SIMCor

In-Silico testing and validation of Cardiovascular IMplantable devices

Call: H2020-SC1-DTH-2018-2020 (*Digital transformation in Health and Care*)

Topic: SC1-DTH-06-2020 (*Accelerating the uptake of computer simulations for testing medicines and medical devices*)

Grant agreement No: 101017578

Deliverable 6.5

Specification and quantification of synthetic boundary conditions

Due date of delivery: 30 June 2023

Actual submission date: 27 June 2023

Start of the project: 1 January 2021
End date: 31 December 2023

This project has received funding from the European Union's Horizon 2020 research and innovation programmeunder grant agreement N° 101017578

Reference

Name	SIMCor_D6.5_Specification-and-quantification-of-synthetic-boundary-conditions_CHA_27-06-2023
Lead beneficiary	Charité – Universitätsmedizin Berlin (CHA)
Author(s)	Leonid Goubergrits (CHA), Jan Brüning (CHA)
Dissemination level	Public
Type	Report
Official delivery date	30 June 2023
Date of validation by the WP Leader	27 June 2023
Date of validation by the Coordinator	27 June 2023
Signature of the Coordinator	Jan Bruening Digitally signed by Jan Bruening Date: 2023.06.27 20:10:21 +02'00'

Version log

Issue date	Version	Involved	Comments
11/06/2023	1.0	Jan Brüning (CHA), Leonid Goubergrits (CHA)	First draft by CHA
23/6/2023	2.0	Wouter Huberts (TUE)	Internal review
26/06/2023	3.0	Jan Brüning (CHA)	Updated version taking into account reviewers' inputs
27/06/2023	4.0	Anna Rizzo (LYN)	Formal review by LYN
27/06/2023	Final	Jan Brüning (CHA)	Submission by the PC

Executive summary

This document provides specifications for the generation of synthetic boundary conditions. Synthetic boundary conditions are required in two instances. First, they can be used to describe virtual patients or datasets to be analysed via in-silico modelling. Second, they can be used to replace or impute missing information for in-silico assessment of real animal- or patient-specific cases. At first, a generic consideration regarding standardisation, relevant quality measures, and file formats is provided. Subsequently, the relevant synthetic boundary conditions for both project use cases, i.e., *transcatheter aortic valve implantation* (TAVI) *and pulmonary artery pressure sensor* (PAPS), are provided. The detailed descriptions of the use cases and data elements are used as dedicated examples for the more generic considerations. Here, the focus will lie on the generation of synthetic surface geometries, i.e., the anatomical boundary conditions, as the generation of functional boundary conditions, such as volume flow rates, was already addressed in *D6.4 - Specification and quantification of subject-specific data-based boundary conditions (CHA, M24)*, both for patient-specific and synthetic cases.

Table of contents

List of figures

Acronyms

Acronym	Full name
AVA	Aortic valve area
CHA	Charité - Universitätsmedizin Berlin
CT	Computed tomography
CO	Cardiac output
HR	Heart rate
LPA	Left pulmonary artery
MPA	main pulmonary artery
MRI	Magnetic resonance imaging
PA	Pulmonary artery
PAPS	Pulmonary artery pressure sensor
RPA	Right pulmonary artery
SSM	Statistical shape model
SV	Stroke volume
TAVI	Transcatheter aortic valve implantation
TUE	Technische Universiteit Eindhoven

Introduction

Virtual cohorts should mimic real animal or patient cohorts as closely as possible so that they ultimately can be used in in-silico assessment of medical devices. For the development and validation of such virtual cohorts, generation of synthetic boundary conditions is an essential prerequisite. In this context, the term 'boundary condition' refers to all data elements required for parameterisation of an in-silico model for device implantation or device effect simulation, be it surface geometries, inflow and outflow information, or material properties. Here, the term boundary conditions includes both functional and anatomical data. Depending on the type and complexity of boundary conditions, this process can either be trivial or challenging. In theory, cohort characteristics can be measured and replicated using different statistical approaches. For example, if the *stroke volume* (SV) of the left ventricle in a real cohort of patients can be well described using a Gaussian distribution, this information can be used to generate synthetic stroke volumes ensuring that the distribution characteristics of the real and virtual cohorts are identical. Other boundary conditions, such as the 3D patient-specific anatomy, might be more complex and cannot easily be described using simple distributions. Here, other approaches, such as *statistical shape models* (SSM) or *machine learning* (ML) based approaches, like geometry-aware generative adversarial networks, can be used to understand, learn, and replicate the complex patterns observed in real patient or animal cohorts.

