Grant agreement no. 823712

CompBioMed2

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

D2.5: Intermediate Report on VVUQ, Exemplar Research Integration and Community Applications

Work Package:	2	
Due date of deliverable:	Month 36	
Actual submission date:	30 September 2022	
Start date of project:	01 October 2019	Duration: 48 months

Lead beneficiary for this deliverable: BSC
Contributors: *UCL, UVA, UEDIN, UNIBO, USFD*

Disclaimer

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

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***

1 Version Log

Version	Date	Released by	Nature of Change
V0.1	02/09/22	BSC	First Draft
V0.5	27/09/22	BSC	Second Draft after internal revision
V1.0	30/09/22	BSC	Final Draft, submitted to the EC
V1.1	17/02/2023	CBK	Updated version following M36 review

2 Contributors

Name	Institution	Role
Jazmin Aguado-Sierra	BSC	Principal Author
Mariano Vazquez	BSC	Co-Author
Gabor Zavodszky	UVA	Contributor
Antonino Amedeo La Mattina	UNIBO	Co-Author
Gavin J. Pringle	UEDIN	Contributor
Alessandra Aldieri	UNIBO	Contributor
Cristina Curreli	UNIBO	Contributor
Jon McCullough	UCL	Contributor
Alberto Marzo	USFD	Reviewer
Blanca Rodriguez	UOXF	Reviewer

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

3 Definition and Acronyms

Acronyms	Definitions
CBM2	CompBioMed 2
VVUQ	Verification, Validation and Uncertainty Quantification
V&V	Validation and Verification
HPC	High performance computing
T&C	Terms and conditions
Q&A	Question and answers
ASME	American Society of Mechanical Engineers
CoU	Context of use

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

4 Public Summary

This deliverable describes the advances within the project regarding a unified strategy for Verification, Validation and Uncertainty Quantification (VVUQ) in Task 2.5 and the exemplar integration research, as well as review the achievements with the community applications codes of the CompBioMed partners.

As described in the previous deliverable D2.3, the First Report on VVUQ Strategies and Community Application Codes, model credibility is 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 modelling and simulation 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 create a unified strategy and analysis to perform VVUQ techniques on the multiple computational codes involved in CBM2. 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 and 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.

This code-specific strategy and risk analyses were established to address the particular needs of each application at the various stages of development within CBM2. The strategy suggested for each code is presented in this deliverable according to their specific level and context of use.

In a second part of this deliverable, we report on the launch of a CBM2 service, as part of the community application codes: the Scalability Service. The CompBioMed Centre of Excellence in Computational Biomedicine now offers free support to organisations in their initial steps towards parallelising existing computational biomedicine applications, or improving the scalability of those applications already parallelised, and then deploying them on high performance computing resources.

This Service is open to all: not only to CompBioMed Core and Associate Partners, but it is also open to individuals and organisations new to CompBioMed.

The CBM2 partners have created close collaborations as a result of the work within this centre of excellence. The exemplar integration research illustrates some of the collaborative work between CBM2 partners. In this part of the deliverable, the results from collaboration between UCL and LRZ regarding exascale visualization methodologies for HemeLB are reported.

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

5 Introduction

This report describes the strategy for the VVUQ technique application codes in Task 2.5 of CBM2, the exemplar integration research, as well as review achievements with community applications.

The aim of Task 2.5 is to develop Verification, Validation & Uncertainty Quantification (VVUQ) strategies and techniques, in line with the FDA endorsed draft ASME V&V-40 standard [5.1/1] applicable to the use cases arising from CompBioMed2, as was described in deliverable D2.3.

The current consensus is that model credibility is 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 V&V10 and V&V20 [5.1/2, 5.1/3] 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 modelling and simulations 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 computational modelling and simulations as a basis for predicting medical device performance, the ASME V&V40 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 computational modelling and simulations to inform a decision. The decision could be internal to an organization or part of a regulatory activity. The risk informed framework 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 modelling and simulation used to address a question of interest. The assessment of model risk depends on the model influence and the decision consequence (Figure 6.1.1/1). Model influence (low, medium or high) is the contribution of the model outcome to the final decision. Typically, model influence is lower when additional sources of evidence are available, and therefore the model relative contribution to the decision decreases. This strongly depends on the COU and needs to be clearly defined. Decision consequence (low, medium or high) is evaluated based on the severity of possible adverse events following an incorrect decision. The combination of model influence and decision consequence establishes the overall model risk, typically using a five-level risk map (Figure 6.1.1/1). This leads to the establishment of the multiple credibility goals for the multiple stages, involving verification and validation of the computational modelling and simulation. In case of a regulatory application, this is proposed by the applicant and reviewed by the regulatory agency, which may require amendments and/or additional credibility activities.

