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 12.2.1

First clinical assessment and validation results for all four disease areas

Due date of delivery: 28th February 2015

Actual submission date: May 15th, 2015

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

Partner responsible for this deliverable: OPBG

Version: 1.3

D12.2.1 - First clinical assessment and validation results for all four disease areas	MD-Paedigree - FP7-ICT-2011-9 (600932)

Dissemination Level:

Document Classification

Title	First clinical assessment and validation results for all four disease areas
Deliverable	12.2.1
Reporting Period	Second
Authors	G. Pongiglione,
Work Package	WP12
Security	RE
Nature	R
Keyword(s)	Clinical validation

Document History

Name	Remark	Version	Date
G. Pongiglione		V1.0	
G. Pongiglione		V1.1	
G. Pongiglione		V1.2	
G. Pongiglione		V1.3	06/04/15

List of Contributors

Name	Affiliation
Tobias Heimann	SAG
Tommaso Mansi	SCR
Olivier Ecabert	SAG
Pieter van Dijkhuizen	IGG
Claudia Mazzà	USFD
Maria Costa	SAG
Enrico Castelli	OPBG
Alex Jones	UCL
Lorenza Putignani	OPBG
Paolo Cappa	URLS

List of reviewers

Name	Affiliation
Edwin Morley-Fletcher	OPBG
Bruno Dallapiccola.	OPBG

Table of Contents

Index of Tables and Figures

Tables

Figures

1. Introduction

Validation is the act of demonstrating and documenting that a procedure operates effectively. Process validation is the means of ensuring and providing documentary evidence that processes (within their specific design parameters) are capable of consistently producing a finished product of the required quality. The aim of the present document is therefore to report on the progress of the validation process started on month 18. As stated in the section 3.1 of Deliverable 12.1 (Introduction and General principles) of the General Outline of the Clinical Assessment and Validation, the validation process must necessarily follow a well-defined logical and chronological course to demonstrate the validity of what has been accomplished. A stepwise approach must therefore be followed and the completion of each step will be preliminary to any further analysis.

At the present state of advancement of the project only the mechanistic model can be addressed by the validation process since it can be accomplished with the amount of data accumulated so far while the validation of the statistical model requires the full upload of the data and a different approach involving features not yet implemented in the system. This part therefore will not be addressed in this document and will be reported in the deliverables due at month 36 and 48.

2. The validation process approach

As previously stated, the validation process must follow a multi-step approach:

1. Partial validation

The first level of validation has consisted in the initial testing and debugging of the mechanistic model. This part is performed mainly by the technical partners in collaboration with the clinicians. It consists in the incorporation of the clinical user requirements as previously defined for each disease area (Deliverable D2.1 Initial requirements analysis document including priorities for the implementation) into the model and the verification that the model itself responds correctly to the modification of one or more variables.

Since the creation of a model is a dynamic process, it was initially foreseen that the User Requirements might be changed in the course of the project, in order to create a more and more functional system. In the section related to each disease area it will be specified whether the user requirements needed to be modified.

Not necessarily at this stage of the project a fully developed model may be available for testing since only a part of the variables identified by the user requirements may have been embedded in the system. In such a case a partially functional model may be available with limited capabilities. Nonetheless, even in this case a limited initial validation can be performed and it is still possible to proceed to the next step.

2. Internal validation

Preliminary to the internal validation the model must be personalized and therefore must be transformed into the model of a specific patient. Eventually, this crucial part will be performed by the clinicians but at present is still completed by the technical partners.

The personalization of the model consists in acquiring the variables that characterize a patient by overlapping the generic model on patient specific MRI or Echocardiographic images. Main characteristic of the internal validation is that it is performed on the same patients utilized to build the model.

Internal validation is divided in two parts:

2.1 Retrospective prediction

This step requires the data from only a few patients, uploaded from one or more clinical institutions at the beginning of the project and for whom a full longitudinal follow-up is present. Clearly, only patients in whom all the variables embedded in the mechanistic model have been measured and entered into the system and available for the personalization of the model can be utilized for the purpose.

On the personalized model the action taken on the patient (i.e.: implant of a cardiac valve, drug manipulation of preload or afterload, etc.) are simulated and the output of the model is tested against the real outcome of the treatment.

2.2 Prospective prediction

A patient for whom a follow up is not yet available is used for personalization and prediction. Obviously, only Good Clinical Practice (GCP) will be the base for any clinical decisions with no consideration for model prediction. A bigger number of patients is necessary for a meaningful prospective prediction as a consequence of the variability among patients and of the effects of the treatment on each of them.

3. **External validation (on patients not utilized to build the model)**
4. **Full-scale validation**
5. **Cross-validation.**

At this time of the project, **points 2.2, 3, 4, and 5 are premature** and will be reported at month 36 and 48 respectively.

3. Accuracy of the prediction

By this term we define how close the predicted modification of any variable is to the real outcome of the patient. The term “accuracy” strictly relates to the concept of “similarity”. The comparison of a value, either from clinical observation or from computer prediction, with the mean value from a normal population can be done in terms of the so called Euclidean distance or in terms of standard deviations from the mean (z value) as it has been elucidated in section 4.4.3 od D12.1. The fact that the project deals with a pediatric population and the observation that in the human body there are variables that can vary much (for instance the anthropometric data) and others that can vary very little (for instance body temperature) emphasize the importance of the z value as the criterion of choice for similarity.

The range of z -1 to z +1 includes more than 50% of the normal population and therefore is a relatively mild criterion of similarity but with which a sufficient number of observation can be included. The evolution of the system and the upload of a great number of cases will allow the narrowing of the search criteria to smaller values of z.

Since the relation that links the variables one another in a model is purely mathematical, the model can then be adjusted in order to obtain the best fit between the pre-existing conditions, the actions taken and the real outcome. If a partially functional model has been utilized, the prediction will necessarily be partial. To estimate model's optimism in internal validation, bootstrapping techniques have been used, that involve taking a large number of samples with replacement from the original sample. Bootstrap method have been shown to be very efficient since the entire data set is used for model development and no new data have to be collected for validation, providing nearly unbiased estimates of predictive accuracy that are of relatively low variance (S.E. Bleeker et al. External validation is necessary in prediction research: a clinical example. Journal of Clinical Epidemiology 56;826:2003).

4. Validation process for each disease area

4.1 Validation for Cardiomyopathies models (WP3-WP8)

The process of validation in the WP3 has started. As reported in the D2.1 the clinical user requirements for the mechanistic model of WP3 have been delivered at month 12 to the technical partners. A detailed list of general and specific clinical expectations for the model have been assessed. In addition, a detailed list of specific variables to be tested has been prepared. In brief the clinical partners expect that for the effect of each of the independent variables - being these the effect of time (i.e. prognosis) and the effect of specific cardioactive medications (affecting either circulating volume, heart rate and/or cardiac afterload)- the model is able to provide (predict) specific changes in the dependent variables provided.

During this second year of the project, as the cardiac model of cardiomyopathies is being implemented, the technical partners have been able to provide relevant feedback to the clinical requirements, defining a specific list of parameters predictable by the model. In addition, a list of tasks, which the model will not be able to predict, has been prepared. The specific list derived from the technical feedback is provided here below, and detailed in the current version of the D2.2.

Patient-Specific Computational Heart Model, Siemens

The following is a list of input and output model parameters based on model design. The accuracy of output parameters and the unicity of input parameters needs to be validated and depends on the available data, model assumptions and the clinical question under consideration (which drives the model assumptions and design).

Model Parameters to Estimate (Input)

The following parameters need to be estimated from images and clinical data in order to obtain a patient-specific heart model.

Anatomy

- Heart shape (from images, using segmentation)
- Fibers (model-based or atlas)
- Tissue substrate like scar, border zone and fibrosis (from images, using segmentation)

Dynamics

- Kinematics (from images, using tracking)

Electrophysiology

- Tissue electrical diffusivity (or conductivity) (from 12-lead ECG or invasive mapping if available)
- Action potential duration (from 12-lead ECG or invasive mapping if available)
- Restitution curve (from 12-lead ECG or invasive mapping if available)
- Heart rate

Biomechanics

- Tissue stiffness (from dynamics and pressure information)
- Maximum active stress (from dynamics and pressure information)
- Relaxation and contraction rates (from dynamics and pressure information)

Hemodynamics (blood flow and whole-body)

- Systemic and Pulmonary Arterial and Venous compliance, resistance (from dynamics, measured volume and flow, and pressure information)
- Valve properties (resistance, inertance, from dynamics, measured volume and flow, and pressure)

Computed Model Parameters (Output)

Anatomy

- Wall thickness, ventricular mass, chamber diameters, sphericity. Baseline directly from images. Acute from model.

Dynamics

- Strain, velocity, displacement. Baseline directly from images. Acute from model prediction.

Hemodynamics (Blood Flow and Whole-Body)

- Ejection fraction, stroke volume, EDV, ESV
- Compliance and elastance

Table 1. Extracted from D2.2 "Revised Clinical Requirements Analysis Document", Appendix.

INPUT VARIABLE	EXPECTED MODELLED OUTCOME VARIABLE CAN BE PROVIDED	CANNOT BE PROVIDED
Impact of Time	Left ventricular systolic/diastolic diameters Left ventricular mass Left ventricular volume Left ventricular sphericity index Left ventricular ejection fraction Left ventricular stroke volume Left ventricular mitral valve dimension Mitral inflow early and late velocities Mitral tissue Doppler velocities Right ventricular tricuspid plane excursion Right ventricular ejection fraction	Mitral valve regurgitant fraction Right ventricular systolic pressure Intra-ventricular systolic synchronicity Inter-ventricular systolic synchronicity Inter-ventricular interaction index Percent regional systolic myocardial deformation
Change caused by medications	Left ventricular stroke volume	Inferior vena cava dimensions

affecting heart rate	Left ventricular cardiac output Left ventricular systolic/diastolic diameters Left ventricular sphericity index Left ventricular ejection fraction Left ventricular stroke volume Left ventricular mitral valve dimension Right ventricular ejection fraction Mitral inflow early and late velocities Mitral tissue Doppler velocities Right ventricular tricuspid plane excursion	Inferior vena cava respiratory variation Mitral valve regurgitant fraction Right ventricular systolic pressure Intraventricular systolic synchronicity Interventricular systolic synchronicity Interventricular interaction index Percent regional systolic myocardial deformation
Change caused by medications affecting pressure afterload	Heart rate Left ventricular systolic/diastolic diameters Left ventricular mass Left ventricular volume Left ventricular sphericity index Left ventricular ejection fraction Left ventricular stroke volume Left ventricular cardiac output Left ventricular mitral valve dimension Mitral tissue Doppler velocities Right ventricular tricuspid plane excursion Right ventricular ejection fraction	Inferior vena cava dimensions Inferior vena cava respiratory variation Left atrial diameter Left atrial volume Mitral valve regurgitant fraction Right ventricular systolic pressure

Accordingly, the model will be able to predict a number of changes in left and right ventricular geometry and function and provide predictive information on the expected hemodynamic effect of cardioactive medication, parameters that clinicians believe will be very helpful in everyday management of patients with CMP (as explained more in details in D2.2 and D13.3). However, it is also clear from the table above, that despite the large number of variables that the model is able to define and predict, the model will not predict a number of significant parameters. For example, the lack of information on left atrial parameters (volume and function) as well as the lack of information on inferior vena cava (dimension and respiratory variation) somewhat represent a prospected limitation to the mechanistic model. In clinical practice, both structures (left atrium and inferior vena cava) are used to define volume load (both in case of hypervolemia and hypovolemic hemodynamic conditions). Thus in the setting of cardiomyopathies (a disease characterized by the lack of ability of the heart to deal with increased volume load) this lack of prediction by the mechanistic model, might represent at present a possible setback on model usability and functionality.

