oxford team with weekly support from Barcelona Supercomputer Center has developed a biventricular model of human cardiac electromechanics, see Figure 6.1.3. The model parameters were calibrated to have validated physiological behaviour, i.e. correct clinical biomarkers of electrophysiology and mechanics, as seen in Table 6.1.2.

Table 6.1.2 Electrophysiological and mechanical biomarkers comparing between literature values and baseline simulation results. Where applicable, imaging methods are detailed in parentheses. LV - left ventricle, RV - right ventricle, EDV - end diastolic volume, ESV - end systolic volume, SV - stroke volume, EF - ejection fraction. SSFP-CMR – steady-state free precession cardiac magnetic resonance, MRI: magnetic resonance imaging, RNV: radionuclide ventriculography. DENSE-MRI: displacement encoded with stimulated echoes magnetic resonance imaging.

Biomarkers	Literature values	Model simulation results
Electrophysiological biomarkers		

*This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement No **823712**"*

QRS duration (ms)	96 ± 9 in men, 85 ± 6 in women	100
QT interval (ms)	350 to 440 s	330
Mechanical biomarkers		
LVEDV (mL)	120, 142 ± 21 (SSFP-CMR), 131 ± 24.5 (tagged MRI)	155
RVEDV (mL)	144 ± 23 (SSFP-CMR)	160
LVESV (mL)	50, 47± 10 (SSFP-CMR), 47.8 ± 12.0 (tagged-MRI)	57
RVESV (mL)	50 ± 14 (SSFP-CMR)	67
LVEF (%)	58, 67 ± 4.6 (SSFP-CMR), 63.1 ± 5.6 (CMR), 62 ± 7 (RNV)	63
RVEF (%)	48 ± 5 (RNV)	57
Peak LV pressure (mmHg)	120, 100-140, 111 ± 4	108
Peak RV pressure (mmHg)	38-40, 67±30	42
Peak longitudinal fractional shortening (%)	16 ± 2 %, ES mid-ventricular mid-wall (DENSE MRI)	11 % shortening from rest, 18 % from end diastole.
Peak wall thickening (%)	33 ± 10 %, radial strain, ES mid-ventricular mid-wall (DENSE MRI)	36 ± 19 % averaged over entire mesh from rest.
Peak torsion angle (°)	peak twist 11.5 ± 3.3° (apex - base) (tagged MRI)	0°

Sensitivity Analysis

Variability in the model parameters

We have aimed to implement state-of-the-art electrophysiological and mechanical models and three-dimensional heterogeneities based on human experimental data at the cellular and organ levels. It was important to investigate the uncertainties in the parameters used to characterise a model, both due to inherent variations extant in the population and also due to coupling effects of various parameters. Building on results from [6.1/22, 6.1/25, 6.1/26], a sensitivity analysis was performed to understand the variation in PV loops and ECG biomarkers to changes in key mechanical model parameters, including pericardial stiffness (K_{epi}), the compliance (C) and impedance (R) of the circulation systems, active tension scaling parameters (T_{scale}) and the linear passive mechanical parameters (K, a, a_f, a_s, a_{fs}). Through high performance computing simulations, we explored large variation ranges (5%--40000%) in these parameters and investigated their electromechanical responses. The largest range of variation was selected to include all uncertainty in model parameters in healthy and diseased conditions. The variation in both individual and combined parameters are considered. Variations in these model parameters influenced mechanical function rather than the electrophysiology and the ECG. The results are effects on ejection fraction (EF), a major clinical biomarker, caused by scaling an individual parameter in Figure 6.1.4 and multiple parameters at the same time in Figure 6.1.5. Figure 6.1.6 shows the combination effects of these parameters, to find the EF is more sensitive to *a* and *K* than other parameters. More results can be found in [6.1/23]. Based on these results we concluded that the effective methods of increasing the EF include decrease in pericardial stiffness (K_{epi}), total arterial resistance (*R*), matrix stiffness (*a*), and increase in active tension parameter (T_{scale}).

*This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement No **823712**"*

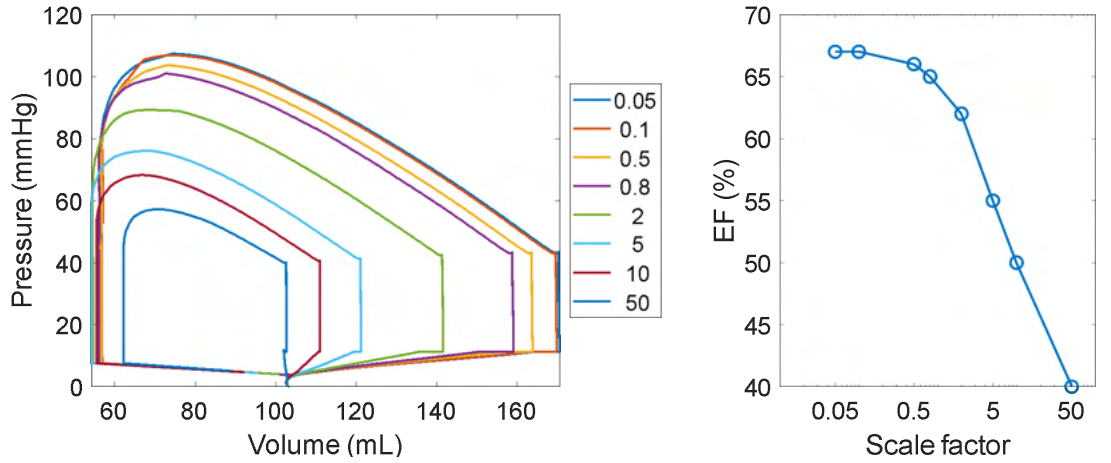


Figure 6.1.1.2 Effects of variation in the pericardial stiffness (Kepi) on PV loops and the LVEF. The legend in the left panel refers to the scale factors multiplying Kepi and the right panel shows their effect on LVEF.

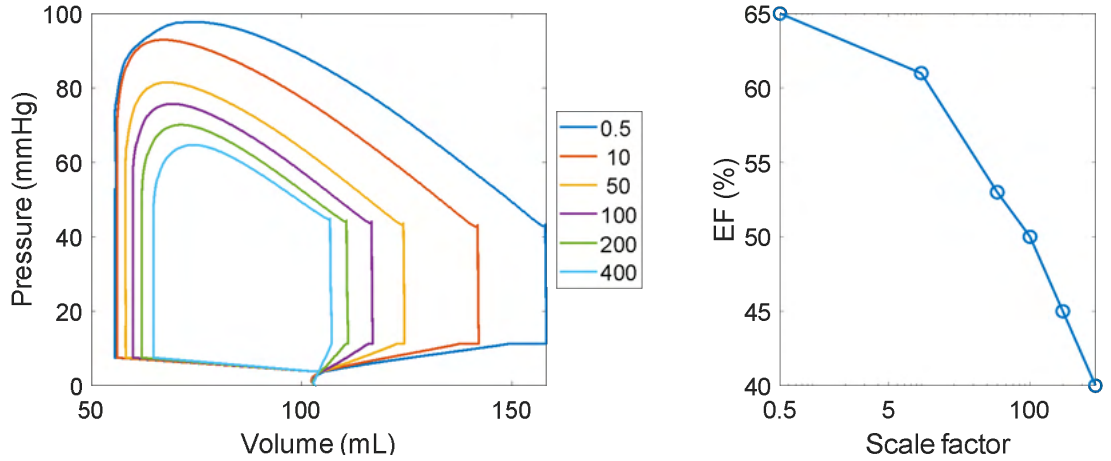
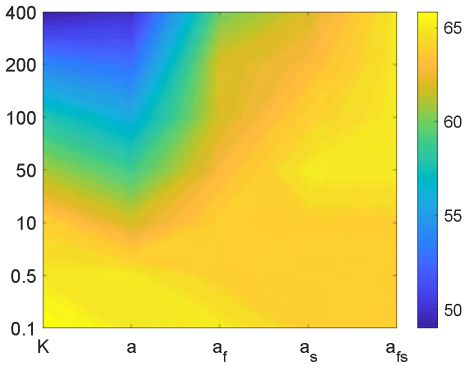


Figure 6.1.1.3 Effect of scaling all linear passive mechanical parameters (K, a, af, as, afs) with same factor: the PV loops and the LVEF. The numbers in legend are the scale factors multiplied on (K, a, af, as, afs), and are used for showing the LVEF on the right.