The activities outlined by the V&V40 to gather evidence and demonstrate model credibility, are reported in Table 5.1. Each activity is evaluated based on the corresponding credibility factors, and the rigor needed for each step is typically assessed based on the COU and model risk. The V&V40 defines a gradation for each activity in the range of (a) to (c), or in some cases (a) to (d), associated with increasing rigor level. For example, if the model is validated by comparison with a single test sample, rigor gradation is (a). While if a statistically relevant number of samples are tested, rigor gradation associated with the activity is (c).

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

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 5.1 Verification, validation and applicability activities reported in V&V40 standard.

This deliverable is intended to showcase the strategies for a subset of the CBM2 codes to perform VVUQ. 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 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 and clinical translation of simulation code, which makes VVUQ mandatory to engage with a range of industries operating across the entire healthcare value chain, from healthcare providers to pharmaceutical and medical device manufacturers, which is part of CompBioMed Objective 6.

Due to the novelty of the ASME V&V40 standard, the application of it is scarce. Only a few publications following the risk assessment framework exist in the literature [5.1/4], therefore the task of applying the ASME standard to new applications, particularly the ones developed within CBM2, involve a considerable step towards more credible simulation results.
The VVUQ strategies for OpenBF, CT2S, BBCT/Bone Strength and Alya Red are described in this deliverable.

Furthermore, information regarding the emergent community application support is reported, which include the new Scalability Service provided by CBM2 for any external user; and the collaborative work between UCL and LRZ to create exascale visualization methodologies for HemeLB simulations.

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

6 Planned activities or Activities Carried out

As a first step we asked each partner developing a code within CompBioMed to report their planned VVUQ activities. The answers were provided in deliverable D2.3, "First report on VVUQ strategies and community application codes".

Starting from these replies, UNIBO engaged with each of these partners to guide them to a redefinition of their VVUQ plans that is consistent with what was exposed in the V&V40 standard, so to better align these codes to their future exploitation. These revised VVUQ plans are provided in this Deliverable D2.5 “Intermediate Report on VVUQ, Exemplar Research Integration and Community Applications”.

The full results of these 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 September 2023.

6.1 VVUQ Strategy per CompBioMed Code

6.1.1 OpenBF solution

Scientific question of interest

Can measurements of blood velocities in proximal cranial vessels inform on the level of perfusion in distal brain territories following an ischaemic stroke event causing complete occlusion of the middle cerebral artery?

Context of Use

The computational model is composed of two parts: a mechanistic representation of the brain circulation, and a probabilistic description of the velocity in important brain arteries, as well as perfusion in distal tissues. The mechanistic model uses as inputs the connectivity of the network of brain arteries and their material properties, as well as the parameters of the windkessel outputs. This model is used to perform a preliminary exploration of the parameter space with the aim of training a Gaussian process emulator in the emulation of outputs of interest, given the model inputs. Outputs of interest are the blood velocities in specific vessels where clinical measurements can be performed, and the perfusion in distal territories. The Gaussian process emulator is used to generate the thousands of data points that are necessary for performing a Sobol sensitivity study on the model: the aim of the sensitivity study is to identify what are the combinations of measurable velocities that better correlates with the distal perfusion. Such combinations can help clinician in understanding what is the perfusion, given the observation of blood velocities in specific locations.

Risk Analysis

Model Influence

We are currently spending effort in validating the model in an animal context. Its application to human patients in clinical context is, at the moment, not in sight. The model gives an indication on the possible status of the patients. We anticipate that the results of the simulations could support clinicians (e.g. neuroradiologists) in deciding how to treat a stroke patient, but it will have a LOW influence in the decision process.

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

Decision Consequence

The model is concerned with stroke patients and what is their perfusion status, which drives the decision of how they should be treated. Incorrect decisions can have serious consequences on the patient, going from severe disabilities to death. The decision consequence is therefore HIGH, as illustrated in Figure 6.1.1/1.