It should be underlined, that the analysis of left atrial structures and inferior vena cava is considered relevant in clinical practice, as no other reliable parameters to predict circulating volume have been established yet. However if the model is able to provide information on the effect of changes in circulating volume by the sole analysis of the ventricles, it would still provide the information needed, despite not analyzing neither the left atrium nor the vena cava. Eventually, a possible mitigation approach would be to enrich the mechanistic model with a statistical model including a number of parameters that the model is unable to predict. This statistical model, used in conjunction with the mechanistic model, might prove to be more efficient and reliable than the use of the mechanistic model alone.

The technical and modelling partners have started to build a mechanistic model of the heart in cardiomyopathies. The model is derived from patients acquired in all three clinical centers, to verify that data provided from each of the participating clinical enrollment centers are usable and provide all the needed information. The model personalization process has started accordingly and preliminary analysis, performed by the technical partners has shown that the parameters provided by the model satisfactorily overlap those of the actual patient. Thus, the first step of internal retrospective validation for the mechanical model in cardiomyopathies has been undertaken. As a general plan, for the time being we will accept a concordance of z +/-1 for each variable tested. In case of a higher discrepancy between the predicted and the actual outcome, the relevant model must be readjusted for a better fit.

According to the technical partners (as detailed in D3.2), thirty complete baseline datasets are considered to be sufficient to build the ‘patient specific’ model. As soon as this number has been achieved, an initial retrospective internal prediction has been performed on two patients.

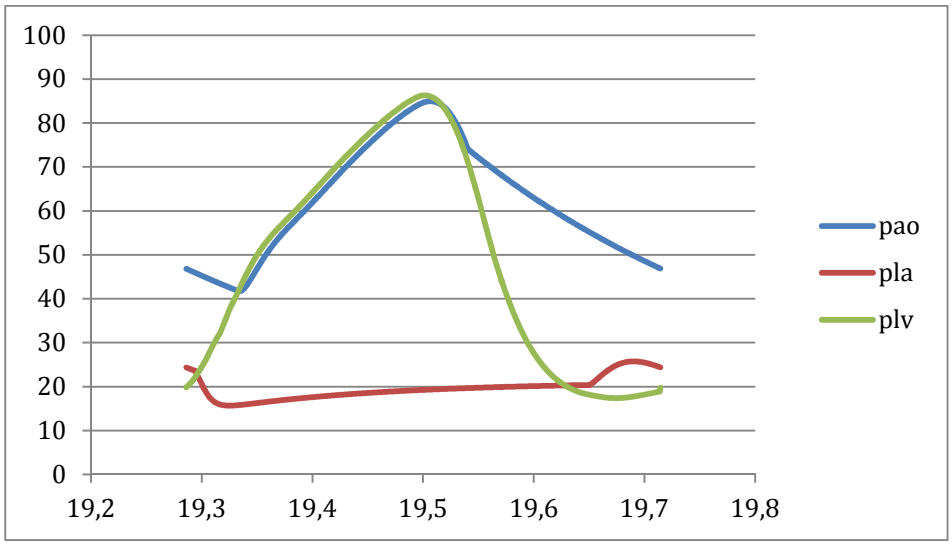
The variable Ejection Fraction (EF) has been retrospectively predicted against the variation of Blood Pressure (BP) and Heart rate (HR) that have been pharmacologically manipulated between baseline and the follow up assessment.

OPBG Patient 011

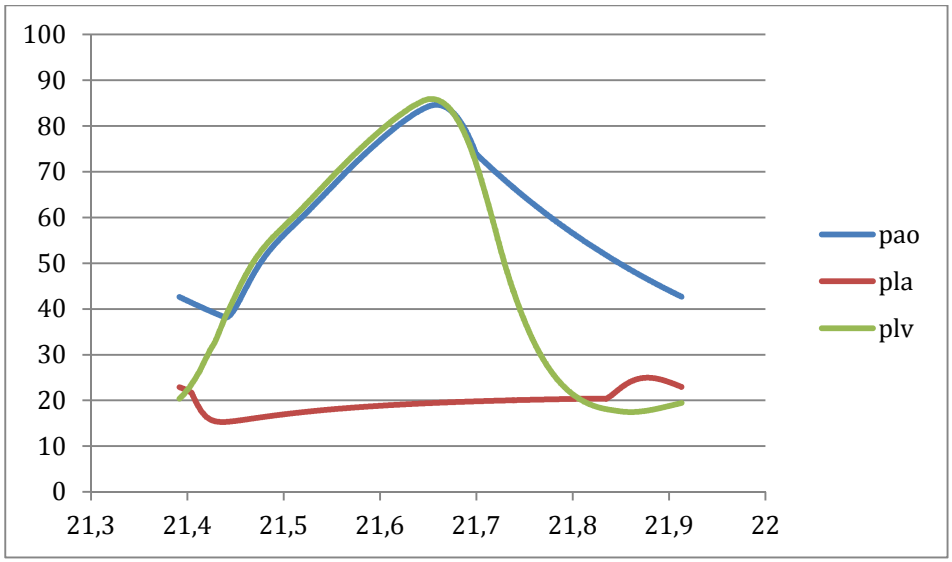
Clinical data:

- baseline HR: 140bpm; BP 90/40mmHg
- follow up HR: 115bpm; BP 111/42mmHg

	Clinical baseline	CFD haemodynamics baseline (personalization)	Clinical follow-up	CFD Haemodynamics follow-up (computation)
EF(%)	30	31.1	40	35



Baseline



Follow up

pao = Pressure Aorta

pla = Pressure left atrium
plv = Pressure left ventricle

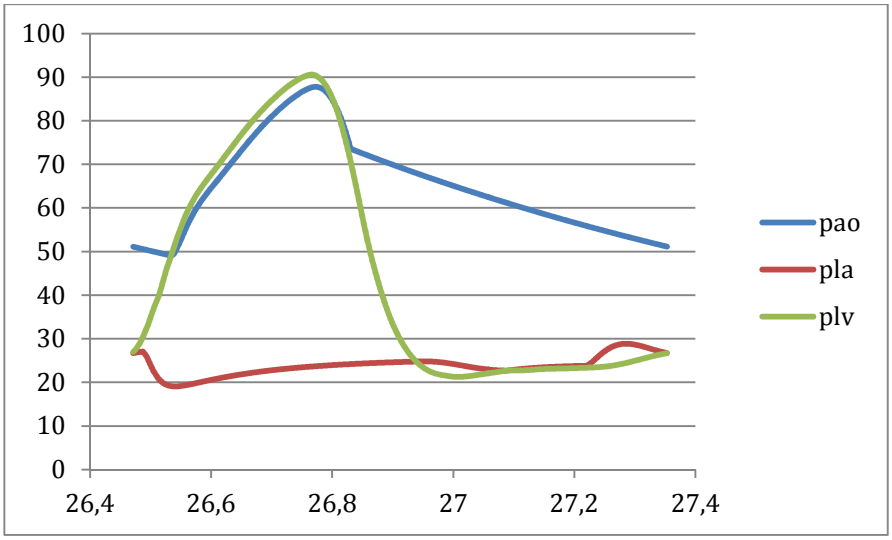
OPBG Patient 019

Clinical data:

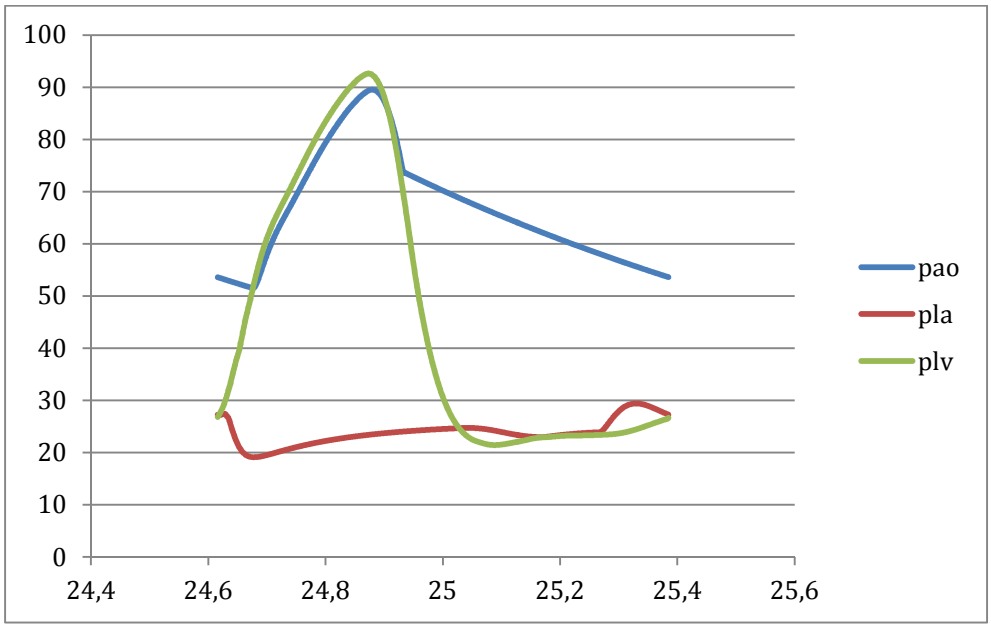
- baseline HR: 68bpm; BP 90/50mmHg
- follow up HR: 78bpm; BP 106/54mmHg

	Clinical baseline	CFD haemodynamics baseline (personalization)	Clinical follow-up	CFD Haemodynamics follow-up (computation)
EF (%)	38	40.5	35	36.8

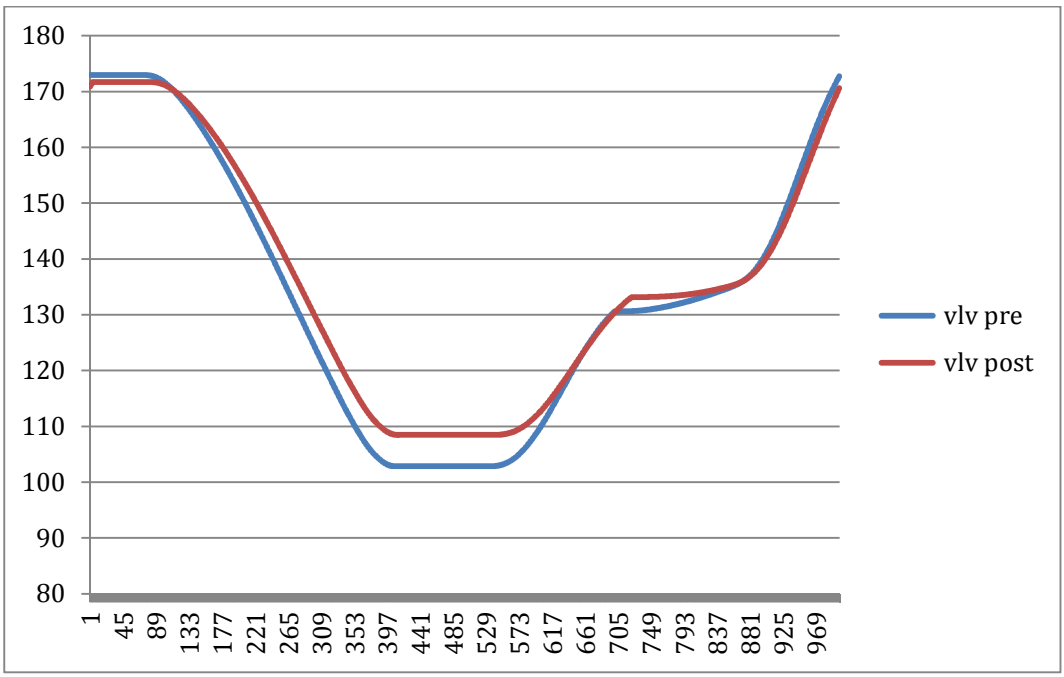
Baseline



Follow up



Change in volume between baseline and follow up



vlv = Volume left ventricle

VALIDATION

Measured and computed trend match

	Clinical baseline	CFD haemodynamics baseline (personalization)	Clinical follow-up	CFD Haemodynamics follow-up (computation)	Measured Trend	Computed Trend
EF OPBG011 (%)	**30**	**31.1**	**40**	**35**		
EF OPBG019 (%)	**38**	**40.5**	**35**	**36.8**		

There is a considerable concordance between measured and computed values that show variations that move in the same direction and with magnitude of the differences within ± 1 z and therefore within the preliminarily accepted accuracy level.

