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 D4.2

Standard operating procedure for data processing for in-silico models

Due date of delivery: 31 December 2021

Actual submission date: 22 December 2021

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 programme under grant agreement N° 101017578

D4.2_- SOPs for data processing for in-silico models	SIMCor – GA No. 101017578

Reference

Name	SIMCor_D4.2_SOPs for data processing for in-silico models_CHA_22-12-2021
Lead beneficiary	Virtual Physiological Human institute (VPHi)
Author(s)	Jan Brüning (CHA), Leonid Goubergrits (CHA), Nina Krüger (CHA), Raphaëlle Lesage (VPH), Wouter Huberts (TUE)
Dissemination level	Public
Type	Report
Official delivery date	31 December 2021
Date of validation by the WP Leader	17 December 2021
Date of validation by the Coordinator	21 December 2021
Signature of the Coordinator	Jan Bruening Digital unterschrieben von Jan Bruening Datum: 2021.12.22 09:26:35 +01'00'

Version log

Issue date	Version	Involved	Comments
06/12/2021	1.0	Jan Brüning, Nina Krüger, Leonid Goubergrits, Raphaëlle Lesage, Wouter Huberts, Silvia Schievano	First draft by CHA
13/12/2021	2.0	Silvia Schievano, Claudio Capelli	Internal review by UCL
17/12/2021	3.0	Jan Brüning (CHA)	Project Coordinator's review
21/12/2021	4.0	Anna Rizzo (LYN)	Final review and formal checking by LYN
22/12/2021	Final	Jan Brüning (CHA)	Submission by PC

Executive summary

This document was developed to provide a *standard operating procedure* (SOP) on clinical data processing for creating virtual cohorts for in-silico models. The SOP is designed for SIMCor partners as well as for the scientific community. Hence, it formulates procedures and workflows for the processing of clinical data (in particular, imaging data) in the view of generating virtual cohorts. Given the heterogeneity of data, general recommendations are provided, rather than precise steps, with specific illustrations based on the SIMCor use case: *transcatheter aortic valve implantation* (TAVI). Standards and guidelines for data quality, data formats and data exchange were applied, when they existed. Confronting the current workflow with the evolving practice as the project continues may lead to updates, thus subsequent iterations are foreseen.

Table of contents

List of figures

Acronyms

Acronym	Full name
CAE	Computer-aided engineering
CSV	Comma separated variables
CT	Computed tomography
DICOM	Digital imaging and communications in medicine
IGES	Initial graphics exchange specification
JPEG	Joint photographic experts group
JSON	JavaScript Object Notation
MITK	Medical imaging interaction toolkit
MRI	Magnetic resonance imaging
NIFTI	Neuroimaging informatics technology initiative
NURBS	Non-uniform rational b-splines
PAPS	Pulmonary artery pressure sensors
SOP	Standard operating procedure
STL	Stereolithography
TAVI	Transcatehter aortic valve implantation
VEC MRI	Velocity encoded magnetic resonance imaging
VMRL	Virtual reality modeling language
XML	Extensible markup language

Cover page of the SOP

SIMCor responsible partner	Charité (CHA)				
Short Title, ID	SOP-SIMCOR-D4.2	**Page**	1	of	X
Title	SOPs for data processing for in-silico models				
Version	1.0	**Created on**	31/12/2021		
Status	draft	**Related SOP**	SOP-SIMCOR-D4.1 SOP-SIMCOR-D4.4		

Purpose and field of application:

This document provides recommendations as well as standardized procedures and workflows for processing of preclinical and clinical imaging data. It focuses on image processing procedures required for generation of virtual cohorts, which is a focal aspect of SIMCor work programme. These virtual cohorts will be used for in-silico modelling of implantable devices aiming at accelerated certification and improvement throughout the development of medical devices. These descriptions are meant to be applicable to a broad range of in-silico models addressing different aspects of medical device performance, either during or immediately after implantation, as well as during long-term function.

Where applicable, existing standards, guidelines and consensus statements for data quality, data formats and data exchange, including standards for statistical data analysis, are referenced.

While data processing is an essential step for virtual cohort generation, it requires well standardized procedures for data acquisition, providing reliable data of good quality. Errors and biases made already during data acquisition will propagate to the data processing procedures as well as all subsequent steps. A dedicated standard operating procedure (SOP-SIMCOR-D4.1) describing procedures and requirements for data acquisition is envisaged in SIMCor. For this reason, this document only provides general recommendations, from which a standardized procedure for a given problem to be investigated must be developed. For those standardized procedures, recommendations regarding documentation provided by SOP 4.2 should be considered.

Furthermore, the deliverables 7.1 and 7.2, describing aspects of definition of model outputs as well as descriptions of simulation models, will showcase the specific requirements of information to be derived from clinical imaging data to be used for virtual cohort generation.

Contributors:

Function	Name and title	Affiliation	Signature and date
Author1	Jan Brüning	CHA	
Author2	Nina Krüger	CHA	
Author3	Leonid Goubergrits	CHA	
Author4	Marius Pullig	CHA	
Author5	Raphaëlle Lesage	VPHi	
Author6	Wouter Huberts	TUE	
Reviewer1	Silvia Schievano	UCL	
Reviewer2	Claudio Capelli	UCL	
Approver1:	Liesbet Geris	VPHi	
Approver2:	PC: Titus Kühne or PM: Edwin Morley-Fletcher	CHA, LYN	

Revision history:

Version	Amendment ID	Version Summary
V1.0		New version

Purpose

This document addresses procedures for processing of image data to be used for the generation of virtual cohorts that are subsequently processed in in-silico studies. Generally, in-silico studies are a wide field, spanning multiple medical disciplines as oncology, neurosurgery, and cardiology, as well as differing complexity and technology, as for example multi-scale and multi-organ models, biophysical, biochemical or data-driven models. Virtual cohorts describe a set of virtual patients or data of relevant parameters and information that is required to run the numerical models intended to be applied to the cohort. Thus, virtual cohorts are adapted to the intended use; application of other models to a previous cohort might not be possible, as relevant information required by those models might be missing.

However, independent of the intended use case or the model to be applied to the virtual cohorts, processing of real data will be required to generate these cohorts in the first place. Also, data processing will be required to obtain necessary information for validation of the virtual cohorts as well as for validation of the in-silico models applied to these cohorts.

In this document, *in-silico* models refer to numerical models that are used to assess the performance of medical devices implanted into the animal or human body, especially within the circulatory system, as well as to quantify the interaction between the medical device and the circulatory system, aiming to minimize adverse events possibly causing device failure or adverse clinical outcomes.

Here, in-silico models can be divided into two groups:

1. Models describing short-term interaction of the blood circulation system vessel with the implantable medical device during the device implantation;
2. Models describing long-term interaction between the medical device and the circulation system.

In general, processing of preclinical and clinical data is primarily necessary to describe the virtual cohort characteristics, which are modelled in a way to mimic real patient that are required by the respective model. Data that could characterize such a cohort and serve for the model may be categorized into the following five groups:

1. Anatomical data, describing the region of interest of the circulatory system;
2. Hemodynamic data, describing flow parameters at the inlets and outlets of this region of interest;
3. Geometric information relevant for virtual positioning and implantation of medical device;
4. Data describing material properties of tissues;
5. Demographic data required to characterize preclinical or clinical cohorts, which, however, are not required for the modelling itself. Nonetheless, this information might be relevant to identify supporting data from literature to mitigate missing information.

Data processing is a vital step as errors and biases introduced during this procedure will affect the virtual cohorts generated using this data and will propagate throughout each model applied to these cohorts. This will affect outcomes, such as predictions of safety and efficacy of a device that is tested in-silico. Thus, data processing procedures must be as standardized as possible, and a rigorous

validation scheme must be employed for the virtual cohort generation as well as for the in-silico models used.

In SIMCor, two different use cases are investigated: *trans-catheter implantation of aortic valve prostheses* (*transcatheter aortic valve* implantation, TAVI) and *implantation of pressure sensors within the pulmonary artery* (*pulmonary artery pressure sensor*, PAPS). In both, i*n-silico* modelling aims to predict the interaction between device, circulatory system, and tissue. However, to evaluate safety and efficacy of these devices using in-silico approaches, different clinical endpoints are of importance and must therefore be predicted. Thus, even in this one project, multiple virtual cohorts as well as models with different data sources and processing procedures will be employed.

Due to the extreme heterogeneity of in-silico projects, it is impossible to provide a *standard operating procedure* (SOP) valid for all those projects or covering all relevant processing requirements and procedures. Thus, this document aims at providing a procedure for establishing a standardized workflow for processing of pre-clinical and clinical data required to generate and validate virtual cohorts, rather than prescribing strict rules to be followed. This description will especially focus on essential steps of clinical image data processing that are commonly associated with errors or uncertainties. Where applicable, the document refers to already existing standards, guidelines, literature, and standard operating procedures.

Additionally, specific exemplary applications of the general procedure and recommendations provided is given, using the two use-cases investigated in SIMCor:

- TAVI: minimally invasive implantation of a biological valve prosthesis into the aortic root via catheterization.
- PAPS: implantation of a pressure sensor into the left or right pulmonary artery aiming to measure the instantaneous blood pressure within the pulmonary artery of patients suffering from heart failure.

