Model Driven Paediatric European Digital Repository

Call identifier: FP7-ICT-2011-9 - **Grant agreement no**: 600932

Thematic Priority: ICT - ICT-2011.5.2: Virtual Physiological Human

Deliverable 12.2.2

Report on Clinical Assessment and Validation Results for all Four Disease Areas (M36)

Due date of delivery: 29/02/2016

Actual submission date: 10/04/2016

Start of the project: 1st March 2013
Ending Date: 28th February 2017

Partner responsible for this deliverable: OPBG

Version: 2.1

Document Classification

Title	Second clinical assessment and validation results for all four disease areas
Deliverable	12.2.1
Reporting Period	Third
Authors	Marcello Chinali
Work Package	12
Security	RE
Nature	R
Keyword(s)	Clianical Validation

Document History

Name	Remark	Version	Date
Marcello Chinali	Preliminary outline of the document	Draft 12.2.1 ver 0	
Marcello Chinali	Contributions from WP 9, 10, 11	Draft 12.2.1 ver 1	
Marcello Chinali	Contribution from WP 8	Draft 12.2.1 ver 2.1	

List of Contributors

Name	Affiliation
Tobias Heimann	SAG
Dominik Neumann	SAG
Lucian Mihai Itu	SAG
Jose Pozo Soler	USFD
Maria Jimena Costa	SAG
Frans Steenbrink	MOTEK
Marjolein van der Krogt	VUMC
Claudia Mazzà	USFD
Luca Modenese	USFD
Olivier Ecabert	SAG
Oliver Pauly	SAG
Pieter van Dijkhuizen	IGG
Gabriele Rinelli	OBG
Paolo Cappa	URLS

List of reviewers

Name	Affiliation

Table of Contents

1) AIM OF THE DOCUMENT

The aim of the Work Package 12 is to assess the accuracy of multidisciplinary derived models in order to improve the identification of markers of outcome prediction and risk stratification, and thus to derive and evaluate personalized treatment models.

The goal is to validate the computational models to assure that they can be personalized by adapting the parameters to the integrated data of a specific patient and to improve the current knowledge and understanding of the disease by simulating different aspects on the evolution of a disease. In addition it also aims at verifying the accuracy of the insights of the effect of a specific therapeutic intervention; being this either pharmacological, behavioral or surgical.

Clinical validation of the models, as stated in the description of work, will be an ongoing process which benefits from the use of the models in different clinical settings thus incresing the number of clinical observations and improving the stability of the clinically derived model. The final aim of WP-12 is to assure that the data repository will be in the condition to be continuously improved, in order to assure the accuracy and stability of the derived models and result into integrated clinical workflows leading to personalized treatment models.

In D12.1 we have described in detail the validation methodology-outline, which has started in the second year of the project and detailed in the D12.1.1.
Given the significant heterogeneity of the different clinical areas and the resulting differences in model definition and expected abilities, the second year validation process required that numerous partners, from both the clinical and the technical areas, proceed on a cogent and well defined validation process. Accordingly, the validation decribed in the present document strictly evolves in parallelel to model production, in a continuous effort of pairing model advancement to validation and clinical utility.

As stated, the current validation process does not seek to override existing clinical and technical procedures, with specific regard to the models validation methodos developed by each of the technological partners.

The aim of the present document is to report on advancement and current status of the clinical validation process for the four disease areas.

2) Validation process for each disease area

2.1. Validation for Cardiomyopathies models (WP3-WP8)

2.1.1 Evaluation of the whole body circulation model

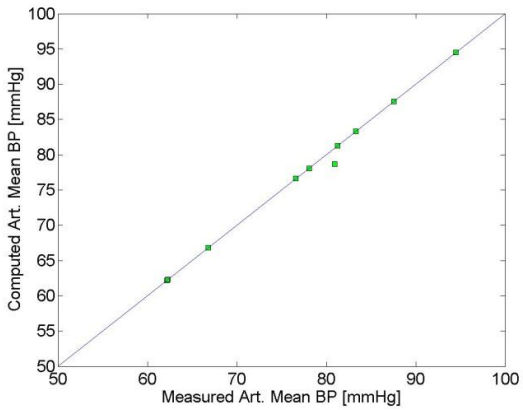
In the third year of the project we have significantly extended the personalization algorithm of the lumped parameter whole body circulation model. Besides the previously used standard set of objectives (12 objectives in total for systemic and pulmonary circulation based on systolic, diastolic and average arterial/venous aortic pressure, interval of time during which the aortic/pulmonary valve is open, and maximum and minimum left/right ventricular volume), we introduced new advanced objectives based on slopes and intervals of time extracted from the pressure and volume profiles. Hence, the number of objectives was extended from 12 to 28. For further details we refer to the concomitant third annual report of WP8.

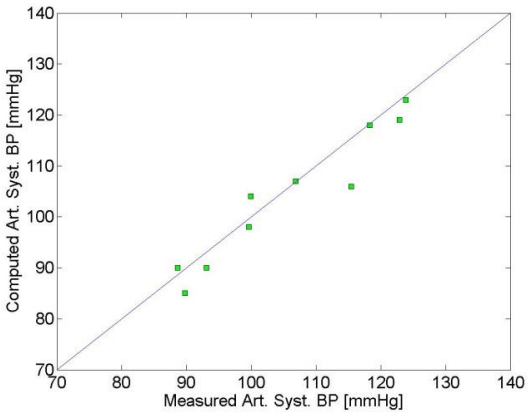
The lumped parameter whole body circulation model and the personalization approach were validated based on the data extracted from 10 patients, both at baseline and at follow-up. For the baseline configuration we have used all types of information that were available (non-imaging data: blood pressure and heart rate, imaging: left ventricular volume), whereas for the follow-up configuration we have only used the new heart rate value. Hence, the whole body circulation model was fully personalized at baseline, and then for the follow-up computations the personalized parameter values were reused and only the heart rate was modified. The rationale behind this approach is:

- Once the model is personalized for the baseline state of the patient, the clinician could test different what-if scenarios, by setting different values of the heart rate, so as to determine e.g. how the ejection fraction of the patient would change (these changes in heart rate could be induced by different medication therapies);
- The follow-up heart rate could be measured by the patient himself, using for example a smartphone, and then used to estimate new values of measures of interest, like the ejection fraction.

2.1.1.a Results at baseline

In the following section we present the results obtained for the personalized blood flow computations at baseline. Since the ejection fraction was determined by the clinicians, we have used as objectives the maximum left ventricular volume determined from the medical imaging data, and the target value of the minimum left ventricular volume was computed from the maximum left ventricular volume and the clinically provided ejection fraction. In Figure 1 is displayed a comparison of measured and computed arterial systolic, diastolic and mean pressure, maximum and minimum left ventricular volume, and the left ventricular ejection fraction. The agreement between the measured quantities and their computed counterparts is very good, with no major outliers. The following Table further displays the correlation and the mean absolute differences between computed and measured quantities: the results further underline that the model was successfully personalized.





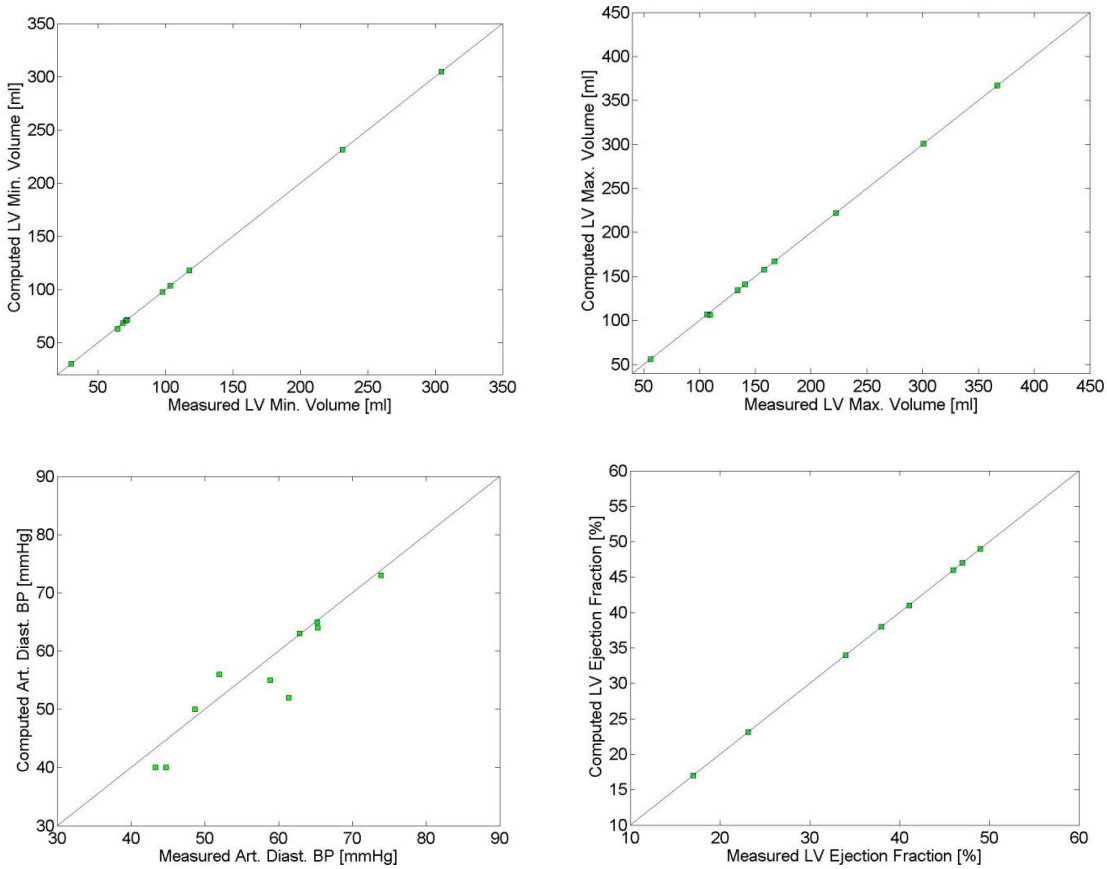


Figure 1: Scatter plots of computed versus measured values of arterial systolic, diastolic and mean pressure, maximum and minimum left ventricular volume, and left ventricular ejection fraction at baseline.

Table: Correlation and mean absolute difference between computed and measured values of arterial systolic, diastolic and mean pressure, maximum and minimum left ventricular volume, and ejection fraction.

Quantity	BP_{syst}	BP_{diast}	BP_{mean}	LV_{Max}	LV_{Min}	LV EF
Correlation	0.962	0.939	0.998	1.0	1.0	1.0
Mean abs. Diff.	2.934 mmHg	2.928 mmHg	0.235 mmHg	0.29 ml	0.16 ml	0.02 %

Results at follow-up

In the following section we present the results obtained for the personalized blood flow computations at follow-up. As mentioned above, all personalized parameters, determined for the baseline computations, were maintained constant and only the heart rate was modified. Hence, no further personalization with respect to any objective value was performed.

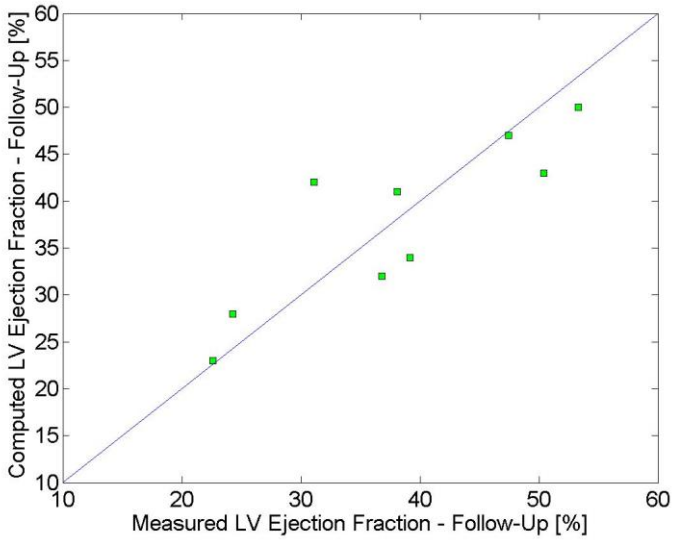


Figure 2 displays a comparison of measured and computed ejection fraction. Overall, the computational model is able to predict well the ejection fraction at follow-up, with a correlation of 0.87 and a mean absolute difference of 4.58%.