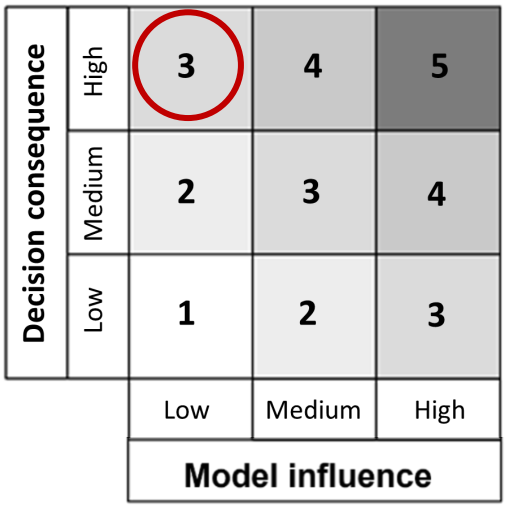

Figure 6.1.1/1. Five-level risk map where the identified OpenBF overall risk has been highlighted as resulting from a low model influence and a high decision consequence.

Table 6.1/2 V&V Computation time requirements of OpenBF.

	Number of Simulations	ShARC nodes	Computation time (hours)
Emulator training	1100	40	6.5
Sensitivity analysis (simulated)	1100	256	650
Sensitivity analysis (emulated)	110,000	0 (1 CPU)	2

6.1.2 CT2S solution

Scientific question of interest

Is the patient at a high risk of hip fracture?

Context of Use

CT2S is an online software-as-a-service tool, used for clinical management of osteoporosis and osteopenia patients. Based on CT2S, the ARF10 tool applies a stochastic biophysics model to estimate the Absolute Risk of proximal femur Fracture over a 10-year period (ARF10) starting from the time of clinical presentation. ARF10 is computed from the patient's height, weight, age, a dual-energy X-ray absorptiometry (DXA) based areal bone mineral density (aBMD) scan and a Quantitative Computed Tomography (QCT) scan of the hip region. The clinician uses the patient's ARF10 and compares it to a population-specific threshold value to determine

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

whether the future risk of the patient to suffer a hip fracture is sufficiently high that the patient should be treated.

Risk Analysis

Model influence
The output of the model has a moderate role in safety decision. Model influence is judged MEDIUM.

Decision Consequence
An incorrect decision could result in severe patient injury or death or other significant impacts. Hence decision consequence is HIGH, as illustrated in Figure 6.1.2/1.

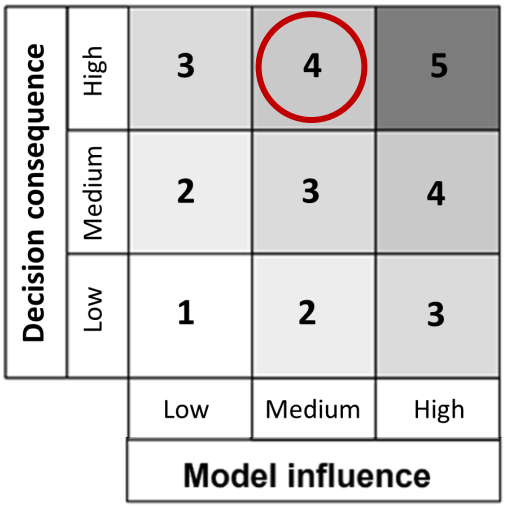

Figure 6.1.2/1. Five-level risk map where the identified CT2S overall risk has been highlighted as resulting from a medium model influence and a high decision consequence.

Verification and validation plan

Below a summarized V&V plan which would represent the starting point for setting up activities for V&V according to ASME V&C-40 for OpenBF and CT2S solutions is provided.

6.1.3 BBCT/BoneStrength solution

Verification and preliminar Validation

Preliminary results were reported in D2.3/6.6.

Scientific question of interest

Which is the optimal dose of a new anti-osteoporosis drug for adults and older adults (from 55 years) according to multi-dose Phase I studies, to be further tested in Phase II studies?

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

Context of Use

BBCT is a methodology where a stochastic biophysics model provides an estimate of the absolute Risk of proximal femur fracture at time zero (ARF0) of a given subject, from their height, weight, and a Quantitative Computed Tomography (QCT) scan of the hip region. This ARF0 is used as a surrogate biomarker of the primary endpoint proximal femoral fracture in place of the measured DXA-based aBMD in multi-dose phase I studies where DXA-based aBMD is currently an accepted outcome.

Risk Analysis

For BBCT, decision consequence was considered low: an inaccurate decision on the efficacy of a new drug in Phase I might relate to patients being either over- or underexposed to the drug in Phase II studies, but that dose would in any case be lower than the maximum tolerable dose and higher than the minimum effective dose, without exposing the patient to any additional risk. The impact on regulatory decision was considered high: the BBCT-based biomarker (ARF0) is proposed as a substitute of aBMD biomarker, which is a surrogate of the fracture endpoint in multi-dose Phase I studies to evaluate the efficacy of different doses for new antiresorptive treatments. As such, the choice of the optimal dose would be based entirely on the BBCT-hip prediction. We are aware that because ARF0 would represent the only evidence used in Phase II studies the impact on regulatory decision might be considered high, but we considered that additional and key evidence for the final regulatory decision will be available from phase III trials. BBCT-hip overall risk was therefore determined to be medium-low for the established CoU, as highlighted in the graphical representation provided by the five-level risk map (Fig. 6.1.3/1).