These specific examples will be highlighted in the main text of the document as follows:

Exemplary use case - specific processing step

This format is used to present practical application of the general recommendations provided in this manuscript.

The processing of any data, especially imaging considered the most information intensive data type, can be subdivided into several procedures starting with data formation, followed by data enhancement, visualization, analysis, and management as shown in *Figure 1*.

This standard operating procedure will also briefly discuss topics regarding design of in-silico studies as well as acquisition of clinical data, as these are relevant aspects defining all subsequent processing steps. However, especially for data acquisition, a dedicated standard operating procedure is provided as part of SIMCor (SOP 4.1). The focus here will be processing of imaging data, exemplified by CT data, used for the generation of virtual cohorts.

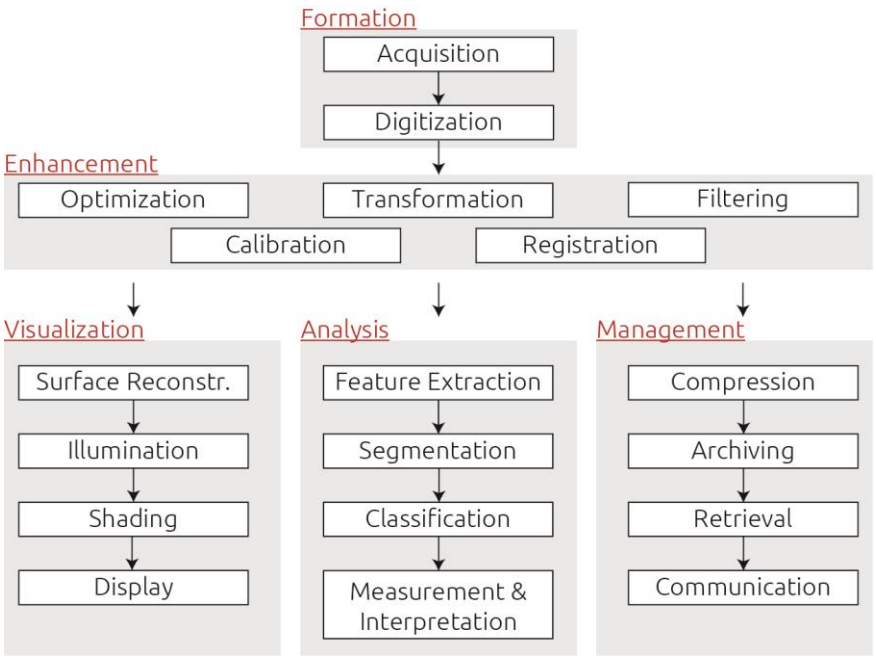


Figure 1: General steps of image data processing (modified from[1]).

[1] https://spie.org/publications/deserno-medical-image-processing?SSO=1

Scope

The SOP addresses researchers from different fields (e.g., engineering, mathematics, information technology, medicine) involved in processing of clinical image data required for generation of virtual cohorts. Due to this document being kept rather general, no specific requirements or skills are needed.

This SOP concerns the (pre)processing of medical imaging data intended to be used for generation of virtual cohorts to which in-silico models are subsequently applied to support development and evaluation of implantable cardiovascular devices. At least part of the described procedures, e.g., pre-processing and filtering of clinical image data, may also apply in the context of data-driven models as for example machine learning models. The proposed steps are illustrated in the biomedical context of cardiovascular systems, but should be applicable to other systems, although some variations may arise, especially in the pre-processing part, due to differences in imaging technologies and contrast-agents.

Responsibilities

No particular responsibilities are specified by this SOP. However, clinical data is usually affected by data privacy regulations as for example the *European General Data Protection Regulation 2016/679* (GDPR). Thus, compliance to those rules should always be ensured. As these rules vary widely with respect to the type of data, the type of processing procedures, as well as national laws, they cannot be summarized in this document. Some aspects, such as transfer and storage of data, will be briefly addressed in *Deliverable 4.1 - SOPs for data acquisition for in-silico models (CHA, M18)*.

Definitions and abbreviations

Virtual cohort: A virtual cohort is a *"set of virtual patients that mimics a real patient or subject cohort with respect to all reasonably expected patient states, patient dynamics and evolution of patient condition which might occur in practice"*[2]. Each member of the cohort is a physiological model, characterized by its own parameter set, that can predict the same endpoint that would be observed in clinical practice. Moreover, the variation between members represents the real population's variability[3]. Three different strategies for virtual cohort generation are discussed by Niederer et al.[4]: 1. One-to-one mapping; 2. Sampling from inferred distributions; and 3. Random variation with acceptance criteria.

Briefly, a virtual cohort created by one-to-one mapping consists of physiological models of many real patients, and the number of virtual patients is thus limited by the availability of complete real patient datasets. The second method can theoretically create an infinite number of virtual patients by sampling a distribution that mimics the parameter variability (and dependencies) inferred from a representative virtual cohort based on one-to-one mapping. The downside of this method is that the

[2] J. G. Chase, J. C. Preiser, J. L. Dickson, A. Pironet, Y. S. Chiew, C. G. Pretty, G. M. Shaw, B. Benyo, K. Moeller, S. Safaei, M. Tawhai, P. Hunter, T. Desaive, Next-generation, personalised, model-based critical care medicine: A state-of-the art review of in silico virtual patient models, methods, and cohorts, and how to validation them. *Biomed. Eng. Online*. **17** (2018), pp. 1–29

[3] M. Viceconti, C. Cobelli, T. Haddad, A. Himes, B. Kovatchev, M. Palmer, In silico assessment of biomedical products: The conundrum of rare but not so rare events in two case studies. *Proc. Inst. Mech. Eng. Part H J. Eng. Med.* **231**, 455–466 (2017).

[4] S. A. Niederer, Y. Aboelkassem, C. D. Cantwell, C. Corrado, S. Coveney, E. M. Cherry, T. Delhaas, F. H. Fenton, A. V. Panfilov, P. Pathmanathan, G. Plank, M. Riabiz, C. H. Roney, R. W. Dos Santos, L. Wang, Creation and application of virtual patient cohorts of heart models. *Philos. Trans. R. Soc. A*. **378** (2020), doi:10.1098/RSTA.2019.0558.

inferred (approximated/sparse) distribution may not be accurate enough to fully capture the physiological parameter ranges and their correlations, possibly resulting in non-physiological combinations of parameters and unrealistic models. In the third strategy, which we use within SIMCor, the parameter sets are randomly generated, and the resulting models are compared against real population measurements, only models that agree with acceptance criteria are retained. Here we add acceptance criteria (a filter) based on advanced sensitivity analysis techniques (see WP7).

The set of real patient information included in a virtual cohort is defined by the intended use for the virtual cohort. This intended use is usually running one or multiple in-silico models using the information provided by the virtual cohort. Thus, the cohort must at least contain all information required for definition of boundary conditions and other model inputs. The boundary conditions can be pressures, velocity profiles or flows, but also proper parameterization of 0D/1D models mimicking the heart or distal vasculature. Additional parameters might be necessary to describe sub-cohorts of specific properties (e.g., paediatric cohorts). The information provided by a virtual cohort can range from simple numerical parameters such as heart rates and blood pressure information, to complex information as for example three dimensional models of parts of the cardiovascular anatomy, as for example the left ventricle. For the latter, sophisticated models for generation of synthetic data entries will be required. Generation of synthetic shapes can be facilitated in multiple ways. Viable strategies include using *statistical or parametric shape models* (SSM) as well as a use of generative adversarial networks[5,6].

[5] Bruse, J.L., McLeod, K., Biglino, G. et al. A statistical shape modelling framework to extract 3D shape biomarkers from medical imaging data: assessing arch morphology of repaired coarctation of the aorta. BMC Med Imaging 16, 40 (2016). https://doi.org/10.1186/s12880-016-0142-z

[6] Heiman T, Meinzer, H.P. Statistical shape models for 3D medical image segmentation: A review. Medical Image Analysis 13, (2009). https://doi.org/10.1016/j.media.2009.05.004

Procedure

Definition of study design and specification of required pre- or clinical data for virtual cohort based in-silico modelling

Rationale: the definition of the in-silico model which is to be applied to a virtual cohort is the focal point, defining all subsequent data processing requirements.

The initial step before a standardized approach to image data processing for generation of a virtual cohort can be defined is to describe the underlying clinical problem as precisely as possible. Virtual cohorts are usually only developed for a specific purpose, which is most often the enhancement of existing real pre- and/or clinical datasets of limited size or variability, or to provide a dataset that can be shared without being affected by data privacy regulations. Furthermore, a virtual cohort will most likely be tailor-made for a focussed question of interest (e.g., medical device under development and certification) as well as a specific in-silico model. The more focussed this question is, the easier it will be to specify the required information that must be derived from medical image data. Here, a vital, initial step might be to clearly specify the endpoints of an in-silico study by answering questions as 'what is the purpose of my model?' and 'which clinical outcome do I want to predict?'[7].