Furthermore, in an in-silico model, a patient, be it real or synthetic, is usually characterised not by only one but by multiple boundary conditions that are not statistically independent, such as the patient-specific anatomy and functional parameters describing in- and outflow boundary conditions of the domain of interest. For example, the SV of a patient is directly described by the end-systolic and end-diastolic volume of the left ventricle. If these correlations or causalities between boundary conditions are known, they can and must be used to generate physiologic combinations of these parameters. If, for example, inflow and outflow boundary conditions were sampled randomly, not taking into account their obvious correlation, fundamental principles, such as the conservation of mass, would be violated.

However, even if data elements are not correlated, generating boundary conditions independently of each other might result in a non-physiologic combination that is not a valid synthetic representation of a real patient. Here, different strategies can be used to identify these patients. Examples of these filtering strategies are described in *D7.6 - Proof of principle of the complete virtual patient generator (TUE, M24)*.

In the following document, first, general considerations regarding the generation of synthetic boundary conditions are provided. This includes aspects of standardisation, quality assurance measures, as well as considerations regarding file formats. Afterwards, the methods that were used to generate synthetic boundary conditions for the two SIMCor use cases, TAVI and PAPS, are described. Here, the document focuses on the generation of the anatomical boundary conditions, i.e., 3D surface geometries, as these are vital both for the device implantation and the device effect simulation. Additionally, the generation of synthetic hemodynamic boundary conditions was already described in *D6.4 - Specification and quantification of subject-specific data-based boundary conditions (CHA, M24)*, as the same strategies for generation of those boundary conditions had to be used for the real and synthetic data sets. Furthermore, several aspects of the generation of synthetic data and synthetic boundary conditions have already been addressed in *D5.6 - Completion of synthetic data creation process (UCL, M18)*, and will therefore only be briefly summarised here.

General considerations

One of the key advantages of synthetic data is that this type of data, which is usually generated using automated procedures, can be very well standardised. Further relevant advantages of synthetic datasets are:

- They are usually not affected by missing data or data sparsity;
- They allow to create a well-balanced dataset that covers the full heterogeneity of the real population;
- They allow to generate synthetic data in specific regions of interest, e.g., scaling up the number of rare cases or focussing on specific anatomical properties.

In contrast, real-world data and even clinical trial data is commonly affected by multiple problems regarding data quality. Often, not all data elements are available for all patients, encumbering or even prohibiting building in-silico models describing these patients. Also, data elements might be acquired using devices from different vendors or processed using different software depending on the clinical centre or physician performing the procedure. This can result in data being available in different data formats, being biased, conflicting with each other or of varying uncertainty.

With synthetic data, usually a robust standard can be elaborated on how data should be generated, not only ensuring that every data element used as synthetic boundary condition for each patient is available, but also that it adheres to the same file types and conventions. Therefore, synthetic data, and especially synthetic boundary conditions, are perfectly suited for automation of in-silico modelling procedures, as less variation in specific aspects of the input data, such as mesh resolutions, method for reconstructing or measuring the data elements, or file types, are to be expected. While these aspects can and must be standardised with process-specific standard operating procedures, full automation is better suited to avoid any deviation.

Standards and quality measures

Currently, no explicit standards for generation of synthetic data exist. The wide variety of data elements, generation methods and potential use cases of synthetic data does encumber the definition of a common standard covering all aspects of this complex topic. However, for standardisation within a specific use case, synthetic data and synthetic boundary conditions are perfectly suited. First, if all models for synthetic boundary condition generation are established, a standardised modelling pipeline using this data can be elaborated, allowing for fully automated data processing. Furthermore, as synthetic data sets are usually not affected, they can be shared openly, allowing to generate standardised data sets for benchmarking and comparison of in-silico models, further enhancing standardisation in this field.

Such a processing pipeline usually will closely adhere to any modelling procedure performed using real patient- or animal data. However, it might also be viable to adjust the modelling pipeline. Data and boundary conditions that are only available in a few patients are usually not viable for setting up a numeric model, as it might constrain application too strongly. As a simplified example, volume flow information might only be available in an averaged manner for all patients, by measurement of the SV and *heart rate* (HR). In other patients, transient measurements of the volume flow rate might be available, allowing transient parameterization of the models. In this case, a model might be designed using only the more generic information, as it allows wider applicability of the model. However, if the rarer transient information can be generated in a synthetic manner, they might be usable. In theory, this approach can also be used to enhance real data sets by providing boundary conditions not available from clinical measurements. This approach was also described in D6.4, where transient volume flow rate information in the pulmonary artery and aorta was calculated using either demographic data (body weight of pigs) or clinical routine hemodynamic information (echocardiography-based measurements of SV and HR).