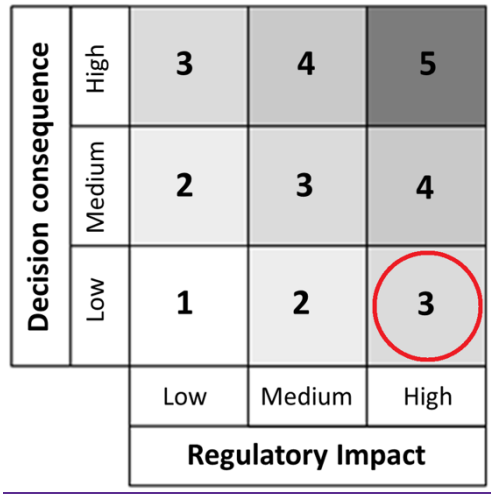

Figure 6.1.3/1. Five-level risk map where the identified BBCT overall risk has been highlighted as resulting from a high regulatory impact and a low decision consequence.

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

Table 6.2/1 V&V plan for CT2S and BBCT solutions according to ASME V&V-40 standard.

Activity	Credibility factor	Rigor Range	Selected	Credibility
Verification				
Code Verification	Software Quality Assurance (SQA) (5.1.1.1)	a-c	b: SQA procedures from the vendors are referenced.	Medium
	Numerical Code Verification (NCV) (5.1.1.2)	a-d	b: multiple benchmark test cases are used to verify the numerical solution.	Medium
Calculation verification	Discretization error (5.1.2.1)	a-c	c: conservation equation balances are checked, and mesh sensitivity study conducted.	High
	Numerical solver error (5.1.2.2)	a-c	c: problem-specific sensitivity study performed on solver parameters.	High
	User error (5.1.2.3)	a-d	b: inputs and outputs verified by practitioner.	Medium
Validation				
Computational model	Model Form (5.2.1.1)	a-c	c: comprehensive evaluation of model form (geometry, choice of equations,…).	High
	Model Inputs			
	Quantification of sensitivities (5.2.1.2.1)	a-c	c: comprehensive sensitivity analysis performed.	High
	Quantification of Uncertainties (5.2.1.2.2)	a-c	c: input uncertainties identified and propagated.	High
Comparator – Observed data	**Test samples**			
	Quantity of test samples (5.2.2.1.1)	a-c	c: statistically relevant number of samples used.	High
	Range of characteristic test samples	a-d	b: samples with range of characteristics near	Medium to High

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

	(5.2.2.1.2)		nominal (to be verified) c: samples representing expected extreme values included. (to be verified)	
	Measurements of test samples (5.2.2.1.3)	a-c	c: all key characteristics measured.	High
	Uncertainty of test sample measurements (5.2.2.1.4)	a-d	c (*in vitro* data).	Medium to High
	Test Condition			
	Quantity of test conditions (5.2.2.2.1)		ND	
	Range of test conditions (5.2.2.2.2)		ND	
	Measurements of Test Conditions (5.2.2.2.3)		ND	
	Uncertainty of Test Conditions Measurements (5.2.2.2.4)		ND	
Assessment	Equivalency of Input parameters (5.2.3.1)	a-c	c: types and inputs equivalent (possible for *in vitro* data).	High
	Output comparison			
	Quantity (5.2.3.2.1)	a-b	b: multiple outputs compared. (to be verified)	High
	Equivalency of output parameters (5.2.3.2.2)	a-c	c: types of outputs equivalent (possible for *in vitro* data). b: types of output similar (for *in vivo* data).	Medium
	Rigor of Output comparison (5.2.3.2.3)	a-d	b-d	Medium

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

	Agreement of output comparison (5.2.3.2.4)	a-c	b-c	High
Applicability				
	Relevance of the Quantity of Interest (QOI) (5.3.1)	a-c	a: QOIs from the validation activities related to those for the CoU c: used for the validation activities equivalent to those for the CoU	Medium
	Relevance of the Validation Activities on the CoU (5.3.2)	a-d	b: partial overlap between the ranges of the validation points and the CoU	Medium

Table 6.3/2 V&V Computation time requirements of BBCT.