TAVI - question of interest

Is it possible to predict the paravalvular leakage (e.g., locations of leakage and leakage volume) during diastole? What is the risk for device thrombosis during systole?

Answering these questions allows specification, or at least constraining, of the region of interest, which must be represented by the in-silico model using either boundary conditions or geometric models. This is already necessary to limit the list of potential information required for the model, as obtaining this information is usually associated with relevant efforts and costs. This is especially true for models requiring *three-dimensional* (3D) anatomies of cardiovascular structures, as for example for generation of synthetic geometries to populate virtual cohorts. The region of interest is most commonly defined by the implantation site of the medical device or the pathology to be investigated. However, often additional regions of the circulation, upstream and downstream of this site, might be of interest, especially in hemodynamic models. This definition must be performed cautiously, as it directly affects the required data for running the in-silico models:

- The required anatomical information must be obtainable from medical image data and all derived parameters;
- The boundaries of the region of interest will specify the required boundary conditions for running the in-silico model.

TAVI - region of Interest

The TAVI procedure describes the implantation of a biological heart valve prosthesis into the aortic valve annulus via a catheter or a delivery system allowing trans-apical access. Depending on the

[7] also referred to as Question of Interest and Context of Use in official standards such as the ASME V&V40

question of interest and the respective in-silico model, the region of interest varies. If the catheter-based deployment is to be modelled, the whole aorta, including the ascending, descending and thoracic aorta towards the access site at the iliac artery will be required. If only hemodynamic information regarding paravalvular leakage is to be investigated, a model containing the aortic root and the proximal part of the ascending aorta might be sufficient.

However, even if the underlying problem is well defined, it might not be possible to specify a final list of required information in an ad hoc manner. It is to be expected that an iterative process, in which the model description as well as the image processing procedures must be adapted continuously, will be necessary. Reasons for this are manifold and include insufficient quality or unanticipated lack of data, or large uncertainty during extraction of information from imaging data. Not all data required for in-silico models are available from real pre-procedural and/or clinical data. For example, tissue properties cannot be assessed from individual patients since tissue samples are required for the respective measurements. In such cases, unavailable data should be replaced by literature data.

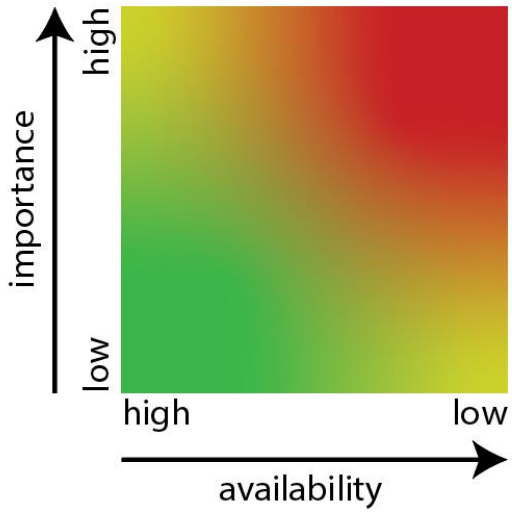


Figure 2: For all data to be used in generation of virtual cohorts or the subsequent in-silico modelling, the trade-off between data availability and importance of data must be evaluated.

Without restraint, an exhaustive list of parameters will come to mind, as the more the available data, the more freedom in model development and validation. However, imaging data from clinical routine is as limited as the information which can be obtained from this routine image data. Also, processing of data is often associated with relevant manual interactions, equalling time and costs. Thus, any initial list of information that might be relevant for generation of virtual cohorts or running an in-silico experiment must be curated.

Here, it is important to weigh two opposing aspects of each parameter against each other: data availability and data importance as shown in *Figure 2*. Data importance heavily relies on the intended in-silico model. Some information might be crucial for the model to run, other data would be useful, and some might be helpful, but not strictly necessary.

TAVI - Importance of Data

High Importance: patient-specific anatomy of the aortic root in which the prosthesis is to be implanted and calcification.

Low Importance: patient-specific information regarding blood rheology (e.g., haematocrit).

Data availability is affected by several aspects, the most obvious being the availability of a method to either measure or calculate the required information. However, even if such a method exists, data availability might be low because that method is not part of clinical routine or the method to extract the required information from routine data is costly or method resolution is too low.

TAVI - availability of data

High availability: ultrasound-derived information regarding the severity of aortic stenosis as for example Bernoulli-based pressure gradients or the aortic valve area.

Moderate availability: *computed tomography* (CT) information of the circulatory system is commonly performed after decision to treat a patient using TAVI.

Low availability: patient-specific velocity profiles in the left ventricular outflow tract obtained using velocity encoded *magnetic resonance imaging* (MRI) is currently no clinical standard procedure and must be acquired purposely for modelling.

Additionally, the purpose of the data within the specific in-silico project must be identified. Here, the following three categories can be defined:

1. Data required for personalization and thus execution of the in-silico model;
2. Data required for validation of the in-silico model;
3. Data required for the validation of the virtual cohort.

This differentiation is important as data required only for validation of either the in-silico model to be used to calculate clinical endpoints, or to validate the virtual cohort generation, are usually not as limiting. In these cases, the required information might be acquired using prospective clinical trials, even though it is not available from clinical routine. However, if multiple or most of parameters that have a high importance for personalization of the model also have a low availability, this might be prohibitive for the real-world application of a model. These aspects will be covered in detail in the SOP 4.1 on data acquisition for in-silico models.

There are several other aspects that are important for the final specification of information to be acquired from imaging data. For example, the uncertainty associated with these parameters should be evaluated. This includes the uncertainty that is associated with measuring the data but also the propagation of uncertainty throughout the models. Here, SIMCor *Deliverable 7.1 - Definition of model output (TUE, M6)* describes different methods that can be employed to identify the most relevant parameters for an in-silico study as well as for virtual cohort generation.

Specification of relevant data types

Rationale: for in-silico modelling it is vital to understand which data types are required for each model and from which data this information can be derived.

As complex and heterogeneous as in-silico models can be, as complex are the types of data that are required for either parametrization or validation of those models. Even if the focus is constrained to in-silico models and virtual cohorts for evaluation of implantable cardiovascular devices, no exhaustive list can be easily provided. The examples of relevant data types provided in this document will be listed with increasing complexity.

In this section, we try to distinguish between the terms 'data' and 'information'. Here, data refers to the unprocessed data, whereas results from data processing, that are used for in-silico model parametrization or virtual cohort generation, are described as information. However, this differentiation is not always possible, especially for types of data that require no or only limited processing (e.g., demographic data as described below).

Demographic patient or subject information: this type of data usually is described by single nominal, ordinal or continuous parameters. Examples for nominal parameters are the patient's sex or ethnicity, whereas aortic stenosis is a common example for ordinal parameters. The patient's age, weight, and height as well as parameters derived from this information, such as the body mass index and the body surface area, are commonly used continuous parameters. This information is usually easily accessible, if documented in the clinical notes. However, those parameters are usually not directly used as boundary conditions of in-silico models but are commonly used to adjust those conditions. For example, the body surface area or body weight might be used to calculate an equivalent patient cardiac output, if this parameter is not available otherwise. Furthermore, the patient's age might be used to adjust stiffness of vascular tissues.

TAVI - demographic information

Demographic information that is required for generation of virtual cohorts for simulation of TAVI procedures are for example the patient's age, sex, height, and weight. This information is not only required to generate virtual cohorts that match the real patient cohort with respect to these parameters, but also to generate boundary conditions that mimic the correlations observed in those cohorts (e.g., larger patients usually have a higher stroke volume and cardiac output).

Simple, zero-dimensional, patient-specific information on anatomy and function: similar to demographics, information regarding the patient-specific anatomy and function can be available as simple nominal, ordinal or continuous parameters. Here, the term simple only refers to the information being expressed as a zero-dimensional value. Acquisition of this type of information, however, might be complex. While nominal information might include descriptions of anatomical shapes, as for example the classification of aortic arch anatomies in gothic or crenel types, this type of information is usually difficult to distinguish from demographic patient information. Also, similarly to demographic information, nominal parameters of anatomy and function usually are not used for specification boundary conditions directly. Ordinal parameters might include ratings of functional aspects, where continuous measurements are not feasible. For example, eccentricity and helicity of

flow patterns observed in 4D flow MRI data can be classified using degrees that are obtained by observer-rating. This type of information is also not ideal for specification of boundary conditions but might be used for validation of in-silico models. Finally, continuous parameters of anatomy and function are the most common type of zero-dimensional data of anatomy and function that are commonly used for parametrization of in-silico models. Typical examples for anatomical parameters are diameters and lengths of vessels, which are commonly used in one-dimensional hemodynamic models, as well as pressure, resistance, velocity, and volume flow rate values which are commonly required to specify inlet and outlet boundary conditions of hemodynamic models.

TAVI - zero-dimensional information

Zero-dimensional information on anatomy and function that are required for simulation of TAVI procedures are, for example, the heart rate, the peak-systolic volume flow rate passing through the aorta, and the aortic valve area during systole.

