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 12.2.3

Third Clinical Assessment and Validation Results for all Four Disease Areas: Periodic update at month 51 of Clinical Assessment and Validation Outcomes (M51)

Due date of delivery: 01/06/17

Actual submission date: 26/06/2017

Start of the project: 1st March 2013

Ending Date: 31st May 2017

Partner responsible for this deliverable: OPBG

Version: 2.2

Document information

Document Classification

Title	Third Clinical Assessment and Validation Results for all Four Disease Areas: Periodic update at month 48 of Clinical Assessment and Validation Outcomes (M48)
Deliverable	12.2.3
Reporting Period	M48
Authors	Marcello Chinali
Work Package	12
Security	PU
Nature	R
Keyword(s)	Clinical Validation

Document History

Name	Remark	Version	Date
Marcello Chinali	Preliminary outline of the document	Draft 12.2.3 ver 1.0	24/05/17
Marcello Chinali	Addition of contribution from JIA	Draft 12.2.3 ver 1.1	27/05/17
Marcello Chinali	Addition of contribution from CVD	Draft 12.2.3 ver 1.2	05/06/17
Marcello Chinali	Revision and contribution for Metabolomics	Draft 12.2.3 ver 2.0	11/06/17
Marcello Chinali	Addition of contribution from CMP	Draft 12.2.3 ver 2.1	15/06/17
Marcello Chinali	Addition of contribution for NMD	Draft 12.2.3 ver 2.2	25/06/17

List of Contributors

Name	Affiliation
Giulia Cinelli	OPBG
Tobias Heimann	SAG
Dominik Neumann	SAG
Jose Pozo Soler	USFD
Maria Jimena Costa	SAG
Frans Steenbrink	MOTEK
Marjolein van der Krogt	VUMC
Claudia Mazzà	USFD
Alexander Jones	UCL
Oliver Pauly	SAG
Pieter van Dijkhuizen	IGG
Gabriele Rinelli	OBG
Thomas Geijtenbeek	TU Delf
Maria Jimena Costa	Siemens Healthhineers
Matjolein van der Krogt	VUmc
JosePozo Soler	USFD
Jaap Harlar	VUmc

List of reviewers

Name	Affiliation
Davide Zaccagnini	LYN
Bruno Dallapiccola	OBG

Table of Contents

1.0 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, has bene an ongoing process which has benefited from the use of the models in different clinical settings thus increasing 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, 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. In the D12.1.2, we have updated the status of the validation and provided initial analysis on baseline models and mid-term evolution prediction.
Given the significant heterogeneity of the different clinical areas and the resulting differences in model definition and expected abilities, the last year validation process required that numerous partners, from both the clinical and the technical areas, proceed on a cogent and well-defined validation process.

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

The aim of the present document is to report on final status of the clinical validation process for the four disease areas, which are hereby described in details. In addition to what previously reported in D12.1.1 and D12.1.2, given its potential in clinical practice and disease understanding, the current document also includes a specific section of validation performed in the Metabolomics and Metagenetics data by WP7.

2.0 Validation for Cardiomyopathies models (WP3-WP8)

2.1 Evaluation of anatomical model

Segmentation of cardiac anatomy is the first step toward modelling cardiomyopathies and to this end; a semi-automated segmentation pipeline was developed in task T8.1. The direct output of this pipeline is a patient-specific cardiac anatomical model over the complete heart cycle, which is difficult to evaluate due to a lack of standardized model parameters and/or representation. However, evaluation can be performed on clinically relevant measurements computed from the model, as listed in Table 1 below. Manual and automated cardiac anatomical

model segmentation was performed for 32 cases at UCL and the measurements described in Table 1 were computed for both methods. The following paragraphs describe the analysis of agreement between the manual and automated anatomical models based on these measurements.

Table 1 Clinically relevant measurements used for anatomical model evaluation

Measure	Computation
End-diastole volume (EDV) End-systole volume (ESV)	Directly computed from the anatomical model
Stroke volume (SV) Ejection fraction (EF)	Indirectly computed from EDV, ESV

Lin's concordance correlation coefficient (R_c) [Lin1989] is a measure of agreement between two continuous units of measurements computed using different methods, where R_c = ±1 implies perfect agreement and R_c = 0 implies no agreement. Figure 1 displays scatter plots for the four measurements described in Table 1, computed using the manual and automatic anatomical model. The figure also includes the value of R_c in each case, which shows strong agreement between the two models based on these measurements.

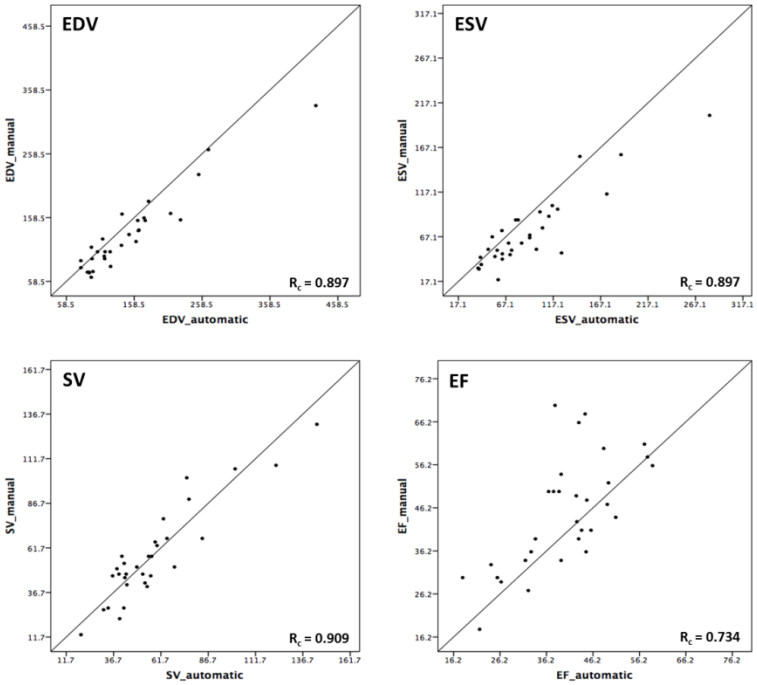


Figure 1 Correlation plots with Lin's correlation coefficient (Rc) for EDV (top-left), ESV (top-right), SV (bottom-left), EF (bottom-right)

The Bland-Altman plot [AltmanBland1983] provides further graphical evaluation of agreement between two models by plotting differences between pairs of computed measurements against their averages. Figure 2 shows the Bland-Altman plots for each of the four measurements described in Table 1, which confirm the high degree of agreement between the two models across all measurements.

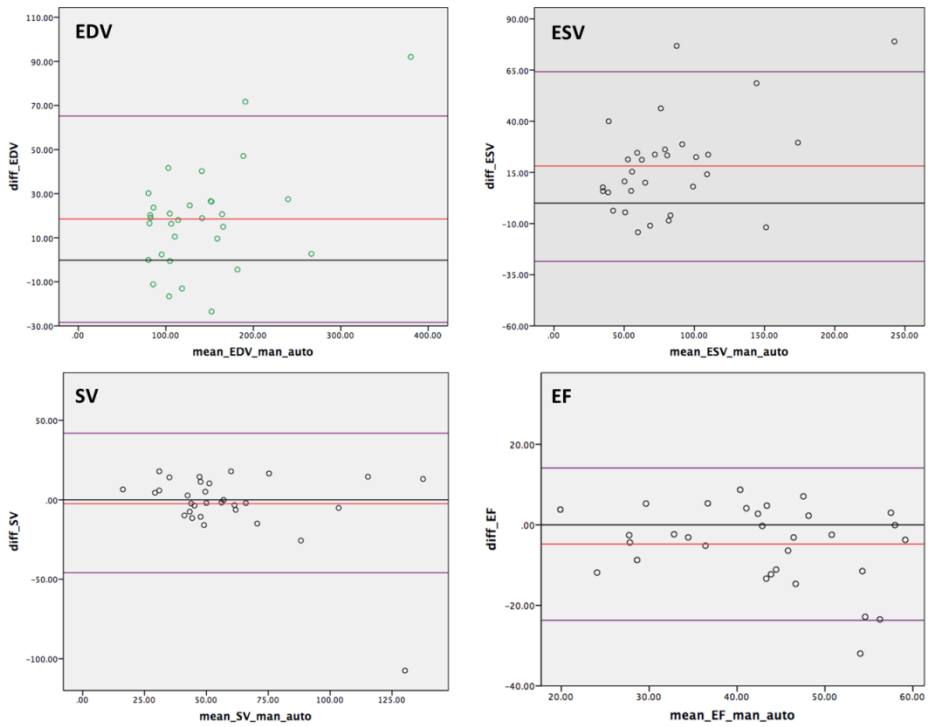


Figure 2 Bland-Altman plots with the mean (red line), limits of agreement (purple line) and zero difference (black line) for EDV (top-left), ESV (top-right), SV (bottom-left), EF (bottom-right)

The differences between pairs of measurements were found to be normally distributed except for EF, where a negative skew was observed. Thus, a standard t-test could be used to check for systematic bias. The results are depicted in Table 2 and indicate that the automated model tends to overestimate EDV and ESV. This could be due to the fact that the model was trained on healthy adult hearts and adapted to work with paediatric hearts with various cardiomyopathies.

Table 2 Systematic error analysis using t-test (*may not be reliable due to negative skew in distribution)

Measure	t-test mean	t-test std. dev.	p-value	Comment
EDV	18.4	23.9	0.000	systematic bias
ESV	18.3	23.9	0.000	systematic bias
SV	-2.0	22.4	0.550	no systematic bias
EF*	-4.78	9.66	0.10	no systematic bias

2.2 Evaluation of electro-mechanical model

As described in the fourth year annual report of WP8, we applied the electro-mechanical computational model personalization pipeline to a large number of patient datasets during the final months of MD-Paedigree. Among the 124 complete datasets (DHZB: 40, UCL: 55, OPBG: 29) that were available to SHC, the personalization

pipeline was able to generate successfully 107 patient-specific electro-mechanical heart models. Only 17 cases could not be fully processed due to various issues including insufficient MR image quality (unable to generate reasonable segmentation) or failure of one of the personalization steps. Overall, the high success rate of this multi-centre study indicates good versatility and robustness of the developed tools.

In order to validate the modelling tools and personalization pipeline, an important step is to test the ability of the model to accurately capture the physiology of the patient. To this end, we compared the outputs of the personalized models to the available clinical data. Several important clinical variables were analysed for all processed datasets in order to assess the goodness of fit (Figure 3).

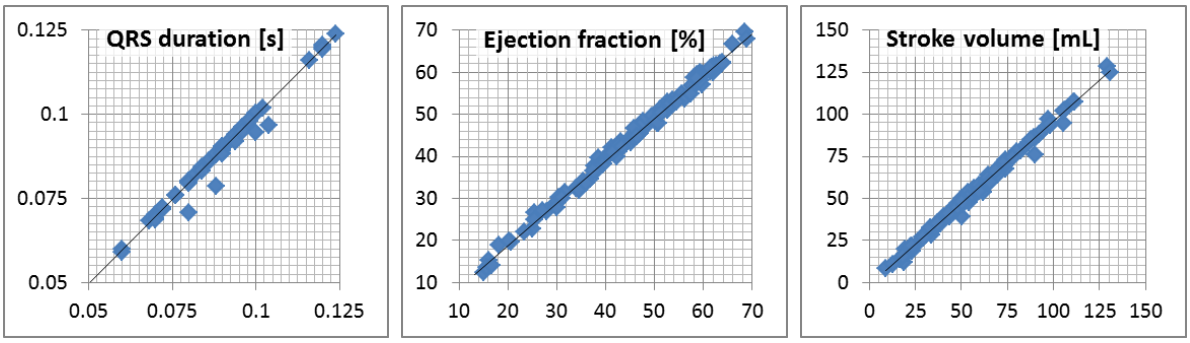


Figure 3: Measured (x-axis) vs. computed (y-axis) clinical variables. From left to right (with Pearson correlation coefficient PCC): QRS duration from 12-lead ECG (PCC: 99.1%), left ventricular ejection fraction (PCC: 99.7%) and left ventricular stroke volume (PCC: 99.7%).

The population-average "errors" (misfit between measured and computed variables) are listed in the table below (Table 3):

Table 3 Population-average "errors"

Quantity	QRS duration	Electrical axis	QT duration	Ejection fraction	Stroke volume
Mean	0.001 s	15.2 deg	0.003 s	1.15 %	3.25 mL
Std.dev.	0.002 s	17.0 deg	0.004 s	0.86 %	2.16 mL

The results are in agreement with preliminary findings presented last year on 35 patients (D12.2): the goodness of fit between measured and computed variables is high and the degree of misfit is mostly below clinical variability. The only "outlier" is electrical axis, where for about one fourth of the datasets no good fit could be achieved (misfit > 30°). The main reasons are potentially invalid modelling assumptions of the cardiac electrophysiology model (for certain pathologies) and imprecise torso fitting (required for simulated ECG calculation) when no whole-body MR images were available. However, the overall positive results suggest that the personalized electro-mechanical model can accurately reproduce several of the most significant indicators of cardiac physiology, even for such a large and diverse patient population.

2.3 Evaluation of consensus for electro-mechanical models

The patient-specific electro-mechanical modelling pipelines developed by Siemens (SHC) and INRIA have been presented in previous deliverables, see D8.2 for an in-depth description. Each of these pipelines involves

estimation of various parameters, with some relevant parameters described in Table 4. Each of the pipelines was used to compute patient-specific biomechanical models for 34 cases across multiple centers (including 4 cases with follow-up data available) and an analysis of parameter agreement was performed.