We believe that a significant number of patient datasets enrolled will prove to be very useful for both the internal retrospective validation and a tentative external retrospective validation, providing a highly satisfactory 1:4/1:5 ratio between the learning cohort and the validation cohort.

As already stated, at current, the mechanistic model is not yet been developed to answer questions on prediction and/or outcome. Thus before the collection of follow up data, which are considered by the technical partners as necessary to enrich the current model with predictive tools, no validation has started on this specific, albeit clinically fundamental, task. Nonetheless, the CMP workpackages (WP3 and WP8) are very confident that the validation of the predictive models (once started) will prove to be very helpful and clinically meaningful. It should be noted that compared to the other clinical work packages, the WP3 focuses on a fast developing and life-threatening medical conditions. Thus, it represents the only clinical WP in which hard endpoint and actual events are expected and can be used to corroborate and clinically validate the models. As reported in the D3.2 among the 101 patients enrolled so far at baseline, one endpoint (cardiac transplantation) has already occurred, even before the predicted follow-up time, and more events are expected to occur within the next year, significantly contributing to provide data on actual outcomes to which models' simulations and predictions are to be tested against.

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

Clinical validation of the models of CVD risk of obesity in childhood will be challenging and is likely to be beyond the scope of MD-Paedigree. Such validation typically requires long-term follow-up of large cohorts in order to demonstrate which early biomarkers impact significantly on disease risk. For CVD and its associated metabolic disorders, such as type II diabetes, this process usually takes several decades. The sample size in MD-Paedigree is too small for such a process to be studied successfully. Therefore, immediate clinical implications of CVD modeling in the project are likely to be limited. However, understanding and determination of CVD risk can be considered a step-wise process. At an adult level, primary prevention focuses on intermediate risk factors that have established links with later CVD, such as hypertension and hyperlipidemia. Such intermediate factors are strongly associated with obesity even in childhood. Therefore, early biomarkers identified in WP4 may be examined for their relationship with such intermediate outcomes e.g. blood pressure or cholesterol level, as a means to assess their potential importance in the causative pathway of CVD. This is a necessary step in defining the primordial prevention of CVD that has been called for by the American Heart Association recently. Once biomarkers of the preliminary physiological disorders in early CVD development are identified in projects such as this one, they can then be tested in larger longitudinal studies. Thus, it is anticipated that MD-Paedigree will contribute to a better definition of the earliest processes in CVD disease development.

Although progress with data collection in WP4 is now well underway, difficulties with data sharing and data processing are only just being overcome. This, together with a primary focus amongst the clinical partners on increasing data acquisition rates, means that few data are yet in a form where early analyses can be carried out to assess clinical relationships. A complete dataset, including follow-up data, will be required to define CVD risk models before reliable statements about their potential clinical value can be made. Acquisition of follow-up data has not yet begun. To offset the expected lag in prospective data provision, UCL made available a retrospective dataset, comprising 88 cases with full body fat MR and cardiovascular MR data comparable to the prospective data in MD-Paedigree. These data have already facilitated progress with the development and early validation of the part of WP9 that focuses on automated body fat compartment identification and segmentation, leading to new data on hepatic fat fraction in the retrospective dataset. Preliminary validation of this approach has been undertaken by exploring expected relationships between hepatic fat fraction and clinical variables such as alcohol intake. This has demonstrated that the atlas-based shape models of the liver successfully identify and segment the liver from MR images drawn from a broad range of body types. Furthermore, Fraunhofer Institute and UCL have developed a new method for estimation of hepatic fat fraction. Hepatic fat fraction is estimated as the modal signal intensity in the histogram of all voxels drawn from within the segmented liver after signal processing of the histogram to smooth it. This has resulted in a robust method that is not disturbed by fat-water swapping, an image artifact that is common in T2*-IDEAL images in the region of the liver adjacent to the diaphragm. Unlike simple measures such as mean or median signal intensity, the new measure is unaffected by varying degrees of this artifact. Thus, we have developed a straightforward data processing pathway for estimation of hepatic fat fraction that is essentially automatic and can be implemented in large-scale studies without the requirement for the time-consuming manual data processing that is the current state-of-the-art. Our demonstration of expected relationships with clinical variables in the retrospective dataset serves as validation of this new approach.

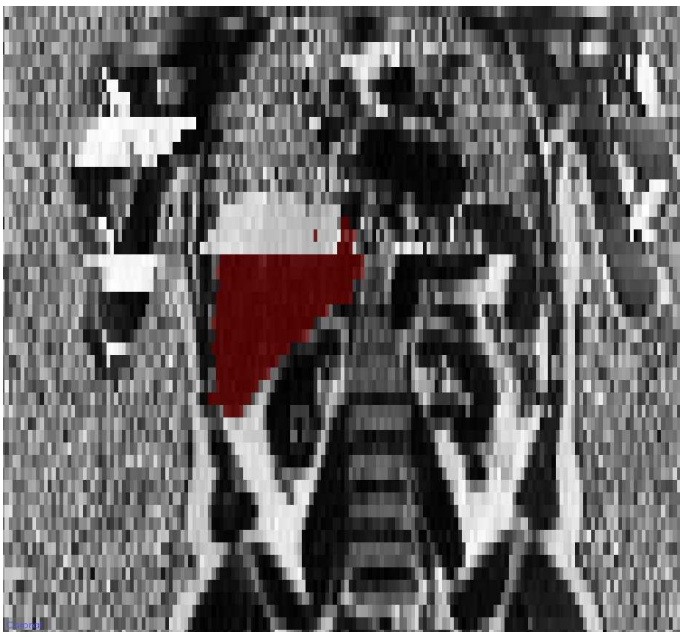


Figure 1. Segmented liver in T2* IDEAL MR data after removal of fat-water swapped part

This figure shows the segmented liver (red) in T2* IDEAL MR data after removal of fat-water swapped parts. It can be seen that the area adjacent to the diaphragm is frequently affected by the fat-water swap artefact.

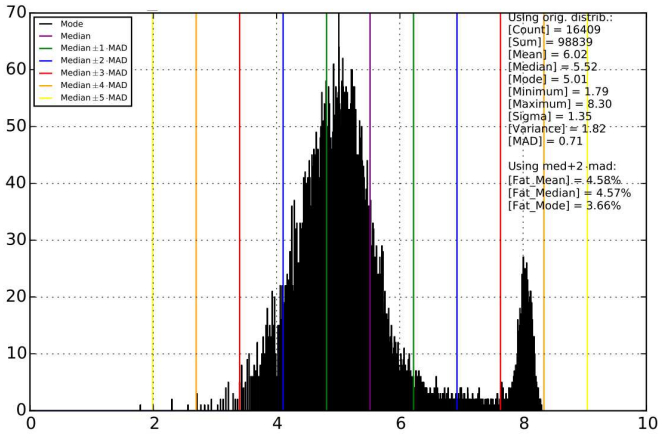


Figure 2. Signal intensities from the automatically segmented liver (1)

This figure shows the histogram of signal intensities from the automatically segmented liver. The second small peak is the result of voxels in the fat-water swap region.

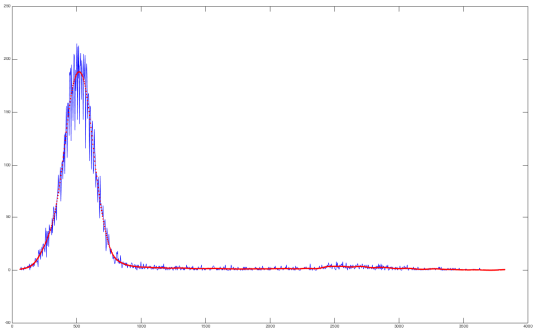


Figure 3. Signal intensities from the automatically segmented liver (3)

This figure shows another histogram of signal intensities from the automatically segmented liver. The red line shows the result of signal processing (smoothing) the histogram. This allows reliable, robust identification of the mode (most frequent signal intensity). This can be transformed into an estimate of liver fat percentage.

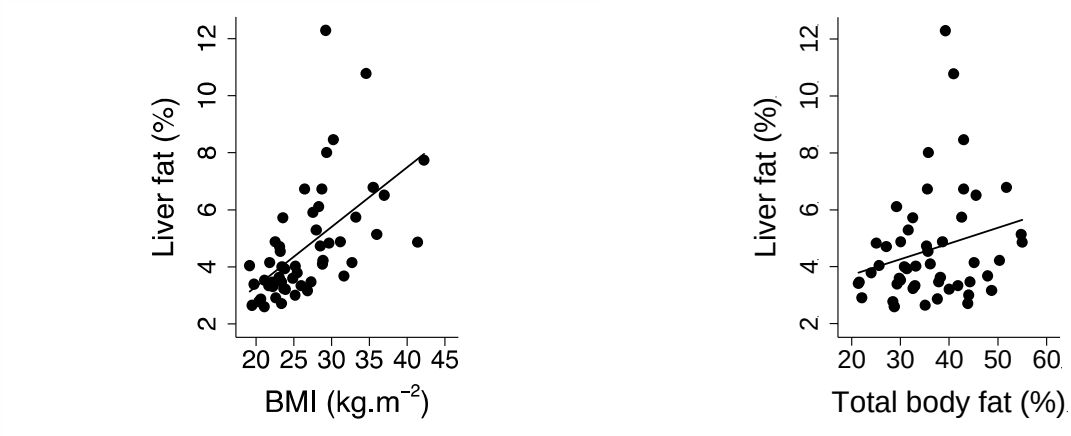
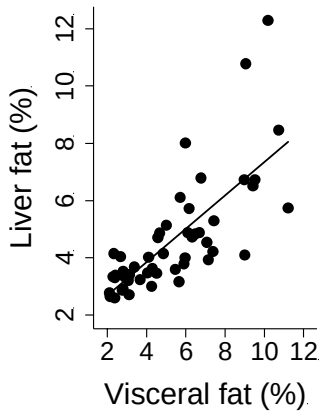
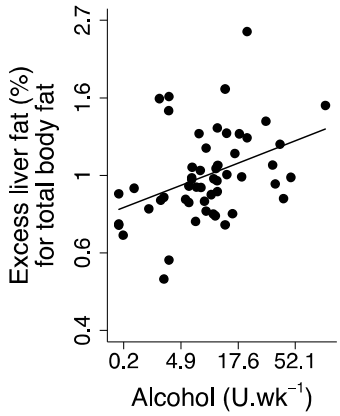


Figure 4. BMI and the liver fat percentage

The figure above on the left shows the relationship between BMI and the liver fat percentage, estimated with the method described above. There is a strong relationship ($r=0.58$, $P<0.0001$), which is also seen with visceral fat percentage ($r=0.71$, $P<0.0001$; figure below left) and to a lesser extent with total body fat percentage ($r=0.23$, $P=0.09$; figure above right).





It is notable that such strong relationships were not seen between these measures of adiposity and liver fat when simpler estimates of liver fat were used (mean or median). For example, the correlation between median liver fat percentage estimates and BMI was 0.15 (P=0.3). This is good evidence that the new method developed for estimation of liver fat using modal estimates after signal processing delivers the most reliable estimates, unaffected by the fat-water swap artifacts. The figure above right shows the relationship between alcohol consumption in the retrospective dataset and liver fat. It is well known that alcohol causes hepatic steatosis but there has not been a robust non-invasive method prior to ours that would allow estimation of hepatic fat fraction in otherwise well subjects. This will allow estimation of early hepatic changes before significant disease arises and is therefore also appropriate for a wide range of conditions that affect hepatic metabolism, including the early effects of obesity in childhood. The relationship shown in the figure demonstrates that significant excess hepatic fat deposition occurs in continuous association with routine alcohol consumption, after adjustment for total body fat percentage (P=0.004).