Complex, patient-specific information on anatomy and function: finally, more complex parameters describing anatomy and function exist. For example, transient information of different parameters, as for example the volume flow rate or the static pressure within a vessel, during the heart cycle. This information can be averaged across multiple heart cycles to provide a robust measurement, or individual heart cycles can be evaluated to also account for differences between consecutive heart beats.

Another common type of data to be used are three dimensional models of anatomies of the cardiovascular system as for example the left ventricular volume, the myocardium, the heart valves, or vessels. This information can also be of transient nature, describing the different anatomical shapes during different phases of the heart cycle, as for example the systolic and diastolic volume of the left ventricle as well as the closed and opened state of heart valves. Additional information that is related to those three-dimensional anatomies can be displacement vectors that describe the point-wise movement of the tissue, as well as maps at the surface describing for example spatially resolved material properties as regions of local calcification.

TAVI - complex information

For simulation of TAVI procedures, the patient-specific anatomy of the aortic root, including the left ventricular outflow tract, the aortic annulus, the aortic sinuses and the proximal part of the ascending aorta are required. Additionally, information on the aortic valve leaflets, calcifications, especially those located at the aortic annulus, as well as the position of the coronary ostia, is required.

Image-data: image data describes anatomical or functional measurements using techniques such as x-ray, CT, MRI, and ultrasound. While these are the most common imaging techniques used in clinical settings, other techniques exist. The term image data refers to these techniques providing images of the patient-specific anatomy. However, modern techniques also allow measurement of cardiovascular function, which nonetheless is represented as image data as well. Here, the most relevant examples are Doppler ultrasound, a ubiquitously available method which allows measurement of patient-specific

blood flow velocities as well as velocity encoded MRI, allowing velocity measurements in a spatiotemporal manner. However, also specific methods for measurement of tissue properties and movements, as for example ultrasound elastography[8] and myocardial tagging MRI[9] exist. While medical imaging is well standardized following the *Digital Imaging and Communications in Medicine* (DICOM) standard[10], the extreme heterogeneity of techniques, sequences, scanner manufacturers and settings make this most likely the most complex data type to be used in in-silico modelling and virtual cohort generation.

This data type also has a rather special position in in-silico modelling, as most, if not all previously mentioned data types, are commonly derived from medical image data. While parameters such as pressures and heart rates can also be obtained using other sensor data as for example cuff-pressure measurements, invasive catheterization, heart rate monitors or electrocardiography, many parameters rely on the availability of image data. This is especially true for anatomical information of the patient-specific anatomy, be it three-dimensional shapes, vessel diameters or tissue movements. The methods to obtain such information from medical image data are as heterogeneous as the data themselves and will thus be the focus of the following section regarding data processing.

[8] Sigrist RMS, Liau J, Kaffas AE, Chammas MC, Willmann JK. Ultrasound Elastography: Review of Techniques and Clinical Applications. Theranostics 2017; 7(5):1303-1329. doi:10.7150/thno.18650.
[9] Ibrahim, ES.H. Myocardial tagging by Cardiovascular Magnetic Resonance: evolution of techniques–pulse sequences, analysis algorithms, and applications. J Cardiovasc Magn Reson 13, 36 (2011). https://doi.org/10.1186/1532-429X-13-36.
[10] https://www.dicomstandard.org/

Specification of relevant data formats

Rationale: for all data that is either used for in-silico models or shared between partners of a joint project, file formats should be specified to reduce loss of information.

Another vital step for processing data in view of generating virtual cohorts, is the specification of file and data formats in which the information that was previously identified is stored and shared. Usually, an important aspect of virtual cohorts is that those datasets can be shared freely without any constraints from data privacy regulations. Thus, they are ideal for specification of standardized cohorts on which comparative analyses can be performed. However, to facilitate this, the modeler should not only provide the data but also specify the data formats which are used to provide the cohort's information. This is necessary to facilitate use of the information to be found within a virtual cohort without conversion to other data and file formats, which might be associated with loss of information and biases. Additionally, especially if the virtual cohort is intended to be shared and used for comparative in-silico investigations, well described data formats will ease the use of the data provided significantly.

Medical image data: medical images are usually stored and made available using the DICOM or *Neuroimaging Informatics Technology Initiative* (NIFTI) standards[10]. However, the latter is mainly used for imaging techniques of the brain as for example functional MRI, whereas the former is used for all tissues and organs and modalities. The DICOM standard also is suitable for data which is not necessarily considered imaging data as for example electrocardiography and electroencephalography. In some cases, medical image data will also be available in more common image file types such as the *Joint Photographic Experts Group* (JPEG) format. However, this is strongly unadvisable, as these file types lack the standardized header specification for storing of vital information for subsequent data processing. Examples for such information are the image voxel size, spacing between slices, temporal resolution of images, and device specific setting. While all conventional imaging modalities can be stored using the DICOM standard, 3D ultrasound images must be considered cautiously. While DICOM requires image information to be stored as cartesian, 3D ultrasound data is usually available as polar information. To mitigate this problem, each manufacturer of ultrasound devices has their own strategy. Either the data is stored as a proprietary data format (e.g., Samsung) or the data is stored in proprietary format within private tags inside the DICOM file (e.g., Philips, GE). In both cases, assessing the information is not trivial and usually requires support by the device manufacturer. Also, the strategy is manufacturer specific.

Surface geometries: there is a large quantity of file formats suitable for storing 3D anatomical information. Usually, these file formats describe the surface by specification of boundary vertices and their respective connection to each other. A very common and rather simple file format is the STL standard (abbreviation from standard triangle language, even though other acronyms exist). Here, a 3D model [11]is specified by a list of triangles, each triangle consisting of three vertices, which again are defined by the three cartesian coordinates (x, y, z). Due to its simplicity STL is supported by most software solutions working with 3D surface geometries.

[11] C. Parks. Initial Graphics Exchange Sepcification Volume 2: Application Protocols. NIST Interagency/Unternational Report #6972. https://doi.org/10.6028/NIST.IR.6972

The VMRL (*Virtual Reality Modelling Language*) format specifies 3D surfaces as triangles: two lists are used, one containing the vertices of the surface mesh, the other containing the vertex indices that form each triangle. This avoids redundancy in vertex specification. The VMRL standard also supports specification of colours and transient information (e.g., vertex movement over time). Further common formats are Wavefront OBJ[12] and the *Polygon File Format* (PLY)[13].

These file formats are commonly used for 3D geometries resulting from 3D design, image reconstruction, or 3D scanning. For *Computer-Aided Engineering* (CAE) another approach is used where edges and surfaces are specified as *Non-Uniform Rational B-Splines* (NURBS). Common file types are STEP (of which the format is defined in ISO 10303-21:2016[14]) and the *Initial Graphics Exchange Specification* (IGES), which both are vendor independent. This distinction is relevant, as CAE software usually cannot process the above-mentioned mesh-based file formats.

Due to the heterogeneity of file formats and the specific requirements from different software for pre-processing or simulation, specification of geometric file formats is strongly advised.

Other information: for all other information, various suitable file formats are available ranging from simple text files and comma separated variables (CSV) files, over more complex formats such as *Extensible Markup Language* (XML) and *JavaScript Object Notation* (JSON) to database structures.

[12] https://www.fileformat.info/format/wavefrontobj/egff.htm
[13] http://paulbourke.net/dataformats/ply/
[14] https://www.iso.org/standard/63141.html

Data processing

Rationale: relevant data processing steps and procedures will be described briefly and relevant literature for each procedure for further investigation is provided.

In this section common procedures for processing of image data will be described. As a plethora of procedures, techniques and software tools exists, only an overview of the respective topics can be provided. References describing the different topics in more detail will be provided where available. This section is subdivided following a stepwise approach commonly found in image processing. First, the pre-processing of image data is described. These procedures are necessary to account for heterogeneity often found in clinical image data. For example, image data can be acquired using different devices with different device specific settings and resolutions. Especially for machine learning based processing of image data, well standardized input information is important. However, also for more traditional approaches, strong variations in the input data will inevitably affect the procedure's outcome. The following subsection is then focusing on the aspect that is commonly associated with clinical image processing, the reconstruction of patient-specific anatomical structures and anatomies, commonly named image segmentation or image annotation. Finally, procedures to derive simulation-ready models from these segmentations as well as to obtain hemodynamic or functional parameters from image data will be described.

Image normalization

Rationale: image analysis methods often require input images in standard formats. The applied processing steps need to be documented regarding the changes that they apply to the original image information.

Data driven models especially benefit from or even require normalization of their inputs. The following section focuses on image data normalization.

Important: in each step, it is important to always retain the original image and ideally share the transformation method and parameters. In this way, accumulated information loss can be avoided while the transformation is optimally documented. Also, information loss should be quantified[15]. For calculation of such features, different libraries[16] are available. For further information, the Quantitative Imaging Biomarkers Alliance[17] offers several resources on standardization, but also anonymization and documentation standards.