	Number of Simulations	Cores per simulation	Computation time (core/hours)
Mesh convergence	11,200	4	230,000
Material parameters	11,200		40,000
Angle sampling	61,500		223,000
BCs	44,800		160,000
Anatomical landmarking	8,400		30,000
Total	134,300		683,000

6.1.4 Alya Red Solver

Sensitivity analysis

The preliminary sensitivity analysis work for the LVAD model was reported in D2.3/6.1. The further development of the VVUQ definition is summarised below.

Scientific question of interest

For an apically implanted left ventricular assist device (LVAD), does the selected pump speed produce: (a) complete aortic valve opening; and (b) a cardiac output compatible with life for a range of heart rate and ejection fraction covering a HF patient population?

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

Context of Use

The heart-LVAD computational model may be used by design engineers to assist in the preclinical development of LVAD, by characterizing aortic root, LVAD and intra-LV flows for a given pump speed. The goal of the heart-LVAD computational model is to provide a computational replica of a benchtop experiment for quantitative analysis in parametric explorations. At this point, the heart-LVAD computational model by no means is replacing animal experiments or clinical trials, but augmenting the totality of evidence.

Risk Analysis

Model influence is Low. Although the numerical test will augment the evidence provided by the bench test to aid design, they do not qualify the safeness of the device. This meaning that animal testing and clinical trials are still required during the regulatory submission to prove safety and efficacy of the device.

Decision consequence is High. While the CoU specifies the usage of the model for design iterations, the results could be used to make indirect decisions that affect the patients' health. If the model fails to make accurate predictions for the question of interest, could advice for an operating condition that produce either: (a) low cardiac output or (b) a permanently closed aortic valve. These might lead to thromboembolic events, aortic regurgitation or death.

Overall, the model risk can be thus considered medium (risk of 3 on the 1–5 scale), as observed in Figure 6.1.4/1.

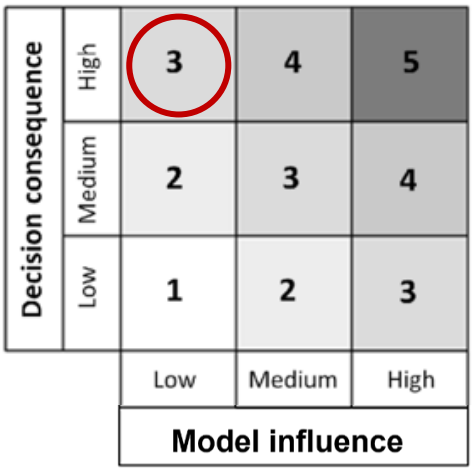

Figure 6.1.4/1. Risk map where the identified LVAD study using Alya overall risk has been highlighted as resulting from a low model influence, but a high decision consequence.

Credibility activities

The detailed credibility includes typical verification, validation and applicability analyses. Among the verification activities are software quality assurance, numerical code verification, convergence and solver parameter analysis. Validation is performed by exploring the influence of model assumptions on the simulation results in comparison to experimental bench data. Furthermore, sensitivity and uncertainty quantification analyses on the input variables are also part of the validation strategy. At every step of the validation, the simulation results are

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

compared to the experimental data obtained with a cardiac simulator. The applicability analyses assess the relevance of the validation results for the context of use.

The approximate computation time employed for this study is reported in Table 6.1.4/1.

Table 6.4.4/1 V&V Computation time requirements of LVAD application using Alya.

	Number of simulations	Cores per simulation	Computation time
Sensitivity analysis	500	256	4 hours
UQ study	500		
Total	1000	256,000	1,024,000 core/hours

6.1.5 HemeLB

We are working to implement VVUQ methodologies with HemeLB in two particular directions. The first is to examine how the algorithmic parameters of the lattice Boltzmann method influence the variability of results of fluid simulations in domains characteristic of common blood vessel geometries. This will help to characterise how simulation parameters and boundary conditions impact the outcomes of some more general scenarios using the lattice Boltzmann method. This campaign was further described in D2.3. The second campaign that we are planning will be particularly focussed on identifying how the choice of flow properties in peripheral vessels can impact the overall nature of blood flow in a large vascular network. This will be used to further develop our simulations of human-scale domains of arteries and veins.

6.1.6 CovidSim

CovidSim was a emergency application employing VVUQ that portrays the capabilities of Compbiomed work and partners, but that it is not part of the core applications that will continue reporting within the VVUQ Compbiomed activities. The main contribution of this code to the VVUQ studies was detailed in D2.3/6.3.