A more detailed analysis reveals that the ejection fraction has changed between baseline and follow-up exam for seven out of the ten patients included in the study. Out of these seven patients, the model has predicted correctly the direction of change in ejection fraction for five patients.

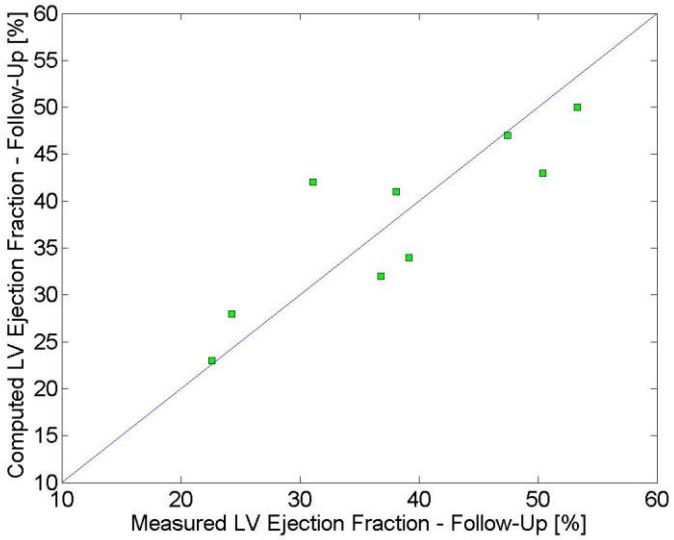


Figure 2: Scatter plots of computed versus measured values left ventricular ejection fraction at follow-up.

The present follow-up study only takes into account the long-term effect of heart rate on the ejection fraction. There are other factors which may lead to a change of the ejection fraction, like disease progression, change of the state of the patient between baseline and follow-up, etc. Furthermore,

since a lumped parameter model was employed, which provides results in a timeframe of minutes, biomechanical, electrophysiological and hemodynamic properties, which influence the ejection fraction, have only been modelled at a reduced scale. A full scale modelling may further improve results.

2.1.2 Evaluation of the electro-mechanical cardiac modelling pipeline

The fully-automatic electro-mechanical simulation and personalization pipeline (WP8) has been finalized during the second year of the project. A major effort in year three was on extensive patient processing to test its robustness against the different patient phenotypes, degrees of pathologies, and varying data quality. The results based on the first 35 autonomously personalized MD-Paedigree patients, including patients from all three clinical centres (OPBG: 17, UCL: 17, DHZB: 1), indicated good versatility of the proposed methods. For technical details we refer to the concomitant annual report of WP8.

In order to validate the modelling tools and personalization pipeline, and to assess the utility of the resulting personalized models, an important step is to test the models' ability to accurately capture the patients' physiologies at baseline. To this end, we compared the personalized models and their outputs to the available clinical data, both quantitatively (stroke volume, ejection fraction, ECG features, etc.) and qualitatively (e.g. cardiac motion observed in cine MRI versus simulated motion).

2.1.2.a Quantitative Evaluation

First, we assessed the quantitative goodness of fit between personalized model outputs and measured data on a case-by-case basis and at population level. To this end, we compared various clinically significant quantities of interest for all processed patients, see

Table 1 and the plots in Figure 4:

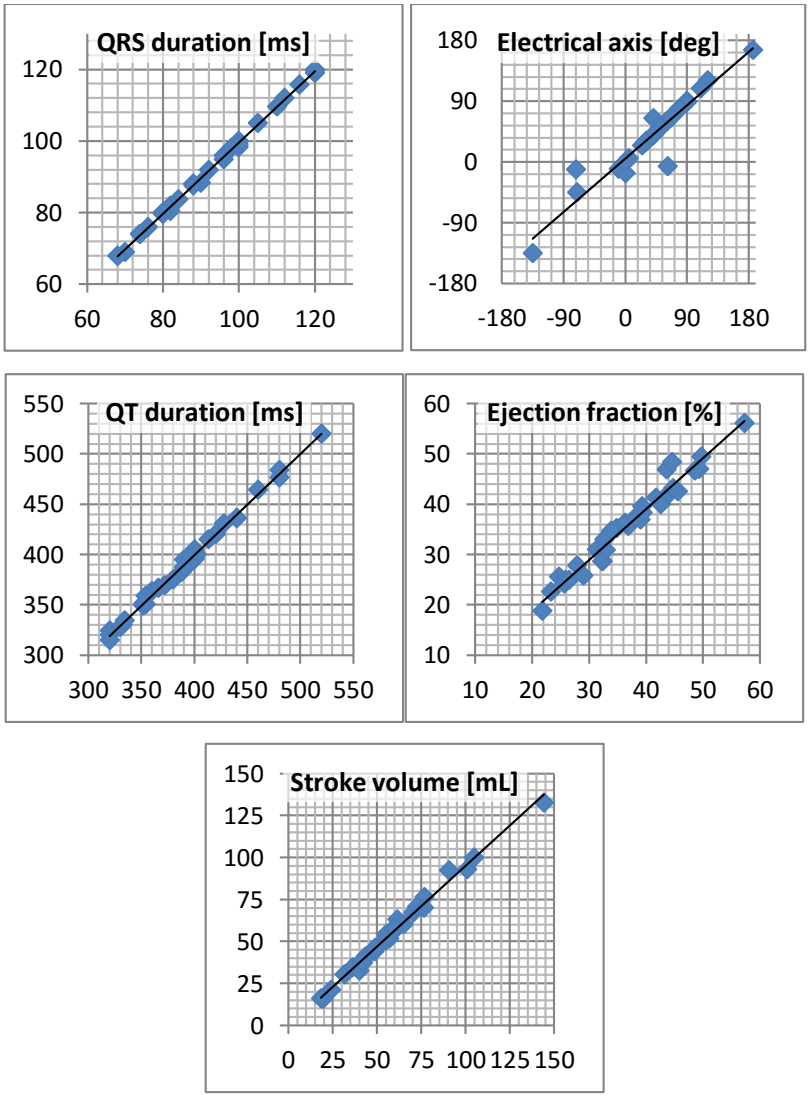


Figure 4: Measured (x-axis) versus computed (y-axis) quantities of interest. First row depicts quantities related to EP derived from 12-lead ECG and the personalized model, respectively. The second row focuses on important markers derived from the coupled electro-mechanical model (y-axis) and compares them against measurements derived from cine MRI (x-axis).

Table 1: Mean misfit and standard deviation: measured versus computed quantities of interest.

Quantity	QRS duration	Electrical axis	QT duration	Ejection fraction	Stroke volume
Mean	0.35 ms	6.61 deg	2.93 ms	1.61 %	3.49 mL
Std.dev.	0.51 ms	16.1 deg	1.62 ms	1.14 %	2.22 mL

As one can see, the agreement between computed and measured values is high and the degree of misfit is mostly below clinical variability. Some outliers (although not severe) can be observed for the

electrical axis, which we are planning to investigate in the future. Potential reasons include failure of a personalization component or modelling assumptions that may be invalid for a subset of the patients.

2.1.2.b Qualitative Evaluation

The quantitative analysis above gives crucial clues on model performance, yet it does not cover the complete spectrum of information provided by the imaging data and the models. For instance, even when cardiac dynamics is captured perfectly at a global scale (e.g. stroke volume), the dynamic, time- and regionally-varying motion might not match well. In an effort to facilitate the review of the modelling performance in this regard, we developed an intuitive web-based platform, called CaSiReView, with the goal to provide a light-weight, yet comprehensive overview of the personalization results with automatic comparison to the available data (see WP8 annual report for more details). A screenshot of the prototype and example qualitative comparisons results for some patients is shown in Figure 5.

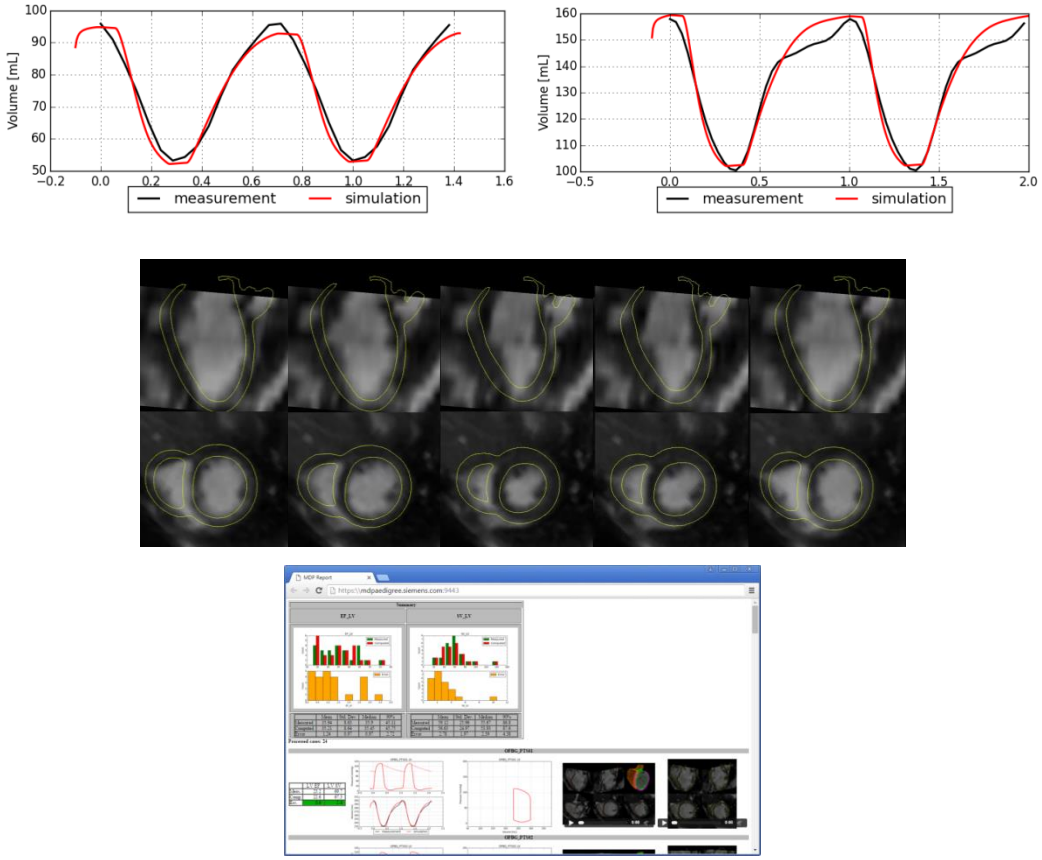


Figure 5: Top: Measured (black) versus computed (red) left ventricular volume curves for two patients indicating excellent global volume fit over time (2 heart cycles). Bottom left: Personalized model overlaid on baseline cine MRI data for one patient. Bottom right: Screenshot of CaSiReView, the tool which was developed

to generate these case-based reviews (including quantitative and qualitative results with plots and videos) for evaluation and validation purposes.

The majority of models match very well to the available baseline data, but in some cases we observed discrepancies in regional motion patterns. We will investigate those cases individually; however, to what extent such discrepancies will have an impact on the predictive power of the model is topic of our active research for the upcoming months.