Table 4 Description of relevant parameters estimated by biomechanical modelling pipeline

Model	Parameter	Description
Siemens	s0_LV / s0_RV	maximum active contraction force for myocardium cells in ventricles
	HO_factor	related to passive myocardium tissue properties ("stiffness")
	krs_LV / krs_RV	myocyte relaxation rate
	katp_LV / katp_RV	myocyte activation rate (controls how fast the maximum force s0_LV/s0_RV is reached upon initial activation)
INRIA	Contractility	maximal myocardial contractility
	Stiffness	passive myocardial stiffness
	Peripheral Resistance	peripheral resistance parameter of the Windkessel model
	Compliance	arterial compliance parameter of the Windkessel model
	Central Venous Pressure	mean venous pressure parameter of the Windkessel model

Correlation within parameters estimated by each method is shown in Figure 4 and displays expected patterns based on the parameter definitions and interactions described in previous deliverables.

	s0_LV	s0_RV	HO_factor	krs_LV	krs_RV	katp_LV	katp_RV
s0_LV	1.00	0.36	-0.40	0.07	-0.12	0.16	0.12
s0_RV	0.36	1.00	-0.25	-0.05	0.07	0.16	-0.07
HO_factor	-0.40	-0.25	1.00	0.29	0.39	-0.37	-0.32
krs_LV	0.07	-0.05	0.29	1.00	0.46	-0.52	-0.67
krs_RV	-0.12	0.07	0.39	0.46	1.00	-0.34	-0.42
katp_LV	0.16	0.16	-0.37	-0.52	-0.34	1.00	0.73
katp_RV	0.12	-0.07	-0.32	-0.67	-0.42	0.73	1.00

	Contractility	Stiffness	Peripheral Resistance	Compliance	Central Venous Pressure
Contractility	1.00	-0.11	-0.04	0.21	0.32
Stiffness	-0.11	1.00	-0.03	-0.28	0.38
Peripheral Resistance	-0.04	-0.03	1.00	-0.63	-0.35
Compliance	0.21	-0.28	-0.63	1.00	0.20
Central Venous Pressure	0.32	0.38	-0.35	0.20	1.00

Figure 4 correlation measures between parameters estimated within each biomechanical model with high (green) and moderate (yellow) correlations indicated.

However, more interesting are correlations between similar parameters across models. These are the contractility and stiffness parameters as highlighted in Table 4. The scatter plots in Figure 5 show the correlation for standardized values of contractility (top) and stiffness (bottom) between the Siemens and INRIA models. The contractility measurement shows a strong correlation with a Pearson's correlation coefficient of 0.81, while the stiffness measurement exhibits weaker correlation with a Pearson's correlation coefficient of 0.31. The weak agreement between the two models for the stiffness parameter can be explained by the fact that this parameter is interpreted and estimated very differently for both models. The agreement to various degrees for both these parameters is encouraging and further investigation is required to discover more such similarities between the two models.

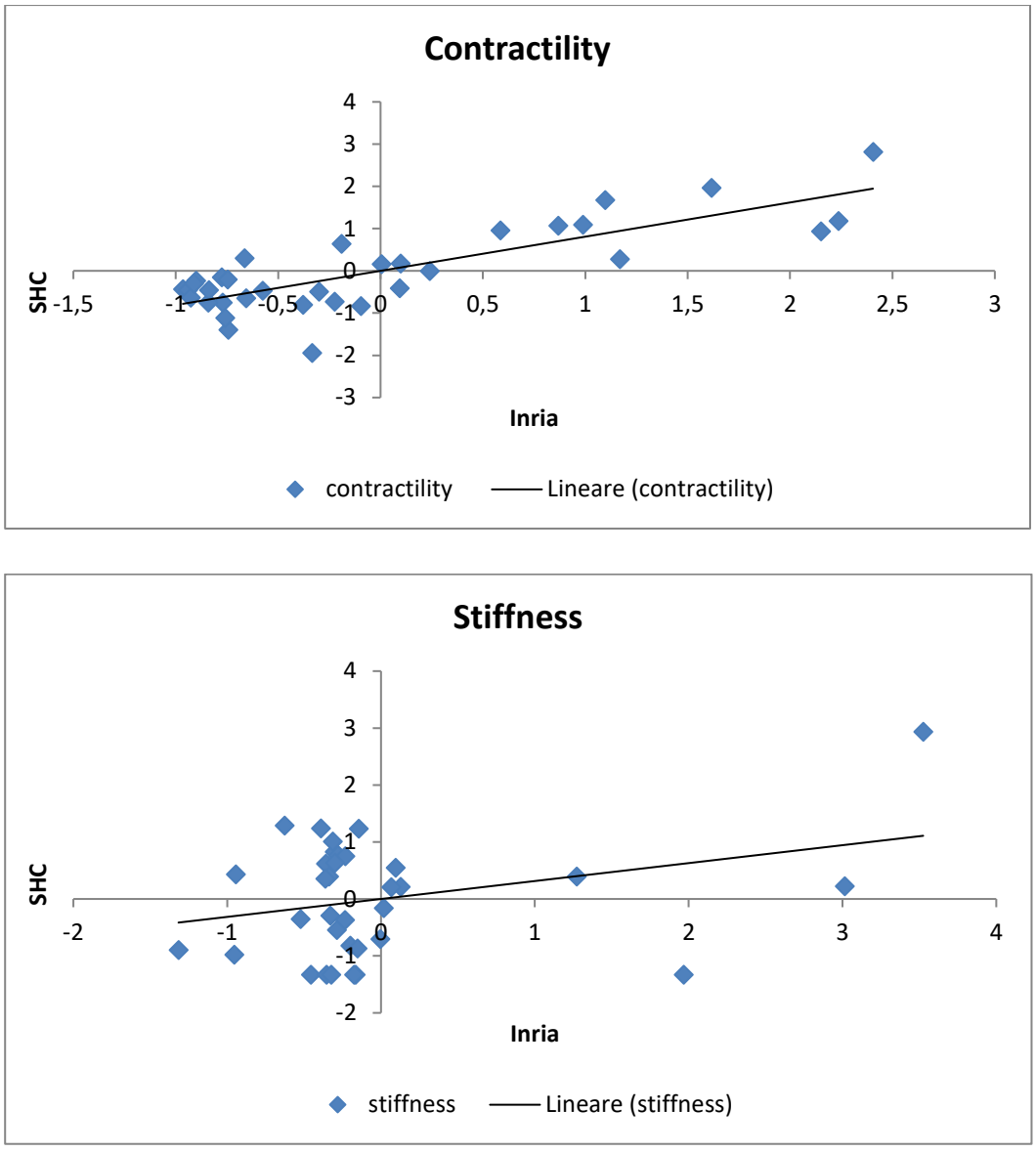


Figure 5 scatter plots for contractility (top) and stiffness (bottom) parameters estimated by the SHC and INRIA models.

Follow-up data for four cases was analysed for agreement in the trend of the contractility parameter between the two models since it exhibited strong correlation in the cross-sectional analysis. The trends are displayed in Figure 6 and are in agreement 75% of the time. Various factors including intervention or another medical event could be responsible for disagreement in the trend for P19. This experiment is encouraging and indicates that further rigorous analysis should be performed with extensive follow-up data, which is being processed by both Siemens and INRIA at this time.

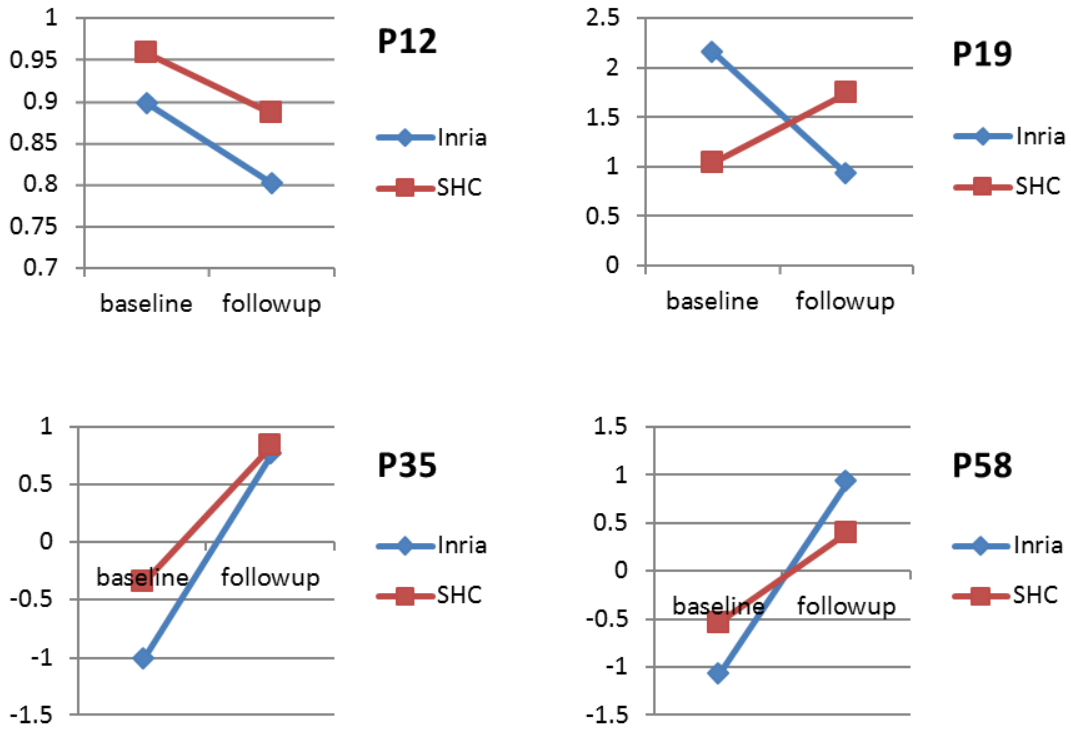


Figure 6 Comparison of trend for the contractility parameter estimated using both models at baseline and follow-up across 4 available cases.

2.4 Evaluation of the whole body circulation model

In the fourth year of the project we have further refined the personalization algorithm of the lumped parameter whole body circulation model (see deliverable D8.3 for a detailed description of this model). The same set of 28 objectives are used (12 standard and 16 advanced), and the initialization procedure used to determine the initial values of the parameters to be personalized was improved so as to reduce the number of iterations required for convergence.

The lumped parameter whole body circulation model and the personalization approach were validated based on the data extracted from 18 patients, both at baseline and at follow-up (8 new patients compared to the dataset used last year, see D12.2.2). The patient datasets stem from all three clinical partners (10 OPBG, 4 DHZB, 4 UCL). 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.

2.4.1 Results at baseline

In the following, we present the results obtained for the personalized blood flow computations at baseline. Since the ejection fraction was determined by manual measurements performed by local 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. **Errore. L'origine riferimento non è stata trovata.** displays 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.

Table 5 further displays the correlation and the mean absolute differences between computed and measured quantities: the results further underline that the model was successfully personalized.

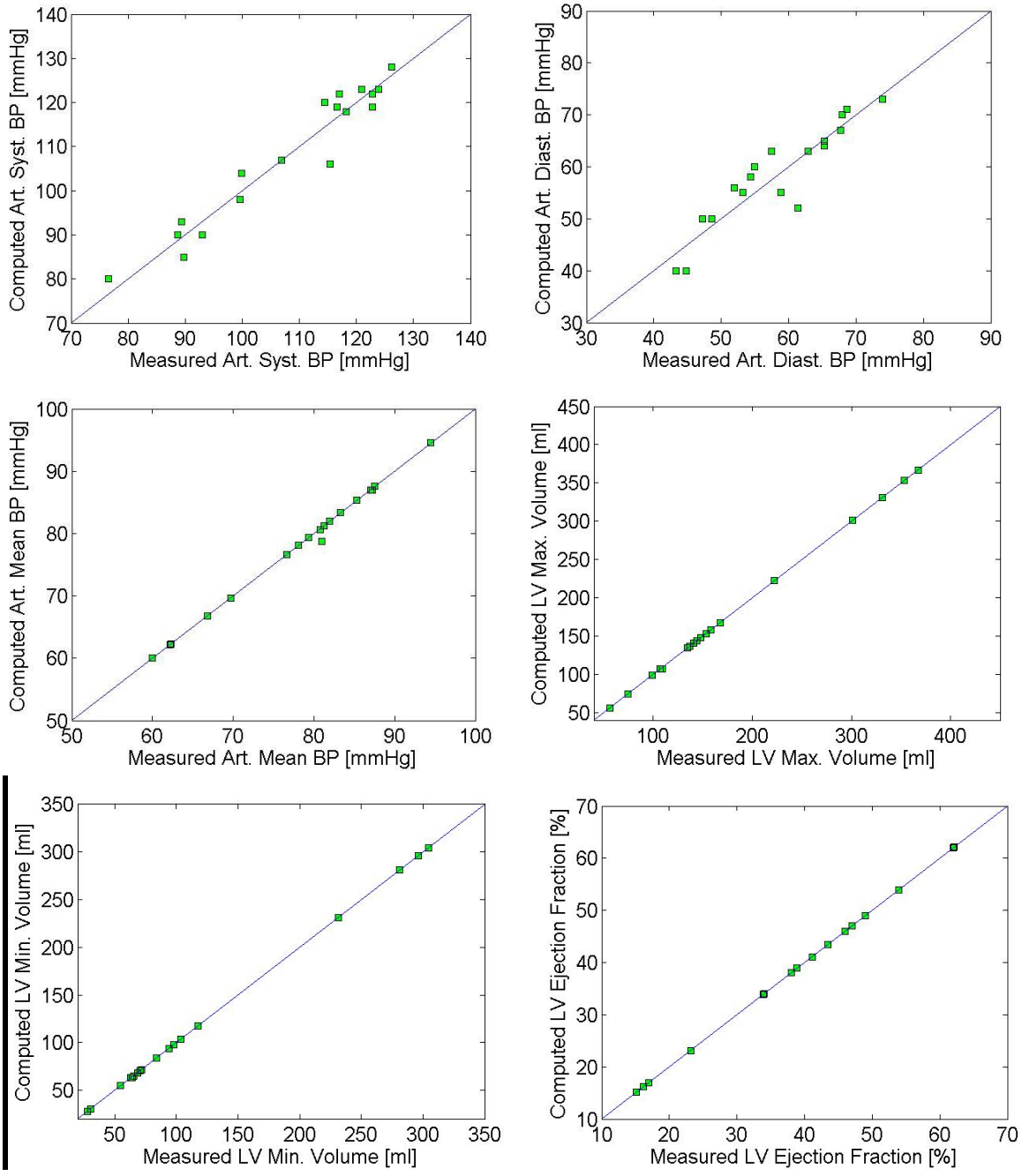


Figure 7 Scatter plots of computed versus measured values of arterial systolic, diastolic and mean pressure, maximum and minimum left ventricular volume, and left ventricular ejection fraction at baseline.