If the in-silico model is changed to accommodate for more suitable synthetic boundary conditions, several aspects must be considered. First, in a hybrid data set containing both real and synthetic boundary conditions, transparent description of which data elements have been measured and which were generated in a synthetic manner. Second, the effects of changing the in-silico model to use other types of boundary conditions must be evaluated thoroughly, as it might compromise comparison and validation of the virtual cohorts generated using the synthetic boundary conditions.

To ensure that the synthetic data is of sufficient quality, several approaches are possible, following the tiered, three-level validation scheme envisaged also for the virtual cohort generators in SIMCor. Here, distributions of generated synthetic boundary conditions should be compared against the real boundary conditions for training. For that, equivalence tests are used to assess whether relevant deviations in distribution, means or standard deviations occur. Furthermore, it is important to also assess the interaction between boundary conditions, as one patient commonly is represented by a set of parameters. As these parameters are not necessarily statistically independent, their correlations must also be reflected by the synthetic data. If this is not the case, non-physiologic combinations of boundary conditions might be represented in the synthetic data sets. However, it is to be noted that even if correlations between different parameters and boundary conditions are maintained in the synthetic data sets, non-physiologic virtual patients might be included in the data sets, which can, for example, be identified by sufficient filtering strategies, as described in *D7.6 - Proof of principle of the complete virtual cohort generator (TUE, M24)*.

In more complex data types, such as synthetic velocity inlet profiles or synthetic surface geometries, these comparisons are not trivial. Here, two approaches can be used to assess the accuracy and usability of the synthetic boundary conditions. First, the complex data set can be described using a reduced set of quantifiable parameters. For example, anatomical geometries can be described using various measurements, such as diameters and volumes, which in turn can be compared between the real and synthetic geometries. This description can also be performed using a statistical shape model, allowing more abstract description of geometries using shape modes and shape weights (see D7.6). Second, real boundary conditions of a patient can be replaced using synthetic boundary conditions that were chosen to match closely. For example, a synthetic geometry close to the patient-specific one (e.g., by minimising an error norm on the shape weights) can be used in the numerical model the boundary conditions are intended for. If all other patient-specific boundary conditions are kept, the isolated effect of the synthetic boundary condition can be evaluated.

As previously explained, synthetic data is ideally suited to generate well-balanced datasets, covering the entire variation observed in real clinical data sets. Therefore, they are ideally suited to quantify biases resulting from these variations in a systematic manner, which might be more difficult in real samples, where specific configurations of input parameters might not be reflected, due to the usually small sample sizes.

Formats

For the synthetic boundary conditions, the same file formats and data structures as described in *D6.2 - Database for anatomy and function based on preclinical and clinical data (CHA, M12)* will be used. This allows to directly use the synthetic boundary conditions within the processing pipelines elaborated using patient-specific or animal data. As a general note, it is recommended to choose non-proprietary, open file types for all synthetically generated data sets, to enhance their uptake when made available.

TAVI use case

Within the project, three different approaches for generation of synthetic cohorts including surface geometries have been employed, resulting from the data processing being performed at two clinical centres (CHA, UCL), as well as the integration of synthetic geometries into the virtual cohort generator by TUE. These approaches are described in the following sections. Where applicable, references to D5.6, D6.4, and D7.6 are made, in which most aspects of the generation of synthetic boundary conditions already have been described. In these instances, only a brief overview is provided.

Synthetic anatomy - CHA

At CHA, the patient-specific anatomies of >100 TAVI patients were reconstructed from CT data in a semi-automatic manner, as described in D5.6 and D6.4. From these geometries, various geometric parameters describing the aortic root were measured, including diameters of the left ventricular outflow tract and the sinotubular junction, as well as the aortic valve opening area. These patient-specific measurements were used as input for a Copula-based model, to learn and replicate the boundary value distributions of all geometric, demographic and functional parameters, as well as their interaction (i.e., correlations). Using this approach, synthetic demographic, functional, and anatomical parameters could be created by sampling on these distributions. At this point, these virtual patients were only characterised by numerical values, describing the synthetic values for age, height, weight, diameters, or the aortic valve area and no complex boundary conditions were generated yet (see Figure 1 and 2). However, these parameters mimicked the patient-specific distributions very well, and also all correlations found within the patient-specific data set were matched sufficiently well, with the largest absolute deviations in correlations observed between parameter pairs being below 0.05.