*This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement No **823712**"*

Figure 6.1.1.4 The EF when scaling linear passive mechanical parameters individually: each of (K, a, a_f, a_s, a_{fs}) is individually multiplied by scale factors (0.1, 0.5, 10, 50, 100, 200, 400) and the colour shows the LVEF

Variability in fibre orientation

The helix orientated fibres in the ventricular wall modulate the cardiac electromechanical functions. Experimental data of the helix angle through the ventricular wall have been reported from histological and image-based methods, exhibiting large variability. It is still unclear how this variability influences electrocardiographic characteristics and mechanical functions of human hearts, as characterized through computer simulations.

One of our research aims is to explore the effects of the range and transmural gradient of the helix angle on electrocardiogram, pressure-volume loops, circumferential contraction, wall thickening, longitudinal shortening, and twist. We run simulations of the state-of-the-art computational human biventricular model [6.1/23] with five models of helix angle variation, parameterized by the helix angle range from endocardium to epicardium their transmural gradient through the left ventricular wall and featured by proportion of circumferentially-oriented fibres with helix angle between ± 22.5° across the myocardium.

Through this UQ study with detailed results in [6.1/24], we have found: 1) Both electrocardiographic and mechanical biomarkers are influenced by these two factors, through the mechanism of regulating the proportion of circumferentially-orientated fibres (helix angle between ±22.5°); 2) With the increase in the proportion of circumferentially-orientated fibres, the T-wave amplitude decreases, circumferential contraction and twist increase while longitudinal shortening decreases.

Variability among patients

Using the cardiac electromechanical biventricular simulation techniques developed in this project, we aim to reproduce the complex and variable phenotypes observed in clinic and to explore the underlying mechanisms. For instance, different post myocardial infarction patients often display complex ECG phenotypes at different healing stages, such as ST segment elevation, and T wave inversion. As shown in [6.1/23], we are able to reproduce the evolution of ECG characteristics from hours to weeks post infarction, as well as the inter-subject variability between transmural and subendocardial infarction.

Summary and future work

We have developed computer models of the human heart to unravel the multiscale and multi-physical mechanisms. The accuracy, efficiency, stability and usability of these models rely on selection of a larger set of parameters. Following a comprehensive testsuite BSC developed for verification and validation on individual scale and physics, we at Oxford calibrate our 3D models of the human hearts such that they can output physiological clinical biomarkers. Following this validation—a baseline model was built, we run many simulations to quantify the uncertainty to the variance in the model parameters and clinical measurements (the fibre orientation from DTMRI). For the moment, we can only analyze the principal parameters based on experience and some trial simulation. Also, each simulation is time-consuming, taking about 1000 core hours for simulation of a heartbeat. Therefore, a better VVUQ toolkit will be developed via collaboration with BSC and other CBM partners. This would enable the sampling of the

*This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement No **823712**"*

parameter spaces with a minimal number of trial simulations, the analysis of the effects of combination of parameters, and the performance of the VVUQ automatically and graphically.

6.2 HemeLB (University College London)

A complete understanding of the VVUQ characteristics of a given simulation study, not just in biomedicine but in numerical research generally, requires the knowledge of several pieces of information. On one hand, one must know the uncertainty in the measurements of the physical system against which the simulation results are being judged. The availability and reliability of these can vary depending on the complexity and scale of the physical system as well as how invasive the process of recording observations is on its operation. On the other hand, one must also have an understanding of the uncertainty inherent within the numerical algorithm being deployed. This can be evaluated against problems with repeatable boundary conditions and analytical solutions. Gaining an understanding of such characteristics will enable more reliable predictions to be made from simulations when input parameters require variation.
Several of the cardiovascular applications in CompBioMed are built on codes using the lattice Boltzmann method. Whilst many publications throughout the literature have examined the impact of parameter variation (particularly the relaxation parameter τ used in single and two relaxation time collision kernels and the lattice spacing necessary for grid independent results) on the accuracy or stability of solutions, we are not currently aware of a rigorous study specifically dedicated to UQ for the lattice Boltzmann method. With a significant upcoming allocation of compute time on ARCHER2, we plan to conduct such a study with HemeLB using the VECMA Toolkit [6.2/1]. This work will specifically represent a solver level study of HemeLB and the lattice Boltzmann method more broadly. This work however will also begin to inform solution level studies conducted with the applications. The results of this work will not only inform the wider cardiovascular studies conducted with CompBioMed applications but should become a critical resource for a simulation method that is becoming widely used in several areas of science and engineering.

6.3 CovidSim Analysis (University College London)

Throughout the global pandemic caused by the SARS-CoV-2 virus, epidemiological modelling has been responsible for informing government decisions on the implementation of lockdowns and other social restrictions. In some situations, the codes used to generate such predictions may be closed-source and have not been subjected to detailed scrutiny, particularly from the perspective of VVUQ. In the UK, such a code - CovidSim [6.3/1] - was used to inform the implementation of the first national lockdown in March 2020. Subsequent to this, the code was made publicly accessible and members from UCL conducted a rigorous UQ analysis of the code to understand its parametric uncertainty, model structure uncertainty and scenario uncertainty. Special focus was given to assessing the robustness of the model to its input parameters and trying to quantify how the code amplifies uncertainties from the inputs to the outputs. This represented a rigorous solver level analysis of the CovidSim code. In this study, it was observed that CovidSim required 940 inputs to generate its predictions. These were examined using a dimension-adaptive sampling method to identify the most influential subset (found to contain

*This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement No **823712**"*

19). The analysis conducted demonstrated how significantly the output of the model can change when settings for these key parameters are adjusted (see Figure 6.3.1).