Table 5 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	0.969	0.919	0.999	1.0	1.0	1.0
Mean abs. Diff.	3.001 mmHg	2.964 mmHg	0.155 mmHg	0.195 ml	0.106 ml	0.14

2.4.2 Results at follow-up

In the following 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.

Errore. L'origine riferimento non è stata trovata. displays a comparison of measured and computed ejection fraction. As can be observed, the ejection fraction prediction performs well except for two cases, for which the error is large. The reason for the discrepancy is currently being investigated together with the clinical partners.

If the two outliers are removed, the computational model is able to predict well the ejection fraction at follow-up, with a correlation of 0.892 and a mean absolute difference of 4.78%. A more detailed analysis reveals that the ejection fraction has changed between baseline and follow-up exam for thirteen out of the sixteen remaining patients included in the study. Out of these thirteen patients, the model has predicted correctly the direction of change in ejection fraction for ten patients.

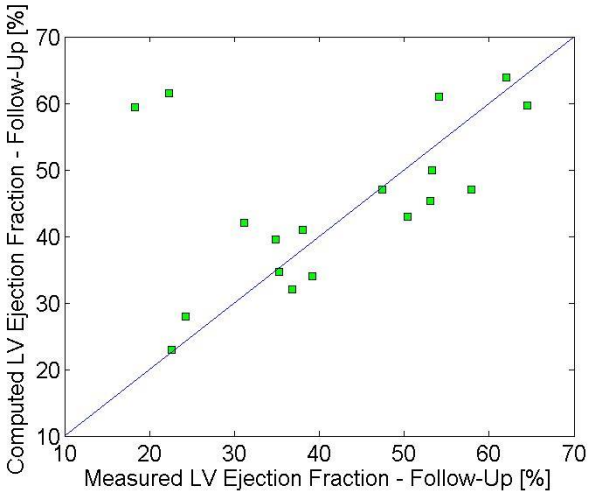


Figure 8 Scatter plots of computed versus measured values left ventricular ejection fraction at follow-up.

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.

2.5 Evaluation of 3D Heamodynamics Model

The 3D haemodynamics computation model built in this project can generate very detailed spatio-temporal information of intra-cardiac velocities and absolute pressures, however care is required to establish quantities that are robust to noise, reconstruction and modelling variations. For example, while averaged computed quantities like stroke volume, EDV or ESV have been matched to their corresponding measurements from 4D flow PC-MRI with an excellent accuracy, accurately matching pointwise quantities like velocity is a task of a completely different complexity. With this in mind, both globally-averaged quantities like kinetic energy (KE), as well as regional quantities, like the newly-developed 7-region velocity averages (see deliverable D8.3), were considered. Qualitative analysis (not presented here) also established visual correlations between haemodynamic structures like the mitral toroidal vortex or the end-diastolic vortex.

Unfortunately, the available data sample (6 patients: 5 OPBG and 1 GOSH) was too small to carry significant statistics on it. The limited size of the sample was due to either the lack of concurrent availability of both 4D flow PC-MRI as well as 3D echo data, or (in a lower percentage) to poor 3D echo data quality, which precluded a meaningful segmentation of the left cardiac valves.

Analysing the available data, we noted a good correlation between the peak systolic KE between PC-MRI and CFD computations (average 2.0±1.6mJ vs 1.7±1.5mJ, Pearson correlation r = 0.94). A lower correlation was observed between the peak diastolic KE between PC-MRI and CFD computations (average 1.3±1.0mJ vs 1.2±0.9mJ, Pearson correlation r = 0.32). This is likely due to the extra flow complexity introduced by the mitral valve apparatus and papillary muscles, which interfere much more with the flow during diastole.

The Bland-Altman plots in Figure 9 show that the mean differences between the KE from computations and data are relatively small: 0.3mJ over a mean of 1.8mJ for systole, respectively 0.1mJ over a mean of 1.3mJ for diastole. KE indices may prove valuable in patient stratification [Kanski2015], although they have not found a validated clinical use yet.

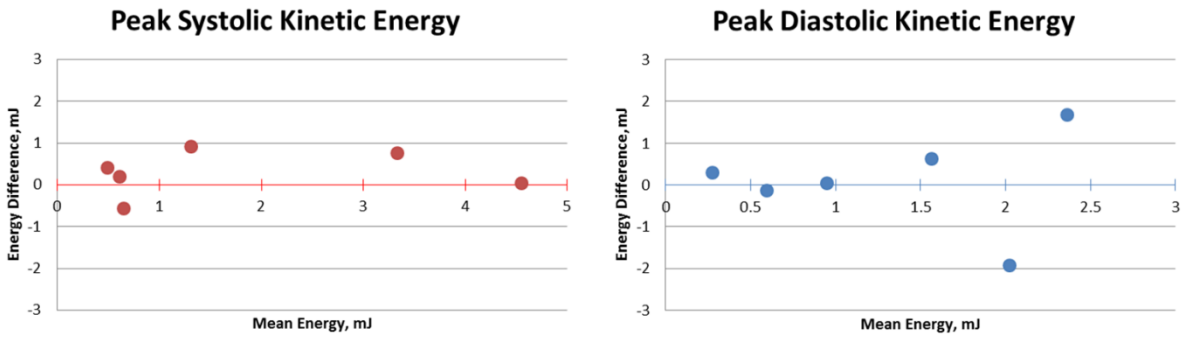


Figure 9 Bland-Altman plots of the differences vs averages of the Kinetic Energy inside the left ventricle, as computed from MR averaged velocities and from CFD averaged velocities. Data comes from six patients, systolic peak KE on the left, diastolic peak KE on the right.

Particularly interesting was the analysis of regional haemodynamics, which was performed using the 7-region intra-

cardiac segmentation from D8.3. A main feature of ventricular haemodynamics has been described (e.g. [Kilner2000]) to be the sling-like trajectory that the blood particles follow from diastole to systole. Our 7-region map is built having in mind both cardiac morphology features as well as this hemodynamic "fast-line" for ventricular flow, which is shown as the dashed curve in Figure 10. Averaged regional velocities were computed for both CFD and PC-MRI (Figure 11), and the differences between at either peak systole or peak diastole them were less than 6cm/s on average. To put things in perspective, maximum velocity was usually in the order of 150cm/s.

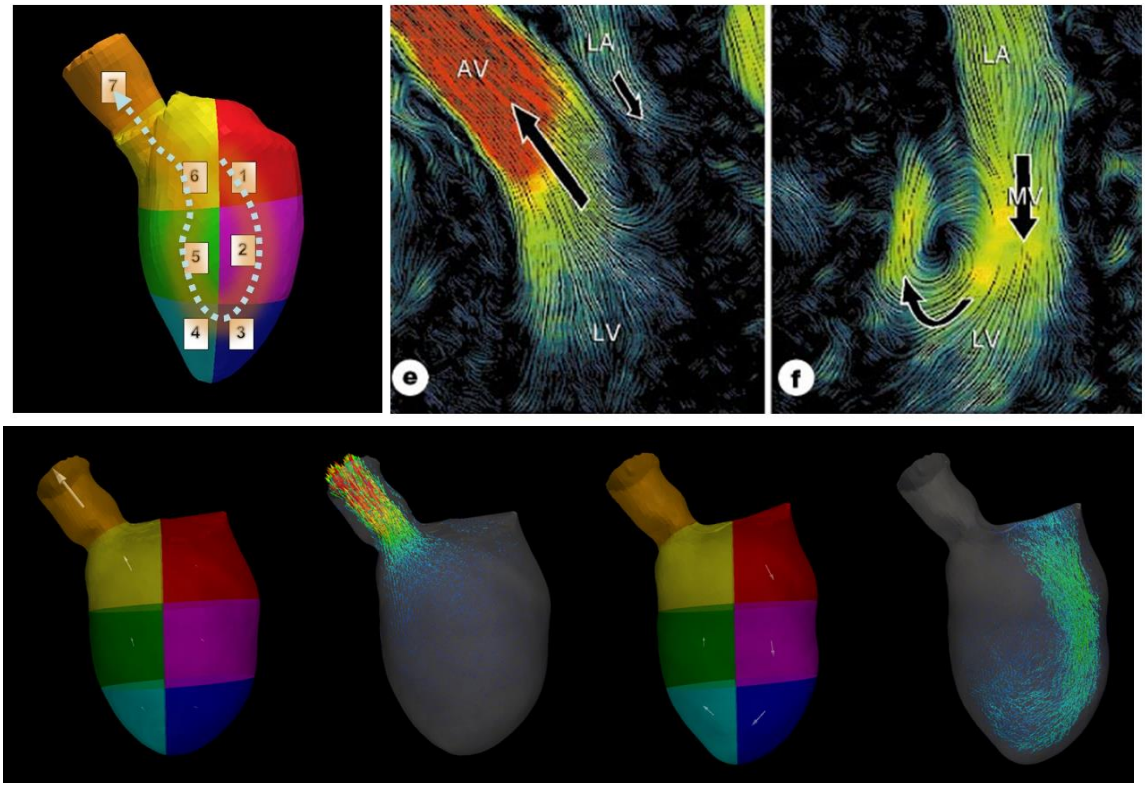


Figure 10 TOP: our novel flow-aware 7-region map of the left ventricular (LV) cavity is shown in the left image. The numbering of the segments follows the natural direction of blood flow inside the LV at peak systole and diastole, which resembles a sling-like trajectory (other images). BOTTOM, from left to right: CFD velocities at peak systole region-averaged and full 3D, CFD velocities at peak diastole region-averaged and full 3D.

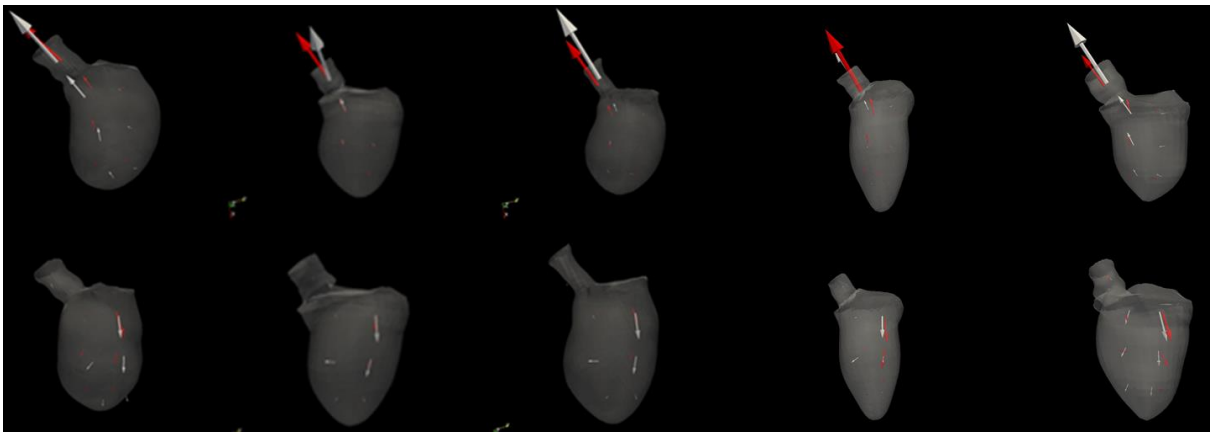


Figure 11 Regional velocity comparison between CFD (red) and PC-MRI (white) for five OPBG patients. Peak systole is shown on the top and peak diastole is shown on the bottom. The mean differences are 6cm/s for both peak systole and diastole.

As a conclusion, the 3D haemodynamics model is able to produce data that reasonably matches averaged information based on the measured PC-MRI flow velocities, while providing extra haemodynamics knowledge like intra-ventricular pressures or haemodynamic stresses. This offers a promising avenue to the extraction of clinically relevant haemodynamic features.

2.6 Statistical model on CMD

In statistics, logistic regression, or logit regression, or logit modelis a regression model where the dependent variable (DV) is categorical. This article covers the case of a binary dependent variable that is, where it can take only two values, "0" and "1", which represent outcomes such as pass/fail, win/lose, alive/dead or healthy/sick. Cases where the dependent variable has more than two outcome categories may be analysed in multinomial logistic regression, or, if the multiple categories are ordered, in ordinal logistic regression.

In this report, we will talk about the concentration of ejection fraction in a cohort, and then present a visualization of the parameters. Principal component analysis will allow us to seperate the variables and the patients in the cohort depending on the relevant axis. We will apply the logistic model on the variable EFm_LV after having it transformed into binary variable. The validity of the models will be argued by tests and eventually, we will explain the process of cross-validation and model selection.

List of Abbreviations

Electrophysiology

- c_myo: related to myocardial electrical conduction velocity
- c_LV / c_RV: fast left / right endocardial conduction velocity (a simple way to "mimic" Purkinje network)
- APD: related to action potential duration

Mechanics

- s0_LV / s0_RV: maximum active contraction force for myocardium cells in left / right ventricle
- HO_factor: related to passive myocardium tissue properties ("stiffness")
- krs_LV / krs_RV: myocyte relaxation rate
- katp_LV / katp_RV: myocyte activation rate (controls how fast the maximum force s0_LV/s0_RV is reached upon initial activation)
- EFm_LV : related to left ventricular ejection fraction

Comparison of EFm LV with average amount 50%

Here we use t-test, called also Student test to compare our set to the average amount of left ventricular ejection fraction which is 50.