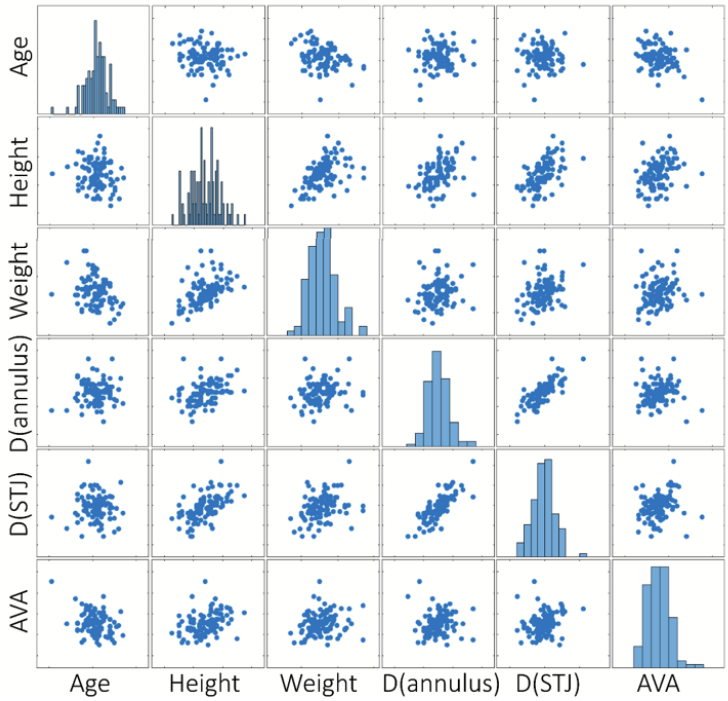

Figure 1: Overview of distributions and correlations between demographic parameters (age, weight and height) and the geometric parameters (diameter of the aortic annulus, diameter of the sinotubular junction and the aortic valve area *(AVA)).*

For generation of synthetic boundary conditions, we used a parametric model of the aortic root including the ascending aorta, which was also the basis for the segmentation. The baseline parameters of this model were fully sampled, spanning the entire range observed in the real data sets, resulting in more than 1 million possible combinations (see D5.6). Anatomical parameters of interest could

directly be calculated from these baseline parameters. Therefore, generation of all possible surface geometries was not necessary. For each synthetic patient, the parametric model with the least deviations (L1-norm) was identified and the corresponding surface geometry was selected (see Figure 3).

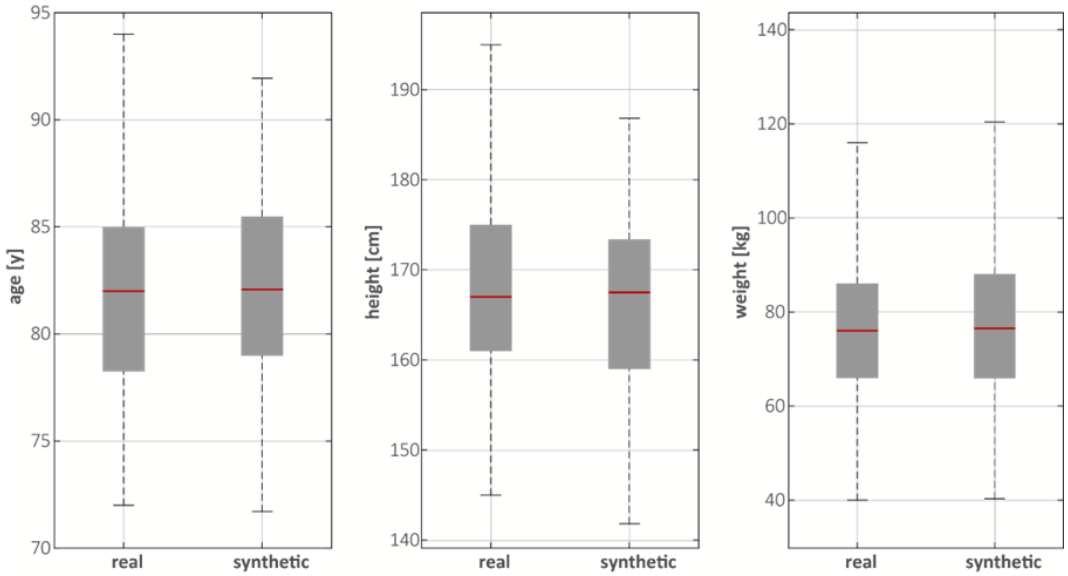

Figure 2: Comparison of the distributions of the demographic parameters age, height, weight, between the real data set and a synthetic data set, which was generated using a Copula-based approach.

Figure 3: Illustration of the selection of suitable synthetic surface geometries (blue) from a large database of potential candidates (red), matching the geometric parameters identified using the Copula-based generation approach.