2.2 Validation for CVD risk in obesity models (WP4-WP9)

2.2.1 Image-based patient characterization

In the context of WP9, our first goal is to create multi-modal patient representations by extracting personalized patient parameters from different sources of information such as clinical, blood tests or imaging (e.g. MRI). Concerning parameters extracted from imaging, some characterizes the cardiac function and result from computational models developed within WP8 that are adapted to the heart of obese children, and others characterizes the fat distribution within the body using a dedicated MRI protocol. Concerning the models from WP8 adapted to the heart of obese children, validation is already being conducted within the context of WP8. Concerning the fat segmentation, the fully automatic active-shape-based liver method developed by Fraunhofer within WP9 has been evaluated on all available prospective UCL datasets.

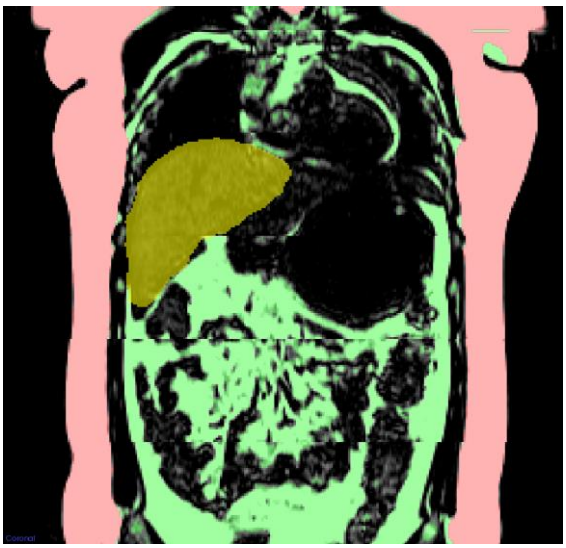


Figure 6: Image showing hepatic, internal and subcutaneous fat respectively in yellow, green and red.

Figure 7: Bland Altman plot of the first fully automatic calculated fat estimates using the tool provided by Fraunhofer compared to fully manual obtained fat estimates (ground-truth) created by UCL. Bland Altman plot comparing the fully automatic calculated SAT fat estimates provided by Fraunhofer with fully manual obtained SAT fat estimates (ground-truth) created by UCL.

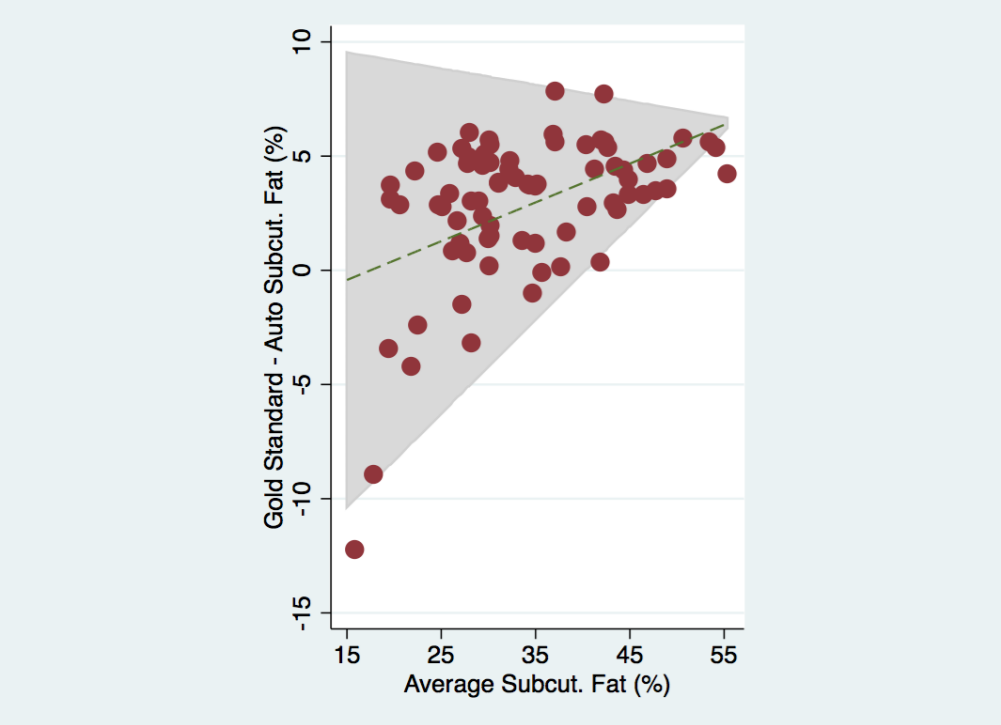


Figure 7: Bland Altman plot of the first fully automatic calculated fat estimates using the tool provided by Fraunhofer compared to fully manual obtained fat estimates (ground-truth) created by UCL.

It can be seen that that the automatic calculated fat estimates contain a positive bias. The reason could be that for the manual estimated SAT fat the image data was cropped the lower leg below the knee and everything above the neck. Further, arms were manually removed from the data. In comparison, the automatic estimated SAT fat contains areas below the knee and above the neck. A further bias is introduced, because automatic arm removal failed sometimes and SAT segmentation of the lower leg produces over-segmentations which contain bony structures.

2.2.2 Case-based reasoning

Once a large dataset of patients with their corresponding multi-modal representation becomes available, case-based reasoning can be used to retrieve similar patients, perform stratification and risk assessment. To this end, we develop a prototype called CaseReasoner that consists in a learned representation for patient data, an associated similarity measure to compare patient representations and finally, a database of reference patients associated with relevant information for risk assessment.

In WP9, we propose to learn a multimodal low dimensional representation of patient data by extending deep generative models to the multi-modal context, either by using stacked RBMs or autoencoders. The resulting patient signatures can be binarized and compared using a Hamming distance.

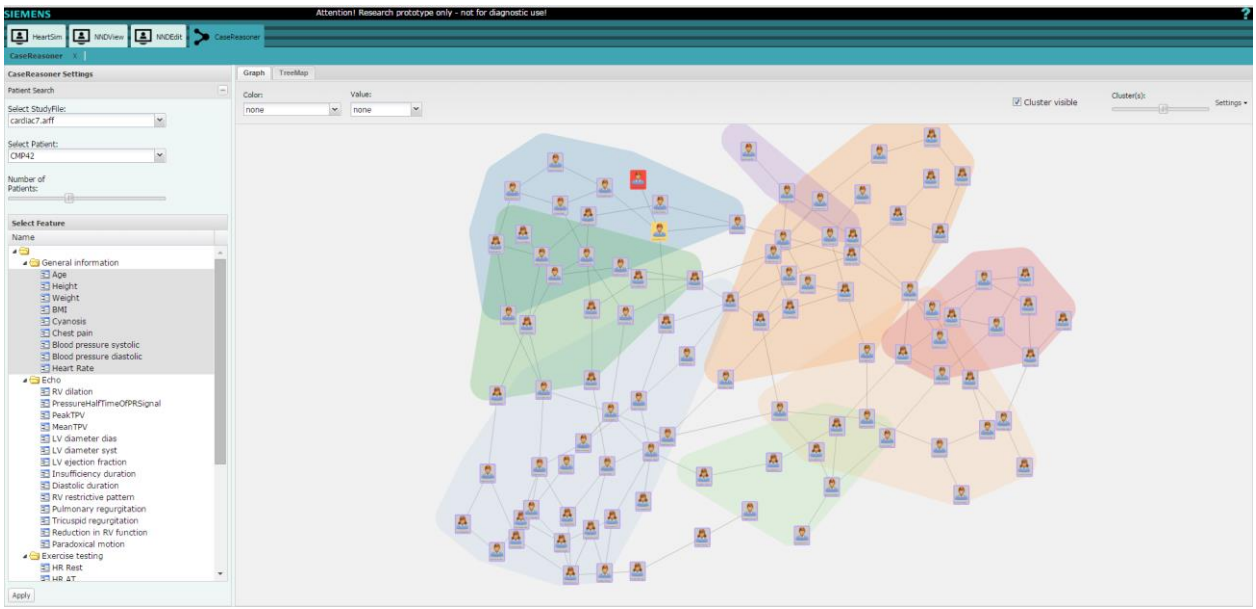


Figure 8 - Patient stratification using CaseReasoner

To perform the technical validation of such case-based reasoning system, our plan is to conduct cross-validation experiments using the different datasets we collect and process in the context of WP9. More precisely, we plan to perform multiple randomized 2-folds cross validation strategy. For each cross-validation run, the patient dataset is randomly split into 2 subsets or folds used as a training set and test set. Using the training set, the deep autoencoder is pre-trained without using any supervision or with supervision using pseudo-targets. Afterwards, either unsupervised or supervised back-propagation can be additionally performed to fine-tune the network. In the supervised case, intermediate outcome information such as resting systolic blood pressure or glucose response to meal or intervention can be used to drive the back-propagation. Once encoded using this binary representation, a new incoming patient from the test set can be compared to all reference patients within the reference database by using Hamming distance. Based on the retrieved nearest neighbor patients, inference can be done by aggregating (intermediate) outcome information among the retrieved patients. Quality measures such as sensitivity and specificity can be then derived from such cross-validation.

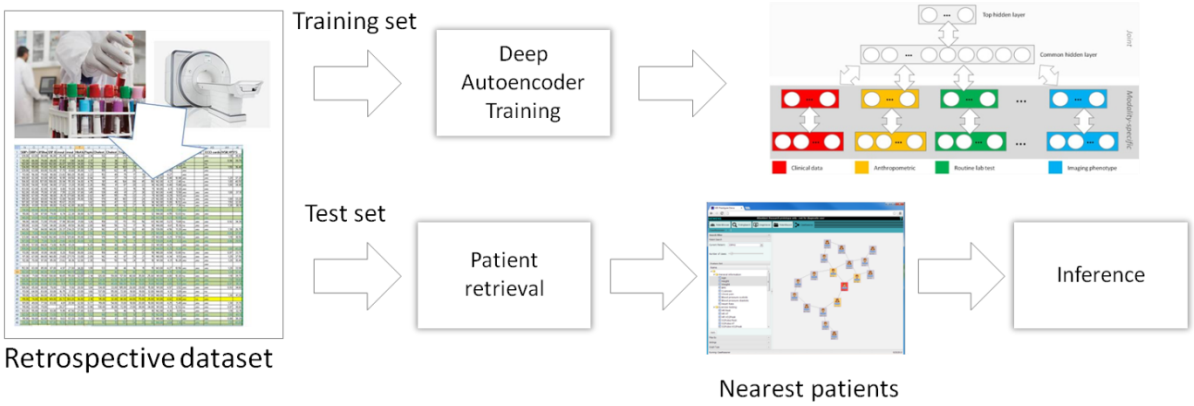


Figure 9: Cross validation experiment for validating CaseReasoner

At the present time, 46 patient data have been collected from OPBG (baseline), and 54 patient data from UCL for the prospective study, and additional 69 patients from a retrospective study. 9 patients from OPBG, 24 patients from UCL (prospective) and 20 patients from the retrospective study have been processed through the segmentation pipeline. The corresponding circulation models have been successfully computed for all patients from the prospective studies (9 + 24). The tool has been tested on the 9 + 24 patient data from the prospective studies. In a fraction of them, segmentation did not perform very well so that manual corrections and or recomputation are required. Resulting parameters needs to be formatted and analyzed prior integration within the file share.

Clearly, the data collection and processing is not advanced enough to start with cross-validation experiments on real data. For this reason, we conducted first proof-of-concept experiments to analyze the properties of autoencoders for dimensionality reduction in a controlled setup by using synthetic data generated from Gaussian mixture models. Note that following experiments are also described in D9.3.

First a simple three-layered auto-encoder with 3 neurons (input layer) – 2 neurons (hidden layer) – 3 neurons (output layer) was considered. For the input and output neurons the linear activation was used, while for hidden ones the sigmoid activation was chosen. The whole network was trained using backpropagation and linear synthetic 3-D samples randomly generated from a multivariate Gaussian distribution, where all 3 clusters of points have been completely separated (Figure 10). The same dataset was used as an input for the PCA for a fairly comparison. The results are shown in Figure 11. Both methods were implemented to project one hundred of 3-D samples on 2-D space.