Voxel size normalization: especially machine learning algorithms might highly benefit from a unification of the voxel size over all input images. This requires selection of a suitable interpolation method for resampling. Increasing the voxel size reduces image resolution, which leads to information loss. However, also reduction of the voxel size can lead to information loss, as the interpolation method can introduce artifacts. Thus, it is advisable to choose a voxel size close to the original voxel size to minimize information loss. Here it should be noted that the pixel spacing and the spacing between

[15] https://www.rsna.org/uploadedFiles/RSNA/Content/Science_and_Education/QIBA/Obuchowski-Algorithm-Comparison-Methods-Stat-Methods-Med-Res-2014-0962280214537390.pdf
[16] https://pyradiomics.readthedocs.io/en/latest/#
[17] https://www.rsna.org/research/quantitative-imaging-biomarkers-alliance

slices often differ, and an isotropic voxel size might not always be ideal. For subsequent processing or analysis of the image, the original image should be considered.

Special attention needs to be paid to resampling label masks (see Annotation & Segmentation), as most interpolation methods produce floating values. A suitable approach requires:

1. Converting each label to float;
2. Applying resampling as for the image data;
3. performing thresholding to get homogeneous integer label.

The changes in volumes of the label masks should be documented.

Image dimension normalization: if images of similar dimensions are required, several approaches are available. One option is resampling to result at the desired image dimensions. This will retain the image extent but will change the voxel size. Thus, the above suggestions apply. Changing the image extent, while retaining the voxel size can be achieved by either cropping or padding the image. Cropping reduces the image to a sub image of specific dimensions. Here it needs to be ensured that all relevant information is still available within the sub image. With padding the original image dimension is increased by adding artificial voxels around the original image. Those can be filled with different methods[18], e.g., a constant value or reflecting the border voxels of the original image.

In addition to retaining the original image, the amount of change per direction and the filling method should be documented. Plotting the old versus the new histogram of intensity levels can point out potential problems, e.g., for voxel intensity normalization.

Voxel intensity normalization: the goal of voxel intensity normalization is to improve accuracy of the model by reducing variance in the voxel values. This is important when working with a multitude of scanners or modalities, but also common practice for single-scanner or single-modality datasets. There are two common methods for voxel intensity normalization: min-max normalization and z-score normalisation. In min-max normalization, all values are scaled between the values 0 and 1.

Z-score normalization depicts the process of centering the pixel values to a distribution with a mean of 0 and a unit standard deviation. LeCun et al.[19] showed that with a centred zero mean and a unit variance models converge faster.

Oftentimes, it is worth investigating all methods empirically on the underlying dataset. Apart from that, voxel intensities might be clipped to regions of interest, e.g., to only depict bone structures. The following properties should be documented to be able to reconstruct the changes: contrast, entropy, intensities (min/max statistics or a histogram).

Standardization of spatial orientation: for some applications it might be required to standardize the spatial orientation of the images, e.g., along an annulus plane. Also, in this case it is advisable to retain the original image and save the parameters of the transformation for each case. For the transformation itself, a resampling will likely be required, and the impact of this transformation should be assessed.

[18] https://docs.opencv.org/4.5.3/d3/df2/tutorial_py_basic_ops.html#:%7E:text=making%20borders%20for%20images%20(padding)

[19] LeCun, Y.; Nottou, L.; Orr, G. B.; Müller, K. Efficient BackProp. Neural Networks: Tricks of the Trade, 9-50. DOI: 10.1007/3-540-49430-8_2

Annotation & segmentation

While the pre-processing procedures specified above are common for medical image processing, this term is usually associated with the measurement or reconstruction of anatomical information from medical image data directly. Here, the degree of complexity to be derived can differ widely from specification of landmarks, marking the location of anatomical structures within images, via identification of contours to reconstruction of complete three-dimensional segmentation masks of anatomical structures. Methods for image annotation and segmentation vary widely from fully manual procedures, via semi-automated approaches to fully automated methods that ideally do not require any user input.

TAVI - Examples for image annotation and segmentation

Landmarks: position of the coronary ostia for assessment of occlusion risks during TAVI implantation.

Contours: cross-section of the aortic annulus for estimation of appropriate device size.

Segmentation masks: segmentation of the complete lumen of the left ventricular outflow tract, the aortic root as well as the ascending aorta to obtain a 3D model of the patient specific anatomy for virtual device implantation.

Manual and semi-automatic methods for image segmentation: the basic principle for manual or semi-automatic image segmentation is to label parts of the image that belong to the anatomy of interest using different tools. For example, if the contour of the aortic annulus is to be reconstructed, a contour could be drawn within the medical image, following the Hounsfield information in CT images. Also, all image pixels in a 2D plane lying within the aortic annulus could be selected (see *Figure 3*). While the latter approach is usually bound to the discrete resolution of the images, drawing of contours is usually independent from the actual image resolution. For drawing of contours, various approaches exist depending on the tool used. Common examples are freehand drawing, definition of individual points that are subsequently combined via splines as well as drag and drop definition of elliptic contours. The curve in *Figure 3* was drawn using the spline-based method.

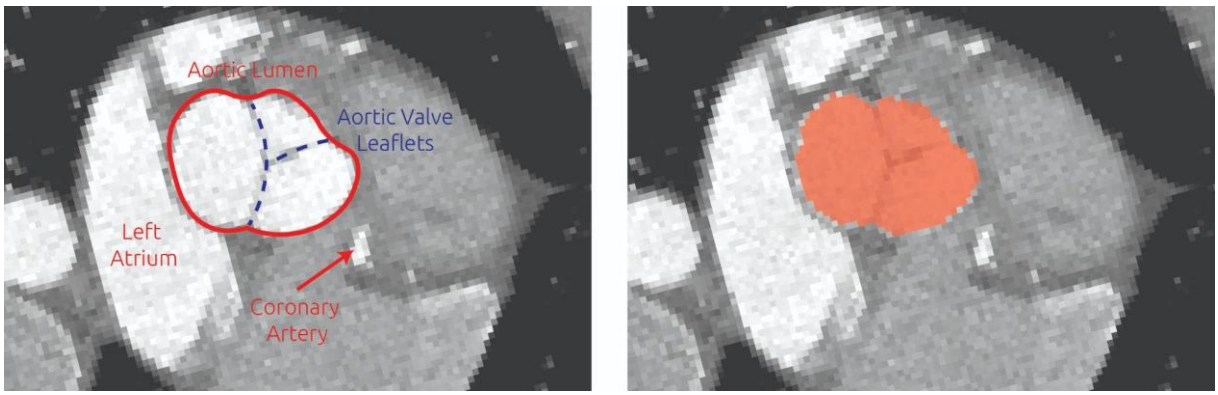


Figure 3: Examples of manual 2D segmentation strategies for the aortic annulus. Either a contour separating the aortic lumen from the surrounding tissue can be drawn (left) or the voxels belonging to the aortic lumen can be labelled.

3D information is usually segmented by generating label masks for each volume slice. Thus, for each image voxel a label is defined. In the most trivial situation two, labels are required, one for the current

voxel belonging to the anatomy of interest, one for the voxel not being of interest. However, multiple labels could be generated to distinguish between different anatomical structures. For example, one label field for each atrium and ventricle and one label field for the aorta and pulmonary artery, to distinguish all large vessels and blood chambers of the heart (see *Figure 4*).