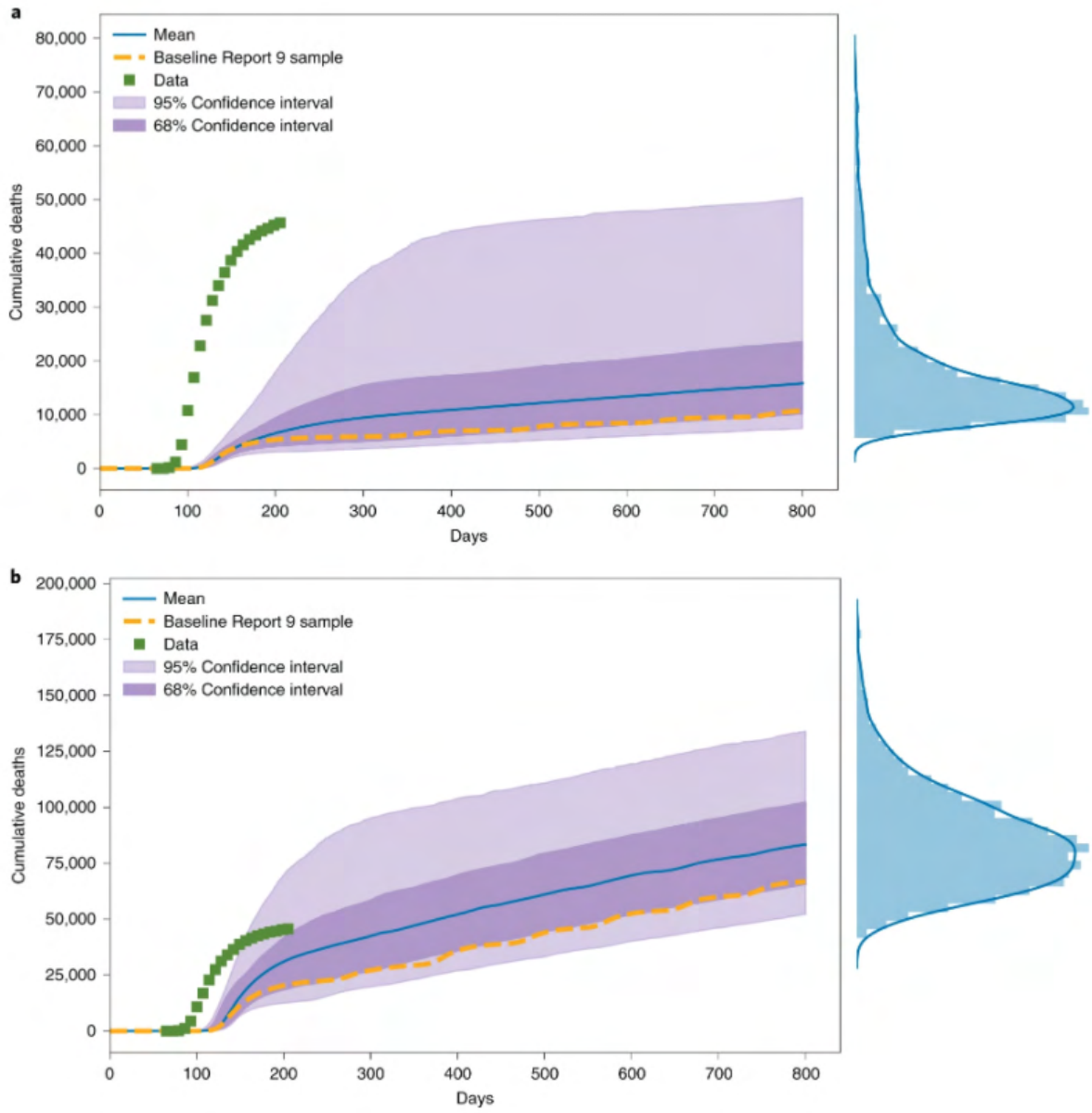


Figure 6.1.1.1 Cumulative deaths predicted by the CovidSim model for two different sets of scenario parameters – the R_0 of the virus and the trigger levels for implementation of interventions based on intensive care admissions. In (a) $R_0 = 2.4$ and ICU on/off triggers of 60/15 cases whilst in (b) $R_0 = 2.6$ and ICU on/off triggers of 400/300 cases. Variation of the results is estimated by variation of other model variables within the given suppression strategy. The green dots represent the observed cumulative deaths in the UK. 'Report 9' is the document using CovidSim that informed the UK Government decision to impose the March 2020 national lockdown.

It was observed that in scenario (a) the key 19 variables accounted for almost 80% of the model variation seen in Figure 6.3.1, in (b) this rose to 90%. Of this subset the 3 most influential variables accounted for 50% and 67% of respective model variation. The study found that CovidSim amplifies the input uncertainty by 300% depending on the chosen scenario of interventions.

*This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement No **823712**"*

It should also be emphasized that the authors of the CovidSim report used by the UK Government did not claim that their parameterization at the time would be able to match the death count data of the coming months. The main message was that it would "…be necessary to layer multiple interventions, regardless of whether suppression or mitigation is the overarching policy goal", and it also showed that doing nothing at all would have disastrous consequences.
The findings of this work [6.3/2] exemplify how sensitivity analysis and uncertainty quantification can help improve model development efforts, and in this case support the creation of epidemiological forecasting with quantified uncertainty.

6.4 BAC and MD (University College London)

Classical molecular dynamics has been in widespread use across many areas of science, from physics and chemistry to materials, biology and medicine. The method continues to attract criticism due its oft-reported lack of reproducibility which is in part due to a failure to submit it to reliable uncertainty quantification (UQ). Here we investigate the uncertainty arising from a combination of (i) the input parameters and (ii) the intrinsic stochasticity of the method controlled by the random seeds [6.4/1]. To illustrate the situation, we make a systematic UQ analysis of a widely used molecular dynamics code (NAMD), applied to estimate binding free energy of a ligand-bound to a protein. In particular, we replace the usually fixed input parameters with random variables, systematically distributed about their mean values, and study the resulting distribution of the simulation output.

A well-known application of molecular dynamics involves the prediction of the binding affinity of a lead compound or drug candidate with a protein target, which is of central importance in drug discovery and personalised medicine. The binding affinity, also known as the free energy of binding, is the single most important initial indicator of drug potency, and the most challenging to predict. There are various approaches to estimate the magnitude of the binding free energy. The choice of which computational method to use is influenced by the desired accuracy, precision, time to solution, computational resources available, and so on. Even today, all these methods remain prone to sizeable errors and are deemed unreliable for decision-making [6.4/2].

In the last few years, we have developed an ensemble-based protocols for free energy calculations, termed "enhanced sampling of molecular dynamics with approximation of continuum solvent" (ESMACS) [6.4/3]. Ensemble approaches lead to increased reliability and reproducibility, with tighter control of standard errors. Here we perform a binding affinity calculation and an uncertainty quantification study using ESMACS approach, applied on a molecular complex of the bromodomain-containing protein 4 (BRD4-BD1) and the tetrahydroquinoline (I-BET726) ligand.

The parameters used in MD simulations are usually calibrated to reproduce one or more available measurements from experiments, calculations from quantum mechanics, or both. While NAMD has a large number of inputs (175) the majority of them are not relevant for forward UQ, as they do not directly influence the solution. Using expert knowledge, we selected a subset of 14 parameters which are known to have an impact on simulation behaviour, to which

*This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement No **823712**"*

we assigned uniform input distributions. After eliminating these inputs, the listing was reduced down to 25 parameters. These remaining parameters can be classified into two groups:

Group 1: "Physical parameters" which control the thermodynamics of the equilibration and binding processes; these essentially refer to the duration, the temperature and the pressure of the simulations (e.g. *setTemperature*, *BerendsenPressureTarget*, *time_sim1*).

Group 2: "Solver parameters" which affect the algorithm used to compute the solution of the molecular dynamics equations; they modify the actual physics solved as well as the accuracy of the resolution (e.g. *initTemperature_eq1*}, *timestep*, *cutoff*).

From the physical parameters we selected a total of 4 parameters based on our experience with MD: temperature, pressure, equilibration duration and sampling duration. Solver parameters were more numerous; there are 21 in total. However, 11 of these parameters are discrete variables which may not be suited for adaptive sampling methods, depending on the method used. Moreover, adding these parameters would drastically increase the cost of the UQ campaign. Because of their influence on the solver behaviour, we do not expect the discrete-valued parameters to have a strong impact on the binding affinity.

For the 14 remaining inputs, we choose uninformative uniform distributions to reflect our lack of knowledge in the most-likely values of these inputs, with bounds at ±15% from their nominal values. Only the temperature is also varied in a reduced range ([280K,320K]) for physical reasons.