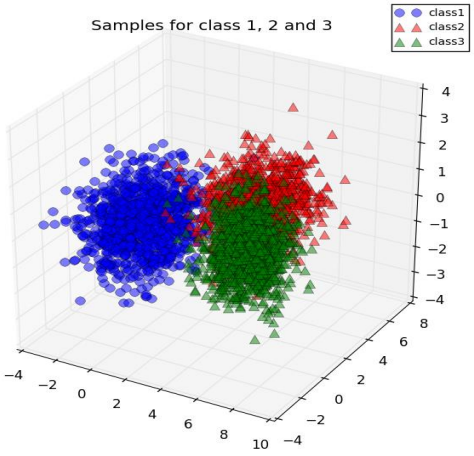


Figure 10: Samples drawn from a multivariate Gaussian distribution

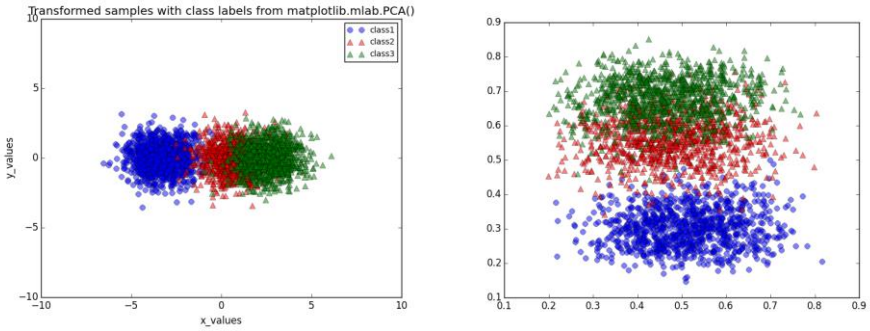


Figure 11: Representation of the representative 2-D features (reduced feature sub-space) for PCA(left) and Auto-encoder (right).

Since the purpose of dimensionality reduction is visualization of highly dimensional data we increased the input dimensionality to 10, resulting in a 10-D feature space. An auto-encoder was trained to reconstruct the input and return a 2-D feature space witch best represents the data (Figure 6).

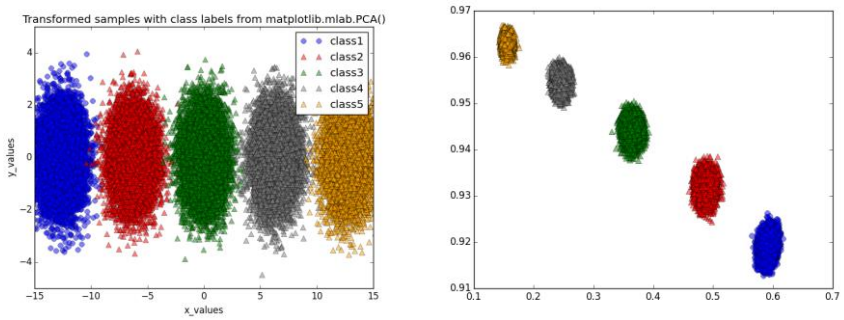
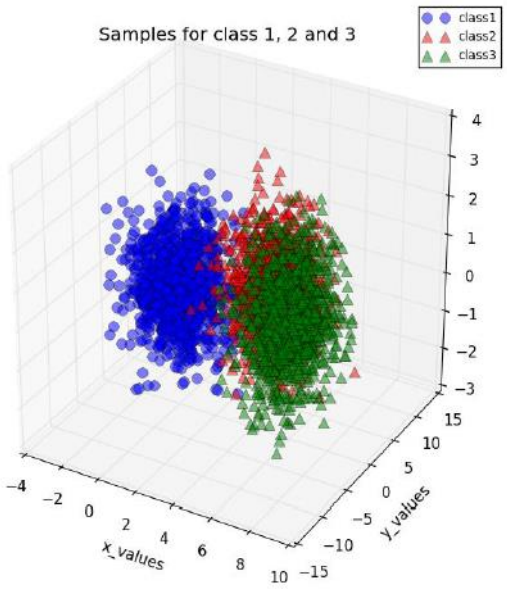


Figure 12: Representation of the selected 2-D features from the 10-D space (reduced feature sub-space) for PCA (left) and Auto-encoder (right).

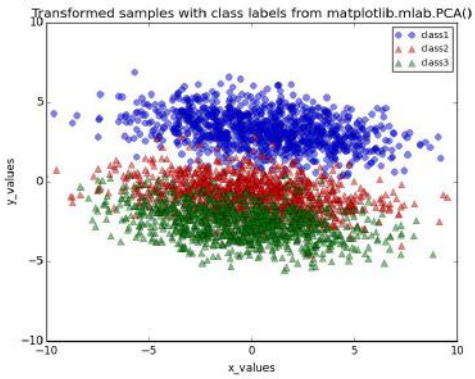
Some other synthetic datasets were generated in which the mean and variance parameters of the Gaussian distribution were modified to capture the behavior of the auto-encoder in different situations (Figure 12). It has been shown that auto-encoders successfully projected the data onto a lower dimensional surface being able to find the directions along which the data has maximum variance. Traditionally, dimensionality reduction depended on linear methods such as PCA, which finds the directions of maximal variance in high-dimensional data. By selecting only those axes that have the largest variance, PCA aims to capture the directions that contain the most information about the inputs, so we can express as much as possible with a minimal number of dimensions. The linearity of PCA, however, places significant limitations on the kinds of feature dimensions that can be extracted. Autoencoders overcome these limitations by exploiting the inherent nonlinearity of neural networks. In future work, we plan to evaluate the abilities of autoencoders to cope with non-linearity, and this, on toy examples first.



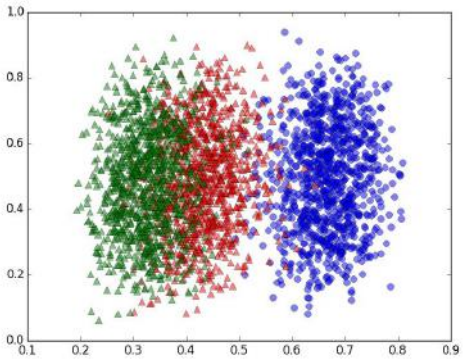
Cluster1:
- mean [0,0,0]
- cov [1,0,0], [0,10,0], [0,0,1]

Cluster2:
- mean [4,0,0]
- cov [1,0,0], [0,10,0], [0,0,1]

Cluster3:
- mean [6,0,0]
- cov [1,0,0], [0,10,0], [0,0,1]



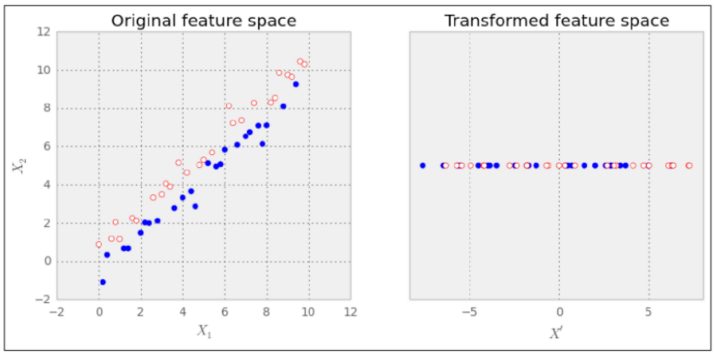
PCA



AE

Figure13: 3-D to 2-D reduction when the standard deviation was increase significantly along the y axis using PCA(bottom-left) and Auto-encoder (bottom-right).

As mentioned previously, PCA being a linear method, it cannot cope with data showing non-linear structure. Another weakness of PCA, is when it it's being applied to special classification problems. In the following example the classes are not distributed accordingly to the axis with the highest variance but the second highest variance. Since PCA chooses the highest variance direction, it clearly fails in this situation. As dimensionality reduction is an unsupervised task, we don't provide any clues regarding the class labels and thus PCA cannot do better.



Regarding the comparison between stacked auto-encoders and PCA we consider the following 2-D dataset, where samples are arranged following a circle path. Samples distributed on the circle are separated into two different classes, the red and the blue one respectively. First, we take a look at how well auto-encoders can reconstruct their original output compared to PCA using 2 dimensions. The reconstruction by the auto-encoder is visibly better than the PCA output, which is a very promising result.

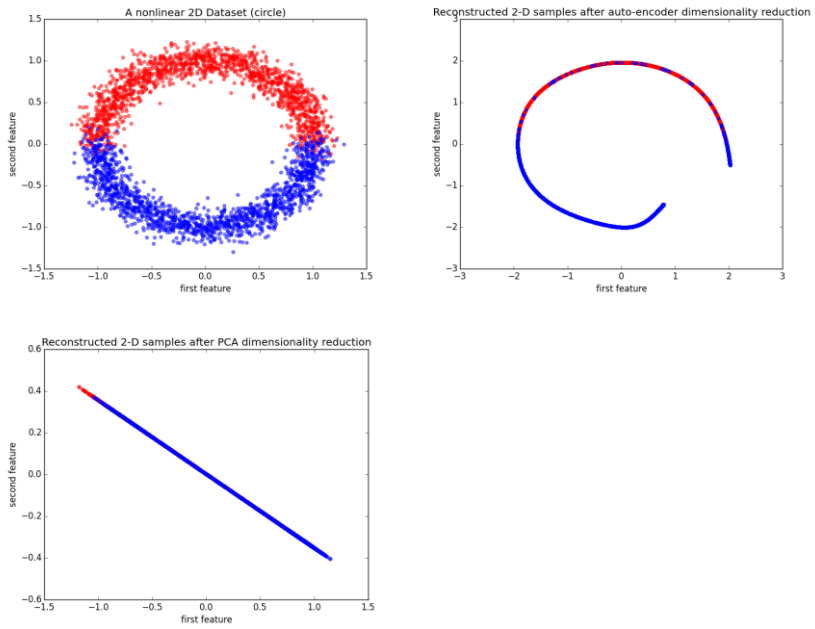


Figure 14: Reconstruction by the auto-encoder

2.3 Validation for the JIA models (WP5-WP10)

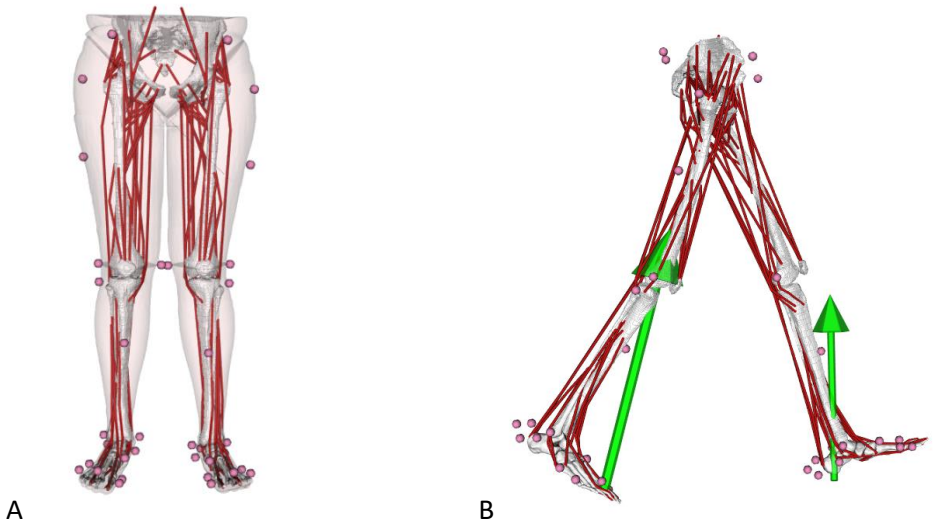
2.3.1 Modelling status

The models developed by WP10 are patient specific musculoskeletal models of the lower limb of children affected by juvenile idiopathic arthritis. Patient specific bone geometries and joint definitions are included in the computational models. In details, the operations involved in building the patient specific models are the following:

1) Virtual palpation of anatomical landmarks (i.e. identification of points in a multimodal display interface where bone reconstructions and MRI are visible at the same time) on the bone geometrical models obtained from the MRI images. A software called NMSBuilder was used for this operation (Valente et al., 2014).
2) Registration of a generic atlas of muscle attachments (Arnold et al. (2010) onto the patient specific bone geometries, using an affine transformation defined by the registration of specific bony landmarks (Ascani et al., 2015).
3) Refinement, by manual adjustment based on the MRI images, of the muscle paths obtained by directly connecting the muscle attachments estimated at the previous step.
4) Calculation of the inertial properties of each segment.