6.1.7 PlayMolecule

PlayMolecule (available at https://playmolecule.com/) is a drug discovery web service that provides a comprehensive set of applications aimed at accelerating and improving drug discovery workflows. Currently the applications on PlayMolecule range from protein, nucleic-acid and small molecule preparation [6.1.7/1], protonation and parameterization [6.1.7/2], molecular dynamics simulations [6.1.7/3], adaptive sampling algorithms [6.1.7/4], and various machine learning predictors [6.1.7/5-7]. Among these applications, we have developed tools aimed at better estimating the binding affinity of a given compound to a protein target. This is the case of K_{DEEP}[6.1.7/6], a 3D convolutional neural network (CNN) that takes as input a voxelized representation of a protein-ligand complex and returns a binding affinity prediction. Importantly, we apply an integrated gradients method to identify which atoms or interactions are more relevant to the binding affinity estimate. This improves the interpretability of the model. Moreover, it helps to validate the network, as we can evaluate if the atoms identified as most relevant match the expectations of a chemist (atoms involved in salt bridges, hydrogen bonds, etc.). Our results suggest that CNNs attribute more importance to classical protein-ligand interactions, such as hydrogen bonds or pi-stacking. This interpretability method, which we call Glimpse [6.1.7/8], is available at https://www.playmolecule.com/Glimpse/.

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

6.1.8 TorchMD

The accurate description of molecular potential energy is critical for computational chemistry and molecular simulations. However, the current simulation engines' compatibility with neural network potentials (NNPs), which have gained a lot of interest in the field, is limited. This hinders the development and testing of new NNPs.
To address this issue, we have developed TorchMD [6.1.8/1], a new molecular simulation application based on the high-performance deep learning python library PyTorch. TorchMD allows researchers to perform all necessary operations to simulate and analyze molecular potential energy using a Torch backend. It includes various types of computations such as bonds, angles, dihedrals, Lennard-Jones and Coulomb interactions, as well as the ability to run standard all-atom simulations using the Amber forcefield.
In combination with TorchMD we have developed TorchMD-NET [6.1.8/2], a PyTorch-based [6.1.8/3] machine learning framework for training and developing NNPs. This framework implements graph-based and transformer models to train NNPs, and has achieved state-of-the-art accuracy on benchmark datasets. With the integration of TorchMD and TorchMD-NET, researchers can efficiently perform simulations and improve the accuracy of NNPs in molecular dynamics.
The equivariant transformer [6.1.8/4], a model based on an attention mechanism, is the most accurate NNP currently available in TorchMD-NET. It has achieved state-of-the-art accuracy on some standard benchmark datasets [6.1.8/5,6]. However, recent advances in the literature show that further mathematical constraints on the network can increase accuracy and efficiency. Currently, a new model is being tested that uses a different mathematical framework while still satisfying these constraints, reducing the computational requirements of traditional NNPs. The equivariant transformer serves as a prime example of the cutting-edge research being done using TorchMD and TorchMD-NET.
TorchMD intends to provide a simple to use API for performing molecular dynamics using PyTorch. This enables researchers to improve in force-field development as well as integrate seamlessly neural network potentials into the dynamics, with the simplicity and power of PyTorch.

6.1.9 HemoCell

The full VVUQ methodology will be reported in the upcoming D2.7, following the results already reported in D2.3/6.5.

6.2 Emergent Community Application Support

6.2.1 Scalability Service

"Task 2.4: Emergent Community Application Support" aims to enable new applications developed within the project or beyond in the wider biomedical community to exploit the CompBioMed HPC infrastructure. To this end, this task has produced the Scalability Service, a free technical service to parallelise and/or scale clients' biomedicine applications.

The technical service offered includes a wide range of support, from informal discussions about efficient use of parallel platforms in general, to code review with a view to offering improvement advice on one hand, to taking the source code and profiling and suggesting improvements, or even working closely with the client and adapting the source code on their behalf.

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

In the last year, the Scalability Service has been completed and recently became active [6.2/1]. The Service includes a detailed application form, a lighter web-form version, and an expert group email: support@compbiomed.eu; documentations: T&Cs, Data Privacy, Data Security, Data Management, and a Corporate SLA; and access to our public Slack channel #scalability, hosted by "In Silico World" Community of Practice, which provides a safe space to share scaling Q&As. The Scalability Service also includes a live, useful overview for programmers with ideas on improving the performance of parallel applications for supercomputers [6.2/2].