Other validation work in this project will proceed as data allow. The cardiovascular models will leverage parallel development in WP3. The models will have the following inputs:

Model Parameters to Estimate (Input)

The following parameters need to be estimated from images and clinical data in order to obtain a patient-specific heart model.

Anatomy

- Heart shape (from images, using segmentation)
- Fibers (model-based or atlas)

Dynamics

- Kinematics (from images, using tracking)

Electrophysiology

- Tissue electrical diffusivity (or conductivity) from 12-lead ECG
- Action potential duration from 12-lead ECG
- Restitution curve from 12-lead ECG
- Heart rate

Biomechanics

- Tissue stiffness (from dynamics and pressure information)
- Maximum active stress (from dynamics and pressure information)
- Relaxation and contraction rates (from dynamics and pressure information)

Hemodynamics (blood flow and whole-body)

- Systemic and Pulmonary Arterial and Venous compliance, resistance (from dynamics, measured volume and flow, and pressure information)
- Valve properties (resistance, inertance, from dynamics, measured volume and flow, and pressure)

The cardiac model will provide a range of parameters as below:

Computed Model Parameters (Output)

Anatomy

- Wall thickness, ventricular mass, chamber diameters, sphericity. Baseline directly from images. Acute from model.

Dynamics

- Strain, velocity, displacement. Baseline directly from images. Acute from model prediction.

Hemodynamics (Blood Flow and Whole-Body)

- Ejection fraction, stroke volume, EDV, ESV
- Compliance and Elastance

The model parameters that are of particular interest are those related to LV strain and diastolic function. The earliest abnormalities thought to occur in obesity are alterations in LV strain related to increased cardiac output demands and the pro-hypertensive effects of obesity. The increased volume load also places a filling demand on the heart that is represented by altered diastolic function. Therefore, measures of ventricular elasticity will be central to understanding the early effects of obesity in childhood.

4.3 Validation for the JIA models (WP5-WP10)

The JIA work package (WP5 and WP10) aims to develop a patient-specific biomechanical ankle model to estimate joint loading (joint reaction force; JRF). Standard modelling practice involves scaling a generic model to a patient's segment lengths. However, JRF is sensitive to muscle moment arms, and patient-specific moment arms can differ significantly from a scaled generic model – particularly in a pathological juvenile population. Therefore, in the current project, patient-specific models are developed, using clinical gait analysis (CGA) and magnetic resonance imaging (MRI) data, acquired at baseline, after 6 months and after 12 months.

To date, patient-specific ankle models have been created using data from three different patients (Table 2).

Table 2: Patient data for the three JIA patients. The Gait Laboratory gives a code corresponding to the two laboratories used to collect data.

	Patient 1	Patient 2	Patient 3
Age (years)	15.9	12.9	9.5
Height (m)	1.45	1.53	1.37
Mass (kg)	50.0	64.2	40.6
BMI	23.8	27.2	21.5
Gait Laboratory	IGG	OPBG	IGG

Constructing the patient-specific model involves segmenting bones, cartilage, muscles and tendons using the MRI data. This has been done manually, involving individuating the abovementioned structures in each slide of the MRI. The validation of this model is now being performed in terms of sensitivity of the outputs to the variation of its inputs, with specific focus on muscles and tendons origins and insertions. This analysis showed (Table 3) that patient-specific placement of muscle paths is significant in the determination of joint forces and muscle activation patterns in the more proximal joints of the lower-limb and that the placement of muscle paths also has a large effect on the ankle joint reaction force. Therefore, the adjustment of muscle paths according to the patient's geometry will provide a very significant improvement from the current modelling practice of scaling a generic adult's geometry to create a model.

This analysis also showed that the role of the Achilles tendon and of the tibialis posterior and peroneous longus are particularly crucial within the JIA application. Tenosynovitis of tibialis posterior and of peroneous longus, in fact, frequently influence muscle functioning in these patients. The prediction of the effects of the tenosynovitis might be facilitated by a patient-specific modelling approach. The next focus of this WP will be on the processing of the longitudinal data, which will help validating this hypothesis.

	Achilles	Peroneus Longus			Tibialis Anterior			Tibialis Posterior		
	I	via1	via2	via3	via1	via2	I	via1	via2	I
Patient 1										
Anterior	10.6	-0.6	-2.8	-0.6	-0.6	-3.0	-2.9	-1.3	5.2	-0.6
Posterior	-8.2	-0.6	2.5	-2.2	-0.6	3.0	3.3	0.7	-4.3	-0.6
Superior	-1.7	-0.6	-3.8	-0.6	-0.6	-3.1	-2.4	-1.0	-1.5	-0.6
Inferior	1.4	-0.6	5.3	-2.0	-0.6	4.1	2.2	-0.6	1.5	-0.6
Lateral	13.4	-0.6	-3.4	-0.9	-0.6	-3.8	-5.5	1.0	8.2	-0.6
Medial	-12.0	-0.6	3.8	0.8	-0.6	3.9	5.8	-1.0	-5.4	-0.6
Patient 2										
Anterior	9.6	1.2	-2.0	1.3	1.2	-2.7	-2.8	-1.0	4.4	1.2
Posterior	-7.9	1.2	1.5	-1.2	1.2	2.4	2.2	2.0	-3.8	1.2
Superior	-1.5	1.2	-2.5	1.3	1.2	-2.5	-1.5	-1.0	2.1	1.2
Inferior	1.6	1.2	2.6	1.1	1.2	2.4	1.5	2.0	-1.5	1.2
Lateral	11.5	1.2	-3.5	1.3	1.2	-4.5	-4.6	2.8	4.8	1.2
Medial	-11.9	1.2	3.8	1.1	1.2	3.8	3.9	-2.1	-3.5	1.2
Patient 3										
Anterior	12.9	-0.8	-1.0	-0.8	-0.8	3.2	-2.3	-0.9	7.8	-0.8
Posterior	-9.9	-0.8	-0.9	-0.8	-0.8	-3.0	2.4	0.6	-6.5	-0.8
Superior	2.0	-0.8	-2.6	-0.8	-0.8	-2.6	-1.6	-0.9	-1.6	-0.8
Inferior	-2.4	-0.8	-2.4	-0.8	-0.8	-3.1	-1.0	-0.6	0.9	-0.8
Lateral	-9.3	-0.8	-1.3	-0.8	-0.8	-3.4	3.3	-1.0	-5.0	-0.8
Medial	9.4	-0.8	-2.3	-0.8	-0.8	3.8	2.0	0.9	6.7	-0.8

Table 3: Maximum value of percentage change in ankle joint reaction force Values (mean over three gait trials) have been calculated in the stance phase of gait – original muscle position value subtracted from

perturbed muscle position value. Muscles are included that have a mean percentage change of greater than or equal to 0.5% in at least one perturbation in one patient. The colour scale is based on the absolute values and ranges from 13.4 (the maximum value with the highest level of shading) to 0 (with a white background colour). Via points are indicated as “via1, via2, and via3”, whereas the insertion points are indicated as “I”.

Since the construction of the models is partly operator-dependent, the next step of the validation will focus to the inter- and intra-operator repeatability of the model construction.

The model has been devised to provide a number of biomarkers that can be used to 1) discriminate and quantify differences between ankles JRF estimated for the two limbs of the same subject 2) predict disease progression in the ankles of JIA patients. In agreement with the clinical partners, the following questions have been highlighted as the key outcomes needed from the biomarkers:

- Do biomechanical alterations affect the extension of arthritis in the lower limbs?
- Do biomechanical alterations affect structural damage progression?

Several quantities have been included in the model that may serve this purpose. Given the early stage of the data collection, data has not yet been analysed on the disease progression over a long timescale. Even if we are still at a speculative stage, considerable thought has been put into analysing the potential parameters that can be used to predict the progression of JIA through mechanical effects. Later in the project, these values will be analysed in relation to disease progression measures, provided by the clinicians. A very large set of parameters will also be included in order to detect any unforeseen interactions.

All of these models will yield a patient-specific estimate of the JRF. These predicted values will be validate by means of a comparison with the few experimental measurements done on cadavers available in the literature. The JRFs themselves will be potential predictors for the outcome of the disease in terms of structural damage progression. These predictions will be internally validated using collected follow up data (specifically the juvenile arthritis damage index [JADI] and a follow up MRI performed after 2 years) in the study.

4.4 Validation for NND models (WP6-WP11)

Siemens AG has proposed a novel method for the automated extraction of subject-specific muscles, bones and skin of the pelvis and legs from 3D MR images, and applied it to a mixed population of normally developed children. The results of this work, as well as the preliminary validation of its outcomes, have been already presented in greater detail in D11.1.

Initially, three dimensional (3D) meshes for 54 individual muscles, 12 bones, fore, middle and back feet as well as the whole skin of one pediatric subject were manually annotated. For each patient, 30 landmarks were annotated as well for qualitative assessment of the results.

For the automatic segmentation method, we have addressed the articulation complexity separately, thus reducing the number of DOF while producing reliable results. Our fully automatic approach is a combination of individual affine transformations for each leg segment, split at the joints, and non-rigid deformation, to effectively and efficiently reduce the total degrees of freedom and search space and avoid the deep local minima that arise in matching complex articulated subjects such as the human leg.

We validate this approach quantitatively by measuring the distance between automatically and manually defined coordinates of landmark sites, as well as by measuring overall mesh similarity between the automatic results and the manually annotated ground truth.

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

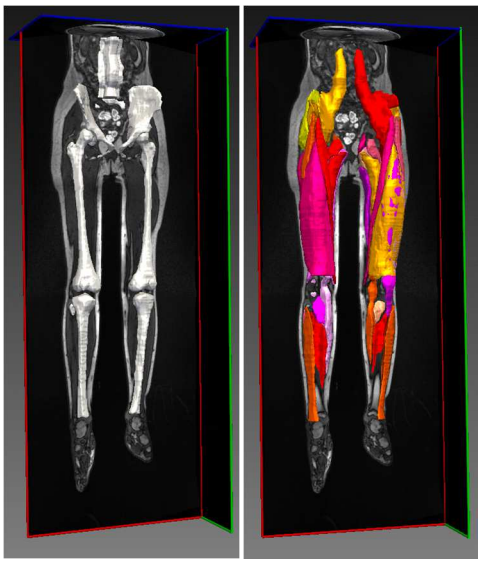


Figure 5: Manually annotated 3D structures and for the reference subject.

We applied our approach to a set of 14 3D MRI scans of healthy children in the ages of to 8 to 15 years old. Child number 9 has been chosen as reference, due to the fact that the image showed the fewest artifacts and the clearest structures.

For a quantitative measure of the results, a set of 30 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 2 below.

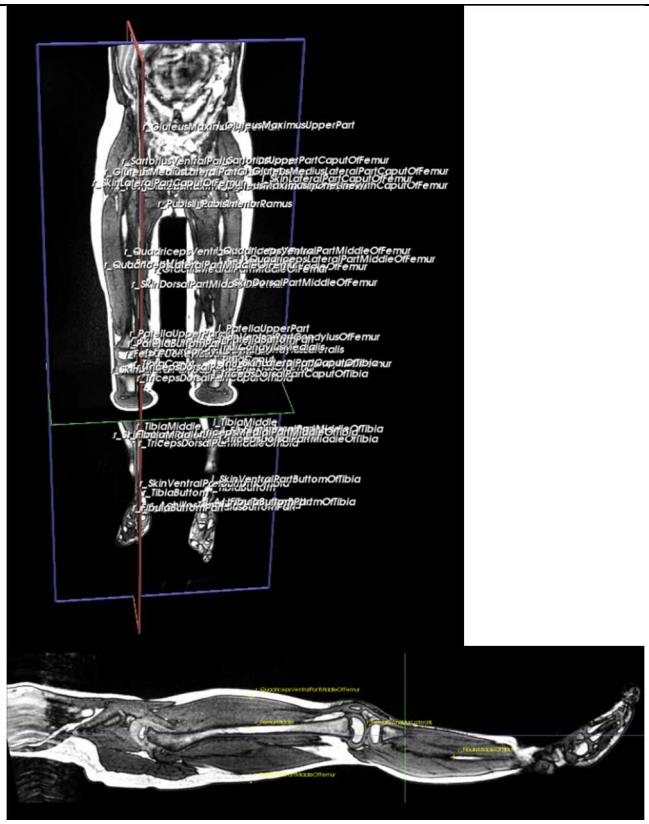
l_AchillesTendonBottomOfTibia l_FemurCaput l_FemurCondylusLateralis l_FemurCondylusMedialis l_FibulaBottomPart l_PatellaUpperPart l_PubisInferiorRamus l_SkinVentralPartBottomOfTibia l_SkinVentralPartCondylusOfFemur l_TalusBottomPart l_TibiaBottom l_TibiaCaput l_TricepsDorsalPartCaputOfTibia l_TricepsDorsalPartCondylusOfFemur r_FemurCaput r_FemurCondylusLateralis r_FemurCondylusMedialis r_FemurTrochanterMajor r_FibulaBottomPart r_PatellaBottomPart r_PatellaUpperPart r_PubisInferiorRamus r_SkinLateralPartCaputOfFemur r_SkinVentralPartBottomOfTibia r_SkinVentralPartCondylusOfFemur r_TalusBottomPart r_TibiaBottom r_TibiaCaput r_TricepsDorsalPartCaputOfTibia r_TricepsDorsalPartCondylusOfFemur	

Table 4: Landmarks used to measure registration quality.