5) Creation of the joints connecting the bodies and definition of their axes using selected anatomical landmarks.
6) Fusion of the ankle and foot model to a model of the lower limb. The latter model, according to the available data, can be either a generic scaled model or a patient specific model obtained from month-6 lower limb MRI data.
7) Registration of the markers from the gait lab and the markers visible in the MRI scans in order to associate the gait data to the anatomical model.
8) Simulations of the patient's gait and estimation of the muscle and joint contact forces.

Once the models have been produced (Figure 1 A), they can be used to simulate the patient's gait employing the kinematics and kinetics data collected in the gait lab (Figure 1 B).



A B

Figure 15: (A) Month-6 models with month-12 patient specific ankles, including virtual markers. (B) A frame of a simulated walking trial.

2.3.2 Current application

Currently, processed data were available for one single patient at the three time points specified in Table2. The biomechanical models were produced, gait simulations generated and contact forces acting at the ankle joint estimated (Figure 2) at all time steps.

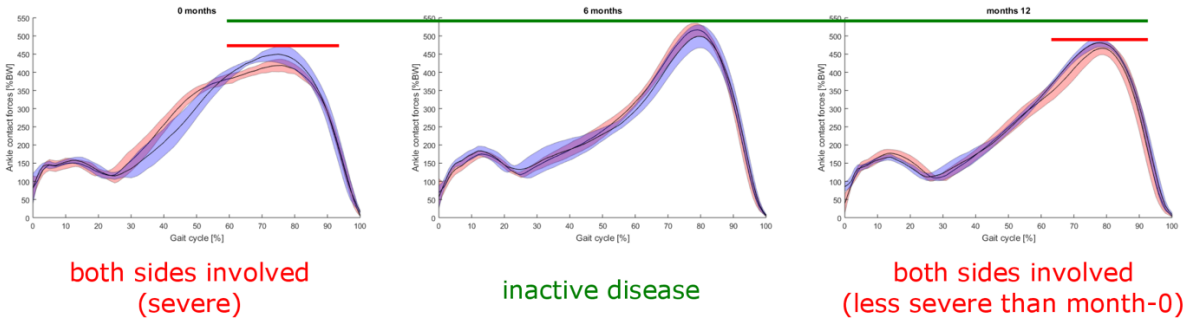


Figure 16: Ankle joint contact forces estimated for right (red) and left (blue) side at the three considered timepoints.

Calculated ankle contact forces were consistent with the clinical condition reported for the patient (severe disease involvement for both ankles at month-0, inactive disease at month-6 and bilateral involvement, but less severe than month-0, at month-12), highlighting a protective strategy of the patient towards the painful joints.

These initial results are encouraging in the perspective of answering the first of the three questions of clinical interest.

Table 2: Magnitude of the contact forces (mean±standard deviation) acting at the patient ankle at the three time points considered. The number of trials considered in the simulations are specified in brackets.

Ankle side	Ankle peak contact forces [%BW]		
	Month 0	Month 6	Month 12
Right	419±17 (6 trials)	517±18 (7 trials)	466±17 (5 trials)
Left	450±17 (5 trials)	499±18 (6 trials)	481±17 (6 trials)
Involvement	Both ankles (severe)	None (Inactive disease)	Both ankles (less severe than month-0)

2.3.3. Validation of the experimental and modelling pipeline

The described pipeline has been assessed through a repeatability extensively described in D10.4. The study was designed to investigate the repeatability of three critical steps of the modelling pipeline

used to generate patient specific models of the foot and ankle joint: virtual palpation of bony landmarks, manual adjustment of muscle paths and the definition of foot joint axes.

The virtual palpation was found to be a repeatable operation, both intra- and inter-operator, and allowed for a refined subset of bony landmarks to be determined from the generic atlas initially developed. The manual adjustment of muscle attachments, even if performed by experienced operators using multiple MRI sequences, has been shown to be a highly operator-dependant step of the current pipeline. These results affect the biomarkers that will be extracted using the biomechanical model in the measure that the lack of repeatability affects ankle contact forces. Assessing the sensitivity of the articular loading at the tibio-talar joint to perturbation of the muscle attachments it was found that misplacement of the path of muscles with larger physiological cross sectional area and moment arm can affect contact forces. Still, even in the worst case scenarios, the sensitivity to overestimated uncertainty in the anatomical modelling was never much larger than 10%. Most of the errors were related to the Achilles' tendon, that from now on will be individualised with particular care. It is realistic to expect that in most cases sensitivity to anatomical uncertainty will never exceed 2-3%. Considering that subject-specific models of this complexity in any case produce predictions that are never more accurate than 90%, this level of sensitivity seems perfectly acceptable for the purpose. Finally, the identification of joint axes appeared to be less critical than the other two steps, although the introduction of semi-automatic tools in joint definition would represent a further improvement.

2.3.4 Validation of the biomechanical models

Validation of musculoskeletal models is a challenging task, because these models are used to estimate forces occurring within the human body, which cannot be directly measured. Traditionally, validation is undertaken as a qualitative comparison between predicted muscle activations and recorded electromyographic (EMG) signals (experimental muscle activations). The same validation procedure was applied for the developed model, comparing the calculated activations with the shape of EMG signals experimentally collected during the gait trials (Figure 25 and Figure). The raw EMG signals were rectified and low-pass filtered.

Another option for validating musculoskeletal models is comparing the values of contact forces predicted by the model against values measured by instrumented prostheses. However, these prostheses are implanted in adult patients undergoing total hip or knee joint replacement because they suffer of osteoarthritis, so the measured contact forces can hardly be used to validate a pediatric model and *a fortiori* a disease one. Although we did not perform a direct validation based on the

calculated joint contact forces, it was verified (Prinold et al., 2015) that our estimates were similar to the results available in the literature from previous simulations (Procter and Paul, 1982).

2.4 Validation for NND models (WP6-WP11)

Medical treatment to possibly regenerate or halt nerve and muscle degeneration in combination with rehabilitation and surgical procedures will hopefully revolutionise the way patients are treated and improve their quality of life. Accurate modeling of the legs and their function is an important step in this direction.

In deliverables D11.1 and D11.2/D11.3, SAG presented a novel method to extract anatomical structures from MRI images of healthy and ill children's legs. Complementarily, in deliverables D11.2 and D11.3, USFD presented a method to estimate from the segmented geometries, the location of bony landmarks, ligaments and muscle/tendons insertion, attachments, and lines of action. In addition, USFD presented a method to compute bone and muscle volumes, and segment-specific reference frames, to be used for the personalization of OpenSim models.

In this document the different technical steps are validate by each of the involved parners. In particular SAG validate the adaptation of the method to extract subject-specific muscles, bones and skin of the pelvis and legs, from MRI images of paediatric patients from 3 neuromuscular disease groups, namely CMT, DMD and CP. USFD validate the accuracy of the complete template anatomical model and of the mesh morphing method used to estimate the landmarks, attachements and lines of action.

2.4.1 Validation of the MRI segmentation

For each of the 3 disease groups, a set of 3-dimensional (3D) meshes for 54 individual muscles, 12 bones, fore, middle and back feet as well as the whole skin of paediatric patients were manually annotated at SAG. For each of the 3 disease groups, 25 landmarks were annotated as well for qualitative assessment of the results.

A qualitative validation between automatically defined muscles and bones and the geometry observed in the subject's medical image data corroborate the quantitative validation.

We validate this approach quantitatively by measuring the distance between automatically and manually defined coordinates of landmark sites.

This approach followed by visual inspection and, if needed, correction to the extracted structures, can dramatically reduce the time required for defining 73 lower limb muscles and bones. Using the proposed method, defining MR-based musculoskeletal models can become a more time efficient and more accurate alternative to rescaling generic models.

2.4.2 Basic Method

The method developed at SAG and used on healthy children's MRI images underwent special adaptations for each disease group. Figure 17 illustrates the basic pipeline, which consists of the following steps:

1. First, one atlas is built for each disease group (CP, CMT, DMD) and individual meshes are manually annotated for every muscle, bone and skin of the lower limbs (label A in Figure 17).
2. Second, landmarks for the joints are detected in each target patient, and will be used to separate both legs into individual, non-articulated segments.
3. For each leg segment, personalized affine transformations will be computed to map the specific disease atlas segment to the corresponding segment in the target image.
4. All affine transformations are then combined in a multi-affine approach, whose result provides both plausible and quite reliable pose information about the target (label B in Figure 17).
5. The result of the previous step is used to initialize non-rigid deformation, whose result will model more precisely the target individual's musculoskeletal anatomy, body-fat compositon, etc. (labels C and D in Figure 17).

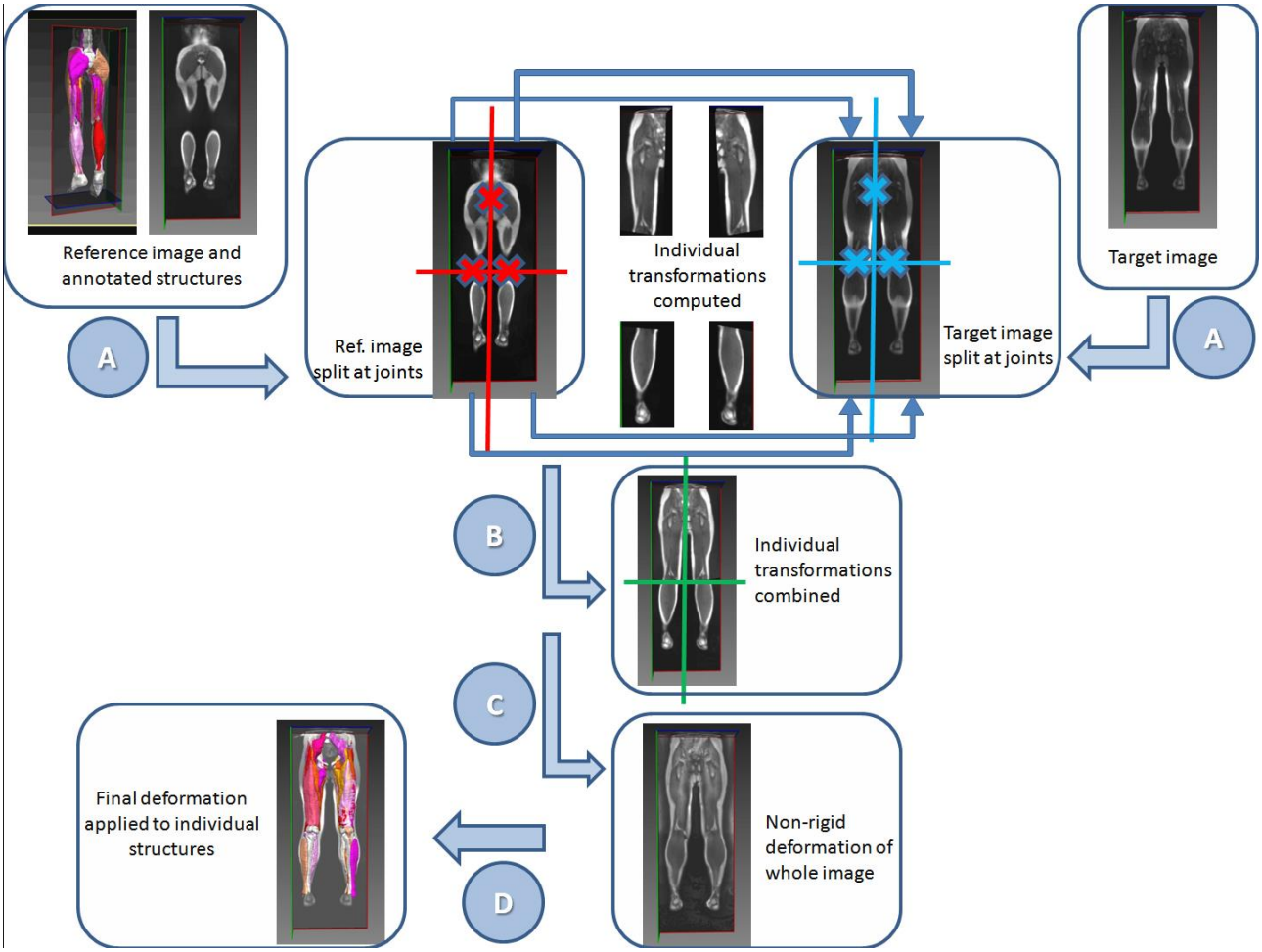


Figure 17: *Overview of the bone, muscle and skin extraction method (SAG).*

2.4.3 Disease-specific adaptations of the method

Extensive tests were performed to assess the viability of the original atlas used with healthy children's images in pathological cases.