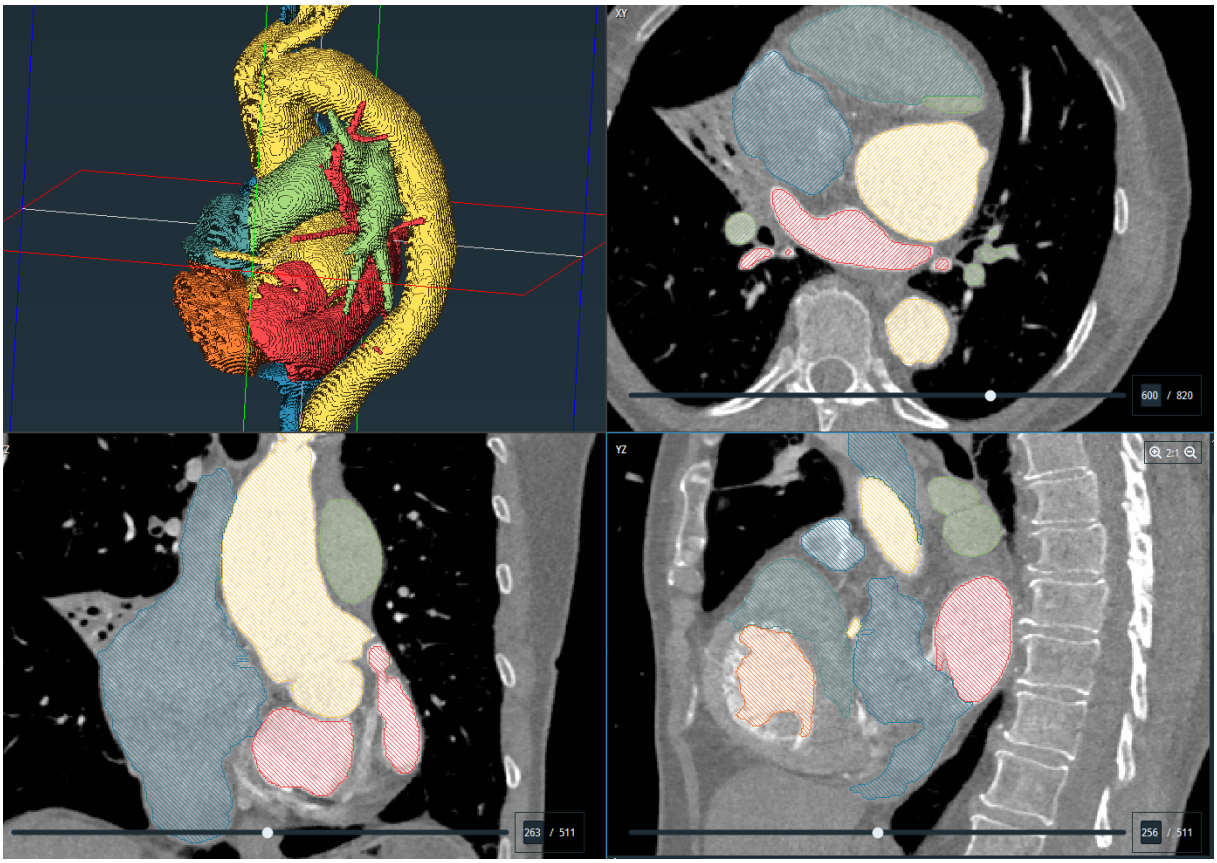


Figure 4: Examples of a manual 3D segmentation of the human heart's blood chambers and large arteries. Each region of interest was reconstructed using another label. The different labels are colour coded. In the top left image, the 3D segmentation is shown, whereas the other 3 panels show slices of the 3D image data with the respective segmented regions.

While the main concept of 3D image segmentation - labelling the image voxels belonging to the structure of interest - is rather straightforward, the segmentation process is usually very labour intensive. The simplest, yet most time-consuming approach is the complete manual segmentation. Here, all voxels must be highlighted manually within the image data with appropriate tools. Common examples for those tools are brushes, freehand contouring (lasso), specifying elliptic or rectangular regions by drag and drop. Using these tools, the 3D image data must be processed slice by slice. Ideally, all three cartesian directions (see *Figure 4*) must be examined at least once. Effectively, the operator is painting in the relevant structures. This approach is, obviously, highly operator-dependent and should always be performed by or at least under supervision of people experienced in image-segmentation and with knowledge of the relevant anatomy. Also, intra-, and inter-operator biases in segmentation results should be examined.

Most software solutions for image segmentation also provide tools to speed up this manual image segmentation procedure. Examples for those tools are (see also *Figure 5* for illustrations):

- Masking: by definition of upper- and lower greyscale thresholds, only voxels within the thresholds are eligible for segmentation. This approach speeds up the manual selection of voxels, as less caution is required at specific regions.
- Magic wand: this tool is also often implemented in most image processing software. An upper and lower grayscale threshold is specified. Then, a pixel is selected. All pixels that are connected to the starting point that also lie within the specified thresholds are automatically selected. To avoid selection of unwanted regions, manual barriers can often be specified. This algorithm is also often implemented as a 3D tool, allowing automatic segmentation of structures that are well distinguished from surrounding tissue.
- Region growing/blowing: here, a seeding point is set within the image data. Starting from that point, neighbouring voxels are selected based on either their absolute grayscale values or the local grayscale gradients. These tools can also operate in 3D and can thus be used for almost fully automatic segmentation of structures that are well distinguished from surrounding tissue. Other implementations follow a drag and drop approach, where the starting point is set with a click, and the strength of the region growing is specified by dragging a distance from the initial point. The advantage of this method over the magic wand approach is that it does not follow an 'all or nothing' approach. Two neighbouring voxels of similar intensity but belonging to different tissues (e.g., two parallel vessels) can already result in the magic wand to select both tissues. The region growing speed however is based on the local image information and will be reduced in regions with high grayscale gradients.
- Calculation of gradient images: the grayscale gradients used for the region growing method can also be used for direct image segmentation. Here, the main principle is that at tissue boundaries high gradients are to be expected. For example, a blood vessel that is perfused by a contrast agent will feature high Hounsfield units in CT images, whereas the surrounding tissue will feature lower values. By calculation of the image intensity gradients, additional information for separation of different tissues or structures can be provided.

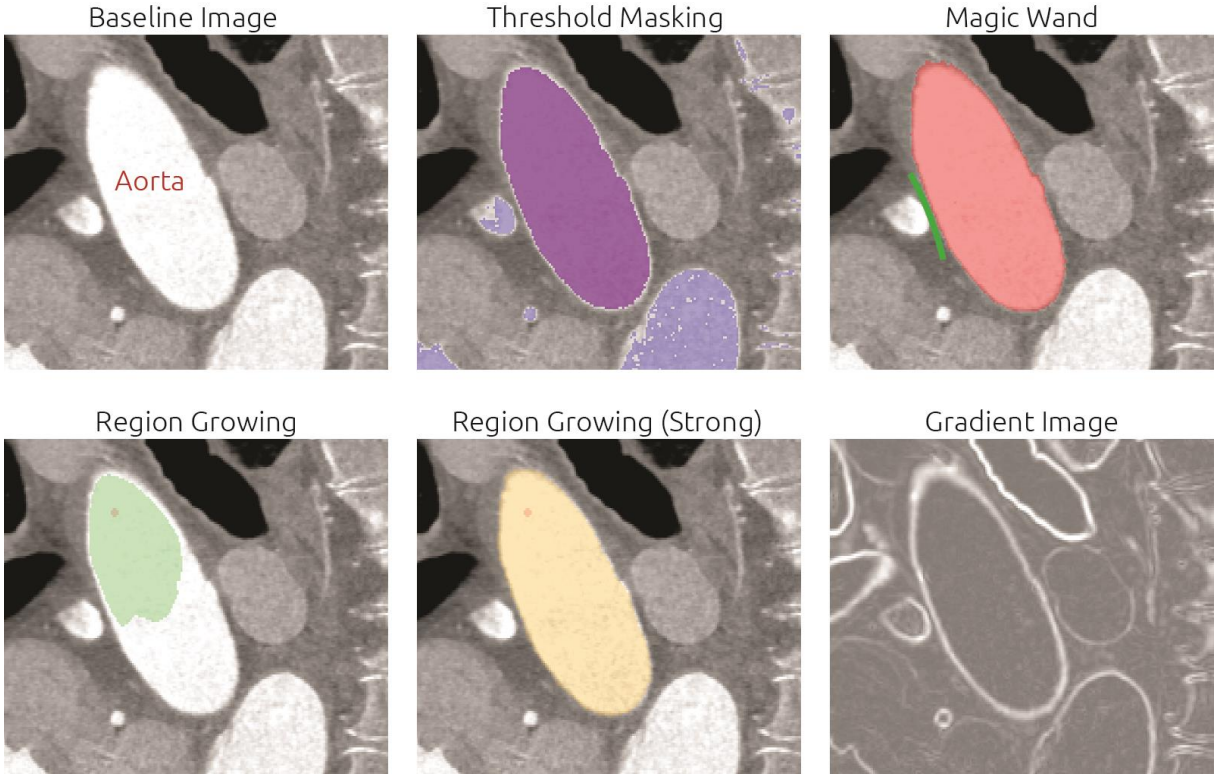


Figure 5: Examples of semi-automatic tools for support for image segmentation. In the upper left panel a sagittal CT slice is shown that contains image information of the aorta. By masking (top, middle) a given grayscale threshold can be predefined. Only voxels within this threshold (blue) are eligible for image segmentation (purple). Magic wand tools (top, right) allow automatic selection of all connected voxels that lie within a predefined grayscale range. To prevent selection of regions with similar intensities, manual barriers for the tool can be specified (green line). Region growing algorithms (bottom left and middle) are similar to the magic wand tool. However, they propagate the segmentation step by step and thus provide more control of the selected region. Also, gradient images (bottom, right) can be calculated which provide additional information on tissue boundaries.

The previously mentioned examples for manual image segmentation are based on image segmentation using the software Amira distributed by ThermoFisher[20]. This is one of many software tools available for image segmentation. While similar strategies can be employed in many other software tools, the specifics might vary widely. Common free software solutions for image segmentation are 3D Slicer[21] and the *Medical Imaging Interaction Toolkit* (MITK)[22], whilst commercial software options include Materialise Mimics[23] and Simpleware, Synopsys[24].

Automatic methods for image segmentation: image segmentation and landmark detection can be automated using methods of Machine Learning / Deep Learning. For images typically convolutional neural networks[25] are used. Willemink et al.[26] offer an extensive overview on preparation of medical image data for machine learning.