Hypothesis of the test :

$_0$ = The difference between the mean of the sample and the average value is equal to 0 $^{\{}$ = The difference is lower than 0

```
##
## One Sample t-test
##
```

```
##  data:  CMD$EFm_LV

##  t = -3.7285, df = 65, p-value = 0.0004063

##  alternative hypothesis: true mean is not equal to 50

##  95 percent confidence interval:
##   41.10937 47.31162

##  sample estimates:

##  mean of x

##      44.2105
```

Given the p-value=0.0004063 lower than the threshold significance level = 0.05 the null hypothesis $_0$ must be rejected and we should accept that indicates that the difference is lower than 0. The test result is :

- 95% of EFm_LV values are between [41.10,47.31]

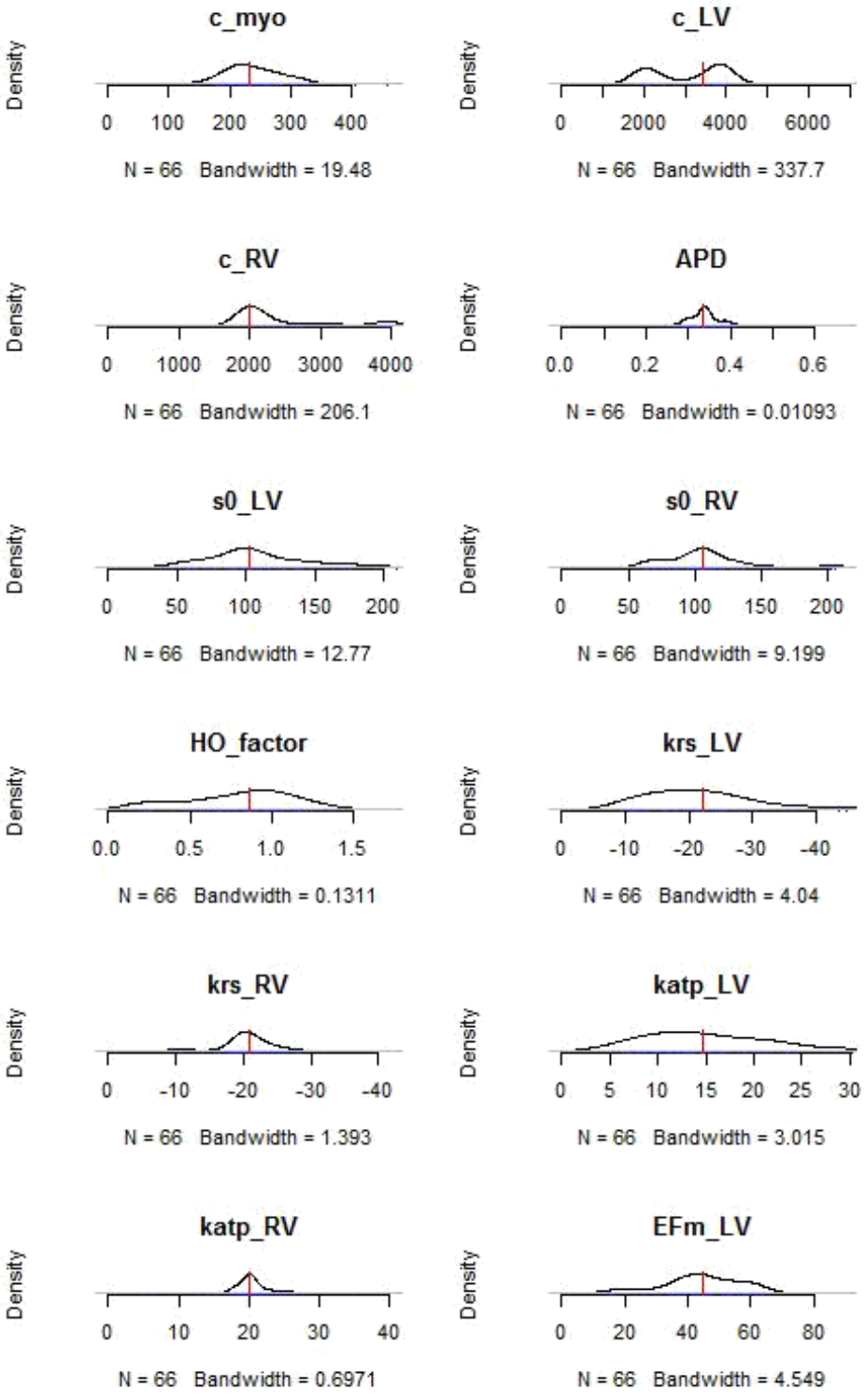
Densities of electrophysiology and mechanic variables

Here the electrophysiological and mechanical variables are plotted using a nonparametric estimation method, it is a method of smoothing the density using a kernel function.

$$\hat{} \; 1 - {}_{h}( ) = \overline{{}_{h} \sum ({}_{h})} \qquad \overline{\qquad\qquad}$$
$$=1$$

- the kernel function,
- the variable estimated
- h a parameter called bandwidth
- the size of the variable
- $_1$, … a sample of a random variable

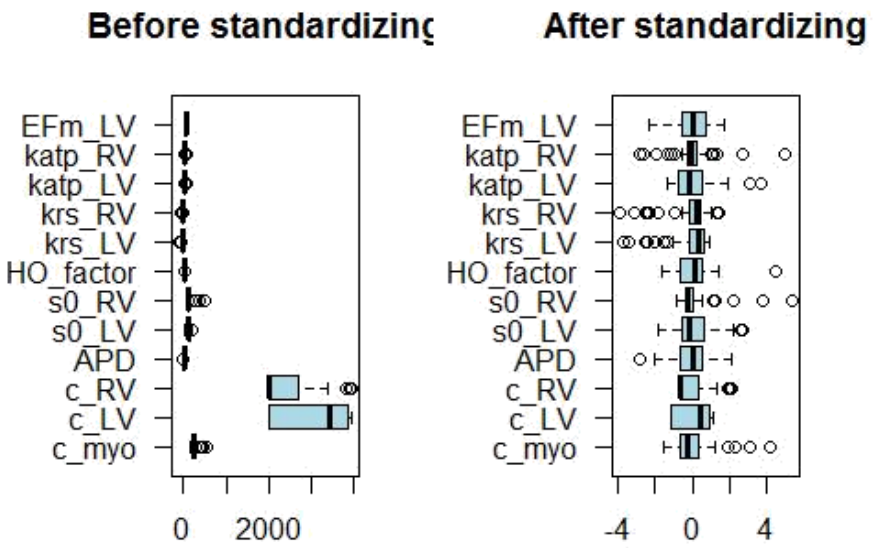
The red line indicates the median, it is the number that separates the variable into two groups of the same effective.


c_myo
Density
0 100 200 300 400
N = 66 Bandwidth = 19.48
c_LV
Density
0 2000 4000 6000
N = 66 Bandwidth = 337.7
c_RV
Density
0 1000 2000 3000 4000
N = 66 Bandwidth = 206.1
APD
Density
0.0 0.2 0.4 0.6
N = 66 Bandwidth = 0.01093
s0_LV
Density
0 50 100 150 200
N = 66 Bandwidth = 12.77
s0_RV
Density
0 50 100 150 200
N = 66 Bandwidth = 9.199
HO_factor
Density
0.0 0.5 1.0 1.5
N = 66 Bandwidth = 0.1311
krs_LV
Density
0 -10 -20 -30 -40
N = 66 Bandwidth = 4.04
krs_RV
Density
0 -10 -20 -30 -40
N = 66 Bandwidth = 1.393
katp_LV
Density
0 5 10 15 20 25 30
N = 66 Bandwidth = 3.015
katp_RV
Density
0 10 20 30 40
N = 66 Bandwidth = 0.6971
EFm_LV
Density
0 20 40 60 80
N = 66 Bandwidth = 4.549

Note :

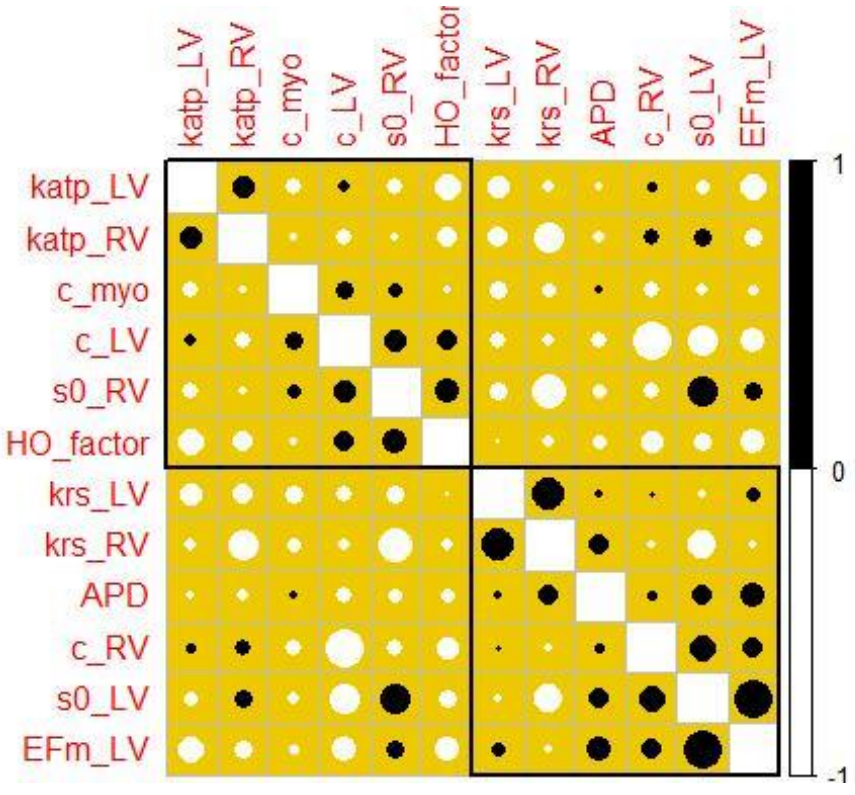
- The red line represents the median (ie 50% of patients on the left side and 50% on the right)
- The bandwidth is a measure of how closely you want the density to match the distribution.

We'll start by doing a standardization on the variable. In this case, the analyzed table is centered and reduced by column, which corresponds to the following transformation.



Principal Component Analysis (PCA)

Correlations



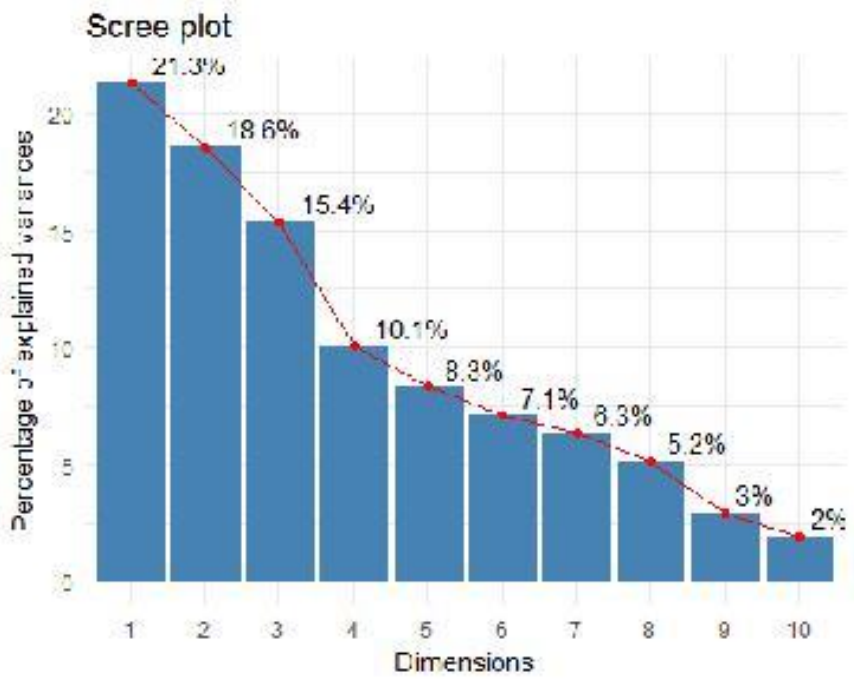
- Positive correlation between EFm_LV and s0_LV, APD, krs_LV.
- Negative correlation between EFm_LV and HO_factor, katp_LV, katp_RV.
- Using a hierarchical clustering function we have two groups of variables :

  - group 1 : katp_LV, katp_RV, c_myo, c_LV, s0_RV, HO_factor
  - group 2 : krs_LV, krs_RV, APD, c_RV,s0_LV, EFm_LV

In the majority of the cases, we notice that group 1 is negatively correlated to group 2. Which means that a sligh positive variation in a variable from the group 1 results in a sligh negative variation in the group 2.

Eigenvalues of the correlation matrix / Choice of number of axis

We aim here to obtain the maximum inertia conserved with the minimum of factors (axis). For this we use Kaiser criterion that leads to retain the axis explaining the maximum of inertia.