The tests revealed that an atlas used for these purposes must be flexible enough to accommodate disease-specific deviations, such as femoral version.

The method that we proposed in D11.1 was finely adapted to each of the 3 disease groups (DMD, CMT and CP), as described in D11.2 and D11.3. We created disease-specific atlas with detailed corresponding structures and adapted the structure extraction methods according to each disease group's specificities.

The original method's parameters were also extensively tuned for each group of patients. The evaluation results are shown in the Results section.

2.4.4. Application

We manually annotated one reference patient image for each disease group, and applied our approach to a set of 3D MRI scans of paediatric patients in the ages of to 8 to 15 years old.

For a quantitative measure of the results, a set of 25 landmarks has been annotated on all patients. The distance between these landmarks in the target image and those in the transformed reference is computed.

The quality-test landmarks are listed in **Table 1** below.

l_FemurMiddle
r_FemurMiddle
l_FemurCaput
r_SkinVentralPartCondylusOfFemur
l_SartoriusVentralPartCaputOfFemur
l_SkinVentralPartCondylusOfFemur
l_QuadricepsVentralPartMiddleOfFemur
l_TricepsDorsalPartCaputOfTibia
l_FemurCondylusLateralis
l_QuadricepsLateralPartMiddleOfFemur
l_PatellaUpperPart
l_GluteusMaximusUpperPart
l_TricepsDorsalPartCondylusOfFemur
r_FemurCondylusLateralis
l_AchillesTendonBottomOfTibia
r_GracilisMedialPartMiddleOfFemur
l_TibiaCaput
l_SkinDorsalPartMiddleOfFemur

r_SkinLateralPartCaputOfTibia
l_FemurCondylusMedialis
l_TibiaBottom
r_FibulaMiddleOfTibia
l_GluteusMaximusInOneLineWithCaputOfFemur
r_PatellaUpperPart
l_GracilisMedialPartMiddleOfFemur
l_FemurMiddle
r_FemurMiddle
l_FemurCaput
r_SkinVentralPartCondylusOfFemur
l_SartoriusVentralPartCaputOfFemur

Table 1: Landmarks used to measure registration quality at SAG.

The distances between the landmarks have been measured and the overall results on these landmarks, as well as the results per disease group, are shown in **Figure 2**

2.4.5 Results of validation

The overall landmark distance was 8.298 mm.
The average per disease group was 8.406 for CMT patients, 7.246 for CP patients and 9.118 for DMD patients.

Overall, the method seems to perform quite well around the femora, as can be seen both quantitatively and qualitatively in **Figures 5 and 6**.

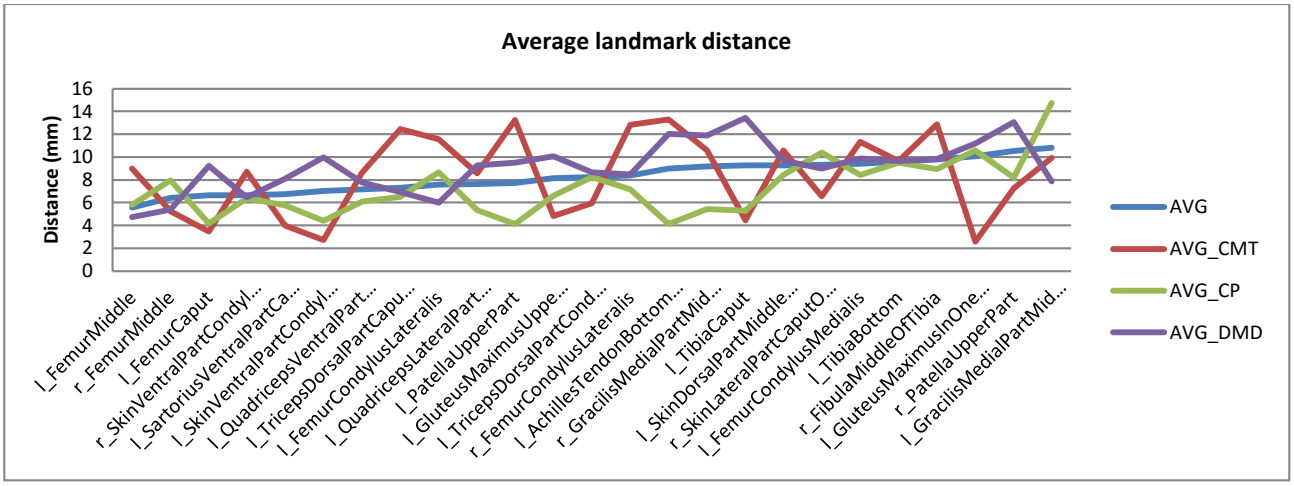


Figure 18: Landmark distance averages per disease group and overall.

Figure 19 illustrates some qualitative resultson different patients.

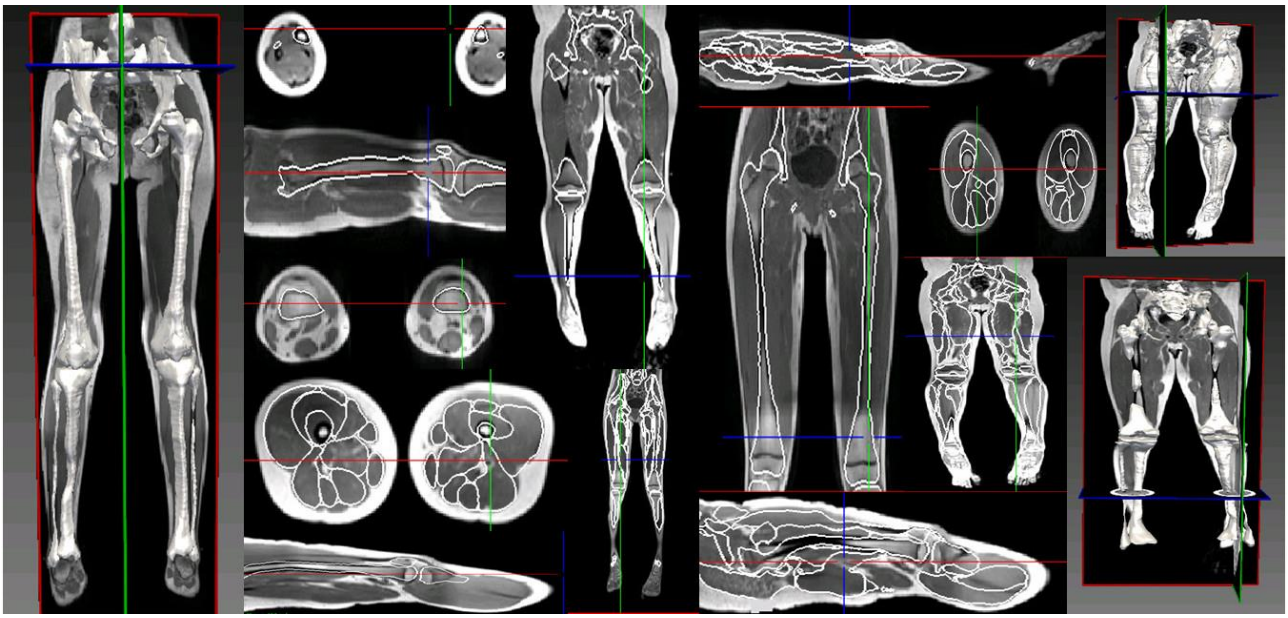


Figure 19: Qualitative results on children with CP, CMT and DMD.

2.4.6 Validation of the complete anatomical model

As presented in deliverable 11.2 and 11.3, from the segmented bone and muscle models, the geometric parameters required for personalizing the biomechanical model must be extracted. This

extraction requires some estimations based on features of the bones and muscles geometry. The most demanding parameters are the origins, insertions, and lines of action of the muscles and tendons. These attachment points are not directly visible in the MR image modality acquired for this project. Thus, USFD designed a strategy to extrapolate their location from the segmented geometries. First, a complete anatomical template model including bony landmarks, ligament and muscle-tendon attachements and lines of action, in addition to bones and mucles geometry, was built. Second a mesh morphing technic, developed by USFD for the STREP project MySpine (http://cordis.europa.eu/project/rcn/97394_en.html), was adapted to the deformation of the template to match the segmented geometries, generating the patient-specific complete anatomical model. The validations performed for the assessment of the accuracy of both elements are presented below.

2.4.7 Complete anatomical template model

The template model was generated by completing the MRI-extracted geometries of the healthy child used as atlas for the MRI segmentation. The bony landmarks, joint centres, ligament and muscle-tendon attachments and lines of action were adapted from the TLEM2.0 model (Carbone et al., 2015), which is publicly available for research from the TLEMsafe project webpage (http://www.utwente.nl/ctw/bw/research/projects/TLEMsafe).

The TLEM2.0 data provided includes the bone geometries, landmarks, and lines of actions, but not the muscle geometries. Because of this, the morphing of the integrated TLEM2.0 model to the geometries of the MRI atlas child was expected to present some anatomical arrors in the resulting attachments and lines of action.

The resulting model was revised for possible anatomical errors. The attachments and lines of action were visually inspected for each individual muscle. Following standard anatomical atlas and the clinical advice of VUMC, several anatomical errors were indentified. They were then manually corrected to follow the expected configuration with respect to the involved bone and muscle geometry. Figure 4 illustrates these manual corrections with three examples. Two iterations were performed for this correction and feedback process. After them, the final template was evaluated and accepted as anatomically correct by the clinical parterns in VUMC.