The parametric configurations of the simulations, hence not the random seeds, are iteratively refined in directions where a variance-based error metric is largest. Each iteration creates an ensemble of model evaluations, which we executed in parallel on the SuperMUC-NG supercomputer at the Leibniz-Rechenzentrum (LRZ) in Germany. We limited our study to the consumption of a budget of 2,000,000 CPUhs, which were allocated for this work. The computations were orchestrated using the VECMA Toolkit (VECMAtk) [6.4/4], and specifically the EasyVVUQ library [6.4/5-6]. Ensembles are chosen to contain a (large) number *N* of replicas such that adding one more replica does not change the statistical properties of the ensemble. The embarrassingly parallel computations of ensembles is particularly suited for modern supercomputers. As NAMD is compute intensive, our strategy consisted of repeated refinement of the sampling plan until our computational budget was depleted. This occurred at 63 samples from the joint input probability distribution function in the reduced temperature range (123 samples in the full temperature range). For each sample, 25 replicas are simulated (using the same 25 seed values every time), each replica constituting an individual microstate. Their ensemble average corresponds to the thermodynamic macrostate. As a result, 1575 (3075 in the full temperature range) ESMACS workflow executions are completed for the purpose of this analysis. The use of an ensemble of replicas is standard in the field of UQ, in which a sufficiently large number of replicas are run concurrently from which reliable statistics can be extracted. Indeed, because molecular dynamics is intrinsically chaotic, the need to use ensemble methods is fundamental and holds regardless of the duration of the simulations performed. The number of replicas necessary in the ensemble varies from one system to the other and must be determined by direct investigation. Our previous studies show that, starting from reliable initial structures such as those obtained from high resolution crystallography experiments with extensive equilibration (each replica was separately equilibrated for 2ns in the case of small

*This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement No **823712**"*

proteins of approximately 150 amino acids), accurate and reproducible results can be achieved from ensemble simulations consisting of 25 replicas with 4ns production runs [6.4/4].

The study shows that the aleatoric uncertainty induces significant variations of the predicted binding energies. The standard deviation associated with the aleatoric uncertainty amounts to two-thirds of that associated with epistemic uncertainty. It should be noted however that the amount of epistemic uncertainty is directly linked to the assumed variance of the input distributions, such that the ratio of aleatoric to epistemic uncertainty changes with the input distribution of the parameters.

The distributions of properties predicted using classical molecular dynamics are commonly assumed to be Gaussian. The assumption, however, needs to be assessed in many cases, particularly when long-range interactions are involved. Our simulations show that there is a significant non-zero probability of observing moderately (positively) skewed distributions. The excess kurtosis is consistently negative, meaning that compared to a normal distribution, the tails are shorter and thinner. Overall, these results imply the presence of non-normal distributions. Finally, we note that skewness and kurtosis appear uncorrelated with the box size, while they are linearly correlated with the temperature.

In summary, we have performed a sensitivity analysis of binding free energy calculations using VECMAtk and EasyVVUQ library [6.4/4]. The study reveals that, out of a total of 175 parameters, just six dominate the variance in the code output. Furthermore, we show that binding energy calculation damps the input uncertainty, in the sense that the variation around the mean output free energy is less than the variation around the mean of the assumed input distributions, if the output is ensemble-averaged over the random seeds. Without such ensemble averaging, the predicted free energy is five times more uncertain.

With the above conclusions, it is evident that we need to perform ensemble simulations to obtain reliable results with proper UQ. Therefore, another area of interest that we are currently working in is the development of workflow manager middleware that enable automatic handling of complicated heterogenous workflows involving large number ensembles with different types of calculations performed concurrently. In this regard, we have substantial experience with RADICAL CyberTools (RCT) [6.4/7-8] that is already employed on a variety of supercomputers. It has been developed by our partners in the US. There is continuous improvement in its performance as well as applicability. It now allows concurrent usage of CPUs and GPUs on the same node for different computational tasks. Its further development is ongoing. In addition, we are now developing a European workflow manager named QCG pilot job manager (QCG-PJM) [6.4/9]. We are currently testing and benchmarking it on all major supercomputing centres including PSNC, LRZ, ARCHER2 and SURFSara. We have identified three applications covering each of the three compute patterns defined in CompBioMed2 for this task. Successful implementation of such middleware on HPC clusters would substantially facilitate performing complicated workflows on HPCs at large scale. This is especially true for BAC applications that includes ESMACS and TIES.

6.5 Hemocell (University of Amsterdam)

HemoCell is an open-source (www.hemocell.eu) high-performance code to simulate dense cellular suspensions. The main application area of HemoCell is pathologic blood flows around

*This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement No **823712**"*

thrombi or in relation to diabetes. The advancement in these applications require a high-fidelity constitutive model to be able to compute the accurate trajectory and deformation of the cells. The current constitutive model for the most numerous cell species in blood (red blood cells) has been validated by comparing to experimental measurements in both single cell and many-cell scenarios [6.5/1]. The form of the mechanical model equations is motivated by the biological spectrin-link structure of the cells and contains three fundamental parameters that are fitted to experimental results. These represent the strength of the spectrin-links (κl), which effectively define the Young-modulus, the bending stiffness of the cell membrane (κb), and the viscosity ration of the inner cytoplasm (Λ) compared to the suspending medium (blood plasma).
As a first step we have investigated the model sensitivity of these parameters. Sensitivity Analysis (SA) is the study of how the uncertainty in the output of a model can be apportioned to different sources of uncertainty in the model input. The aim of SA is thus to quantify the sensitivity of the model output to its input parameters rather than the uncertainty of the model output as is done in Uncertainty Quantification (UQ). There are several reasons for using SA. SA can, for example, be used for testing the robustness of a model to confirm if the model is not too sensitive regarding only a few parameters, which can cause unreliable model output. SA can also be used to find which parameters or regions in the space of parameter values are critical for the dynamics of a model. If some parameter values are found to contribute very little to the output of a model, the parameters can be fixed, or the model can be simplified accordingly. SA is often done alongside UQ, since both practices can yield important results used for model validation.

The sensitivity analysis was performed on the HemoCell RBC (red blood cell) model to gauge for indications of possible unidentifiability for the RBC model parameters. For the analysis the variance-based Sobol method for Global Sensitivity Analysis has been applied [6.5/2]. The results of the sensitivity analysis are shown in Figure [6.5.1]. The SIs of the link force coefficient have values close to one. Hence, this parameter is an important model parameter. On the contrary the bending force coefficient has low values of SIs. The viscosity contrast variability slightly affects the model output with the SIs values up to 0.12 for high values of the shear stress.

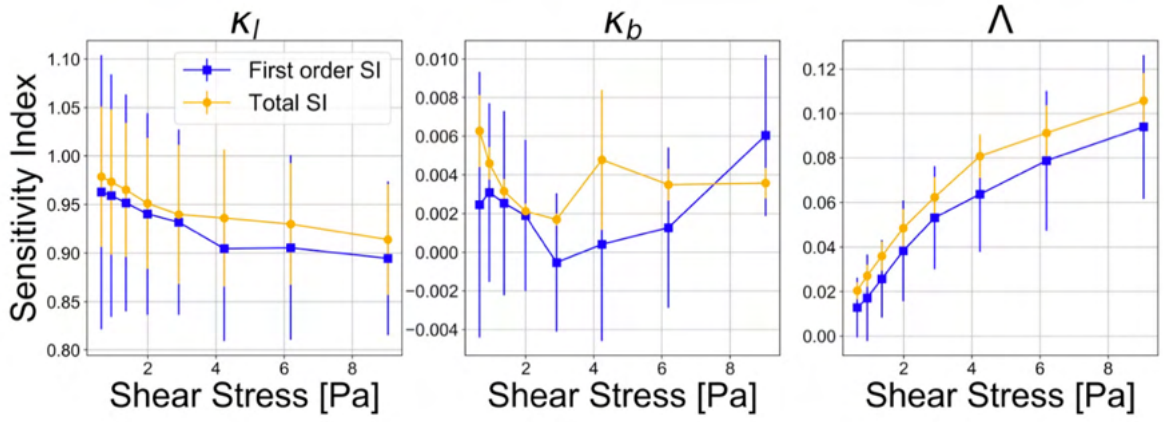


Figure 6.1.1.1 SA results with the HemoCell RBC model, where κl, κb, and Λ are the link force coefficient, bending force coefficient, and viscosity ratio, respectively [32].

In the continuation of this work we apply Bayesian Inverse UQ, where we sample from the posterior distribution from the discrepancy of the model response with ektacytometry data using a prediction error model. We also include the same SA and IUQ analysis on diabetic, stiffened RBCs.