While this selection step resulted in minor deviations between the anatomical parameters generated by the Copula-based method and those observed in the synthetic surface geometries, both the distributions of anatomical parameters and the correlations between parameters still agreed perfectly with the real data set (see Figure 4).

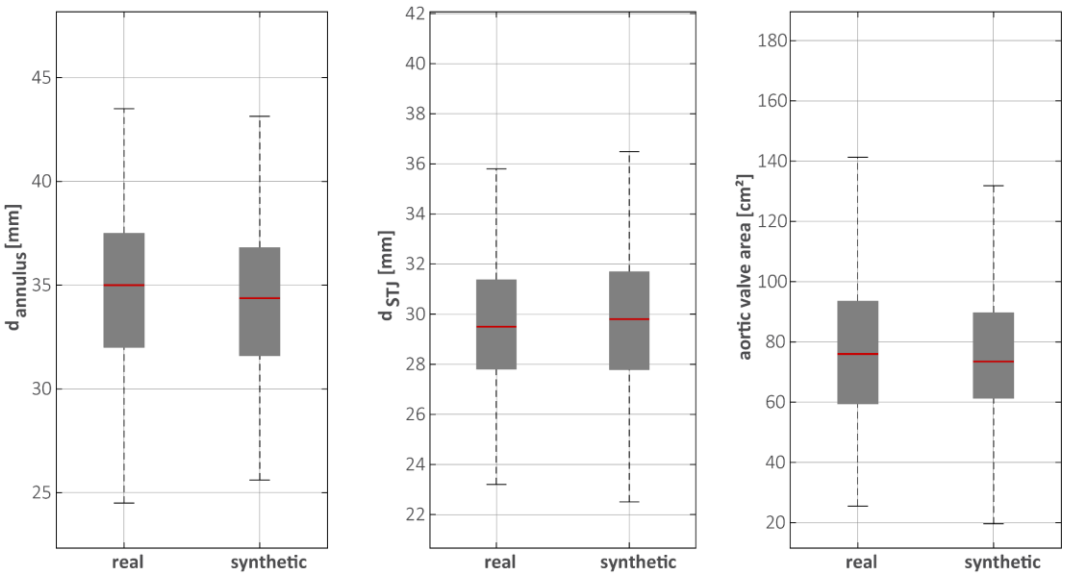

Figure 4: Comparison of the distributions of the aortic annulus diameter, the diameter of the sinotubular junction, and the AVA of the real and synthetic dataset. The values were calculated after selection of best-fitting surface geometries, indicating that this procedure did not result in any relevant biases between the real and synthetic data sets.

Synthetic anatomy - TUE

TUE used the patient-specific reconstructions of the aortic root provided by CHA to generate an SSM of this region of interest. Using this approach, the shape weights of the SSM and their correlations to other parameters, such as demographics and functional parameters, can directly be considered during generation of the synthetic anatomies. The overall approach for sampling of the synthetic data sets, however, is similarly based on learning and reproducing boundary value distributions of the different parameters using Copulas, as the CHA approach. The main difference is that the use of shape weights directly within this step does not require sampling the entire possible parameter space and can, in theory, directly produce only geometries within the physiological envelope. While this approach reduces the probability of generation of boundary conditions outside the physiological envelope, filtering strategies are nonetheless necessary to identify any remaining erroneous cases. Furthermore, filtering strategies can be applied to generate sub-cohorts with specific properties, such as the specific demographic characteristics (e.g., age, sex, weight), but also anatomical and functional aspects, such as a convergent or divergent anatomies of the left ventricular outflow tract, or specific degrees of aortic stenosis severity. While demographic information of course can also be controlled within a real clinical trial by specifying inclusion and exclusion criteria, some more complex anatomical traits require some initial assessment, which might prohibit using these properties in real clinical trials, due to high costs as well as legal and ethical barriers. This filtering approach is detailed in D7.6.

PAPS use case - porcine data

Synthetic porcine anatomy

The generation of synthetic anatomical boundary conditions, i.e., synthetic surface geometries of the pulmonary artery, was already described in detail in D5.6 for the animal cohort. Briefly, a centreline-based SSM of the pulmonary artery was generated, including the main, left, and right pulmonary artery, the branching vessels of the left and right pulmonary arteries, using available retrospective data. Here, a simplified approach was used assuming the same amount of branching vessels on the left and right pulmonary artery to ensure the same topology, which is a prerequisite for SSM.

This approach allowed us to generate a large data set of about 4000 synthetic geometries by sampling of the shape weights using their respective distributions. To subsequently select suitable cases out of this dataset, an interactive tool was created in MATLAB (see Figure 5). The main application of this tool is currently to create a synthetic representation of the chronic animal trial performed in SIMCor, while simultaneously allowing to explore the real and synthetic datasets and assess their similarities with respect to different parameter distributions. Beside the ability to generate a matched virtual cohort, the tool also allows to find multiple synthetic cases matching a real case, therefore generating larger cohorts with the same parameter distributions.