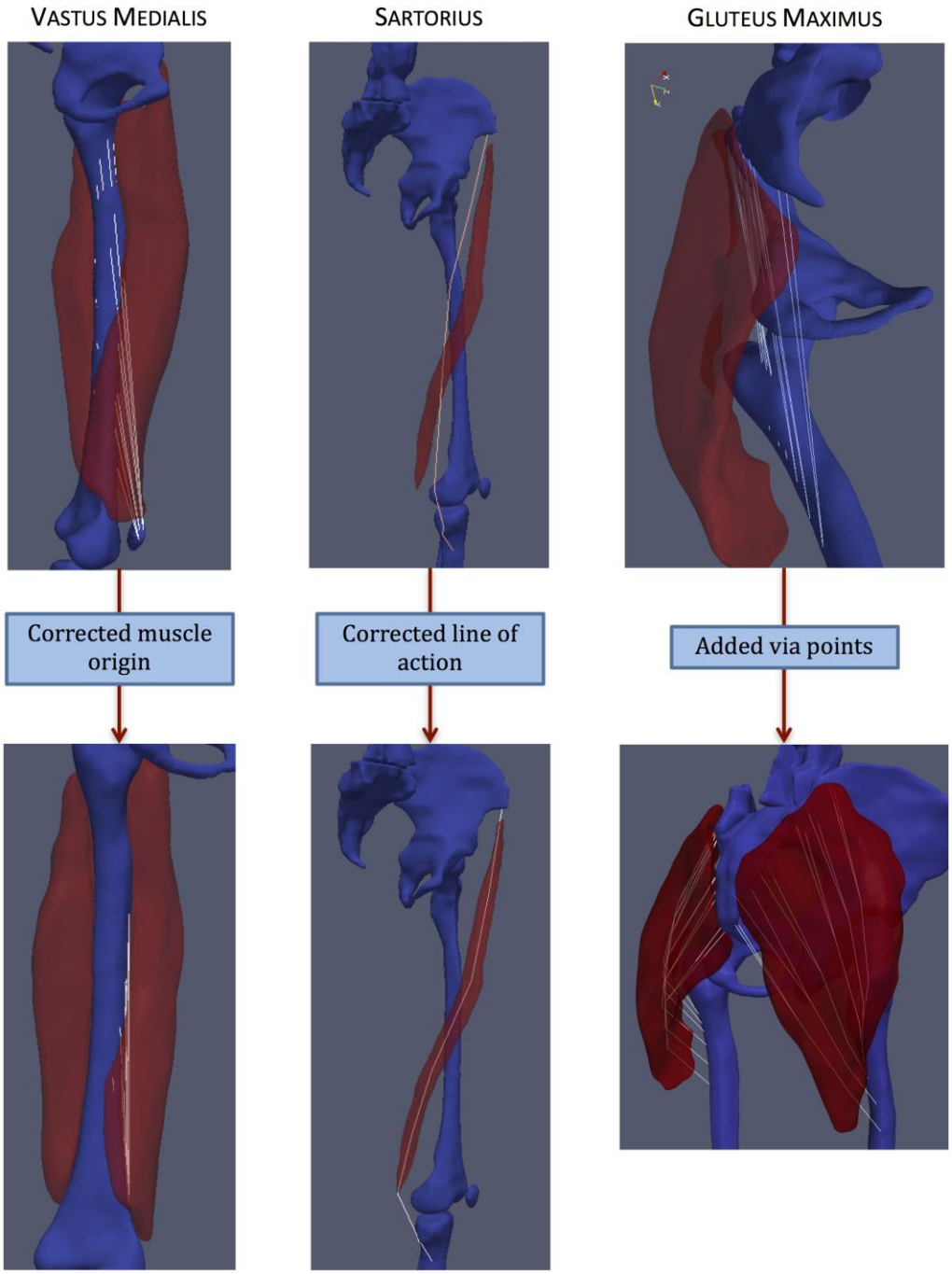


Figure 20: Three examples of the manual corrections of muscle-tendon lines of action performed during the evaluation of the anatomical accuracy of the complete MDPaedigree template for the lower limbs.

2.4.8 Mesh morphing for patient-specific complete anatomical model

The patient-specific complete anatomical model is obtained by mesh morphing of the complete template to the segmented geometries. The applied mesh-morphing algorithm has been adapted

from the algorithm developed for the European STREP project MySpine (http://cordis.europa.eu/project/rcn/97394_en.html). It allows the personalization of volumetric or surface model templates, including the extrapolation or interpolation of structures not directly extracted from the medical images, for biomechanical simulations (Malandrino et al., 2015, Castro-Mateos et al., 2016).

The developed morphing algorithm has been applied to the set of healthy and diseased children for which the MRI has been segmentated until today (Feb. 2016). This includes cases from the 3 diseases considered: Cerebral Palsy (CP), Charcot-Marie Tooth (CMT), and Duchene Muscular Dystrophy (DMD). Figure 5 shows three examples of the resulting complete anatomical model, one for each disease group.

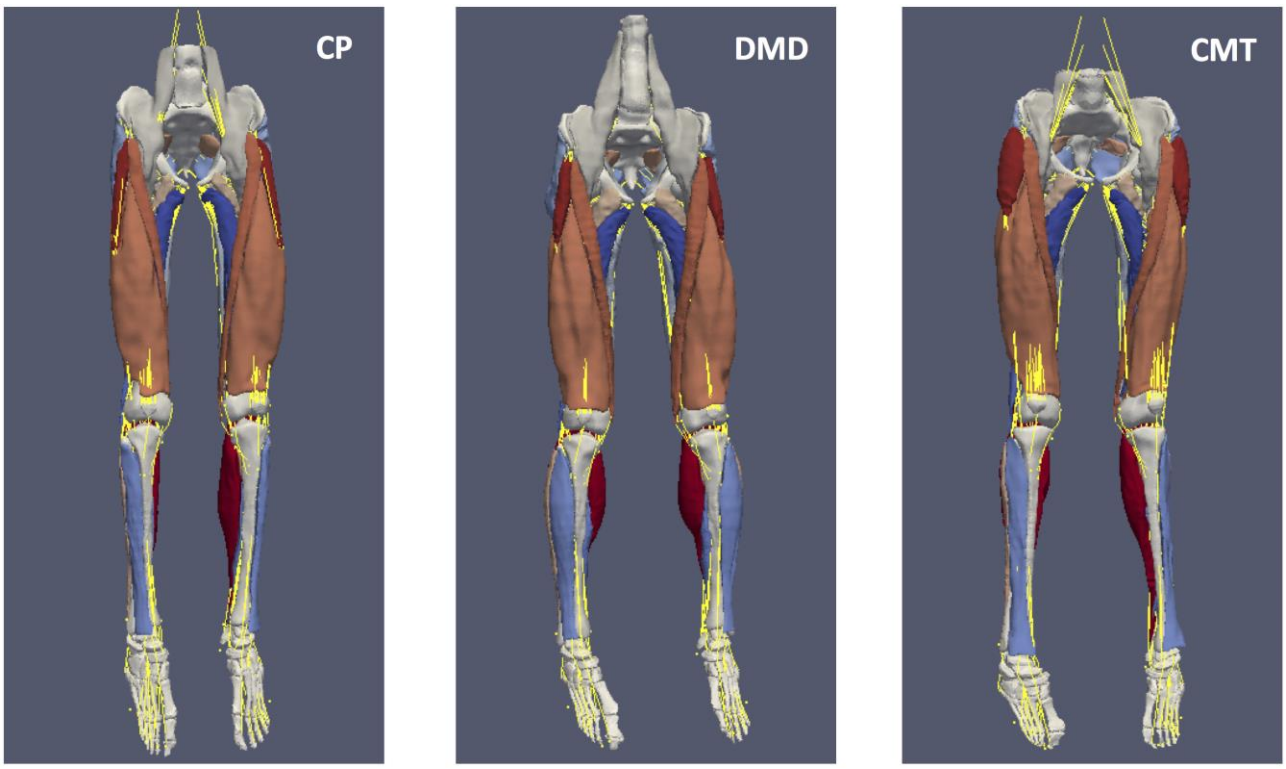


Figure 21: Three examples of patient-specific complete anatomical model, one for each of the diseases (CP, DMD, and CMT). They have been obtained by the morphing of the template anatomical model to the geometries segmented from MRI. Detailed visual assessment was performed for all the cases.

All the cases were visually assessed for its qualitative correctness. No problem was found in any of the cases in the final processing. Some problems detected for a pair of cases, actually revealed that the transferred data from SAG to USFD had some format corruption, involving missing shapes or

misspelings. This helped to refine the transference protocol to be more robust. Once the input segmentated model was corrected, the morphed complete anatomical model presented no qualitative error.

Surface-to-surface distance is a standard measure of the similarity between two surface models. In order to quantitatively measure the fiting accuracy of the patient-specific morphed template to the originally segmented geometries, the surface-to-surface distance between both shapes has been computed for all the cases. Figure 6, displays the distribution of the surface-to-surface distance across the complete model stratified by disease group. No important differences are observed. The morphing seems to give slightly larger error for CP and CMT children than for DMD children, which have a similar behaviour than the healthy volunteer. In any case the errors are of the order of 1mm, which are negligible in comparison with the segmentation errors.

Even if the global errors are small, it could be still possible that some individual bone or muscle had a larger error. For this reason, the mean surface-to-surface distance has been computed per anatomical element. Figure 7 shows these errors per anatomical element, also stratified by disease group. No important difference is observed for any disease or anatomical error. The only point to remark is the relatively larger error for the gracilis muscle in the volunteer child. However, even for this case, the mean error is around 1mm.

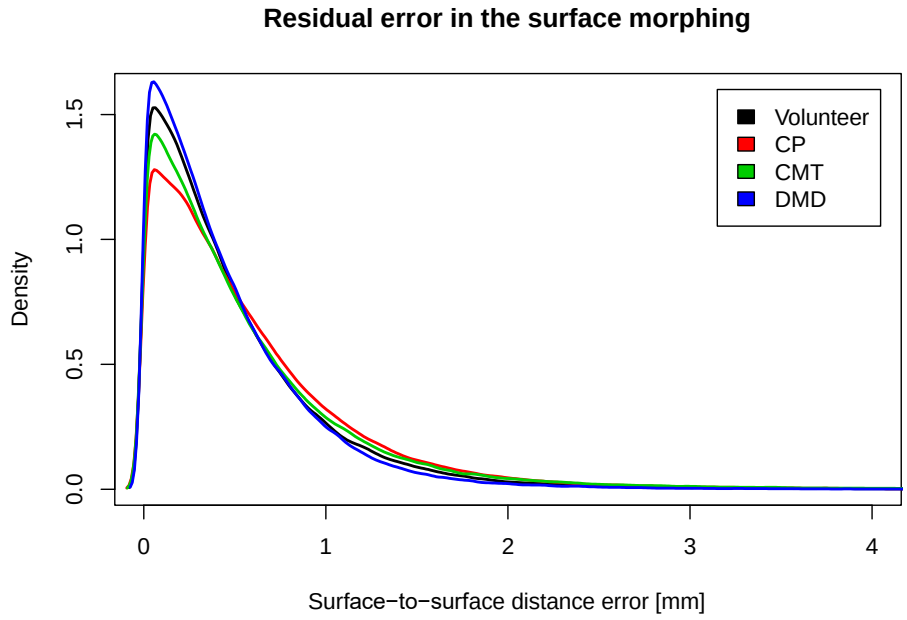


Figure 23: Distribution of the surface-to-surface distance error between the morphed model and the segmented geometries, across the complete model. The error has been stratified by disease group..