*This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement No **823712**"*

6.6 CT2S/BoneStrength VVUQ plan (University of Sheffield/UNIBO)

We identified a VVUQ strategy and a credibility plan following the endorsed ASME V&V-40 standard [6.6/1] aimed at the credibility assessment for BoneStrength *In Silico* Trials solution, based on CT2S/ARF0 Digital Twin solution. Specifically, the steps to be accomplished within the verification and validation phases, in relation to the computational model features and hypothesized Context Of Uses (COUs) have been established, together with the necessary credibility levels.

6.6.1 Verification

6.6.1.1 Code Verification

CT2S/BoneStrength solution relies on Ansys Mechanical APDL (Ansys Parametric Design Language) computer program. Ansys Mechanical APDL has been extensively used in different engineering fields for commercial purpose since 1970 and it is today widely used in the medical device industry as well. It is certified with the internationally accepted quality standard ISO 9001:2008 and is continuously tested by the developers as new capabilities are added. A Quality Assurance test case library (collection of benchmark studies where Mechanical APDL solutions are compared with known theoretical solutions or experimental results) is available for users to ensure that the software is functioning correctly and produces reliable results. Therefore, 1) Software Quality Assurance (SQA) documentation and 2) quality assurance documentation provided by the Vendor will ensure Ansys code verification in our case.

6.6.1.2 Calculation Verification

We have identified the main steps which will be followed for the achievement of the calculation verification, which will allow us to determine the influence of the model discretization.

1) A general and preliminary calculation verification will be first performed in terms of forces and moments equilibrium and basic nonlinear contact analysis check (e.g., amount of penetration). Also, the time step will be set to find the most efficient and stable solution process as a compromise between increment size and convergence.

2) In order to estimate the discretization error associated with solving the computational problem at a finite number of spatial grid points, a mesh converge analysis will be performed taking into account the spatial discretization effects on the quantities of interest. The mesh refinement will follow ASME V&V10 recommendations, with a refinement factor higher than 1.3. Moreover, numerical uncertainty due to discretization will be evaluated using the Grid Convergence Index (GCI) applied based on Richardson extrapolation theory used for discretization too.

3) A mesh quality check (e.g., shape, aspect ratio, element Jacobians) will be also performed.

4) Because the FE-based bone strength assessment relies on heterogeneous material properties, a sensitivity analysis will also be carried out aimed to assess the influence of the number of different Young's moduli assigned to the model elements. Hence, simulations will be performed based on the binning of the model elements into the highest number of different Young's moduli, which will allow to identify the uncertainty due to the material discretization.

*This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement No **823712**"*

5) The numerical solver error will have to be mandatorily assessed. Therefore, a sensitivity analysis will be performed taking into account solver parameters. In particular, attention will be paid to the contact problem nonlinearity. Contact tolerance and stiffness will be considered, and different cost functions will also be evaluated using Newton-Raphson optimization.
6) The user error will be addressed by an external peer review aimed at the key input and outputs verification.

6.6.2 Technical Validation

6.6.2.1 Computational model

6.6.2.1.1 *Model form*

The key assumptions involved within the form of the FE model for the bone strength assessment are related to the imposed Boundary Conditions (BCs). Hence, we will evaluate the impact of the BCs definition.
In particular, the effect of uncertainties when defining the BCs will be studied. Primarily, the application of a full constraint at the trochanter will be compared to the definition of a non-linear contact in the same region in terms of the effects on the quantities of interest. Besides, uncertainties related to the identification of specific locations for the BCs definition, such as the knee center or the node for the impact load application, will also be quantified.
Additionally, USFD conducted a study on muscle force uncertainty influence on femur strain prediction during gait (please find it attached to this document).

6.6.2.1.2 *Model inputs*

Model results sensitiveness will be evaluated considering the quantity of interest related to the strength definition with respect to the following key input parameters:

1) geometric model resulting from the CT images segmentation procedure

2) density and Young's modulus calibration parameters

3) area considered for the strain averaging in the post-processing phase.

6.6.2.2 Comparator

The reliability of CT2S/BoneStrength solution relies on validation activities performed *in vitro* on cadaver femurs [6.6/2-5]. These latter were performed to validate the principal strains-based fracture prediction based on FE analyses, which is part of the BoneStrength solution. There, excellent agreement between FE outcomes and experimental evidence was found. Details on the results can be found in the Annex.

6.6.3 Clinical Validation

The *in silico* prediction based on BoneStrength will then need to undergo a proper clinical validation aimed to demonstrate its validity and predictive capacity. Hence, the mode

*This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement No **823712**"*

predictions will be compared to the clinical observation in an adequate number of patients/time points.
CT2S has already been demonstrated to stratify fractured and non-fractured patients significantly better than the current standard of care (areal bone mineral density) in a retrospective cohort of 100 pair-matched Caucasian post-menopausal women [6.6/6-7]. Its validation will be further completed using a collection of femur CT scans available at the Rizzoli Orthopaedic Institute (linked third party of UNIBO). The CT scans, which comprise the thigh region, had originally been collected to provide CT-based surgical planning of a total hip replacement procedure, but under an informed consent that included also secondary use for research.

6.7 PALABOS (University of Geneva)

UNIGE addresses VVUQ from different viewpoints: *a priori* verification, improvement of the numerical model, a posteriori validation of the numerical results, sensitivity analysis, and continuous integration of the open-source project based on automatic code testing.
On the aspect of verification, Palabos and its npFEM plugin fully resolve the blood flow, meaning that both the blood plasma and the red blood cells are fully resolved with an accurate non-linear finite element solver. The fluid flow is resolved with the Lattice Boltzmann method and two-way coupled with the deformable blood cells solver through the immersed boundary condition (IBM), forming a complex system of fluid-structure interaction. Both the fluid solver and the finite element solver deliver mesh convergence guarantees.
These convergence guarantees are however not always applicable in practice, as typical current use scenarios of the software on non-exascale hardware operate in a coarse-grained, numerically under resolved regime in order to represent a meaningful domain scale. Moreover, additional sources of uncertainties present themselves in these kinds of simulations. For instance, the most common fluid-structure coupling approach used in the literature, the immersed boundary method, is not a fast-converging method and has the issue of overlapping kernels (no clear distinction is made for areas inside and outside the boundary). Furthermore, other types of boundary conditions usually used to model outer walls (e.g. the artery walls) typically break the parametric behavior of the lattice Boltzmann model. In this case, simulation results are not robust and may typically vary with the values of an ad hoc parametrization of boundary conditions, even when non-dimensional numbers are set.
For this reason, UNIGE research approaches the problem from a fundamental model verification viewpoint: the errors introduced at coarse and asymptotic regimes are systematically analyzed. In particular, accurate boundary conditions are validated in their coarsest operating conditions. This research leads to new types of boundary conditions that can appropriately simulate fluid in the narrow gaps between red blood cells while respecting the parametrization of the numerical model. These accuracy considerations are embedded in a research context focused on computational efficiency and high-performance computing.
Validation and sensitivity analysis apply to the modeling approach used to represent the physics of deformable red blood cells, as well as to the setup of numerical simulations used to determine statistical blood properties. To validate the modeling approach, a series of numerical experiments with varying red blood cell parameters and various flow factors were cross-checked against data available from *in vitro* counterparts and *in silico* studies. Furthermore, the statistics of platelets transport in blood flow was evaluated, to guarantee that the simulations operate at

*This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement No **823712**"*

a numerical resolution for which the result is converged. This validation and sensitivity analysis leads to strong confidence about the validity of current UNIGE models. Further robustness of the tools and the numerical investigations is achieved through the development of new fundamental approaches to reduce the model uncertainties, and to diminish the sensitivity to both numerical resolution and numerical parameters. This allows us to move far beyond the validation range with more reliability.
All model developments are integrated into the Palabos code, in which the newly added code and the existing Palabos code base through a Continuous Integration process. This process, which is based on extensive regression testing, increases workflow reliability, and protects against accidental error injection. Red blood cell model developments have at this stage already been integrated into Palabos Continuous Integration, to which they contributed new regression tests relevant to this field. Theoretical developments of new boundary conditions are intended to be integrated into the Continuous Integration system of Palabos in a near future.