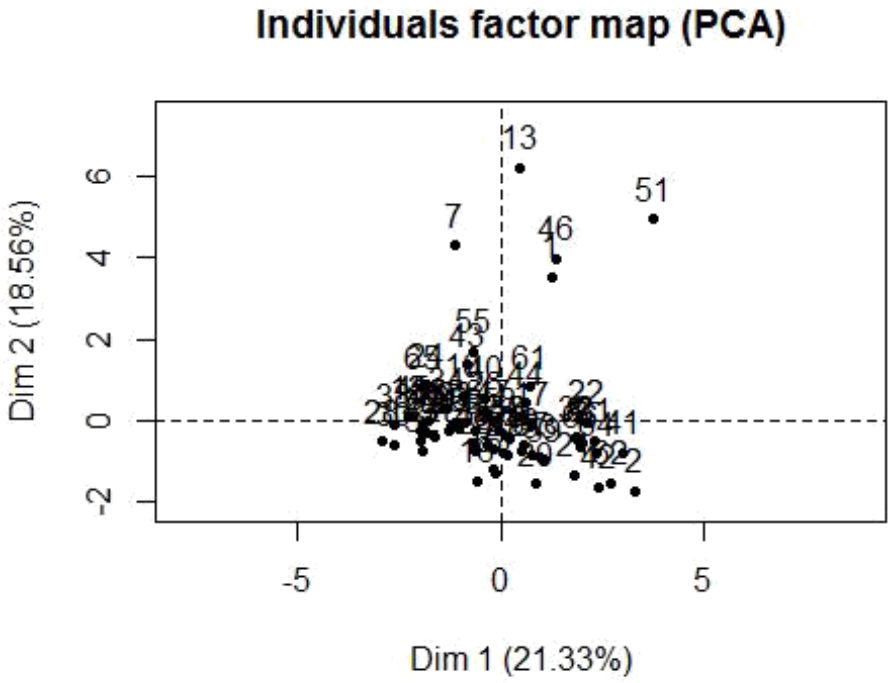


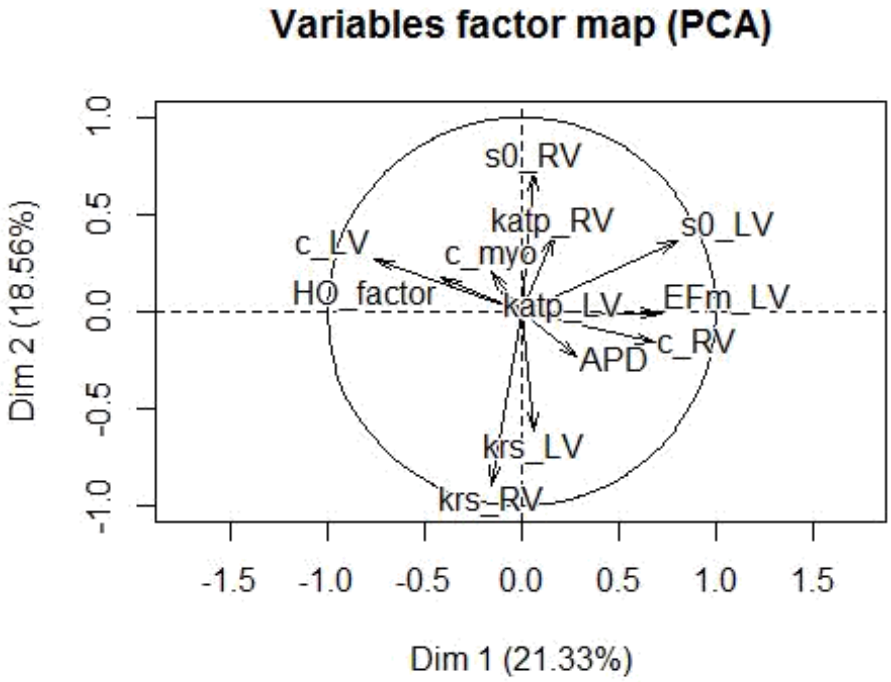
The first five axis represent 73.7% of the available informations. We realize here that if we can only stick to the first two factors with 39.9% of information which is a low quantity of information.

Using the table above and Kaiser criterion we select 5 axis, which retain 73.7% of the total inertia. The axis 1 retains 21.3% of the total inertia, however, for axis 2 which holds 18.6%, axis 3: 15.4%, axis 4: 10.1% and axis 5: 8.3%. The selection of 5 axis is relevant, and leads to an explained rate of inertia of 73.7%, which is a good result. It may therefore be interesting to study it as well. Except that the

disadvantage is that we have 5 axis on which we will represent our data which is not practical.

It is known that the variables most contributing to the formation of axis 1 are those whose coordinates on this axis are close to 1 in absolute value. To identify significant contributions, we use the vector below.





We compare the values of the PC1 (first Principal Component), the coordinates of the first factorial axis to the average contribution $\frac{1}{n}$ = 0.2672 (26.72%), the sign is giving the meaning of contribution :

(-)	(+)
c_LV	s0_LV
HO_factor	EFm_LV
()	c_RV

The first axis opposes EFm_LV,s0_LV, c_RV and APD to HO_factor and c_LV (passive myocardium tissue properties and left endocardial conduction velocity).

The variables APD and c_myo are very close to the average contribution and we can integrate them into the interpretation of the axis if they go in the direction of the interpretation that can be made without it. Mostly the axis 1 opposes left endocardial conduction velocity to active contraction force for myocardium cells in left ventricle.

Similarly, the individuals contributing most to the formation of axis 1 are those who's coordinates on this axis are the highest in absolute value.

We can get an idea of the quality of representation of a variable on an axis by directly reading the graph : a well-represented variable is close to the edge of the circle of the correlations and close to the axis.

We compare these values which represent the coordinates of individuals on axis 1, to the root of the first eigenvalue $\sqrt{2.5999606}$ = 1.6124. The sign is giving the meaning of contribution.

There are therefore two classes of patients:

(-)	(+)
28,30,38,52,45,33,29,35,54,39,24,65,60,15	51,13,46,1,41,22,64,62
23,16,25,11,10,5,56,36,26,6,27,57,19,43	
47,9,58,49,20,37,53,44,59,7,4,61,3,42	

We compare the values of the column PC2 (first Principal Component), the coordinates of the first

factorial axis to the average contribution 26.72%, the sign is giving the meaning of contribution :

```
##                Dim.1      Dim.2
##  c_myo -0.15714139  0.2048186
##  c_LV  -0.75563466  0.2717191
##  c_RV   0.69053331 -0.1618105

## APD     0.28719337 -0.2352504
##  s0_LV  0.80444774  0.3647747

##  s0_RV  0.07066679  0.7209244
```

(-)	(+)
krs_RV	s0_RV
krs_LV	katp_RV
c_RV	(s0_LV)

Given the difference in contribution between these three variables, the high contribution of s0_RV (23.33%), katp_RV (6.92%) and s0_LV (5.97%) positively and krs_RV (35.80%), krs_LV (17.24%) negatively.

The second main component can be considered as essentially related to maximum active contraction force for myocardium cells in right ventricle and myocyte relaxation rate.

Patients :

To identify individuals with a significant contribution, we use the table below. We compare these values which represent the coordinates of individuals on axis 2, with the root of the second eigenvalue

$\sqrt{2.752}$

```
##      Dim.1      Dim.2
## 1 1.2694076  3.5337027
## 2 3.3125464 -1.7395853
## 3 2.0284842 -0.5494104
## 4 1.8531803 -0.4403119
## 5 -0.6383108 -0.2675103
## 6 -0.1490152 -1.2928868
```

There are therefore two classes of patients

(-)	(+)
2,3,4,5,6, 8, 9,10,11,12,14,15,16,17,18,19,20,21,63,64	1,7,13,46,51,55
28,29,30,31,32,33,34,35,36,37,38,39,23,24,25,26,27,22,61,62	
40,41,42,43,44,45,47,48,49,50,52,53,54,56,57,58,59,60,65,66	

Regressions

We want to explain **EFm LV**, that corresponds to left ventricular ejection fraction, by continious measurements as electrophysiological and mechanical variables. We have demonstrated above that MRI left ventricular ejection fraction is a normal variable, so we could fit a generalized linear model using a gaussian family that describes the error distribution and link function used in the model.

The model we used here is represented by the system below :

$$\sim (\quad)$$

$$\{ \quad (\quad) = (\quad) = \quad + \quad + \cdots +$$
$$0 \quad 1 \ 1$$
$$1 -$$

Where *is the law of Y, that represents the MRI_LVEF.* is the link function that links between the expected value and the explanatory variables. * = ($_0$, $_1$) is the vector of the parameters to estimate.

Univariate regression

In statistics, univariate regression is a model with a single explanatory variable. It concerns two-dimensional sample points with one independent variable and one dependent variable. Conventionally, the x and y coordinates in a Cartesian coordinate system in order to predict the dependent variable values as a function of the independent variable. Univariate regression refers to the fact that the outcome variable is related to a single predictor.

Univariate analysis was applied to all electrophysiological and mechanical variables of the patients, in order to predict MRI left ventricular ejection. We find that only the variables s0_LV related to maximum active contraction force for myocardium cells in left ventricle and HO_factor related to passive myocardium tissue properties are highly significant.

First model (active contraction force for myocardium cells)

We represent our model by the system of equations below :

$$\begin{cases} \sim ( ) \\ ( ) = + 1 + \end{cases}$$

```
##
```

```
## Call:

## glm(formula = CMD$MRI_LVEF ~ CMD$s0_LV, family = binomial(link = "logit"),

##     data = CMD[, 3:13])
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.9964  -0.5351   0.3714   0.6676   2.4579
##
## Coefficients:
##                 Estimate Std. Error z value Pr(>|z|)
## (Intercept)      5.96700    1.42858   4.177 2.96e-05 ***
## CMD $s0_LV      -0.04594    0.01202  -3.820 0.000133 ***
## ---

## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

##
## (Dispersion parameter for binomial family taken to be 1)
##

## Null deviance: 84.020 on 65 degrees of freedom ## Residual
deviance: 57.798 on 64 degrees of freedom
## AIC: 61.798

##
## Number of Fisher Scoring iterations: 5
```

The p-values of $_0$ (Intercept) and **s0_LV** are highly significant their p-value respectively equal to 2.96e-05 and 0.000133 lower than 0.05 (5%). This means that the intercept and **s0_LV** are highly significant for the univariate model. The column Estimate gives the estimation of the parameters of the model, here we have $_0$ = 5.96 and $_1$ = −0.045.

$$[\ \ ] = 5.96 - 0.045\times\ \ \_$$

And we note that there is a negative influence of active contraction force for myocardium cells in left ventricle on MRI left ventricular ejection.

Wald test

We want to test here the hypothesis :

$$\begin{cases} H_0 : \beta_1 = 0 \\ H_1 : \beta_1 \neq 0 \end{cases}$$

And The test statistic is :

$$W = (\hat{\beta}_1 - 0)^2 \left[\widehat{Var}(\hat{\beta}_1) \right]^{-1} (\hat{\beta}_1 - 0) \sim \chi^2_1$$

with degrees of freedom 1 estimates are involved.

```
## Analysis of Deviance Table

##
## Model: binomial, link: logit

##
## Response: CMD$MRI_LVEF
##
## Terms added sequentially (first to last)

##
##
##                  Df Deviance Resid. Df Resid. Dev  Pr(>Chi)

## NULL                              65     84.020
## CMD $s0_LV  1   26.222             64     57.798     3.043e-07 ***
## ---

## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The p-value here is lower than 0.05, which means than we should accept the alternative hypothesis, that the parameter depending on the variable **sO_LV** is significant in our model. By applying the likelihood ratio test we have the same results, and the variable **s0_LV** still significant.

Second Model (passive myocardium tissue properties)

We represent our model by the system of equations below :

$$\begin{cases} \sim (\ ) \\ (\ ) = {}_0 + {}_1 \times \ \_ \end{cases}$$

```
##
```

```
##  Call:
##  glm(formula = CMD$MRI_LVEF ~ CMD$HO_factor, family = binomial(link = "logit"),
##      data = CMD[, 3:13])
```

```
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.9361  -0.7751   0.5065   0.7615   1.7104
##
## Coefficients:
##                  Estimate Std. Error z value Pr(>|z|)
## (Intercept)        -1.877      0.785  -2.391 0.016818 *
## CMD $HO_factor      3.387      1.002   3.381 0.000723 ***
##  ---
##  Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
##  (Dispersion parameter for binomial family taken to be 1)
##
```

```
## Null deviance: 84.020 on 65 degrees of freedom ## Residual
deviance: 68.877 on 64 degrees of freedom
##  AIC: 72.877
```

```
##
##  Number of Fisher Scoring iterations: 5
```

The p-values of intercept is highly significant itd p-value is less than 2e-16 (highly significant i.e lower than 0.05). Concerning the variable **HO_factor** its p-value is equal to 0.000723 lower than 0.1(10%) we can admit that it is highly significant. The column Estimate gives the estimation of the parameters of the model, here we have ${}_0 = -1.877$ and ${}_1 = 3.387$.

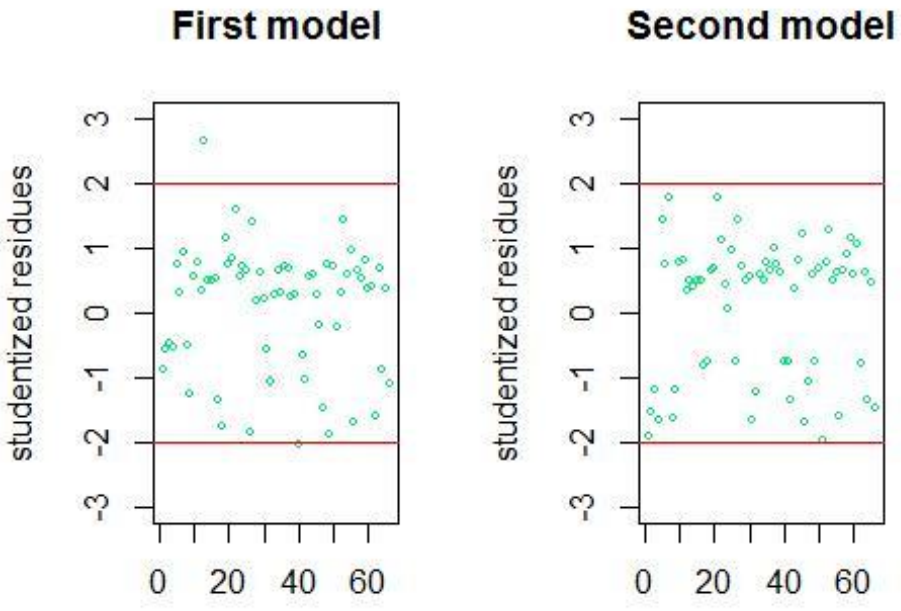
$$[\ ] = -1.877 + 3.387 \times \ \_$$

Wald test for the second model

```
## Analysis of Deviance Table

##
## Model: binomial, link: logit
##
## Response: CMD$MRI_LVEF

##
## Terms added sequentially (first to last)

##


##
##                      Df Deviance Resid. Df Resid. Dev                Pr(>Chi)
## NULL                                    65     84.020
## CMD $HO_factor     1   15.143            64          68.877 9.968e-05 ***
## ---

## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The p-value here is lower than 0.05, which means than we should reject 0 and accept h . The variable HO_factor is very significant in our model. By applying the likelihood ratio test we have same results, and the variable HO_factor still very significant.