[20]https://www.thermofisher.com/de/de/home/industrial/electron-microscopy/electron-microscopy-instruments-workflow-solutions/3d-visualization-analysis-software/amira-advanced-image-processing-quantification.html
[21]3D Slicer documentation on image segmentation: https://slicer.readthedocs.io/en/latest/user_guide/image_segmentation.html
[22] MITK documentation: https://docs.mitk.org/2021.02/UserManualPortal.html
[23] https://www.materialise.com/de/medical/software/mimics-innovation-suite/products-services/mimics
[24] https://www.synopsys.com/simpleware/software.html
[25] https://www.deeplearningbook.org/contents/convnets.html
[26] https://pubs.rsna.org/doi/full/10.1148/radiol.2020192224

Quantification

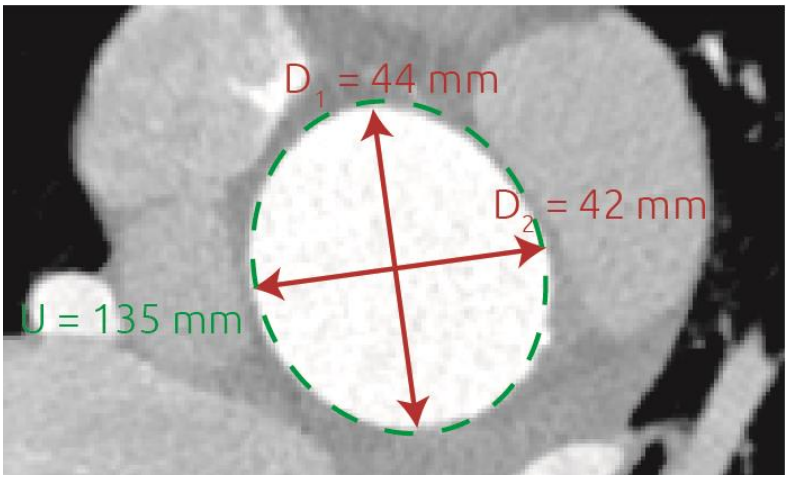


Figure 6: Simple anatomic parameters as diameters and circumferences can be directly assessed from medical image data without need for image segmentation.

The aim of data processing is not only to reconstruct the patient-specific anatomies to be used as boundary conditions in in-silico modelling or definition of virtual cohorts, but also to quantify anatomical and functional parameters. While simple anatomical parameters as for example metrics for size (lengths, diameters) can be measured directly within medical image data, others require a detailed segmentation of the respective anatomy. Finally, functional parameters are commonly calculated from anatomical parameters. For example, if the left ventricular volume during diastole and systole is known, the stroke volume can be calculated as the difference between both values. However, also specific imaging modalities exist that facilitate direct measurement of functional parameters.

Measurement of simple anatomical parameters: since for each image voxel in medical images, the cartesian coordinates are usually well defined (this might not be valid for 3D ultrasound data, see chapter 3), measurement of lengths and diameters in image data is straightforward and usually does not even require image segmentation, as long as the voxel size is known and considered correctly. For example, for measurement of the diameter of a vessel the distance between both opposing sides of the vessel cross section can be measured (see *Figure 6*). However, using segmentation information or even 3D geometries for assessment of those parameters is still favourable, as it is difficult to determine whether a 2D image slice is oriented perfectly perpendicular to the vessel orientation.

Especially if parameters are measured using manual interaction, the respective procedures must be well specified (e.g., how was the evaluation plane specified, how are the diameter measurements specified).

Measurement of complex anatomical parameters: measurement of complex parameters, as for example the volume of anatomical structures, usually requires 3D segmentation of these structures. However, especially for cardiovascular applications, methods exist to estimate the ventricular volumes from a set of planar measurements[27] and the assumption of the ventricle to be of elliptic shape. Nonetheless, accurate assessment of the volume of more complex anatomical structures can usually

[27] Carly Jenkins, Stuart Moir, Jonathan Chan, Dhrubo Rakhit, Brian Haluska, Thomas H. Marwick, Left ventricular volume measurement with echocardiography: a comparison of left ventricular opacification, three-dimensional echocardiography, or both with magnetic resonance imaging, European Heart Journal, Volume 30, Issue 1, January 2009, Pages 98–106,

only be achieved using 3D segmentation. Here, the simplest approach is to simply sum the volume of all voxels belonging to the respective segmentation mask. If a 3D geometry of the anatomy exists, various geometric parameters can be assessed including aspect ratios (relationship of shortest and longest dimension), sphericity indices, local curvatures, to name only a few examples. Complex anatomical parameters are of specific interest in some research topics. For example, in cerebral aneurysms, the geometric properties of aneurysms are discussed to be associated with the respective aneurysm's risk of rupture[28].

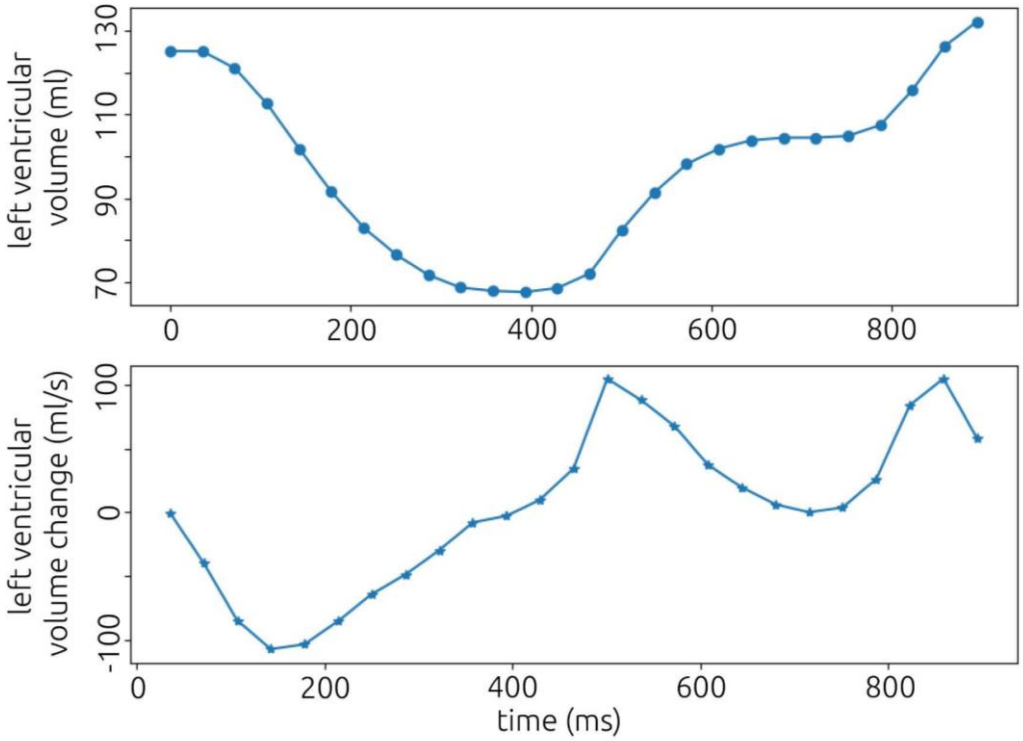


Figure 7: Calculation of the patient specific volume change in the left ventricle from the anatomical information about the left ventricular volume for 25 phases of the heart cycle.

Measurement of functional parameters using anatomical information: some functional parameters can be calculated or at least estimated from anatomical information. The most common example is the calculation of the stroke volume and ejection fraction from the end-diastolic and end-systolic left ventricular volume. While these parameters are arguably describing a change in the anatomy, they also provide relevant information for in-silico models for calculation of hemodynamic. If the left ventricular volume is available for multiple phases of the heart cycle, not only the stroke volume, but also the patient-specific volume flow rate can be calculated (see *Figure 7*).

Direct measurement of functional parameters: depending on the image data available, also direct assessment of functional parameters might become possible. Here, the most common examples for hemodynamic modelling are the assessment of patie

[28] Goubergrits L, Hellmeier F, Bruening J, Spuler A, Hege HC, Voss S, Janiga G, Saalfeld S, Beuing O, Berg P. Multiple Aneurysms AnaTomy CHallenge 2018 (MATCH): uncertainty quantification of geometric rupture risk parameters. Biomed Eng Online. 2019 Mar 25;18(1):35. doi: 10.1186/s12938-019-0657-y. PMID: 30909934; PMCID: PMC6434802.

nt-specific velocities using either Doppler ultrasound or 4D velocity encoded MRI (4D VEC MRI). While the former is ubiquitously available, the latter requires additional efforts with respect to scanning times and costs. Ultrasound velocity measurements are usually directly evaluated at the ultrasound machines or using the respective vendor software. Hemodynamic evaluations using Doppler Ultrasound include measurement of maximum velocities, average velocities, as well as derived information as Bernoulli pressure gradients or the calculation of the aortic valve area using the continuity equation[29].

If *velocity-encoded cine* (VEC) MRI data is available, additional hemodynamic information can be assessed. For example, if planar VEC MRI measurements are performed, the patient-specific volume flow rate can be calculated. Also, the patient-specific velocity vectors can be assessed and used as boundary conditions[30],[31]. From these planar measurements, parameters describing the flow profile, as for example the normalized flow displacement[32], which describes the eccentricity of the blood flow in a vessel, can be calculated and used for model validation. If 4D measurements (x, y, z, time) are available, direct comparison of spatiotemporal flow fields with hemodynamic simulations is possible [Quelle]. Also, 4D measurements allow post-hoc change of evaluation planes and provide in general more information[33]. For example, using appropriate models, the pressure gradient across aortic stenosis[34] or the wall shear stress in vessels can be assessed using this information.