The tool was designed in an interactive manner, to allow the user to select different parameters of interest and their specific importance. For example, if for device implantation simulations a near perfect match in the vessel diameters is considered relevant, the tool will allow to select geometries with small deviations for this parameter, while allowing larger deviations for others. Furthermore, the tool provides an overview of the best matching surface geometries from the synthetic data set, allowing the user to also visually choose a geometry in case multiple well-suited matches are identified based on the geometric parameters alone. The tool is meant as proof-of-concept for a selection procedure based on user-specific requirements and will subsequently be implemented in the VRE or made available as standalone application distributed via the VRE. In the future, the tool will be also adapted to be used for human PAs.

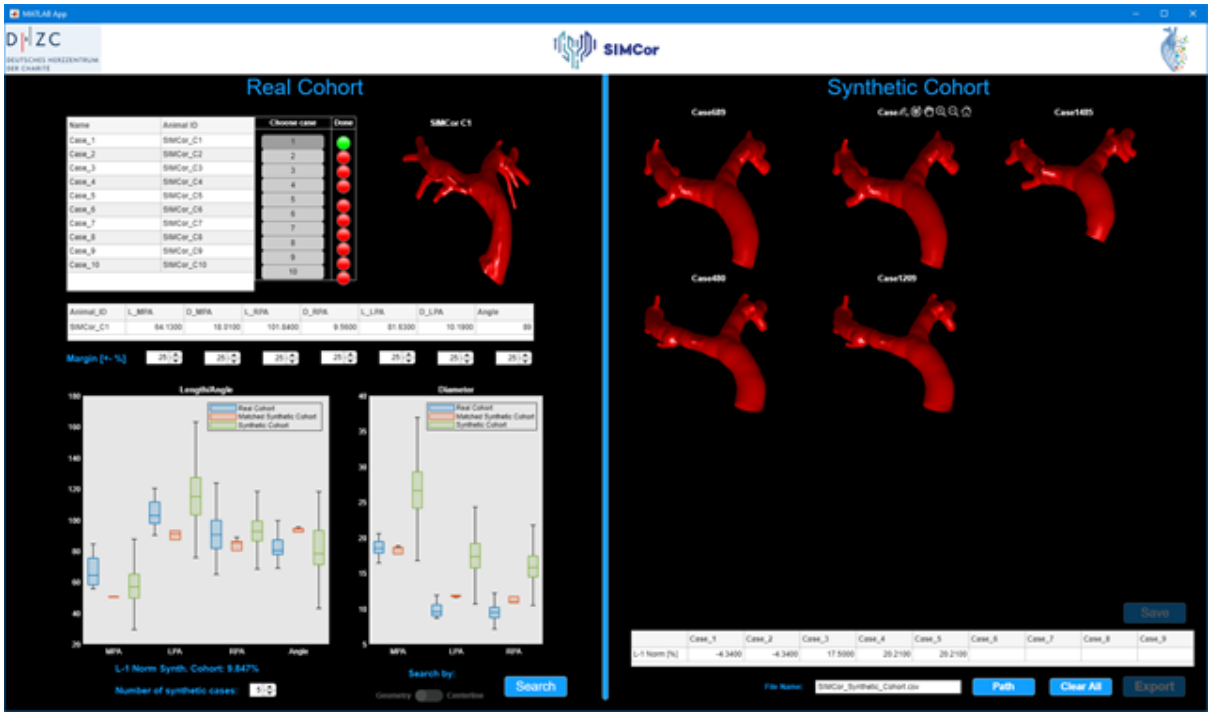

Figure 5: Graphical user interface of the software application for generation of a synthetic cohort of porcine pulmonary artery geometries based on user-inputs.

Synthetic MPA inlet flow rate curves

For the 10 animals investigated in the chronic animal study, no direct information regarding the haemodynamic boundary conditions were available. Flow rate waveforms at the MPA were generated synthetically using a hybrid approach described in D6.4. First, CO and HR were estimated based on the animals' weights, according to scaling laws published earlier[1]. Next, flow rates were synthetically generated based on principal component analysis of earlier published MRI-measurements[2]. Figure 6 shows exemplary 10 flow rate waveforms for the CFD study investigating the effect of the PAPS on the PA hemodynamics.