After obtaining a model, it is necessary to diagnose the regression in order to validate or not the model. From this point of view, residue analysis is very important. It is important to note that in logistic regression, one is mostly interested in residues of deviance. They generally take values that oscillate between -2 and 2. A plot index is usually constructed to detect outliers (outside rows)



Multivariate regression & model selection

Multivariate regression is a technique that estimates a single regression model with more than one outcome variable. When there is more than one predictor variable in a multivariate regression model, the model is a multivariate multiple regression.

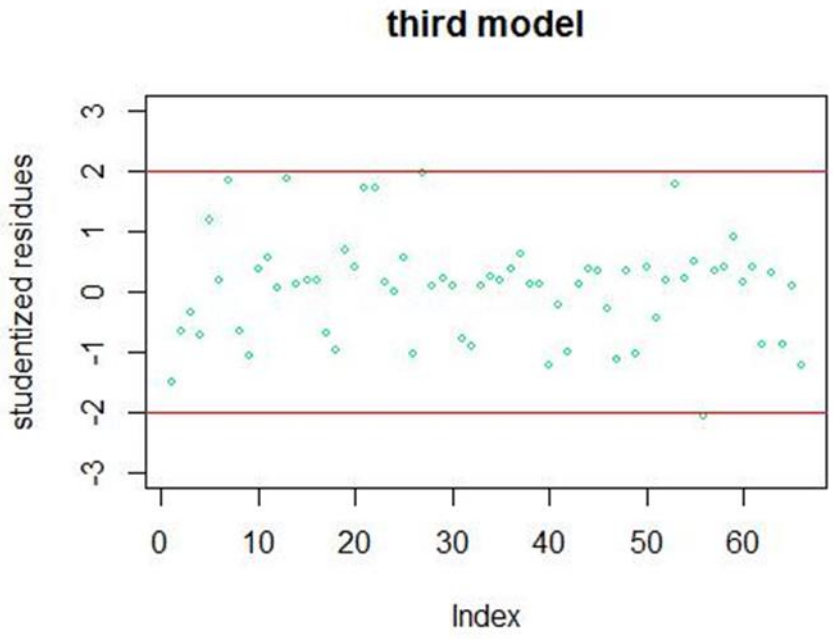
```
##
##  Call:

##  glm(formula = CMD$MRI_LVEF ~ CMD$s0_LV + CMD$HO_factor, family = binomial(link = "logit"),
##      data = CMD[, 3:13])
##
## Deviance Residuals:
##      Min        1Q    Median  3Q        Max
## -2.0100  -0.6422          0.1492  0.3884    1.9116
##
## Coefficients:
##                        Estimate Std. Error z value Pr(>|z|)
## (Intercept)          3.55472        1.60432    2.216 0.026711       *
## CMD$s0_LV          -0.04941        0.01354   -3.648 0.000264 ***
## CMD $HO_factor    4.29273        1.39409    3.079 0.002075       **
## ---

## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)

##
##      Null deviance: 84.020          on 65    degrees of freedom
## Residual deviance: 44.116          on 63    degrees of freedom
## AIC: 50.116
##

## Number of Fisher Scoring iterations: 6
```

The Column estimate gives the estimation of the parameter of the model ($_0$, $_1$, $_2$) We found that the intercept **s0_LV** and **HO_factor** are again significant, with a p-value respectively equal to 0.026711 0.000264 and 0.002075 lower than than 0.5 (5%). and the equation of the model is :

$$[\ \ ] = 3.55 - 0.049\times\ \ \_\ \ + 4.29\times\ \ \_$$

the variable related to maximum active contraction force for myocardium cells in left ventricle influe negatively while the passive myocardium tissue propertie influe positively.

also the suspicious structure. If a suspicious structuring appears, it may be appropriate to add a new variable in order to take into account this structuration. But in this case our model fits.



Wald test , likelihood ratio test and score test

Likelihood ratio test (LRT)

The test statistic is

$$-2\log(\ ) = -2\log(\ \frac{0}{1}\ ) = -2(\ \ -\ ) \rightarrow\ 2_{01}$$

where $_0$ is the maximized likelihood of the reduced model (subject to restrictions on the parameters according to), $_1$ is the maximized likelihood of the full (unrestricted) model, and = ? ? ? $_0$ is the difference of their numbers of parameters.

Score Test

The test statistic is :

$$= \left(\hat{\theta}_0\right) - \left(\hat{\theta}_0\right)^{-1} \left(\hat{\theta}_0\right) \rightarrow \chi^2_{,}$$

where is defined as above. Unlike in the LRT statistic, only restricted estimates are involved. #####

Application of the tests

```
## Analysis of Deviance Table
##
## Model: binomial, link: logit
##
## Response: CMD$MRI_LVEF
##
## Terms added sequentially (first to last)
##
##
##                          Df Deviance Resid. Df Resid. Dev  Pr(>Chi)
## NULL                                        65     84.020
## CMD$s0_LV           1      26.222              64     57.798 3.043e-07 ***
## CMD $HO_factor     1      13.682              63     44.116 0.0002165 ***
```

```
## ---
## Signif. codes:          0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
## Analysis of Deviance Table
##
## Model: binomial, link: logit
##
## Response: CMD$MRI_LVEF
##
## Terms added sequentially (first to last)
##
##
##                          Df Deviance Resid. Df Resid. Dev  Pr(>Chi)
## NULL                                        65     84.020
## CMD$s0_LV           1   26.222              64     57.798 3.043e-07 ***
```

```
## CMD $HO_factor  1   13.682              63     44.116 0.0002165 ***
```

```
##  ---

##  Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

##  Analysis of Deviance Table

##
##  Model: binomial, link: logit

##
##  Response: CMD$MRI_LVEF
##
## Terms added sequentially (first to last)
##
##
##                      Df Deviance Resid. Df Resid. Dev      Rao      Pr(>Chi)
## NULL                                    65     84.020
## CMD$s0_LV         1     26.222          64     57.798  22.9980  1.622e-06 ***
## CMD $HO_factor    1     13.682          63     44.116   9.9418   0.001616 **
##  ---

##  Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Odds-ratio

The odds ratio is used to measure the effect of a quantitative variable or the contrast between the effects of a qualitative variable. The general idea is to reason in terms of probabilities or odds ratio (odds).
The odds (chance) for an individual to get the answer = 1 is defined by:

$$( ) = \frac{( )}{1 - ( )}, \quad ( ) = (= 1 | \ = \ )$$

The odds ratio between two individuals and is $\hat{}$

$$\frac{( )}{( )} \qquad \frac{\hat{}}{( ) - ( )}$$

```
## Waiting for profiling to be done...

##                      OR        2.5 %        97.5 %
## (Intercept)      34.977884    1.8755914  1169.1183349
## CMD$s0_LV         0.951788    0.9222027     0.9737845
## CMD$HO_factor 73.165826       6.2936340  1701.1294488
```

Note

- The odd-ratio of the intercept is usually not interpreted.
- A confidence interval of the OR that contains 1 implies that the variable is not significant. That implies that **s0_LV** and **HO_factor** are highly significant in our model. *When using all the mechanical and electro-physiological variables in the multivariate model. Then the model gives two significant variables :
  - The intercept ($_0$)
  - **s0_LV**
  - **HO_factor**

The influence of s0_LV & HO_factor on the odds ratio

Risk comparison between two individuals with different values on maximum active contraction force for myocardium cells in left ventricle.
We denote the individual 1 and the individual 2 ^. Then the odds ratio concerning the variation of

maximum active contraction force for myocardium cells is :
((_ = 150), (_ = 100)) = 4

individual presents the disease than the individual ^.

^
we may say that it is 72 times more probable that the
((_ = 2), (_ = 1)) = 72

Likelihood ratio test

```
## Loading required package: zoo
```

```
##
##  Attaching package: 'zoo'
##  The following objects are masked from 'package:base':
##
##          as.Date, as.Date.numeric
##  Likelihood ratio test
##
##  Model 1: CMD$MRI_LVEF ~ CMD$s0_LV
##  Model 2: CMD$MRI_LVEF ~ CMD$s0_LV + CMD$HO_factor
##      #Df  LogLik Df  Chisq Pr(>Chisq)
##  1   2 -28.899
## 2   3 -22.058  1 13.682  0.0002165 ***
##  ---
##  Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##  Likelihood ratio test
##
##  Model 1: CMD$MRI_LVEF ~ CMD$HO_factor
##  Model 2: CMD$MRI_LVEF ~ CMD$s0_LV + CMD$HO_factor
##      #Df  LogLik Df  Chisq Pr(>Chisq)
##  1   2 -34.439
## 2   3 -22.058  1 24.761   6.49e-07 ***
##  ---
##  Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

By applying the likelihood test on the model using only **s0_LV** and the model using **s0_LV** and **HO_factor** we find that the second model is better than the first.

Same thing between the model using **HO_factor** and the model using **s0_LV** and **HO_factor**.

We Conclude that the multivariate model is the best to predict MRI left ventricular ejection fraction.

Logistic regression: Goodness of fit assessment with deviance

In statistics, deviance is a quality-of-fit statistic for a model that is often used for statistical hypothesis testing. It is a generalization of the idea of using the sum of squares of residuals in ordinary least squares to cases where model-fitting is achieved by maximum likelihood. Deviance is used for both goodness of fit and model comparison. The deviance for a model M0, based on a dataset y, is defined as

$$D(y) = -2(\log(p(y \mid \hat{\theta}_0)) - \log(p(y \mid \hat{\theta})))$$

```
##  model1    model2    model3
##  57.79777  68.87704  44.11598
```

The smallest deviance corresponds to the model third model which is represented by the equation below:

$$\hat{y} \sim B(n, \hat{p})$$

$$\text{logit}(\hat{p}) = 3.55 - 0.049 \times \text{Age} + 4.29 \times \text{Sex}$$

Confusion Matrix

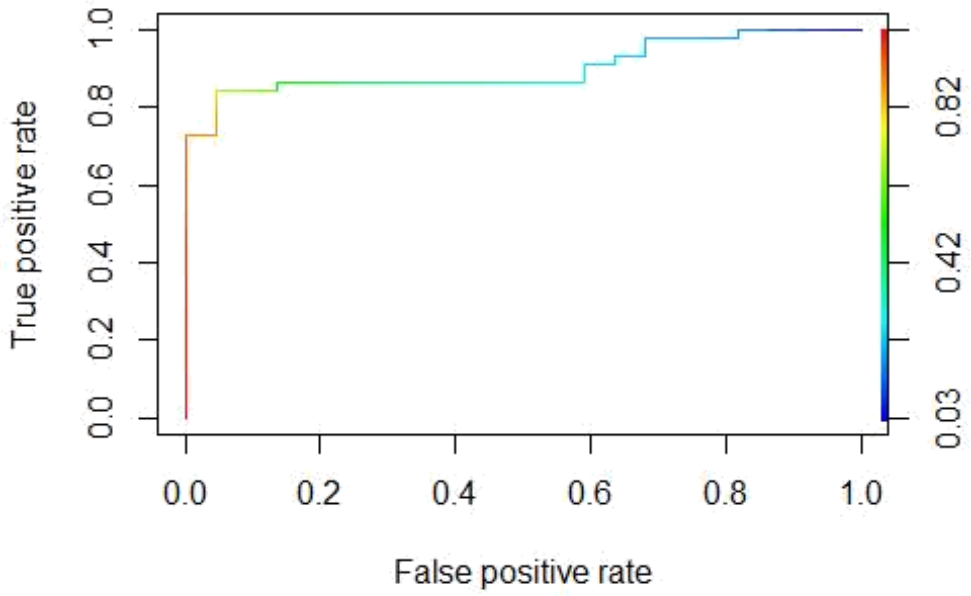
It is nothing but a tabular representation of Actual vs Predicted values. This helps us to find the accuracy of the model and avoid overfitting. It look like the table below :


```
##
```

```
##       FALSE TRUE
##  0      21    1
##  1      8     36
```

ROC curve

Receiver Operating Characteristic(ROC) summarizes the model's performance by evaluating the trade offs between true positive rate (sensitivity) and false positive rate(1- specificity). For plotting ROC, it is advisable to assume p > 0.5 since we are more concerned about success rate. ROC summarizes the predictive power for all possible values of p > 0.5. The area under curve (AUC), referred to as index of accuracy(A) or concordance index, is a perfect performance metric for ROC curve. Higher the area under curve, better the prediction power of the model. Below is a sample ROC curve. The ROC of a perfect predictive model has TP equals 1 and FP equals 0. This curve will touch the top left corner of the graph.

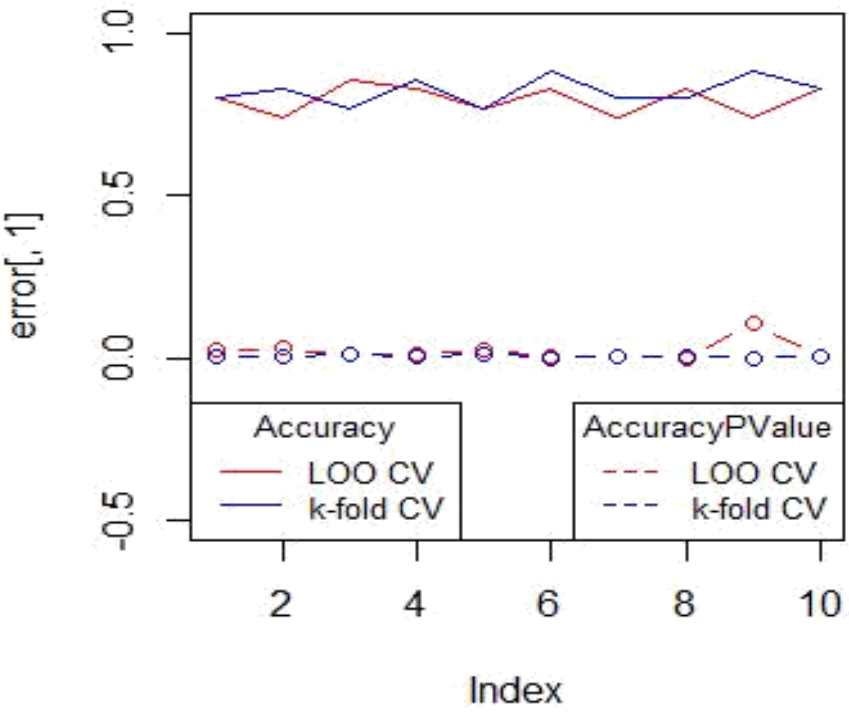


Note: For model performance, you can also consider likelihood function. It is called so, because it selects the coefficient values which maximizes the likelihood of explaining the observed data. It indicates goodness of fit as its value approaches one, and a poor fit of the data as its value approaches zero.