7 Risk Management

The following possible sources of risks have been identified:

a) The internal budget of HPC core-hours might not be enough to cover all the VVUQ analyses defined in Task 2.3.

Probability	Medium
Impact	Medium
Risk assessment	Medium
Mitigation	For now, the consortium is evaluating possible ways to provide access to more core-hours for VVUQ, and partners are encouraged to also apply for external budgets individually. This first deliverable on VVUQ strategy will continue reassessing the situation in detail throughout the course of the project and plan ahead accordingly.

b) Delays in the proposed work due to COVID-related reduced interactions (e.g. reduced mobility, such as on-site visits between partners).

Probability	Low
Impact	Low
Risk assessment	Low
Mitigation	All research tasks using direct experimental data or relying on lab access and mobility have been coping well with the current situation. A prompt migration to online collaborative tools and the surge in the availability of such tools has helped to avoid any delays so far. We continue to keep an eye on the situation as it

*This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement No **823712**"*

	evolves.

c) Insufficient or non-timely input from partners delaying the deliverables.

Probability	Low
Impact	Medium
Risk assessment	Medium
Mitigation	The progress of the deliverables is checked internally on a regular basis. The regular work package meetings and the intra-workpackage teleconferences and discussions facilitate the information exchange and allow the WP leader, or if needed the Project Manager, to step in and act to mitigate the problem.

d) Validation data is incomplete or inaccessible to CMB2 researchers.

Probability	Medium
Impact	Medium
Risk assessment	Medium
Mitigation	Validation has to be performed based on the Context of Use for the specific application. When not enough experimental data is available, a Validation of Unobserved Quantities will have to be employed.

8 Conclusions

This deliverable reported on the VVUQ strategy to be deployed within the CompBioMed project. As a first step, each partner developing a code within the project reported their present planned VVUQ activities. Based on these answers, partner UNIBO will engage with each of these partners and guide them to a redefinition of their VVUQ plans that is consistent with what was exposed in this deliverable, so to better align these codes to their future exploitation. A common and correctly defined strategy will ensure that all partners progress in parallel with the support from other partners that have more advanced applications and more knowledge on the VVUQ process.

In summary, establishing the three-level representation of each code developed within CBM2, and assessing the current status of their VVUQ activities was crucial towards the development of a common strategy that will be key to accomplish the objectives of this Task.

*This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement No **823712**"*

9 Bibliography/References

4.1/1 American Society of Mechanical Engineers, 2018. “Assessing Credibility of Computational Modeling through Verification and Validation: Application to Medical Devices - V V 40 - 2018”. ASME V&V 40-2018, p. 60.
4.1/2 American Society of Mechanical Engineers, 2009. “Standard for Verification and Validation in Computational Fluid Dynamics and Heat Transfer: ASME V&V 20”. The American Society of Mechanical Engineers (ASME).
4.1/3 Morrison, Tina M., et al. "Assessing computational model credibility using a risk-based framework: application to hemolysis in centrifugal blood pumps." *Asaio Journal* 65.4 (2019): 349.
[6.1/1] Benjamin, E. J., Muntner, P., Alonso, A., Bittencourt, M. S., Callaway, C. W., Carson, A. P., Chamberlain, A. M., Chang, A. R., Cheng, S., Das, S. R., et al., 2019. “Heart disease and stroke statistics—2019 update: a report from the american heart association”. Circulation, 139(10), pp. e56–e528.
[6.1/2] Fang, J. C., Ewald, G. A., Allen, L. A., Butler, J., Canary, C. A. W., Colvin-Adams, M., Dickinson, M. G., Levy, P., Stough, W. G., Sweitzer, N. K., et al., 2015. “Advanced (stage d) heart failure: a statement from the heart failure society of america guidelines committee”. Journal of cardiac failure, 21(6), pp. 519–534.
[6.1/3] Jorde, U. P., Kushwaha, S. S., Tatooles, A. J., Naka, Y., Bhat, G., Long, J. W., Horstmanshof, D. A., Kormos, R. L., Teuteberg, J. J., Slaughter, M. S., et al., 2014. “Results of the destination therapy post-food and drug administration approval study with a continuous flow left ventricular assist device: a prospective study using the intermacs registry (interagency registry for mechanically assisted circulatory support)”. Journal of the American College of Cardiology, 63(17), pp. 1751–1757.
[6.1/4] Bartoli, C. R., Zhang, D., Kang, J., Hennessy-Strahs, S., Restle, D., Howard, J., Redline, G., Bermudez, C., Atluri, P., and Acker, M. A., 2018. “Clinical and *in vitro* evidence that subclinical hemolysis contributes to lvad thrombosis”. The Annals of thoracic surgery, 105(3), pp. 807–814.
[6.1/5] Rossini, L., Braun, O. O., Brambatti, M., Benito, Y., Mizeracki, A., Miramontes, M., Nguyen, C., Martinez-Legazpi, P., Almeida, S., Kraushaar, M., et al., 2020. “Intraventricular flow patterns in patients treated with left ventricular assist devices”. ASAIO Journal.
[6.1/6] Lowe, G. D., 2003. “Virchow’s triad revisited: abnormal flow”. Pathophysiology of haemostasis and thrombosis, 33(5-6), pp. 455–457.
[6.1/7] Varga-Szabo, D., Pleines, I., and Nieswandt, B., 2008. “Cell adhesion mechanisms in platelets”. Arteriosclerosis, thrombosis, and vascular biology, 28(3), pp. 403–412.
[6.1/8] Zhao, R., Marhefka, J. N., Shu, F., Hund, S. J., Kameneva, M. V., and Antaki, J. F., 2008. “Micro-flow visualization of red blood cell-enhanced platelet concentration at sudden expansion”. Annals of biomedical engineering, 36(7), p. 1130.
[6.1/9] Tan, K. T., and Lip, G. Y., 2003. “Red vs white thrombi: treating the right clot is crucial”. Archives of internal medicine, 163(20), pp. 2534–2535.
[6.1/10] Chivukula, V. K., Beckman, J. A., Prisco, A. R., Lin, S., Dardas, T. F., Cheng, R. K., Farris, S. D., Smith, J. W., Mokadam, N. A., Mahr, C., et al., 2019. “Small lv size is an independent risk factor for vad thrombosis”. ASAIO journal (American Society for Artificial Internal Organs: 1992), 65(2), p. 152.
[6.1/11] Prisco, A. R., Aliseda, A., Beckman, J. A., Mokadam, N. A., Mahr, C., and Garcia, G. J., 2017. “Impact of lvad implantation site on ventricular blood stagnation”. ASAIO journal (American Society for Artificial Internal Organs: 1992), 63(4), p. 392.