Specific processing for physics-based models

While the previously mentioned processing steps for functional parameters are already relevant for providing patient-specific boundary conditions for in-silico models or virtual cohorts, several models also require patient-specific anatomies. Most tools provide solutions to generate surface geometries as for example triangulated meshes from 3D label masks. A common technique to generate those surfaces is the marching cubes algorithm[35]. However, surfaces generated commonly have a rather coarse surface resulting from the discrete resolution of the image voxels and therefore require subsequent smoothing. Especially for vascular structures, the common assumption is to model the lumen boundary as a smooth surface, as the endothelial layer of the intima is known to be very smooth. For other structures as for example the left ventricle, multiple approaches regarding the surface smoothness exist. While the boundary between myocardium and left ventricular lumen is defined by the trabeculae carneae as well as the papillary muscles, these structures are often ignored[36].

Various algorithms for surface smoothing exist. However, the algorithm should be chosen carefully, as the smoothing procedure can affect the surface geometry. For example, Laplacian smoothing

[29] https://ecgwaves.com/lesson/principles-of-hemodynamics/

[30] Youssefi P, Gomez A, Arthurs C, Sharma R, Jahangiri M, Alberto Figueroa C. Impact of Patient-Specific Inflow Velocity Profile on Hemodynamics of the Thoracic Aorta. J Biomech Eng. 2018 Jan 1;140(1). doi: 10.1115/1.4037857

[31] Goubergrits L, Mevert R, Yevtushenko P, Schaller J, Kertzscher U, Meier S, Schubert S, Riesenkampff E, Kuehne T. The impact of MRI-based inflow for the hemodynamic evaluation of aortic coarctation. Ann Biomed Eng. 2013 Dec;41(12):2575-87. doi: 10.1007/s10439-013-0879-2. Epub 2013 Aug 2.

[32] Burris, N.S., et al., Systolic flow displacement correlates with future ascending aortic growth in patients with bicuspid aortic valves undergoing magnetic resonance surveillance. Invest Radiol, 2014. 49(10): p. 635-9

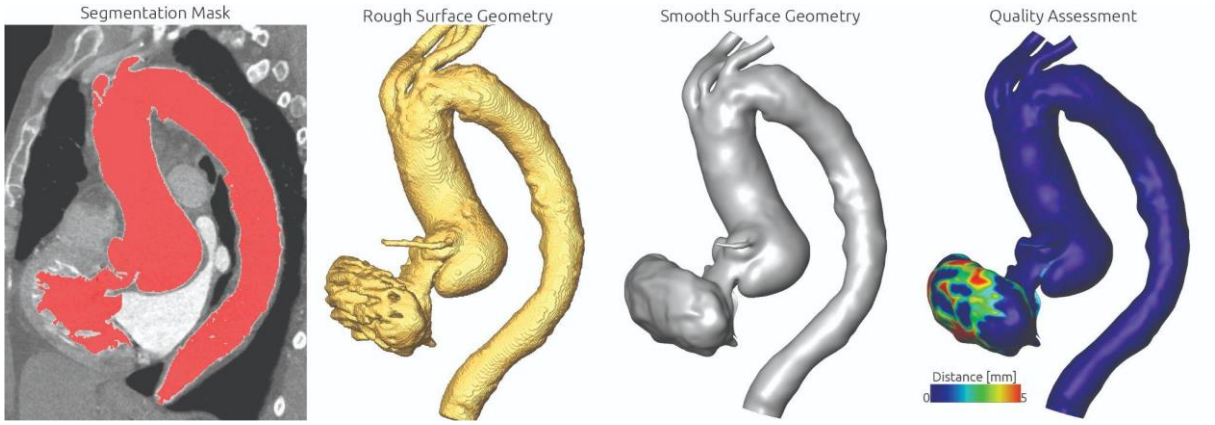
[33] Dyverfeldt P, Bissell M, Barker AJ, Bolger AF, Carlhäll CJ, Ebbers T, Francios CJ, Frydrychowicz A, Geiger J, Giese D, Hope MD, Kilner PJ, Kozerke S, Myerson S, Neubauer S, Wieben O, Markl M. 4D flow cardiovascular magnetic resonance consensus statement. J Cardiovasc Magn Reson. 2015 Aug 10;17(1):72. doi: 10.1186/s12968-015-0174-5. PMID: 26257141

[34] Ha H, Lantz J, Ziegler M, Casas B, Karlsson M, Dyverfeldt P, Ebbers T. Estimating the irreversible pressure drop across a stenosis by quantifying turbulence production using 4D Flow MRI. Sci Rep. 2017 Apr 20;7:46618. doi: 10.1038/srep46618. PMID: 28425452

[35]Xin Wang, Su Gao, Monan Wang, Zhenghua Duan, A marching cube algorithm based on edge growth, Virtual Reality & Intelligent Hardware, Volume 3, Issue 4, 2021

[36] H. Razafindrazaka et al.. Mesh Based Approximation of the Left Ventricle Using a Controlled Shrinkwrap Algorithm. In: Functional Imaging and Modeling of the Heart DOI: 10.1007/978-3-030-21949-9_25.

algorithms are known to result in diameter reduction and volume shrinking, especially in regions of high curvature. Thus, volume preserving smoothing algorithms might be better suited for smoothing of anatomical structures[37]. Finally, while elaborating a process for smoothing of surface geometries, the resulting geometry should always be tested against the original geometry. For example, the distance between the original and the smooth surface can be calculated to identify regions that were



altered too heavily by the chosen algorithm (see *Figure 8*).

Figure 8: To obtain a smooth surface geometry of the left ventricle and the aorta, a voxel mask of the respective structures must be labelled first (left). Using a marching cubes algorithm, a rough surface geometry (second from left) is generated from the 3D voxel mask. From this rough geometry, a smooth surface is generated that is suitable for hemodynamic modelling (third from left). To evaluate the effect of the smoothing procedure and localize regions of strong deformation, the surface distance between rough and smooth surface can be calculated (right).

[37] Kuprat, A., Khamayseh, A., George, D. & Levi, L. Volume Conserving Smoothing for Piecewise Linear Curves, Surfaces and Triple Lines. Journal of Computational Physics. 172, 99–118 (2001).

Algorithmic description

Finally, if all algorithms and procedures for data processing to obtain the data required for running an in-silico model or specification of a virtual cohort are determined, they should be documented in a suitable manner. This documentation procedure is covered in the *Deliverable 4.4 - Guidelines for documentation (IIB, M12)*. Thus, it will only be covered briefly in this document.

The more complex the procedure or tool used to process medical image information, the more parameters can usually be chosen by the operator. Thus, it is important to describe the settings as well as possible to reduce possible operator biases to a minimum. Ideally, this bias should also be assessed and quantified if possible. In general, the less automatic a procedure is, the more detail should be given to the description of each individual step.

First, the purpose of the specific processing procedure should be briefly described. This includes the type of data used as input as well as the type of output that is generated by the algorithm or procedure. Additionally, relevant parameters of this procedure should be stated. Often, a certain freedom in parameters is necessary to account for the heterogeneity commonly observed in clinical information. For example, to obtain a smooth surface geometry of the left ventricle, different strengths of the smoothing kernel or iterations of the smoothing procedure might be necessary depending on the voxel resolution of the image data, even if resampling of the data was performed first. In those cases, ranges for each variable parameter can be provided. If this is not possible, a guideline on how to determine the suitable parameter value for a given case should be provided at least.

Also, methods for quality assurance as well as results from uncertainty and intra- and inter-operator variability should be described. This information will allow any operator to understand the importance of specific processing steps and parameter settings. This information can also include calculations that should be performed after the data processing procedure to obtain quality measures describing the success of the procedure.

TAVI - smoothing of initial surface geometries

To obtain smooth surface geometries for hemodynamic simulations, the initially reconstructed geometries obtained from image segmentation must be smoothed as the discrete resolution of the image voxel results in stepped surfaces. The smoothing procedure is performed using Meshmixer (v3.3, Autodesk). The parameter ‘Smoothing Scale’ should not be set higher than 10, the parameter ‘Constraint Rings’ should always be set to 10. To check the impact of the smoothing algorithm on the resulting surface, the volume of the initial and smoothed surface must be calculated. The volume change should be below 1 percent. Additionally, the distance between two surfaces is calculated for each surface vertex. The average difference between both surfaces must be below the image voxel resolution.

Contingencies

If the members of SIMCor cannot adhere to the content of these SOP instructions for justified reasons, then deviation from the procedure must be documented and appropriate actions must be taken, considering SOP revision if relevant.

Attachments

In future versions of this document, a Checklist for users will be provided. However, this document is not yet finalized.

Publication policy

To SIMCOR partners, please mention the following statement in each of your publications: *"This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101017578"*.