K-fold Cross Validation

The k-fold cross validation method involves splitting the dataset into k-subsets. For each subset is held out while the model is trained on all other subsets. This process is completed until accuracy is determine for each instance in the dataset, and an overall accuracy estimate is provided.

below find the accuracy while the validation using leave one out cross validation (LOOCV) and k-fold to evaluate the model.



2.7 Considerations and Conclusions

With more than 100 processed cases for anatomical and electro-mechanical models, these first stages of the modelling pipeline could be evaluated extensively in the project, and both provided very good accuracies compared to clinical measurements. A rare opportunity was the comparison of two independently developed electro-mechanical models (by SHC and INRIA, respectively), which was conducted for more than 30 cases until the preparation of this report, with more cases being processed until the final review. This comparison suggests that the contraction force (a parameter that cannot be measured clinically) can be estimated relatively reliably by both models. On the other hand, the myocardial stiffness provided by both models showed significant discrepancies. While this is most probably due to the different internal meaning of this parameter for both models, our conclusion is that for now, care should be taken to infer any clinical meaning from this parameter.

Unfortunately, relatively few complete follow-up datasets became available during the project, which made any validation of predictions difficult. Still, the developed whole-body circulation model showed promising results when predicting the ejection fraction for 16 out of 18 cases with high correlation to the clinical measurements. Overall, we regard this application as one of the best opportunities to improve patient care in cardiomyopathies with computational models.

The 3D haemodynamics model has the highest requirements regarding image acquisition, as several complex modalities and protocols need to be available with high quality for each study case. Due to these constraints, we could validate only 6 cases for this model. The results, overall, show that the same general flow patterns can be observed in both measurements and simulations, and regional flows can be estimated well. However, an accurate local (i.e. point-wise) estimation for the entire cardiac flow is still out of reach with the employed methods, and more research is required to get closer to this target.

Arguably, the single-most valuable achievement of the cardiomyopathies modelling area is the proof that patient-specific modelling can be performed on a large-scale with standard clinical data. This is a significant

difference from computational modelling in most other research settings, where special data acquisitions are used in order to provide the highest quality input data, and a lot of time is spent to tweak models for every individual case.

3.0 Validation for CVD risk in obesity (WP4-WP9)

3.1 Introduction

Obesity is a complex disorder and a known major risk factor for the development of cardiovascular disease in both children and adults. In particular, many children, even those with significant obesity, may have normal resting physiological parameters such as blood pressure or fasting glucose, despite evidence from more comprehensive assessment that their cardiometabolic health is compromised. The main aim of WP9 was to build a model able to identify young people with early physiological derangements that signify high cardiovascular risk and that cannot be detected reliably by traditional approaches, by utilizes a complex multi-variable dataset and advanced analytical tools to address this challenge.

3.2 Screening Approach

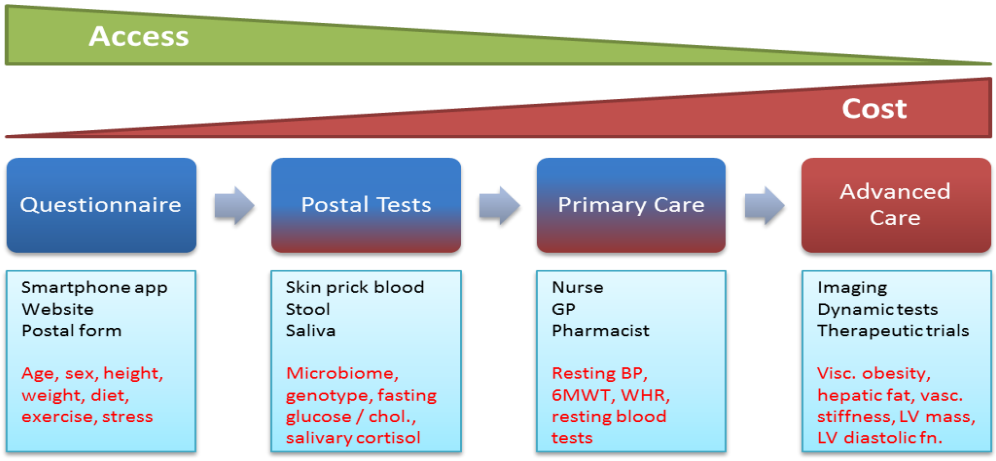


Figure 12 A novel screening approach for identifying young patient at risk

As depicted by Figure 12 , our novel screening approach relies on a sequential strategy where patients at risk are identified through the acquisition of increasingly complex phenotypic measures drawn from very different sources, such as questionnaires, stool and blood samples, clinical assessments and advanced medical imaging. At each stage, predictive models based on deep learning facilitate a risk assessment that determines whether to recall a given patient for the next stage of more advanced examination.

3.3 Validation process

in order to validate the proposed sequential approach, A pilot multi-centric study was conducted to collect a

cross-sectional dataset of approximately 160 patients, including questionnaire, anthropometrics, genetic, clinical as well as imaging data. To further characterize cardiac function and fat distribution within the body, advanced image processing and machine learning tools have been developed yielding each patient's specific parameters. 2 years follow-up examinations were performed among 33% of the cases.
Based on computational modelling approaches developed in the work package dedicated to cardiomyopathy, we have personalized heart models to the MRI data from obese children to extract parameters characterizing different aspects of the cardiac function such as left ventricular mass index and vascular stiffness. In addition, our colleagues from Fraunhofer developed a unique and advanced technology for automated characterization of body fat distribution from MRI data.

For testing the different associations of our screening strategy, we propose to compare 4 types of models: i) a deep neural networks specialized on classification, ii) a multi-task deep neural network as presented in the previous section, iii) a simple logistic regression model using BMI z-score only as input and iv) a random sampling model using the prior distribution (figure 13, see setion 4 of D9.4 for a deeper explanation).

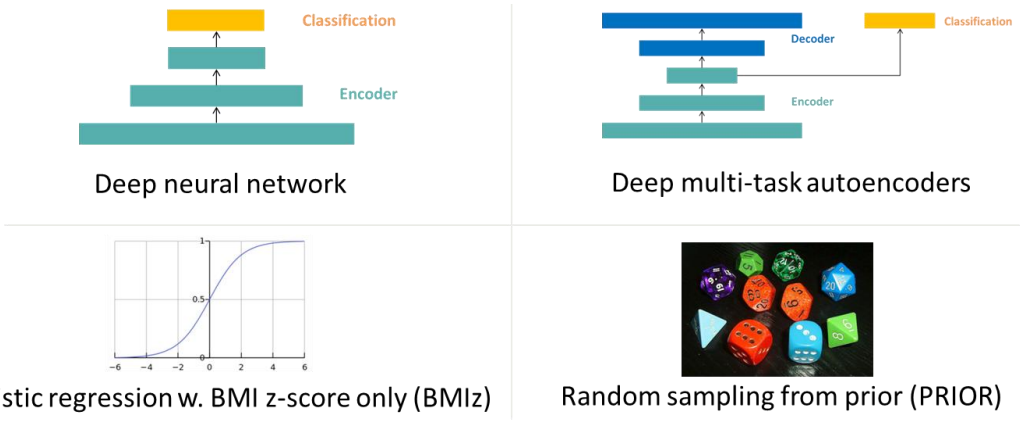


Figure 13 Comparative experiments between deep neural networks, deep multi-task neural networks, logistic regression using BMI z-score only, and random sampling from prior

According to the different stages of our proposed screening, the different modalities, e.g. sources of information, have been structured into five levels: (i) questionnaire, (ii) microbiome, (iii) clinical parameters, (iv) imaging, (v) follow-up (see section 2 of deliverable D9.4).

The experiments carried out for WP9 aimed to evaluate the performance of our models in predicting intermediate outcome by using information and parameters coming from the from the four different levels (level I to level IV) and the meal challenge data.

3.4 Predicting intermediate outcome from questionnaire data (Level I)

In these set of experiments, we evaluated the performance of our models for predicting the different targets described above by using information and parameters coming from the questionnaire:

- **dfs_ncore**: this parameter is a normalized measure that characterizes the amount of sugar and fat consumed by the considered patient,

- **nssec_mother, nssec_father**: this parameter characterizes the social class of the parents,
- **pss**: this parameter is a measure of perceived stress level,
- **tanner**: this parameter characterizes the stage of puberty,
- **caffeine_mg_d**: this parameter measures the amount of caffeine (in mg) per day,
- **ghq**: this set of parameters (_soma, _dysf, _anxiety, _depressed) are taken from classical general health questionnaire and that permits to assess somatic symptoms, anxiety and insomnia, social dysfunction and severe depression [7],
- **wreq**: this set of parameters are taken from the weight-related eating questionnaire as described in [8]
- **act_overall**: this parameter summarizes the level of activity of the patient
- **BMI-z**: this parameter is the body mass index z-score normalized according to the age.

For more details concerning those parameters and how they are computed, we refer the reader to the deliverables from WP4.

Figure 3 summarizes the overall results of the validation of models predicting the different target of interest based on the questionnaire data. The blue bars provides the average of the sensitivity and specificity measure for our best model aggregated over the 3 folds (higher is better). The red bars provides the significance of this results compared to random (p-value, lower is better). A very interesting result is the one for predicting **compliance**. Indeed, our best model provides an average sensitivity-specificity of **73%** with a p-value compared to random of **0.04**.

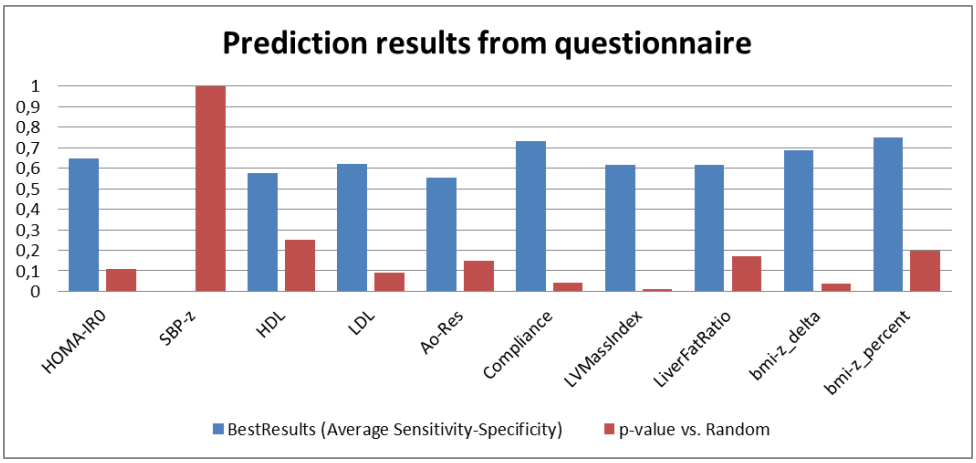


Figure 14 Prediction results using questionnaire data and BMI z-score

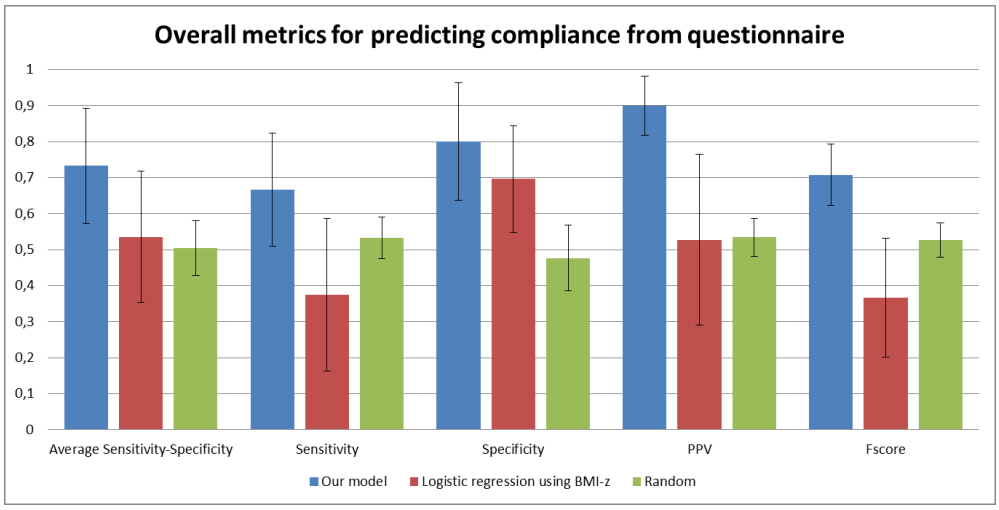


Figure 15 Detailed results for predicting compliance from questionnaire

Details of the results for predicting compliance from the questionnaire, e.g. by providing the different metrics used for evaluation: sensitivity, specificity, positive predictive value (PPV) and f-score. Average metrics for our model, a logistic regression model based on BMI z-score only and the random sampler are depicted in blue, red and green respectively. The standard errors computed over the folds are shown as error ranges on the corresponding bar. Our best model based on classification as main driver for training reaches a **sensitivity of 67%** and a **specificity of 80%.** Given the accessibility of the questionnaire test (Level I), this is very promising as the p-value suggests that it is a significant result. Resulting embedding from the 17 dimensions of the questionnaire data into 2 dimensions. The blue dots and triangles correspond to the training and test samples that have high compliance and thus are at lower risk. The red dots and triangles correspond to the training and test samples having lower compliance and thus being at higher risk. This suggests that our model is able to well stratify patients according to their compliance and this, based on data that is cheap to acquire.