The distances between the landmarks, as well as the DICE and JACCARD indices of the structure meshes for the right leg, have been measured and the results are shown in **figures 6, 7 and 8.**

Results

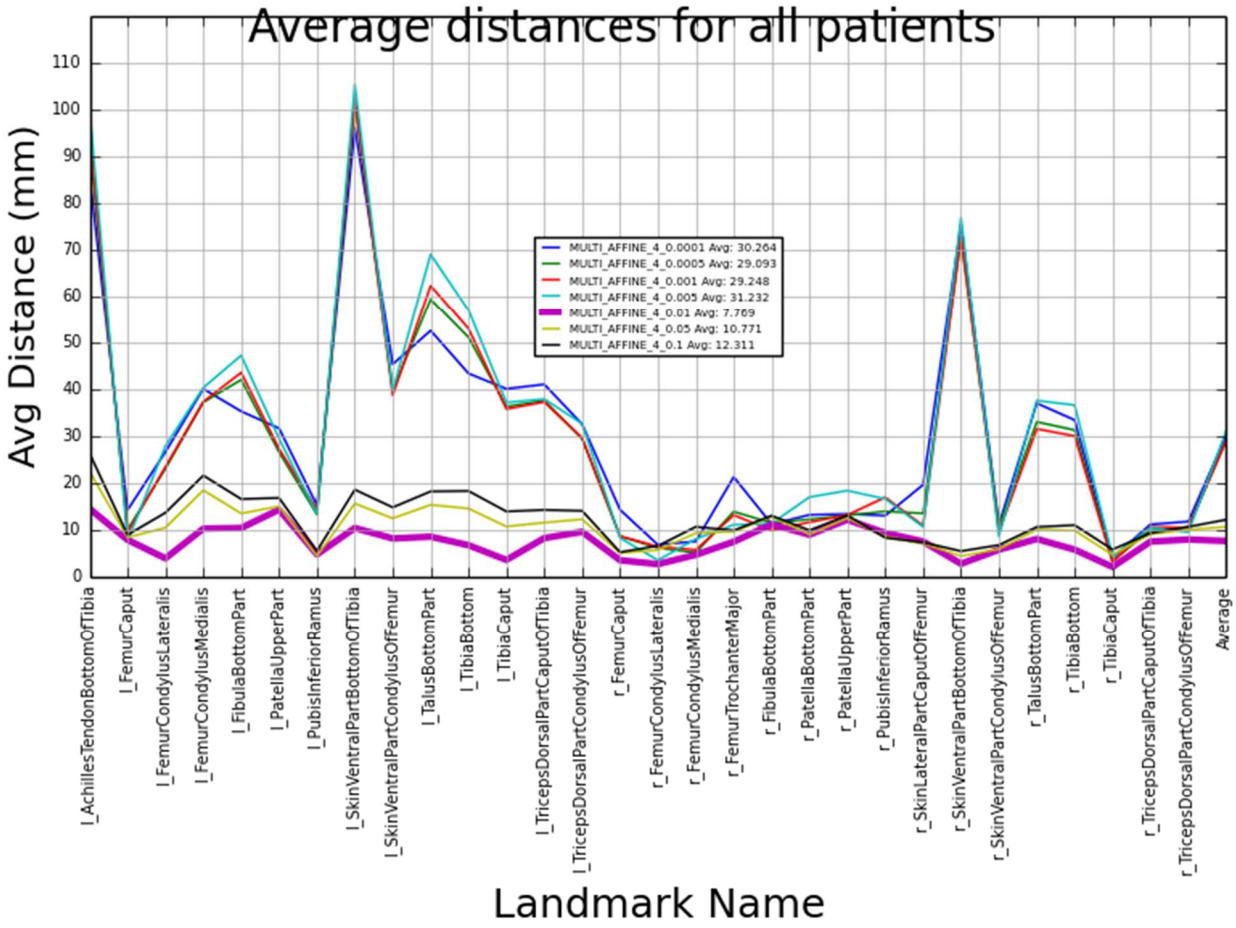


Figure 6: Results on landmark distances for different parameter configurations*. The distances are measured in mm between the registered reference image and the target image, for all images. Each series in the plot represents a different set of parameters to the algorithm, with the best shown in a bold magenta line, resulting from using our approach with a bending penalty of 0.01.*

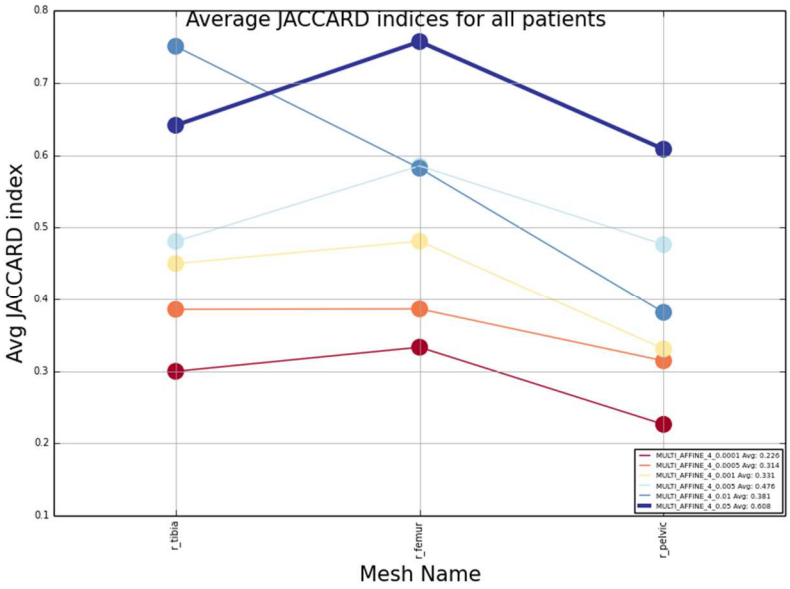


Figure 7: JACCARD similarity indices for the annotated meshes in all patients. *The indices have been computed between the meshes of the registered atlas and those of the target image. Each series in the plot represents a different set of parameters to the algorithm, with the best results provided by using the original images, multi-affine approach with different bending penalties. The best results were obtained with a bending penalty of 0.05, and are shown with a bold blue line in the figure.*

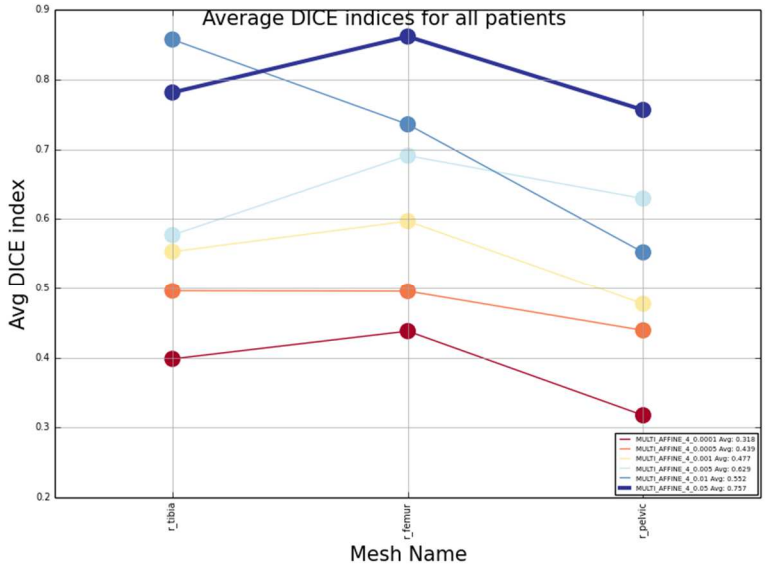


Figure 8: DICE similarity indices for the annotated meshes in all patients. *The indices have been computed between the meshes of the registered reference image and those of the target image. Each series in the plot represents a different set of parameters to the algorithm, with the best results provided by using the original images, multi-affine approach with different bending penalties. The best results were obtained with a bending penalty of 0.05, and are shown with a bold blue line in the figure.*

As evidenced in the landmark distances in Figure 3, an approach that is neither too rigid (bending penalty of 0.1) nor too flexible (bending penalty of 0.0001) produces the best results.

In Figure 6, the results obtained by using a very strong bending penalty (0.1) are clearly sub-optimal. On the other hand, both the landmark distances and the mesh DICE and JACCARD indices (Figures 3, 4 and 5) confirm that very flexible approaches (with very small bending penalties) are not adapted either.

The best approaches seem to use intermediate bending penalties, producing a mean landmark error of 7.7mm. It is interesting to see that more flexible bending result in worse performance around landmarks situated in soft tissues, such as on the skin surface (see Figure 3).

For further quantitative evaluation, the femur, tibia and pelvic bone were manually segmented for the right leg of all patients, and used as ground truth. Given the resulting Jaccard and Dice indices (Figures 7 and 8), it becomes clear that smaller bending penalties result in larger variations from the ground truth, which confirms what was concluded from the landmark distances in Figure 6. The tendency is similar as with the landmarks: a not too rigid, not too flexible approach (bending penalty of 0.05) seems to produce the best results. Since the process of manual annotation of ground truth meshes is still ongoing, we will rely on the best parameters for the landmark distance results, and leverage the best parameters for mesh similarity indices when more meshes are available for validation.

Figure 6 illustrates some of our results. Despite the evident intensity bias in the images, and the differences in posture, body composition and even some motion artifacts, the results are very promising.

Another point illustrated in Figure 9 is the adaptability of our approach to high intensity bias, as well as asymmetries and anatomical differences of the patients. For instance, different knee/leg positions, different musculoskeletal and body/fat composition are also nicely leveraged to produce very encouraging results.

In general, and in spite of the challenge of multi-patient registration, the method accounts quite well for positioning and size differences in the patients in spite of the intensity bias, while still leveraging the smaller, more local anatomical differences.