*This project has received funding from the European Union’s Horizon 2020 research and innovation programme under the Grant Agreement No **823712**“*

[6.1/12] Ong, C., Dokos, S., Chan, B., Lim, E., Al Abed, A., Osman, N., Kadiman, S., and Lovell, N. H., 2013. “Numerical investigation of the effect of cannula placement on thrombosis”. Theoretical Biology and Medical
[6.1/13] Liao, S., Neidlin, M., Li, Z., Simpson, B., and Gregory, S. D., 2018. “Ventricular flow dynamics with varying lvad inflow cannula lengths: In-silico evaluation in a multiscale model”. Journal of biomechanics, 72, pp. 106–115.
[6.1/14] Chivukula, V. K., Beckman, J. A., Li, S., Masri, S. C., Levy, W. C., Lin, S., Cheng, R. K., Farris, S. D., Wood, G., Dardas, T. F., et al., 2020. “Left ventricular assist device inflow cannula insertion depth influences thrombosis risk”. Asaio Journal, 66(7), pp. 766–773.
[6.1/15] Neidlin, M., Liao, S., Li, Z., Simpson, B., Kaye, D. M., Steinseifer, U., and Gregory, S., 2021. “Understanding the influence of left ventricular assist device inflow cannula alignment and the risk of intraventricular thrombosis”. Biomedical engineering online, 20(1), pp. 1–14.
[6.1/16] American Society of Mechanical Engineers, 2018. “Assessing Credibility of Computational Modeling through Verification and Validation: Application to Medical Devices - V V 40 - 2018”. Asme V&V 40-2018, p. 60.
[6.1/17] American Society of Mechanical Engineers, 2009. “Standard for Verification and Validation in Computational Fluid Dynamics and Heat Transfer: ASME V&V 20”. The American Society of Mechanical Engineers (ASME).
[6.1/18] Wong, K., Samaroo, G., Ling, I., Dembitsky, W., Adamson, R., Del Alamo, J., and May-Newman, K., 2014. “Intraventricular flow patterns and stasis in the lvad-assisted heart”. Journal of biomechanics, 47(6), pp. 1485–1494.
[6.1/19] Garcia, M. A. Z., Enriquez, L. A., Dembitsky, W., and May-Newman, K., 2008. “The effect of aortic valve incompetence on the hemodynamics of a continuous flow ventricular assist device in a mock circulation”. ASAIO journal, 54(3), pp. 237–244.
[6.1/20] Corral-Acero, J., Margara, F., Marciniak, M., Rodero, C., Loncaric, F., Feng, Y., Gilbert, A., Fernandes, J.F., Bukhari, H.A., Wajdan, A., Martinez, M.V., Santos, M.S., Shamohammdi, M., Luo, H., Westphal, P., Leeson, P., DiAchille, P., Gurev, V., Mayr, M., Geris, L., Pathmanathan, P., Morrison, T., Cornelussen, R., Prinzen, F., Delhaas, T., Doltra, A., Sitges, M., Vigmond, E.J., Zacur, E., Grau, V., Rodriguez, B., Remme, E.W., Niederer, S., Mortier, P., McLeod, K., Potse, M., Pueyo, E., Bueno-Orovio, A., Lamata, P.: The ‘Digital Twin’ to enable the vision of precision cardiology. Eur. Heart J. 41, 4556–4564 (2020). https://doi.org/10.1093/eurheartj/ehaa159.
[6.1/21] Margara, F., Wang, Z.J., Levrero-Florencio, F., Santiago, A., Vázquez, M., Bueno-Orovio, A., Rodriguez, B.: In-silico human electro-mechanical ventricular modelling and simulation for drug-induced pro-arrhythmia and inotropic risk assessment. Prog. Biophys. Mol. Biol. (2020). https://doi.org/10.1016/j.pbiomolbio.2020.06.007.
[6.1/22] Levrero-Florencio, F., Margara, F., Zacur, E., Bueno-Orovio, A., Wang, Z.J., Santiago, A., Aguado-Sierra, J., Houzeaux, G., Grau, V., Kay, D., Vázquez, M., Ruiz-Baier, R., Rodriguez, B.: Sensitivity analysis of a strongly-coupled human-based electromechanical cardiac model: Effect of mechanical parameters on physiologically relevant biomarkers. Comput. Methods Appl. Mech. Eng. 361, 112762 (2020). https://doi.org/10.1016/j.cma.2019.112762.
[6.1/23] Wang, Z.J., Santiago, A., Zhou, X., Wang, L., Margara, F., Levrero-Florencio, F., Das, A., Kelly, C., Dall’Armellina, E., Vazquez, M., Rodriguez, B.: Human biventricular electromechanical simulations on the progression of electrocardiographic and mechanical abnormalities in post-myocardial infarction. EP Eur. (2021).
[6.1/24] Wang, L., Wang, Z.J.W., Doste, R., Santiago, A., Zhou, X., Quintanas, A., Vazquez, M., Rodriguez, B.: Effects of fibre orientation on electrocardiographic and mechanical functions in a computational human biventricular model. Lect. Notes Comput. Sci. (2021).
[6.1/25] Mincholé, A., Zacur, E., Ariga, R., Grau, V., Rodriguez, B.: MRI-Based Computational

*This project has received funding from the European Union’s Horizon 2020 research and innovation programme under the Grant Agreement No **823712**“*

Torso/Biventricular Multiscale Models to Investigate the Impact of Anatomical Variability on the ECG QRS Complex. Front. Physiol. 10, (2019). https://doi.org/10.3389/fphys.2019.01103.
[6.1/26] Cardone-Noott, L., Bueno-Orovio, A., Mincholé, A., Zemzemi, N., Rodriguez, B.: Human ventricular activation sequence and the simulation of the electrocardiographic QRS complex and its variability in healthy and intraventricular block conditions. Europace. 18, iv4–iv15 (2016). https://doi.org/10.1093/europace/euw346.
[6.2/1] VECMA Toolkit (2021), https://www.vecma-toolkit.eu/
[6.3/1] CovidSim Repository (2020), https://github.com/mrc-ide/covid-sim
[6.3/2]W. Edeling, H. Arabnejad, R. Sinclair, D. Suleimenova, K. Gopalakrishnan, B. Bosak, D. Groen, I. Mahmood, D. Crommelin, P. Coveney, "Model uncertainty and decision making: Predicting the Impact of COVID-19 Using the CovidSim Epidemiological Code", Nat Comput Sci 1, 128–135 (2021), DOI:10.1038/s43588-021-00028-9
[6.4/1] Vassaux, M.; Wan, S.; Edeling, W.; Coveney, P. V., Need for Ensembles to Handle Aleatoric and Parametric Uncertainty in Molecular Dynamics Simulation, *J. Chem. Theory Comput.* **2021,** DOI: 10.1021/acs.jctc.1c00526
[6.4/2] Wan, S.; Sinclair, R. C.; Coveney, P. V. Uncertainty Quantification in Classical Molecular Dynamics. *Philos. Trans. R. Soc. A* **2021**, 379, 20200082.
[6.4/3] Wan, S.; Bhati, A. P.; Zasada, S. J.; Coveney, P. V., Rapid, Accurate, Precise and Reproducible Ligand-Protein Binding Free Energy Prediction. *Interface Focus* **2020,** *10*, 20200007.
[6.4/4] Groen, D.; Arabnejad, H.; Jancauskas, V.; Edeling, W. N.; Jansson, F.; Richardson, R. A.; Lakhlili, J.; Veen, L.; Bosak, B.; Kopta, P.; Wright, D. W.; Monnier, N.; Karlshoefer, P.; Suleimenova, D.; Sinclair, R.; Vassaux, M.; Nikishova, A.; Bieniek, M.; Luk, O. O.; Kulczewski, M.; Raffin, E.; Crommelin, D.; Hoenen, O.; Coster, D. P.; Piontek, T.; Coveney, P. V. VECMAtk: A Scalable Verification, Validation and Uncertainty Quantification Toolkit for Scientific Simulations. *Philos. Trans. R. Soc. A* **2021**, 379, 20200221.
[6.4/5] Richardson, R. A.; Wright, D. W.; Edeling, W.; Jancauskas, V.; Lakhlili, J.; Coveney, P. V. EasyVVUQ: A Library for Verification, Validation and Uncertainty Quantification in High Performance Computing. *J. Open Res. Softw.* **2020**, 8, 11.
[6.4/6] Wright, D. W.; Richardson, R. A.; Edeling, W.; Lakhlili, J.; Sinclair, R. C.; Jancauskas, V.; Suleimenova, D.; Bosak, B.; Kulczewski, M.; Piontek, T.; Kopta, P.; Chirca, I.; Arabnejad, H.; Luk, O. O.; Hoenen, O.; Weglarz, J.; Crommelin, D.; Groen, D.; Coveney, P. V. Building Confidence in Simulation: Applications of EasyVVUQ. Adv. Theory Simul. **2020**, 3, 1900246.
[6.4/7] Balasubramanian V.; Turilli M.; Hu W.; Lefebvre M.; Lei W.; Modrak R.; Cervone G.; Tromp J.; Jha S. Harnessing the power of many: Extensible toolkit for scalable ensemble applica- tions. In 2018 IEEE International Parallel and Distributed Processing Symposium (IPDPS). IEEE, **2018**, 536–545.
[6.4/8] Merzky A.; Turilli M.; Maldonado M.; Santcroos M.; Jha S. Using pilot systems to execute many task workloads on supercomputers. In Workshop on Job Scheduling Strategies for Parallel Processing. Springer, **2018**, 61– 82.
[6.4/9] https://github.com/vecma-project/QCG-PilotJob
[6.5/1] Závodszky, G., van Rooij, B., Azizi, V., & Hoekstra, A. (2017). Cellular level in-silico modeling of blood rheology with an improved material model for red blood cells. Frontiers in physiology, 8, 563.