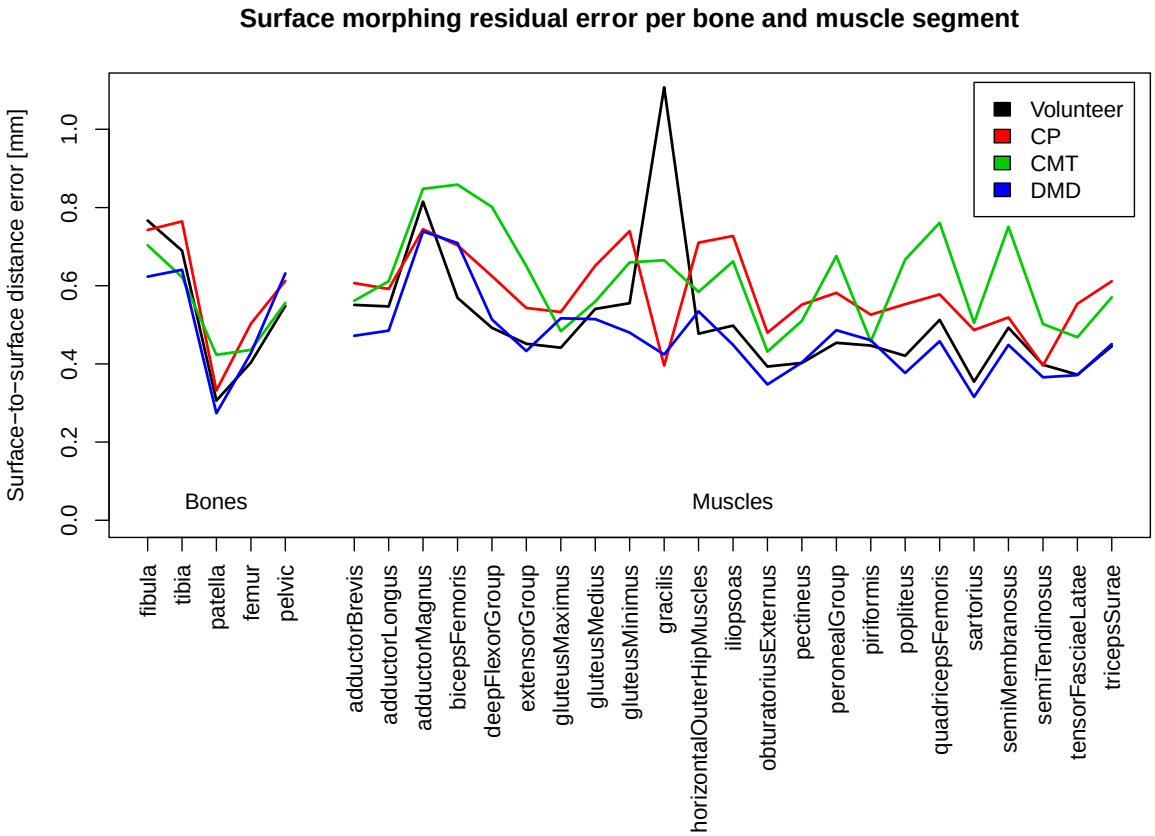


Figure 24: Mean surface-to-surface distance error for each individual bone and muscle, between the morphed model and the segmented geometries. The error has been stratified by disease group.

From these results we can conclude that the morphing algorithm is able to accurately personalize the template complete anatomical model to the patient-specific bone and muscle geometries segmented from MRI.

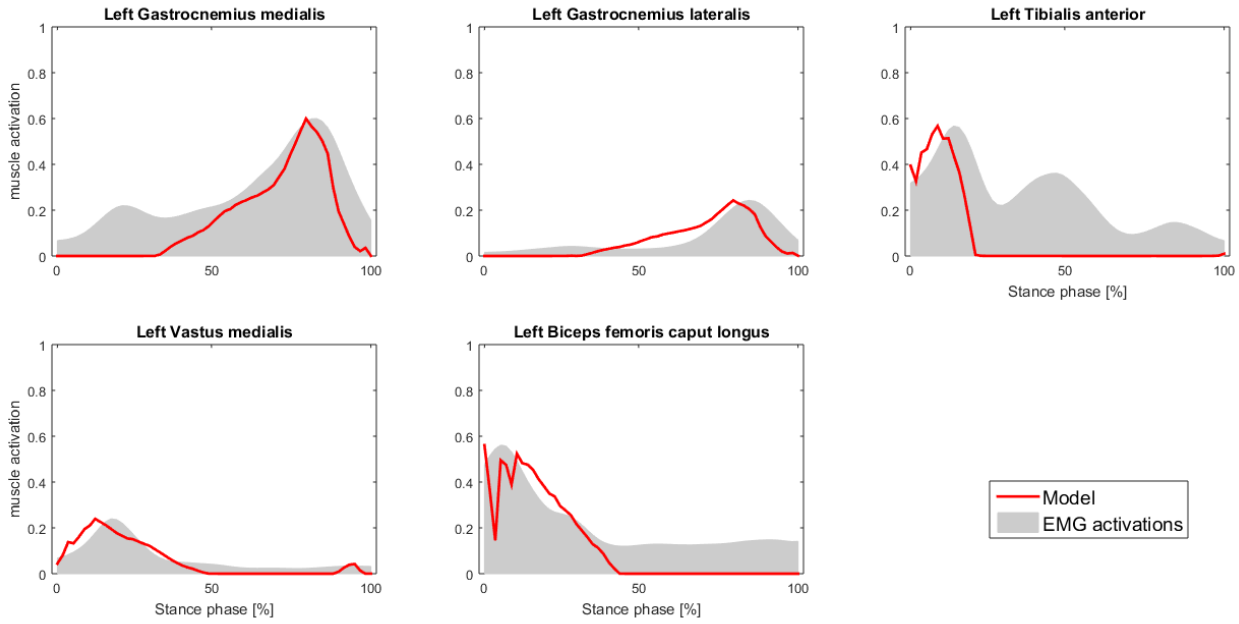


Figure 25: Comparison of EMG signals and computed muscle activation for one walking trial at month-6 (left side).

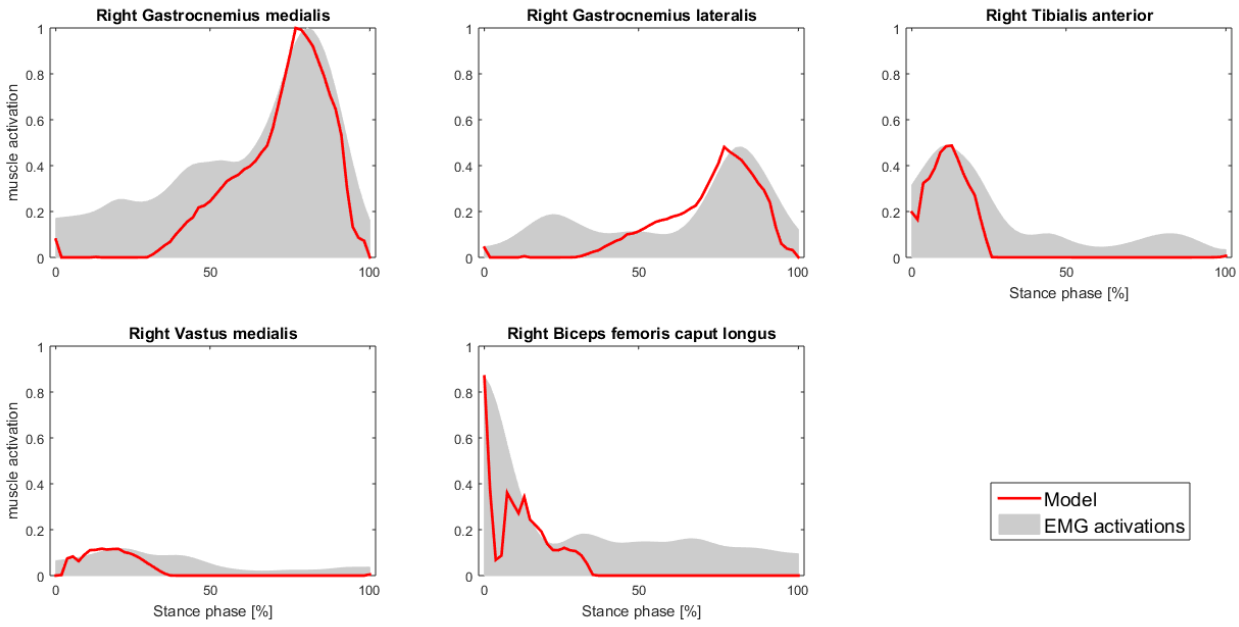


Figure 26: Comparison of EMG signals and computed muscle activation for one walking trial at month-6 (right side).

3. Future work

At current status, retrospective prediction in all four disease areas is close to completion. Delivered models in the four disease areas have been derived and tested in a substantial number of patients included in the study sample, and as detailed in the present document the validation process has so far demonstrated that the current model provide significant accuracy and robustness in the definition of the different diease. The models are now ready to be tested in the whole study population to provide both patient specific models (gained in almonst all disease areas) and to expand its ability on the patinet specific prediction, mainly thanks to the nearly completed data acquisition process and also to the increasing number of patients uploaded into the digital repository.

As detailed in the WP12, the steps of the validation have been defined in order to guarantee a uniform validation approach. The single validation steps for each disease area include a number of processes and analysis which include:

1) Initial testing and debugging of the mechanistic model to assure that the models comply with the clinical user requirements and that the information obtained from the models provide clinically useful information and predictions.
2) Internal validation including:
    a. Data selectivity and specificity, to define to the extent to which it can determine key factors in a complex mixture of data without interference from other components.
    b. Accuracy of detection, related to the reproducibility and repeatability of the models, in order to verify that repeated measurements under unchanged conditions provide the same results
    c. Limits of detection and quantification, in order to establish the lowest amount of useful data that can be used by the models to provide an accurate and precise prediction of single disease definition and progression.
3) External validation

A schematic representation of the status of validation can be seen below:

	Initial debugging	Data specificity	Internal validation	Accuracy	Limits of detection	External validation
CMP-PS						
CMP-PM						
CVD-PS						
CVD-SM						
JIA-PS						
JIA-PM						
NM-PS						
NM-PM						

CMP=cardiomyopathies; CVD=cardiovascular disease; JIA: juvenile idiopathic arthritis; NMD=neuromuscular diseae; PS= patient specific model; PM= predictive model; SM=statistica model. Green square= achieved, Orange square= achievement in progress, Blu square= to be completed during the last year of the study.

Accordingly the validation process in the fourth and last year of the project will focus on patients prediction validation and to the external validation process, as it is well known that the prognostic prediction is almost always better on the data set on which the model has been constructed (learning series) compared to the performance of the same model on new data (testing series).

During the fourth year of the study the WP12 will also focus on the definition of revised clinical workflows for the four disease area. Following the work provided by the WP2 in the deliverables D2.2 and D2.3, clinical pathways will be modified and revised accordingl to the validated impact of models in patient specific disease definition and patient specific disease prediction. Clinical workflow will integrate changes in current workflows derived from the validated potential impact of therapeutic intervention effect.

REFERENCES

1) Castro-Mateos I, Pozo JM, Lazary A, Frangi AF (2016) Automatic Construction of patient-specific biomechanical models of the spine from IVDs and vertebral segmentations. In SPIE Medical Imaging. International Society for Optics and Photonics.

2) Malandrino A, Pozo JM, Castro-Mateos I, Frangi AF, van Rijsbergen MM, Ito K, Wilke HJ, Dao TT, Tho MCHB, Noailly J (2015) On the relative relevance of subject-specific geometries and degeneration-specific mechanical properties for the study of cell death in human intervertebral disk models. Frontiers in bioengineering and biotechnology 3.

3) Arnold, E., Ward, S., Lieber, R. and Delp, S., 2010. A Model of the Lower Limb for Analysis of Human Movement. Annals of biomedical engineering 38, 269-279.

4) Prinold, J. I., Mazzà, C., Di Marco, R., Hannah, I., Malattia, C., Magni-Manzoni, S., Petrarca, M., Ronchetti, A., Tanturri de Horatio, L., van Dijkhuizen, E. H. P., Wesarg, S. and Viceconti, M., 2015. A Patient-Specific Foot Model for the Estimate of Ankle Joint Forces in Patients with Juvenile Idiopathic Arthritis. Annals of biomedical engineering 1-11.

5) Procter, P. and Paul, J. P., 1982. Ankle joint biomechanics. Journal of Biomechanics 15, 627-634.

6) Valente, G., Pitto, L., Testi, D., Seth, A., Delp, S. L., Stagni, R., Viceconti, M. and Taddei, F., 2014. Are Subject-Specific Musculoskeletal Models Robust to the Uncertainties in Parameter Identification? PLoS ONE 9, e112625.