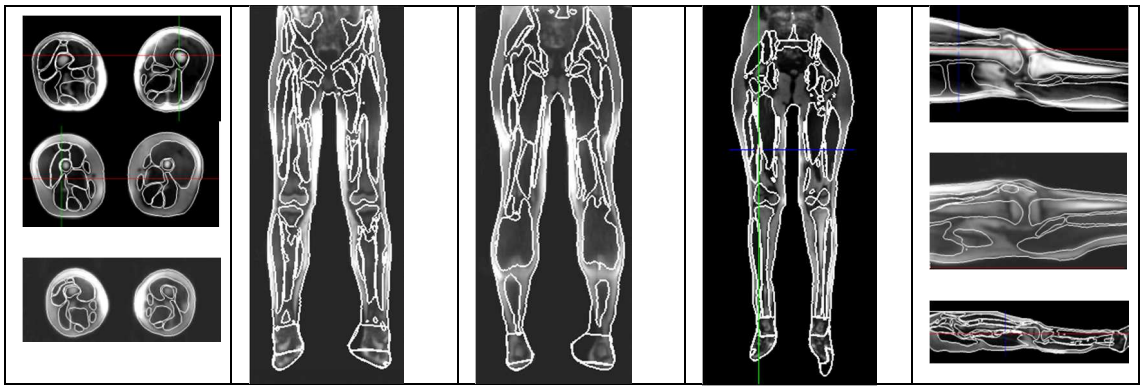


Figure 9: Different results of transformed meshes (white contours), overlaid on the target images.
Particularly evident in these images are the anatomical differences, as well as the intensity bias, which causes the same tissue to have very different intensities within the same volume. Our method has shown very promising results even in these cases. Some errors can be seen around the feet, which can be explained by the movement artifacts at the time of image acquisition.

Motek Medical has developed a new version of the musculoskeletal biomechanical Human Body Model software dedicated for clinical gait analysis. These alterations were implemented based upon the clinical user requirements as defined by wp6. The main requirement was that patients often are not capable to stand in the required calibration position, resulting in inaccurate joint rotations, moments and estimated muscle lengths. New regression methods were implemented in order to ensure accurate joint rotation calculations even if the patients has joints deformations or cannot stand in the required calibration position due to spasms or contractures. For the hip rotation center the regression method of Harrington et al. (2006) was implemented and for the knee and ankle additional markers were added during the initialization to calculate the joint axis.

For technical quality assurance and initial validation of the biomechanical modelling adjustments two typically developing children, age 8 and 11, were subsequently measured in the gaitl abs in Leuven, Rome (OPBG) and Amsterdam (VUmc). Raw motion capture and external force data was collected and re-processed by Motek Medical using the personalized HBM software. Data is now available for the digital repository and further comparison.

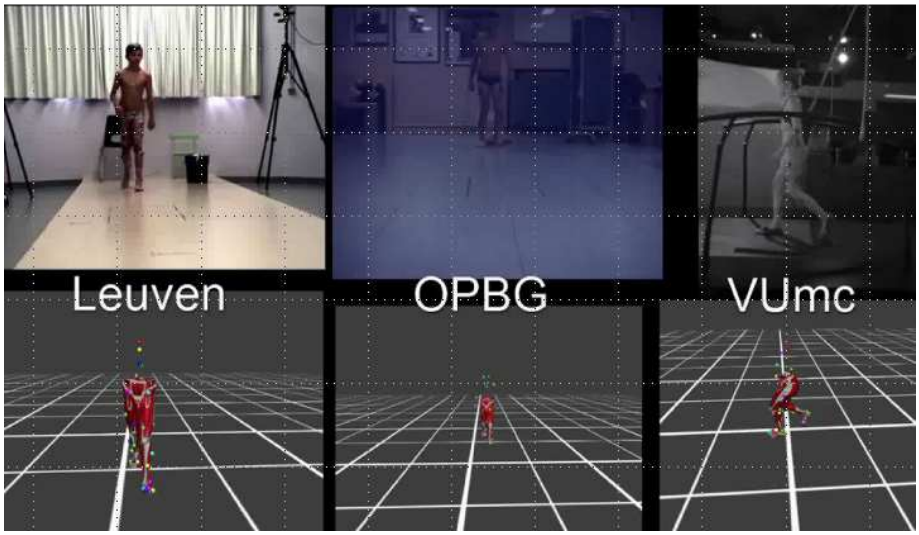


Figure 10: Video images of the TQA recordings in the three different gaitlabs and HBM captures of the recorded data

Delft University of Technology has been working on a forward dynamic model for predictive simulations of pathological gait, based on high-level optimization criteria such as gait speed, gait stability and energy efficiency. The system allows integration of neurological constraints, such as motor noise and neural delays, as well as subject-specific models of spasticity and obligatory synergies, derived from the physical exams performed by the clinical partners. The subject-specific neuromuscolar controller will be used in combination with the personalized musculoskeletal models in order to make fully personalized predictive simulations. These simulations will help clinicians predict the outcome of surgical interventions, such as tendon transfer surgery, spasm reduction using pharmaceuticals, or bone reconstruction – as well as help them gain further insights into the understanding of the pathological gait, and possibly aid in developing new compensation strategies.

For validation, motion capture, force plate and EMG data, will be compared to our simulated kinematics, dynamics and neural commands. This will include data from both healthy subjects and CP children.

Validation for TQA (Technical Quality Assurance) on Clinical Gait Analysis

The aim of technical quality assurance is to: i) evaluate the performance of the equipment in three laboratories (low level); ii) assess the repeatability of the measurement conducted in the three laboratories on two healthy children (high level). For both levels URLS, in charge for the Technical Quality Assurance, developed the protocols and performed measurements to assess the quality of the measurements conducted in the involved labs. The CGA centers involved in the experimental protocol are: i) KU Leuven; ii) VU Medisch Centrum (VUmc); and iii) Children's Hospital 'Bambino Gesù' (OPBG).

Low level

In the low level were compared the performances of the optoelectronic system (OS), force platform (FP) and the signal synchronization between the force platform and the EMG system between three laboratories (S-synchro).

The OS-validation protocol consisted, briefly, in moving the calibration wand for ten seconds inside the measurement volume. From the acquired positions of the markers, distances and angles among markers were evaluated. The values were then compared with the actual distances and angles, which were computed with the wand placed on the origin of optoelectronic system for five seconds during a static acquisition. The repeatability of parameters was evaluated as root mean square error (RMSE) between actual and measured values for each parameter. The performances of the optoelectronic system does not shows significant differences between the three laboratories.

The FP-validation procedure consisted in the application of arbitrary forces to each force platform pushing on it with an ad-hoc system sensorized with a 6-component load cell. Statistical analysis was then conducted to find statistical differences between FPs and between points on the edge and on the middle part of platforms. The obtained values demonstrates data comparability among centers. As regards the signal synchronization between EMG signals and FP ones the higher value observed at KU Leven (65 ms) assures consistent clinical analysis

High level

To perform the high level, the protocol includes the following features:

- two healthy children, age-matched with the range considered in the project, have been recruited;
- data collection was performed on these subjects in all the involved centers (KUL, OPBG and VUA);
- two therapists per center performed the marker placement on each subject (those therapists were the ones who usually performed gait analysis in the centers);
- five walking trials were collected.

The considered variables are:

- Joint angles (Kinematics); ii) Joint moments (Kinetics);
- Timing on EMG signal activation.

In the following figure, as paradigmatic example, the waveform of kinematic and kinetic variables of subject #1 evaluated in the three laboratories were reported.

An overall exam of the kinematic and kinetic variables, demonstrates that the repeatability between operators and between laboratories were always excellent (the coefficient of multiple correlation CMC>0.9) in the sagittal plane and lower in frontal and transverse plane, but always in the range of good repeatability (CMC>0.8), moreover the repeatability between the three centers was lower than the repeatability between OPBG and KULeuven, however always in the range of a good repeatability.

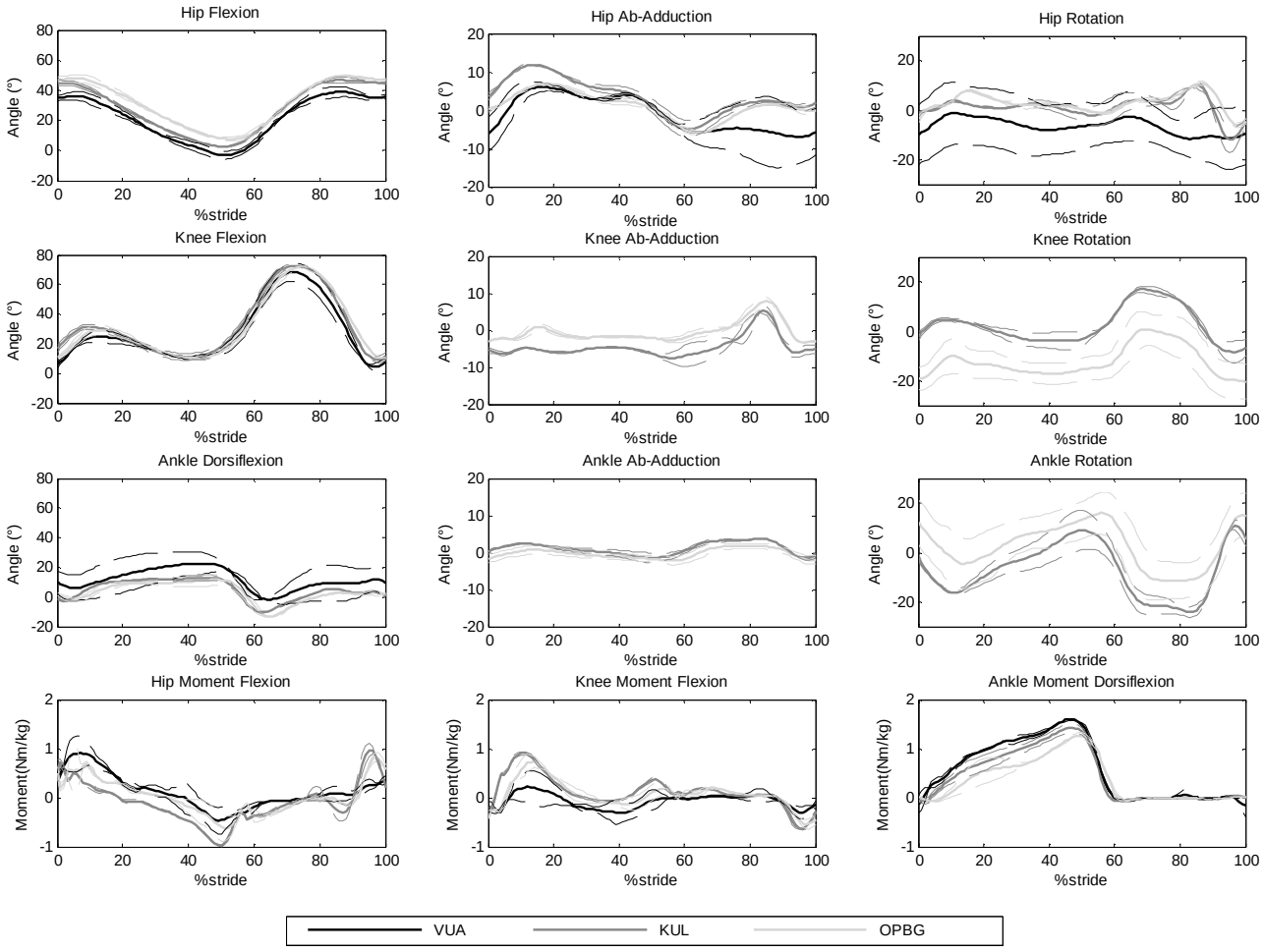


Figure 1*: Mean and standard deviation curve among two operators of kinetic and kinematic variables of subject #1, right side, between laboratories (KUL, OPBG, and VUA).*

As regard EMGs, the activation time of 8 muscles (4 agonist and 4 antagonist) of one lower limb was calculated. Then, paired-sample t-tests were performed *within laboratory*. As an example in the following figure the EMG signal waveforms of right side for subject #1 are reported.