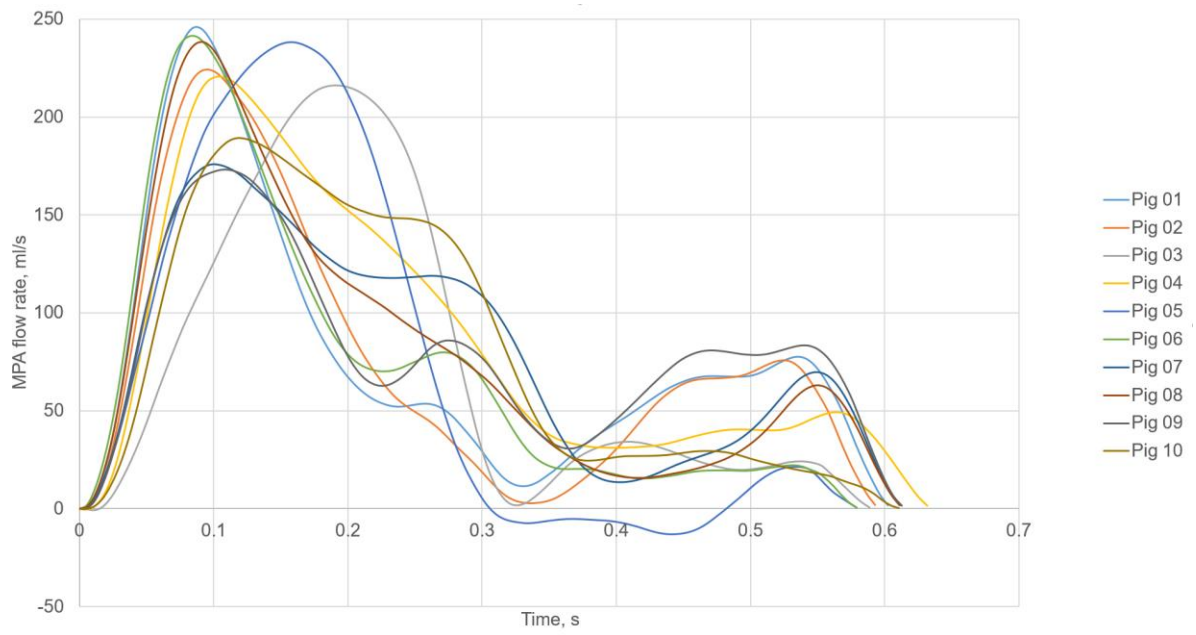

Figure 6: Volume flow waveforms used as inlet boundary condition for all 10 animals of a chronic animal study.

[1] van Essen, G. J., Te Lintel Hekkert, M., Sorop, O., Heinonen, I., van der Velden, J., Merkus, D., et al. (2018). Cardiovascular Function of Modern Pigs Does not Comply with Allometric Scaling Laws. Sci Rep 8, 792

[2] Faragli, A., Alogna, A., Lee, C., Zhu, M., Ghorbani, N., Lo Muzio, F., et al. (2021). Non-invasive cmr-based quantification of myocardial power and efficiency under stress and ischemic conditions in landrace pigs. Front Cardiovasc Med. 8, 689255.

PAPS use case - patient data

The main approach for generation of the synthetic data sets for the human pulmonary artery is similar to the porcine pulmonary artery. The procedure can be summarised as follows. First, the real geometries of the pulmonary artery are characterised using a centreline-based representation. Based on this centreline, all relevant anatomical parameters are calculated. A centreline-based SSM is generated to describe the overall shape variance of the anatomies. This SSM and the information with respect to its shape weights and their distributions is used to generate a large data set of synthetic geometries. Finally, for generation of a synthetic cohort, the most suitable synthetic surface geometries are selected from this large data set. However, the centreline-based SSM had to be significantly altered to allow this approach to also be used for the human pulmonary artery. The reason for this requirement was that the number of smaller vessels branching from the left and right pulmonary artery was more heterogeneous in humans than in animals. Assuming a constant number of branching vessels for the left and right pulmonary artery was therefore not justified.

Consequently, the SSM was adjusted to first only include the main, left and right pulmonary artery. The branching vessels were not directly included in the SSM. Thus, using the SSM, only synthetic geometries of the main vessels can be created using information of the shape weights and their distributions observed in the real human data sets. To also account for the branching vessels, a heuristic model was generated that aims to replicate the variation in the number, diameter, and orientation of the branching vessels. This heuristic model considered correlations found in the real data sets but is to be considered as an addition to the SSM.