This Service was created and is managed by Task 2.4; Task 6.1 and Task 4.5 promotes the Service; T4.3 provides technical effort and an active Helpdesk; and Task 5.2/Task 5.3 assist if the application is to incorporated. Task 2.4 and Task 4.3 have collaborated with both POP CoE and PRACE SHAPE to produce application forms and documentation.

A discussion of what has or are currently exploiting the service, and which facets of the service they employed, will appear in Deliverable 4.3: Report on Uptake of CompBioMed Services, due in month 37.

6.3 Exemplar Research Integration

6.3.1 UCL-LRZ collaborative work

One focus for the UCL developers of HemeLB has been collaborative efforts for visualization with LRZ which has extended previously reported results (D2.2) to datasets of over 7 TB whilst looking to further develop detailed and immersive visualizations of blood flow data (Figure 6.3.1/1) [6.3.1/1, 6.3.1/2]. This work has been presented at SC21 as a finalist in the Scientific Visualization and Data Analytics Showcase and we have continued to develop this platform to generate a wider variety of visualization formats.

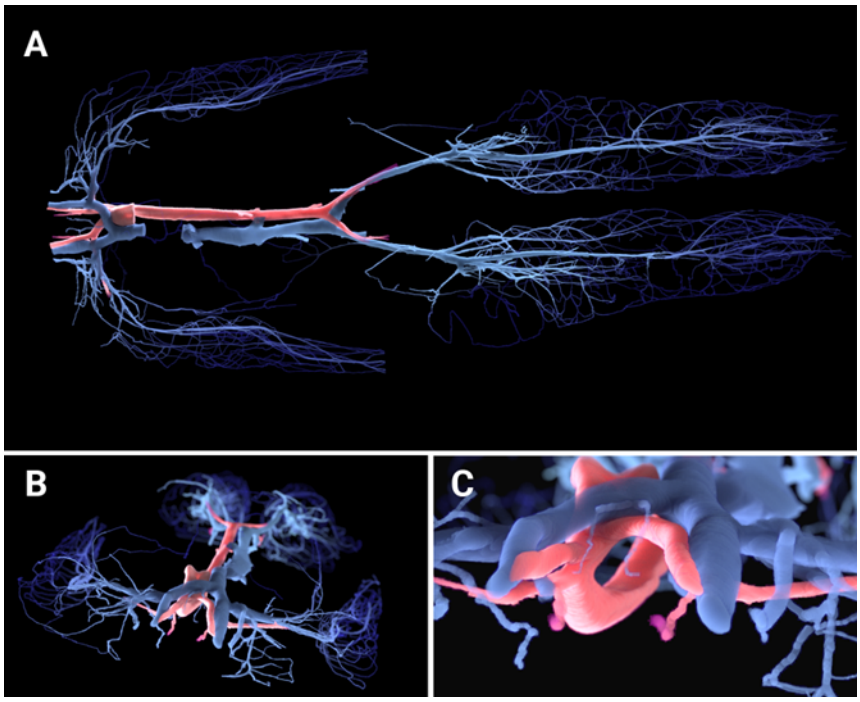

Figure 6.3.1/1: Collaboration between UCL and LRZ has developed the capability to generate detailed and immersive visualizations of the large and complex 3D flow data generated by HemeLB.

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

In collaboration with LRZ we have been able to update and examine the performance of the intrinsic formulation used within the CPU version of HemeLB at full machine scale on SuperMUC-NG. Compared to the original SSE3 implementation, we observed a 48% improvement in walltime and a 62% improvement in parallel efficiency with the AVX implementations when deployed on 6400 nodes (307200 cores). This represents a significant step forward in HemeLB's capabilities for exascale readiness. Looking forward to the increasing diversity of GPU platforms that are to be deployed on exascale HPC, we have been making efforts with ATOS and LRZ to prepare ports of the HemeLB GPU code to enable execution on AMD and Intel GPUs respectively. The AMD port has shown good initial results on small GPU counts. We anticipate being able to test on Intel GPUs within the second half of 2022.

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	The consortium will continue evaluating possible ways to provide access to more core-hours for VVUQ, and partners are still encouraged to also apply for external budgets individually. The VVUQ strategy will continue to evolve so reassessment will be planned accordingly.

b) 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-work package 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 CBM2 researchers.

Probability	Medium

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

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. The VVUQ strategy within CBM2 has progressed satisfactorily. Establishing the CoU's and expected risk in decisions is a key step to advance all codes to follow the VVUQ guidelines to accomplish the objectives of this Task.

A common and correctly defined strategy ensures that all partners progress in parallel with the support from other partners that have more advanced applications and more knowledge on the VVUQ process.