*This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement No **823712**"*

[6.5/2] de Vries, K., Nikishova, A., Czaja, B., Závodszky, G., & Hoekstra, A. G. (2020). Inverse Uncertainty Quantification of a cell model using a Gaussian Process metamodel. International Journal for Uncertainty Quantification, 10(4).
[6.6/1] 'Assessing Credibility of Computational Modeling through Verification & Validation: Application to Medical Devices - ASME'. https://www.asme.org/codes-standards/find-codes-standards/v-v-40-assessing-credibility-computational-modeling-verification-validation-application-medical-devices (accessed Jul. 16, 2021).
[6.6/2] C. Falcinelli *et al.*, 'Multiple loading conditions analysis can improve the association between finite element bone strength estimates and proximal femur fractures: a preliminary study in elderly women', *Bone*, vol. 67, pp. 71–80, Oct. 2014, doi: 10.1016/j.bone.2014.06.038.
[6.6/3] E. Schileo *et al.*, 'An accurate estimation of bone density improves the accuracy of subject-specific finite element models', *J Biomech*, vol. 41, no. 11, pp. 2483–2491, Aug. 2008, doi: 10.1016/j.jbiomech.2008.05.017.
[6.6/4] E. Schileo, L. Balistreri, L. Grassi, L. Cristofolini, and F. Taddei, 'To what extent can linear finite element models of human femora predict failure under stance and fall loading configurations?', *J Biomech*, vol. 47, no. 14, pp. 3531–3538, Nov. 2014, doi: 10.1016/j.jbiomech.2014.08.024.
[6.6/5] E. Schileo, F. Taddei, L. Cristofolini, and M. Viceconti, 'Subject-specific finite element models implementing a maximum principal strain criterion are able to estimate failure risk and fracture location on human femurs tested *in vitro*', *J Biomech*, vol. 41, no. 2, pp. 356–367, 2008, doi: 10.1016/j.jbiomech.2007.09.009.
[6.6/6] P. Bhattacharya, Z. Altai, M. Qasim, and M. Viceconti, 'A multiscale model to predict current absolute risk of femoral fracture in a postmenopausal population', *Biomech Model Mechanobiol*, vol. 18, no. 2, pp. 301–318, Apr. 2019, doi: 10.1007/s10237-018-1081-0.
[6.6/7] M. Qasim *et al.*, 'Patient-specific finite element estimated femur strength as a predictor of the risk of hip fracture: the effect of methodological determinants', *Osteoporos Int*, vol. 27, no. 9, pp. 2815–2822, Sep. 2016, doi: 10.1007/s00198-016-3597-4.

10 Annexes

10.1 Results: VVUQ of CT2S during gait

We investigated the variability of muscle forces on femoral loading using a virtual population [10.1/1], for which the musculoskeletal model (MSKM) is coupled with finite element model (FEM) [10.1/2]. The virtual population was generated based on the **g**ait analysis, CT (0.79 mm x 0.79 mm x 0.63 mm) and MRI (1.08 mm x 1.08 mm x 3 mm) scans of the lower limb of one woman (age, 70.5 years; body mass 61.4 kg; body height, 1.64 m). MRI segmentations were used to build mono-lateral personalised MSKMs of the lower limb [10.1/3]: maximal isometric force was computed as $F_{max} = \sigma V/l_{opt}$, where σ=specific tension61 N/cm^2, l_{opt} = optimal fibre length (estimated from [10.1/3]), and V = muscle volume assigned from samples drawn from uncorrelated normal distributions (Number of samples (N)=100). Distribution mean and standard deviations (SD) were based off MRI measurements taken from 11 elderly women [10.1/1]. These 100 MSKMs (considered here as models of 100 "virtual subjects") were analysed for one walking trial using OpenSim (https://simtk.org/) to obtain muscle forces and resultant joint contact forces (JCFs). Specific gait frames corresponding to the first (P1) and second (P2) peak of hip JCF were identified. Using as input the muscle forces and JCFs at these frames, the

*This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement No **823712**"*

peak principal strains (e1 and e3) at the femoral neck region were predicted for each virtual subject, following CT-based FEM (using ANSYS APDL 19.1) and body-organ coupling models reported previously [10.1/2,4]. The peak element strain energy density (SED) over the full femur volume were also computed at these frames per virtual subject.
Hip JCFs estimated by the 100 MSKM simulations (Figure 10.1.1) varied by up to 0.8 body weight (BW) at P1 and 3.1 BW at P2. The mean ± SD peak e1 and e3 at P1 were 0.37 ± 0.016 % and 0.41 ± 0.016 % respectively, and at P2 were 0.22 ± 0.038 % and 0.27 ± 0.044 % respectively (Figure 6.6.4.1). The mean ± SD peak of the SED was 4.57 ± 0.46 GPa at P1 and 10.73 ± 4.43 GPa at P2.

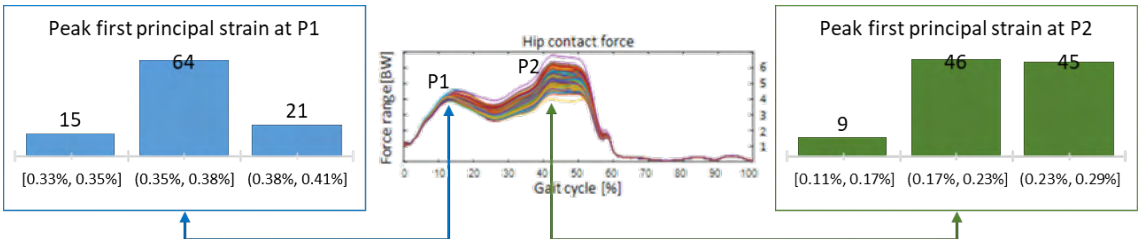


Figure 10.1.1 Right and left, distribution of virtual subjects in respect of the peak first principal strains predicted at the two peaks (P1 and P2) of the hip JCF gait frames. Middle, hip JCFs corresponding to the distribution of F_{max}

Changes in individual muscle F_{max} caused variations in JCF estimates that were larger than those reported by previous studies [10.1/5]. Associated variations in peak principal strain were considerably higher than what has been observed on previous studies based on a much smaller cohort (20 subjects) [10.1/6]. These results indicate that femoral loading is highly sensitive to surrounding muscle functions. In future, techniques such as Gaussian Process Emulator could be used to find a direct link between muscle parameters and predicted femoral strains.

References

[10.1/1] Montefiori et al. Plos ONE 2020.
[10.1/2] Altai et al. Plos ONE 2021.
[10.1/3] Modenese et al. J Biomech, 2018.
[10.1/4] Schileo et al. J Biomech, 42:83-91, 2008.
[10.1/5] Valente et al. Plos ONE, 9:e112625, 2014.
[10.1/6] Kersh et al. JBMR, 33:1999-2006, 2018.

*This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement No **823712**"*