This model has a relevant advantage for the device implantation and device effect simulations to be performed for the PAPS use case in SIMCor: the effect of the branching vessel can be analysed individually. To facilitate this, different configurations of branching vessels are generated for the same synthetic main vasculature (i.e., the same shape of the main, left, and right pulmonary artery). In real datasets, this isolated evaluation is not possible, as another configuration of branching vessels will always be associated with differences in the diameters, lengths, and curvatures of the main vessels. As the branching vessels have been identified as a relevant aspect, especially for the fixation of the PAPS devices, this isolated approach might allow us to assess the importance of the branching vessels and how to account for them in the design of the sensor fixation. An example of a geometry with four different configurations of branching vessels is shown in Figure 7.

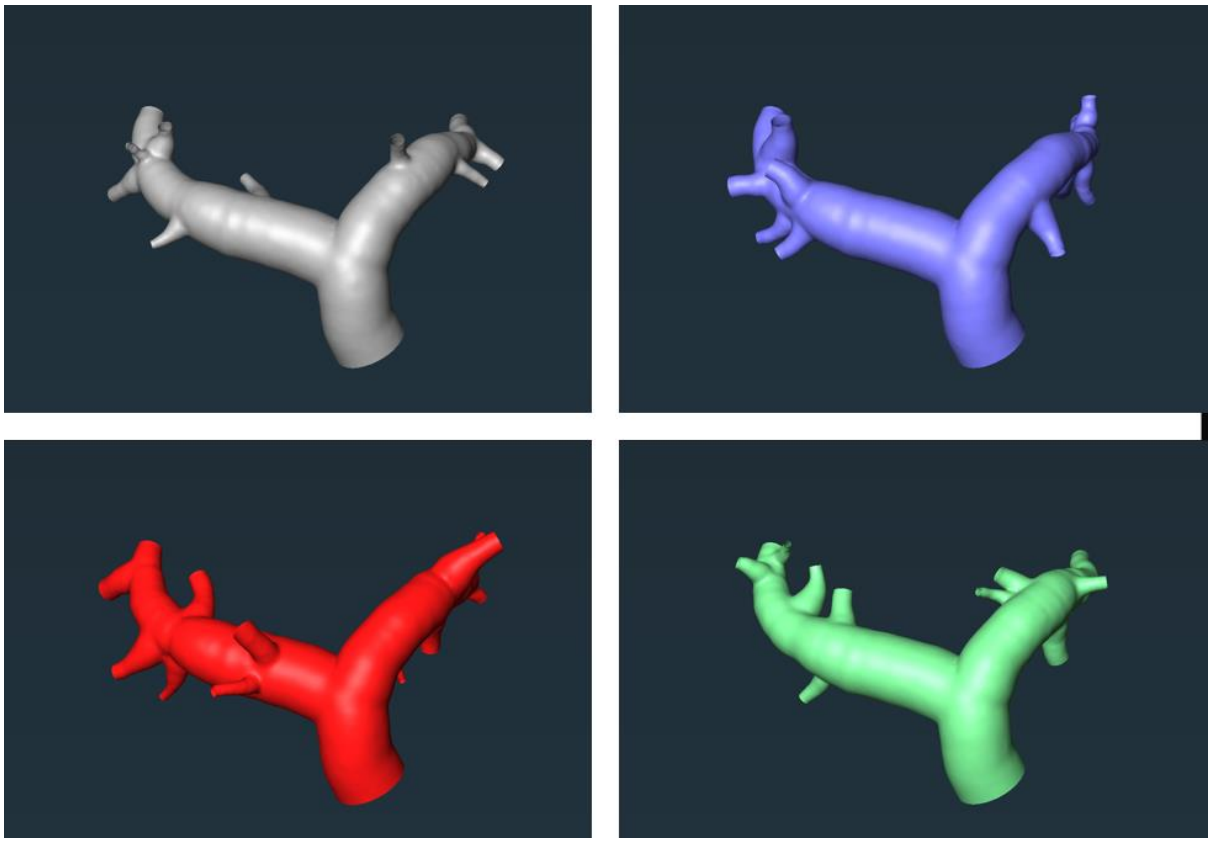

Figure 7: Example of a synthetic main geometry of the human pulmonary artery (main, left, and right pulmonary artery) with four different synthetic configurations of side branches.

Overall, more than 2000 synthetic geometries or 500 different main bifurcation shapes with at least 4 different combinations of branching vessels will be generated. No relevant differences in shape weights and geometric parameters between men and women have been identified. Furthermore, correlations between shape weights, demographic and anatomical parameters were weak in all cases (<0.3). Therefore, we decided against generation of the synthetic parameter distributions using the Copula-based method used for the TAVI use case, as no meaningful correlations were existing that should already be considered during the generation step. Thus, generation of synthetic geometries was achieved by sampling on the shape weight distributions observed in the real data sets.