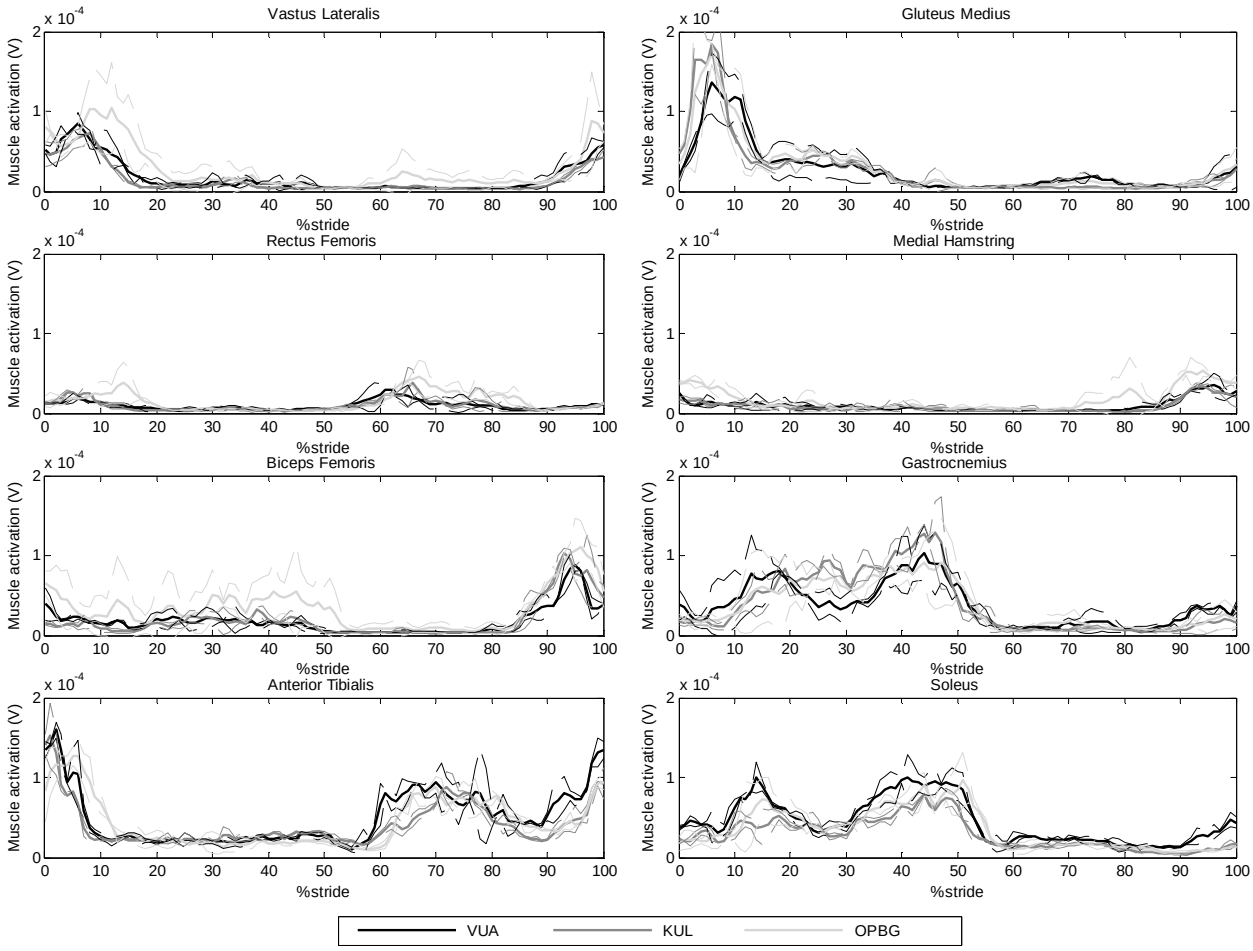


***Figure** : Mean and standard deviation curve between two operators of EMG signal of subject 1 right side between laboratories (OPBG, KU Leuven and VUmc).*

The EMG signals, presented a good repeatability both *within laboratory* as well as *between laboratories*, except for anterior tibialis in subject 1 at OPBG and gastrocnemius at VUmc for subject 1 and in KU Leuven for subject 2. Moreover, statistical differences *between laboratories* were found for gluteus medius in both subjects and in anterior tibialis in subject 2.

As a concluding remark we can state that valid data are collected at the three labs for the kinematic, kinetic and EMG data sets.

4.5 Genetic and metagenomics analysis validation (WP7): the case of metagenomics and metabolomics of gut microbiota studies

4.5.1 Introduction: state of the art of the microbiota determinations in the MD-PAEDIGREE Project

In the context of the Model Driven Paediatric European Digital Repository (MD-PAEDIGREE), besides clinical data, the collection and management of genomic and metagenomic data, may actually complement instrumental, routine laboratory and clinical data as a staple resource for medical research. Clinical data is collected during the course of ongoing patient care and the -omic and meta-omic information may actually complement the electronic health records, providing piece of evidence of the entire spectrum of ontological features of the "patients".

In detail, the entire set of age (*i.e.*, stratification), flare-up conditions, naïve baseline of the pathology manifestation, external perturbations such as diet, antibiotic administration, stress-related symptoms, may be synthetically named by using the term "phenomics", expression of the several phetotyping traits of the patient. Over the past 15 years, many authors have proposed that phenomics - large-scale phenotyping - is the natural complement to genome sequencing as a route to rapid advances in systems biology, preparing the route to systems medicine (Schork, N. J. Genetics of complex disease-approaches, problems, and solutions. Am. J. Respir. Crit. Care Med. 156, S103–S109, 1997; Schilling, C. H., Edwards, J. S. & Palsson, B. O. Toward metabolic phenomics: analysis of genomic data using flux balances. Biotechnol. Prog. 15, 288–295, 1999; Houle, D. *In* The Character Concept in Evolutionary Biology, ed. Wagner, G., 109–140, Academic Press, 2001; Bilder, R. M. et al. Phenomics: the systematic study of phenotypes on a genome-wide scale. Neuroscience 164, 30–42, 2009; Freimer, N. & Sabatti, C. The human phenome project. Nature Genet. 34, 15–21, 2003).

Phenomic-level data are necessary to understand which genomic variants affect phenotypes, to understand pleiotropy and to furnish the raw data that are needed to decipher the causes of complex diseases (obesity, juvenile idiopathic arthritis, cardiopathies). Our limited ability to understand many important biological phenomena suggests that we are not already measuring all important variables and that broadening the possibilities will pay rich dividends. Fundamentally, we can choose to include into this new point of view, additional parameters or "data" such as genomic fingerprinting indexes (*e.g.*, disease-gene candidates, polymorphisms) and metagenomic gene scaffolds (gut microbiome), linked to metabolic activities (metabolome), to provide additional and useful indexes of disease. All genotyping and phenotyping parameters need to be measured by omics and meta-omics technologies; indeed WP 7 actually provide the *added value* to the Project, thanks to technologies for high-throughput phenotyping and genotyping which are fully available in the MD-PEDIGREE Consortium, at the OPBG facilities, and which include conceptual, analytical frameworks, fused to advanced bioinformatic approaches that enable the use of very high-dimensional data.

The individual phenotype, therefore, is the combination of all these *trans-acting* elements, combined to genomic and metagenomic reservoirs, through genetic and epigenetic controls. Once the single microbiota is fully described, a genetic fingerprinting is available to complement the individual genetic reservoir (code), through multi-level meta-omic platforms (metagenomics, metabolomics, metaproteomics). The produced data can be employed at individual and population level, to assist in the design of therapeutic and diagnostic pipelines or, rather, in the disease risk prediction of important disease at early onset, respectively.

4.5.2 Gut microbiota profiling in obese and JIA patients: experimental and bioinformatic pipelines

Gut microbiota affecting factors have been fully considered in the first step of patient recruitment and sample collection (baseline, onset) and progressively they will be considered during follow-up (e.g., flare-up), during MD PAEDIGREE. They have been analyzed for each patients and the associated ontologies or categories of clinical-diagnostic treats will be uploaded onto the maatG database as qualitative and quantitative metagenomics and metabolomics maps, expressed in term of relative abundances of OTUs (operational taxonomic units, phylotypes) and metabolites (volatilome, metabotypes). Up to date, the process of data collection has taken place at three levels:

a. OPBG repository database, with household data processing and storage procedures;
b. NCBI BioProject submission, EBI repository database ;
c. maatG data submission, with the intent to generate a shared platform for model generation.

4.5.3 Detailed specific and comparative microbiota results:

Metagenomics, metabolomics profiling of gut microbiota;

The patient samples were catalogued, barcoded and stored according to OPBG Biobank procedures. The calculation of reads, min and max sequences were computed, with a simultaneous comparison between disease and reference sequences. The category for the metadata analysis was chosen in term of taxonomy ranking, from phyla to genus (Figure 15). The distribution of OTUs were described in term of individual relative abundances and represented by using different criteria, e.g. age stratification. An example of OTUs proportion representation at phylum level for obese and JIA patients (CVD Obesity), is reported in Figure 15, in which patient metagenomes are clustered, compared to reference healthy individual metagenomes (CTRL), in a case-control approach study. The Figure 15 shows the high inter-individual variability of microbiota detected under health and diseased conditions; such a difference, however, does not hamper the possibility to generate microbiota clustering, specifically associated to different phenomic conditions.

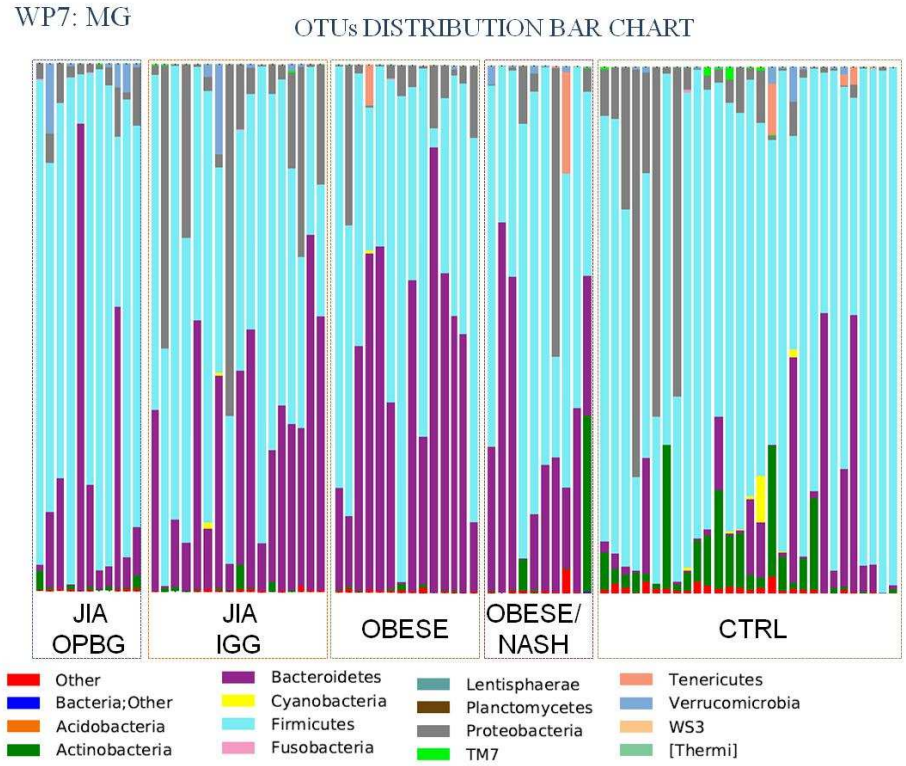


Figure 15. OTUs distribution clustered for patient type and age: semi-quantitative representation

Particularly, besides statistical models (multivariate analyses, parametric and not parametric tests) to assess statistical significance to metagenomic phylotypes and metabotypes, some model validation for microbiota patterns for JIA and obese patients are under optimization.

First of all, qualitative and quantitative metagenomic analyses of gut microbiota OTUs at Phylum, Family and Order levels have been provided, including the bioinformatic elaborations of JIA and obese gut microbiota types, described by weighted/unweighted UNIFRAC and Bray Curtis algorithms, parameters of sequence richness and diversity (Figure 16, 17).