9 Bibliography/References

5.1/1 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.
5.1/2 American Society of Mechanical Engineers, 2020. "Standard for Verification and Validation in Computational Solid Mechanics: ASME V&V 10". The American Society of Mechanical Engineers (ASME).
5.1/3 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).
5.1/4 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.
6.1.4/1 A. Santiago *et al.*, "Design and execution of a verification, validation, and uncertainty quantification plan for a numerical model of left ventricular flow after LVAD implantation," *PLOS Computational Biology*, vol. 18, no. 6, p. e1010141, Jun. 2022, doi: 10.1371/journal.pcbi.1010141.
6.1.7/1 Doerr, S., et al. "HTMD: high-throughput molecular dynamics for molecular discovery" Journal of chemical theory and computation 12.4 (2016): 1845-1852.
6.1.7/2 Galvelis, R., et al. "A scalable molecular force field parameterization method based on density functional theory and quantum-level machine learning" Journal of chemical information and modeling 59.8 (2019): 3485-3493.
6.1.7/3 Harvey, M. J., Giupponi, G., De Fabritiis, G. "ACEMD: accelerating biomolecular dynamics in the microsecond time scale" Journal of chemical theory and computation 5.6 (2009): 1632-1639.
6.1.7/4 Doerr, S. and De Fabritiis, G. "On-the-Fly Learning and Sampling of Ligand Binding by High-Throughput Molecular Simulations" Journal of Chemical Theory and Computation 10.5 (2014): 2064-2069.
6.1.7/5 Jiménez J., et al. "DeepSite: protein-binding site predictor using 3D-convolutional neural networks", Bioinformatics 33.19 (2017): 3036–3042.
6.1.7/6 Jiménez, J., Skalic, M., Martinez-Rosell, G., De Fabritiis, G. (2018). "K deep: protein–ligand absolute binding affinity prediction via 3d-convolutional neural networks" Journal of chemical

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

information and modeling 58.2(2018): 287-296.
6.1.7/7 Skalic., M., Martínez-Rosell, G., Jiménez, J., De Fabritiis, G., “PlayMolecule BindScope: large scale CNN-based virtual screening on the web” Bioinformatics 35.7 (2019): 1237–1238.
6.1.7/8 Varela-Rial, A., et al.”PlayMolecule glimpse: Understanding protein–ligand property predictions with interpretable neural networks” Journal of chemical information and modeling 62.2 (2022): 225-231.
6.1.8/1 TorchMD: A Deep Learning Framework for Molecular Simulations | Journal of Chemical Theory and Computation. https://pubs.acs.org/doi/10.1021/acs.jctc.0c01343.
6.1.8/2 Thölke, P. & De Fabritiis, G. TorchMD-NET: Equivariant Transformers for Neural Network based Molecular Potentials. Preprint at https://doi.org/10.48550/arXiv.2202.02541 (2022).
6.1.8/3. Paszke, A. *et al.* PyTorch: An Imperative Style, High-Performance Deep Learning Library. in *Advances in Neural Information Processing Systems* vol. 32 (Curran Associates, Inc., 2019).
6.1.8/4. Thölke, P. & Fabritiis, G. D. Equivariant Transformers for Neural Network based Molecular Potentials. in (2022).
6.1.8/5. Batatia, I., Kovács, D. P., Simm, G. N. C., Ortner, C. & Csányi, G. MACE: Higher Order Equivariant Message Passing Neural Networks for Fast and Accurate Force Fields. Preprint at https://doi.org/10.48550/arXiv.2206.07697 (2023).
6.1.8/6. Batzner, S. *et al.* E(3)-equivariant graph neural networks for data-efficient and accurate interatomic potentials. *Nat Commun* **13**, 2453 (2022).
6.2/1 https://www.compbiomed.eu/compbiomed-scalability-channel
6.2/2 https://www.compbiomed.eu/rough-guide-to-preparing-software-for-exascale
6.2/3 CompBioMed2: D4.2
6. 3.1/1 E. Mayer, S. Cielo, J. McCullough, J. Gunther, P. Coveney , Visualization of Human-Scale Blood Flow Simulation using Intel OSPRay Studio on SuperMUC-NG, SC21 Scientific Visualization & Data Analytics Showcase (Finalist), https://www.youtube.com/watch?v=Zv2U-Qe6QwU.
6. 3.1/2 J. W. S. McCullough, E. Mayer, S. Cielo, J. Gunther and P. V. Coveney, Immersive visualization of high resolution, 3D blood flow in human-scale vasculatures, Frontiers in Physiology (Under Review).

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