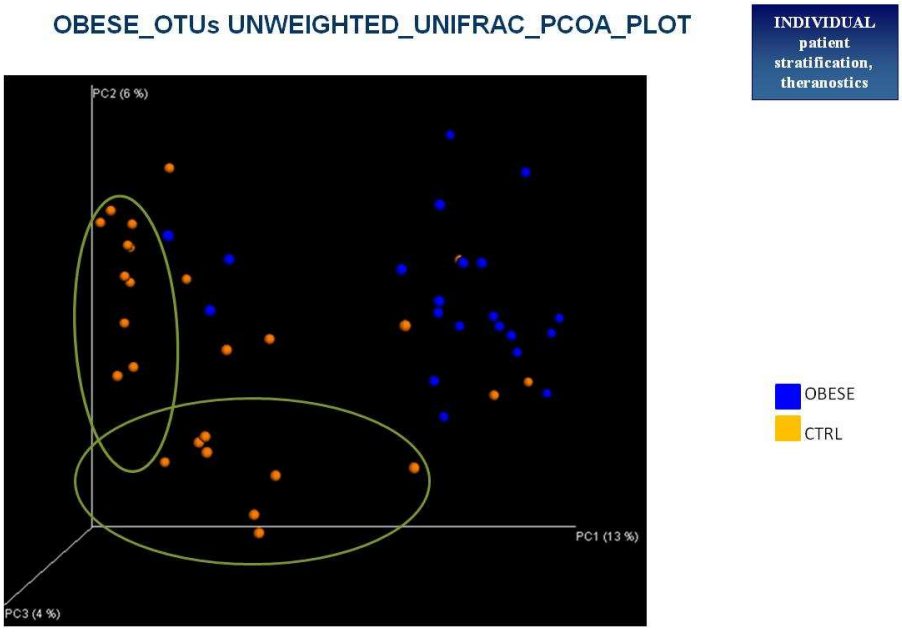


Figure 16. Quantitative evaluation of differential sequence richness in obese *versus* healthy control individuals

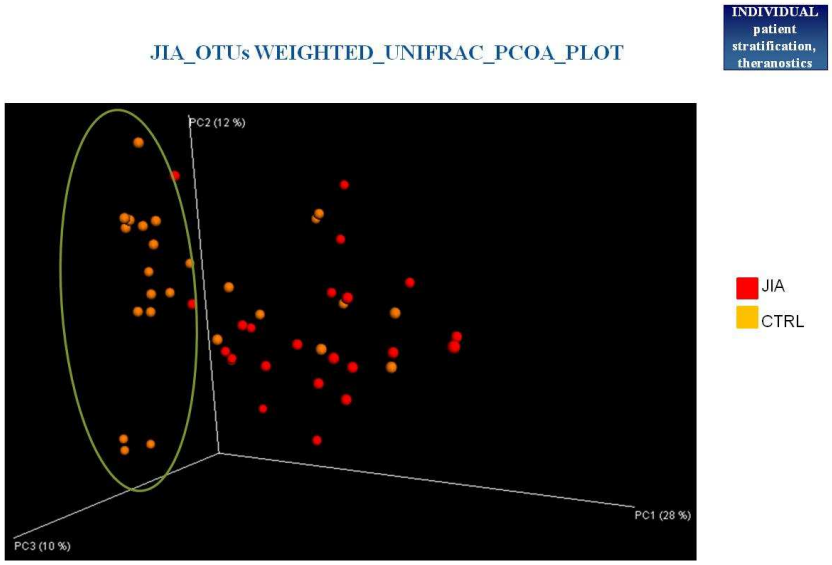


Figure 17. Quantitative evaluation of differential sequence richness in JIA *versus* healthy control individuals.

Remarkably, *comparative analysis of JIA and Obese gut microbiota types have been also performed to infer transversal differences in the microbiota extended genotypes.*

Indeed, metagenomic analyses were integrated and categories for both obese and JIA patients were assessed, compared to healthy individuals. Interestingly, metagenomic clustering resulted different for all the groups also weighting also the OTUs distribution, in addition to sequence diversity and distribution (Figure 18).

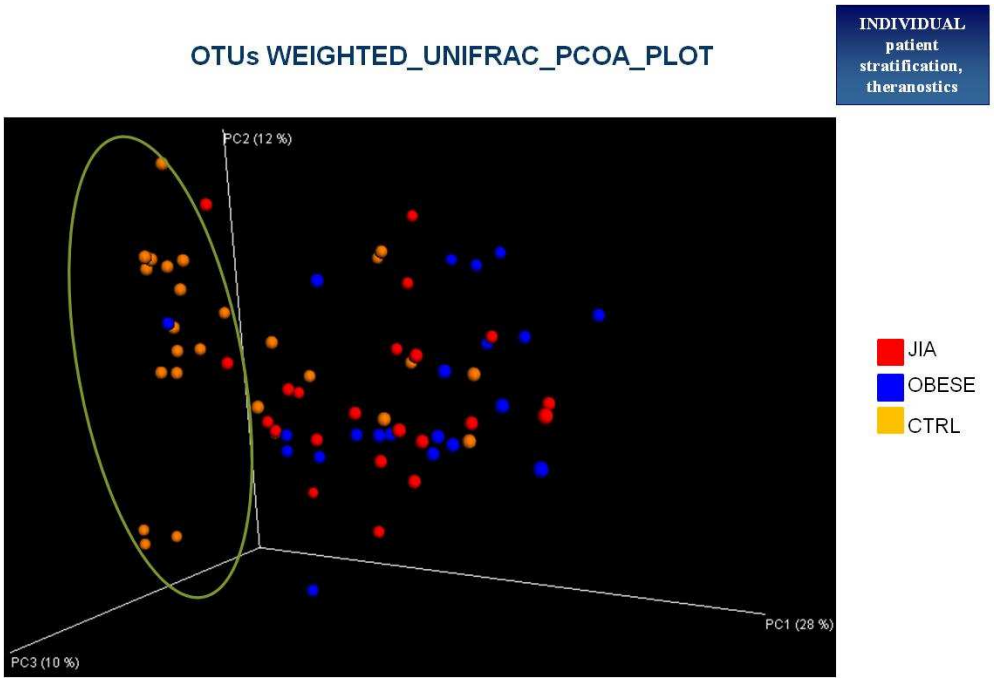


Figure 18. Quantitative evaluation of differential sequence richness in JIA, obese patients *versus* healthy control individuals.

Additionally, *dynamic models* that link clinical phenomena across levels, have been designed and are currently under advancement. Although analyses of genomic data have been successful at uncovering biological phenomena, they are - in most cases -supplementing rather than supplanting phenotypic information. In the WP7 we have identified the operational rationales to integrate phenomic to genomic and metagenomic data by advanced original approaches, such as *Cytoscape*, *Circus* computational pipelines (Figure 19).

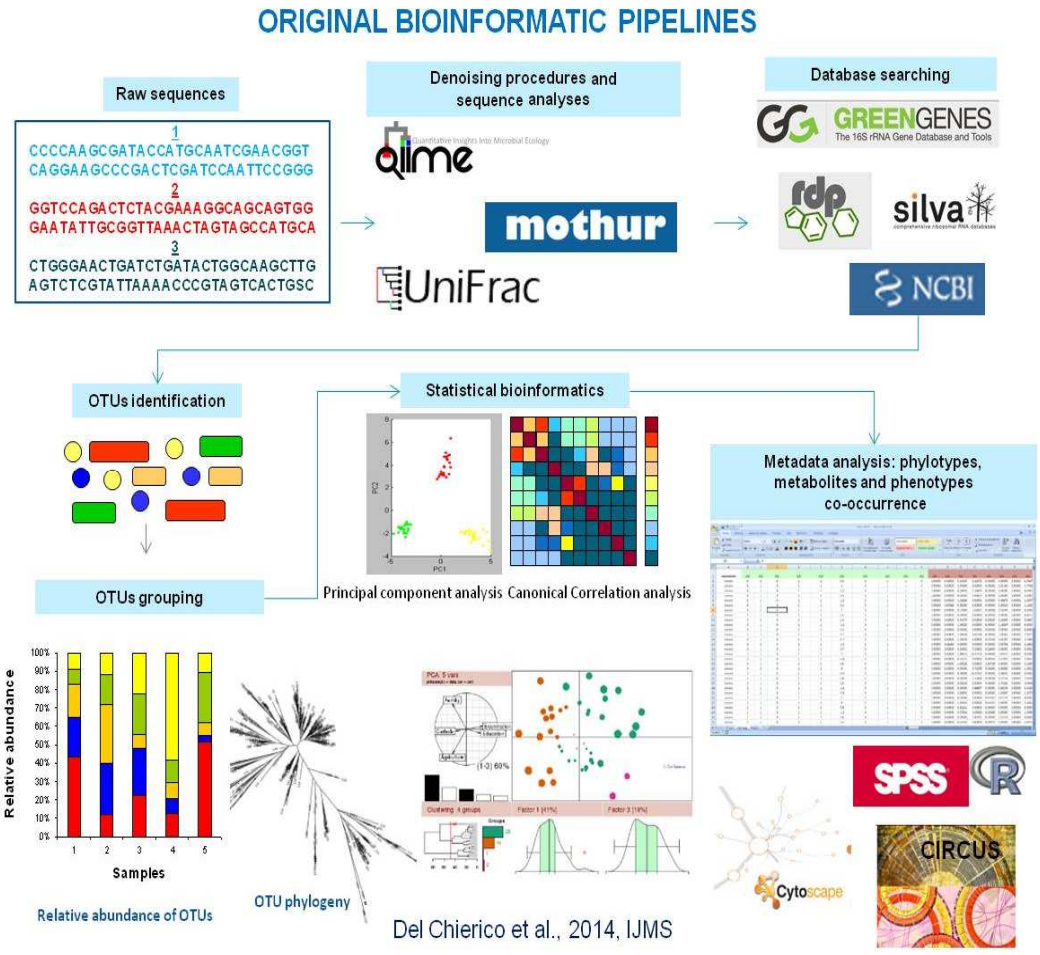


Figure 19. Original bioinformatic pipelines designed and set to generate and process metagenomic and metabolomic data.

To evaluate the role of genomic (assessed by disease-gene or candidate gene analysis) and metagenomic (based on gut microbiota signatures) profiling on the onset and progress of diseases and on their outcomes, *the post-analytical data collection and analysis processes* have represented one of the milestones of the WP7.

Currently we are defining and setting a *mathematical algorithms* to establish a threshold able to assess *validated enterotypes and metabotypes* as a diagnostic report, by using a new representation of relative OTUs abundances and metabolites. The idea is to establish a baseline for healthy individuals (CTRLs) and patients, represented through a *fold-change rank ordering analysis* (FCROA) representation. This analysis provides the differentially expressed metabolites or OTUs in the considered patient group (*e.g.*, JIA, obese) compared to CTRLs.

5. Future work

With the completion of the retrospective prediction in all four disease areas, expected in the next few month, the progressively increasing number of patients uploaded into the digital repository will allow a more complete and challenging validation since the model will be tested in patients not used to develop it (external prediction). It is well known that the prognostic prediction is almost always better on the data set on which the model has been constructed (derivation set) compared to the performance of the same model on new data (validation set) and that, even when information bias and selection bias have been carefully taken into account, model prediction from internal validation is still too optimistic.

It is clear therefore that internal validation per se is no guarantee for generalizability and thus no substitute for external validation. In the following months we will undertake the external validation that aims to address the accuracy of a model in patients from a different but plausibly related population, which may be defined as a selected study population representing the underlying disease domain.

However, while subsystem test procedures are beneficial to evaluating the variation of one variable against one or more input modifications, when it comes to simulation of an actual clinical scenario full scale tests are needed, where all the variables involved both on the input and on the output side are considered and analyzed jointly. Only the population of the database will permit a further refinement of the model allowing a full-scale validation that is closer to the actual patient conditions where all the variables interact with one another.

Furthermore, since the results of external validation are also not always unambiguous and trustworthy, they require a substantial sample size to provide sufficient power to find similar performance. For this purpose we will take advantage of the digital repository and of the tools developed since the Health-e-Child project, namely the MaatG platform and infostructure, the AITION software for data curation and disease signature, the Similarity Search and Case Reasoner tools for the identification of homogeneous groups of patients. Eventually, cross validation between our mechanistic model and a so created statistical model of a homogeneous group of patients is the task that will be dealt with in the next months.