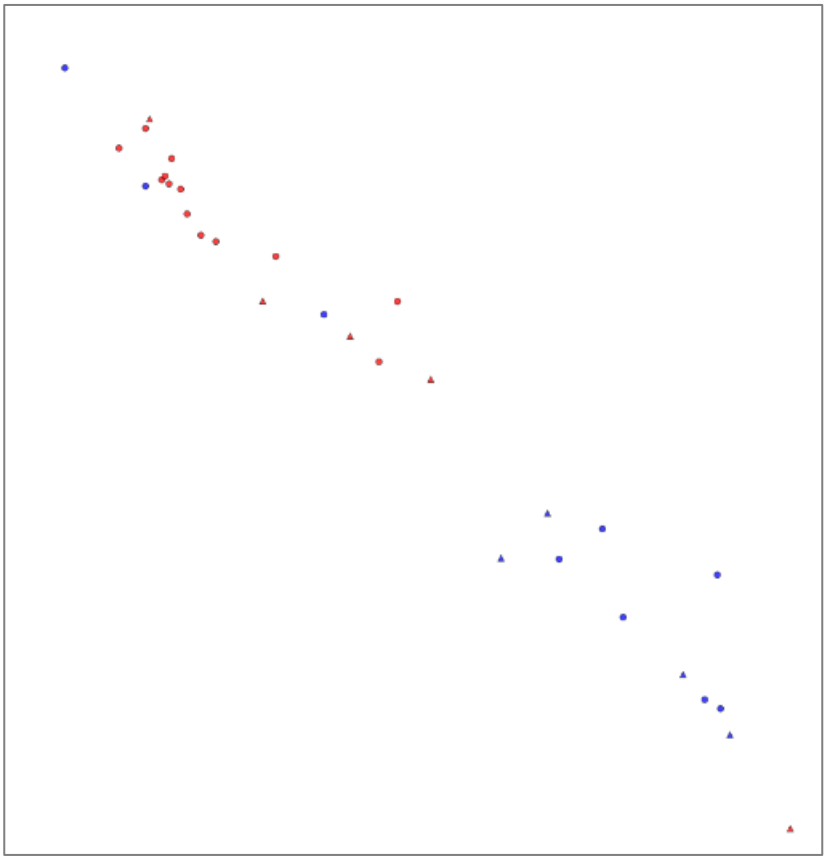


Figure 16 Embedding of the questionnaire samples from 17D into 2D using our approach. The blue dots/triangles correspond samples having high compliance and thus lower risk and the red dots/triangles correspond to samples at higher risk.

3.5 Predicting intermediate outcome from microbiome data (Level II)

In these experiments, we aim at predicting the different targets of interest by using parameters characterizing the populations of bacteriae within the intestinal tract and this, at different levels L2, L5 and L6. Average sensitivity-specificity and the p-value compared to random for models using microbiome information only and microbiome with BMI information respectively. When adding BMI information, models get increased performance and this, with a better significance. Looking at the p-values, there are 2 interesting results here: predicting liver fat ratio, and predicting the change in BMI. For predicting liver fat ratio, our multi-task model using microbiome information at level L2 and BMI reaches an **average sensitivity-specificity of 74%** with a **p-value of 0.05**. For predicting the change in BMI z-score, our best model also based on microbiome L2 and BMI reaches an **average sensitivity-specificity of 70%** with a **p-value of 0,03**.

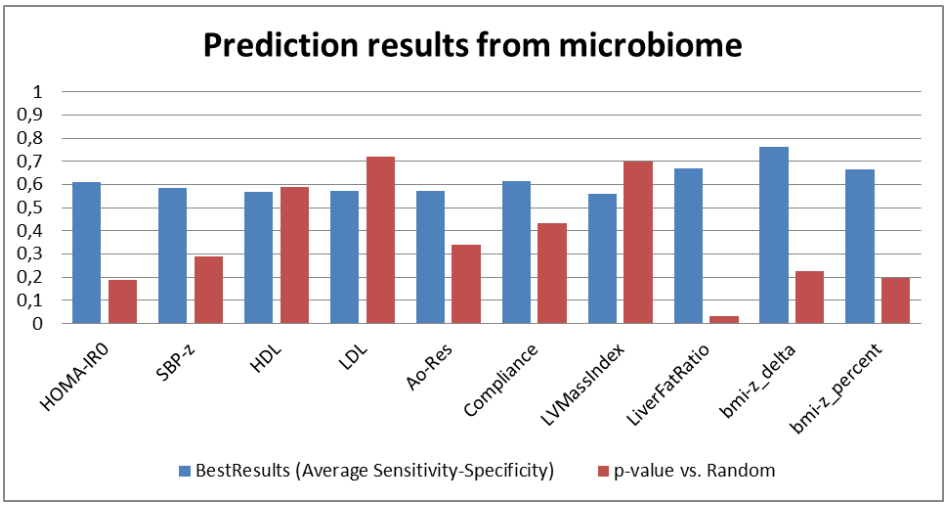


Figure 17 Prediction results of our model based on microbiome information

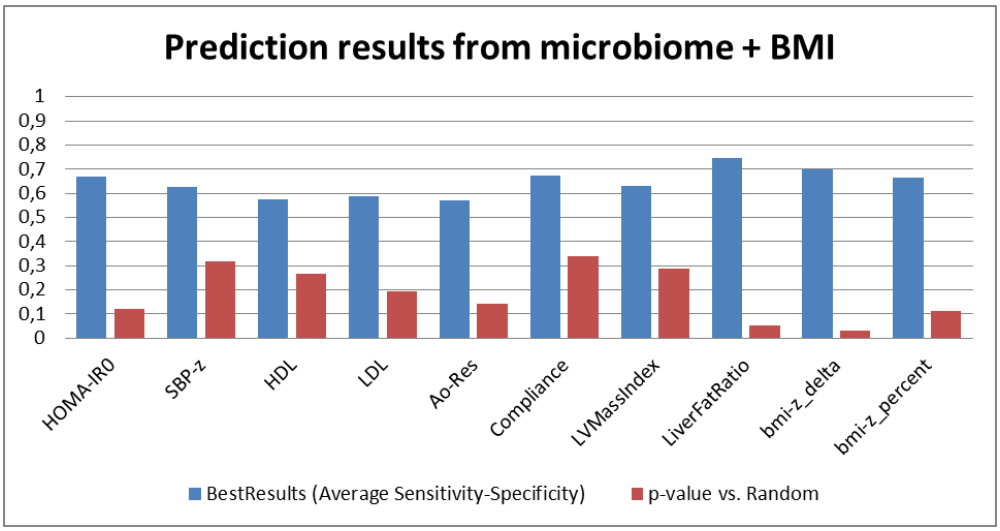


Figure 18 Prediction results of our model based on microbiome and BMI z-score information

Details the results of predicting liver fat ratio from the microbiome and BMI information by providing the different metrics used for evaluation: sensitivity, specificity, positive predictive value (PPV) and f-score. Average metrics for our model, a logistic regression model based on BMI z-score only and the random sampler are depicted in blue, red and green respectively. The standard errors computed over the folds are shown as error ranges on the corresponding bar. Our best model based on multi-task reaches a **sensitivity of 61%** and a **specificity of 88%**. This is a very interesting result that suggests that there might be an association between the microbiome and the liver function. Resulting embedding from the 7 dimensions of the microbiome L2 data and BMI z-score into 2 dimensions. The blue dots and triangles correspond to the training and test samples that low liver fat ratio and thus are at lower risk. The red dots and triangles correspond to the training and test samples having high liver fat ratio (>5%) and thus being at higher risk. This suggests that our model aims at separating patients according to their liver fat ratio.

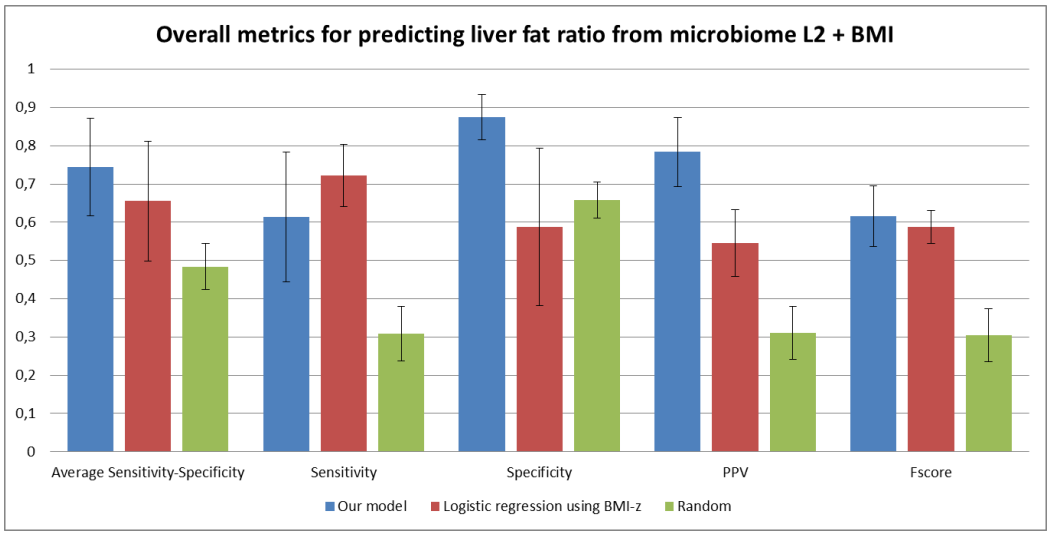


Figure 19 Results of predicting liver fat ration from microbiome L2 + BMI z-score

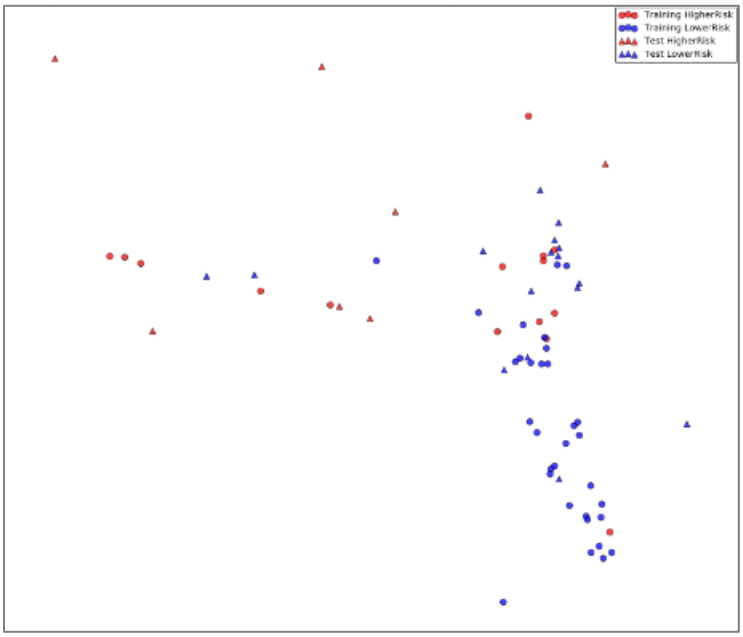


Figure 20 Embedding results provided by our model for projecting microbiome data 7D into 2D

Details the results of predicting change in BMI z-score from the microbiome and BMI information. Our best model based on multi-task reaches a **sensitivity of 61%** and a **specificity of 79%**. This is an interesting result that suggests that there might be an association between the microbiome and the success of a diet.

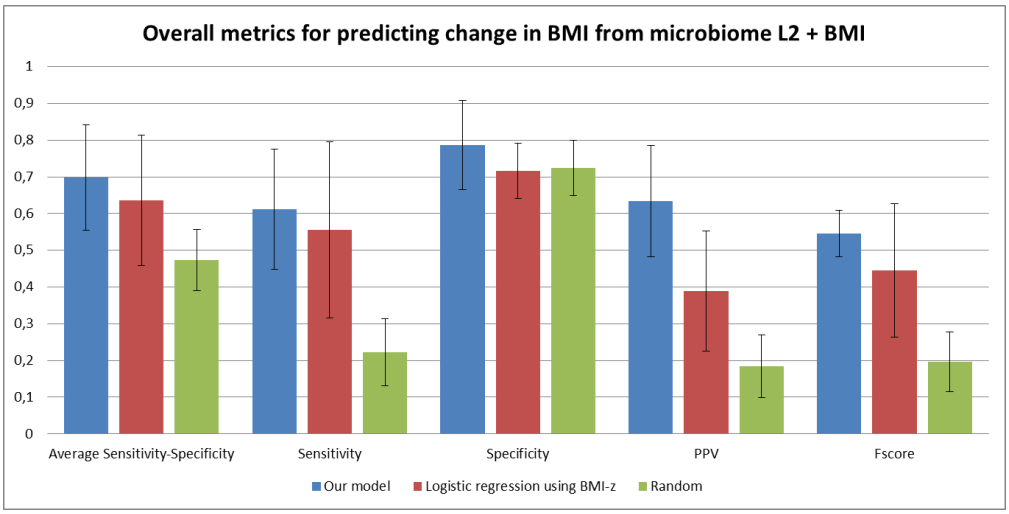


Figure 21 Detailed results of predicting change in BMI from microbiome L2 and BMI information

3.6 Predicting intermediate outcome from genetic data (Level II)

In these experiments, we aim at predicting the different targets of interest by using genetic parameters that characterize the risk carried by certain single nucleotide polymorphisms (SNPs) associated with a specific trait such as blood pressure, lipids, liver fat or left ventricular dimension. Given the SNP data for each patient a risk score belonging to [0, 0.5, 1.0] is computed depending whether the considered SNP carries zero, one or two risk alleles respectively (based on research in literature, ref to deliverable from WP4 for more details). Using this converted data, we performed 3 different types of analysis from simplest to more complex: **(a) Single SNP analysis, (b) Aggregated Risk Analysis, (c) Multivariate Risk Analysis**

A) Single SNP analysis

We have conducted the single SNP analysis for following targets: fasting glucose, fasting insulin, HOMA-IR, HDL, LDL, SBP-z, Ao-Res, Compliance, LiverFatRatio and LVMassIndex. Detailed corresponding p-values are reported in the deliverable D9.4. Note that while in some cases the p-value is low, we found that some risk values are inverted within our data, for instance, that low LDL value was associated with a high risk according to its SNPs. Here we provide a summary table (Figure 22) containing a shortlist of the significant and meaningful associations we found.

SNP	Target	P-value
rs1532085	CholestHDL	0,001151363
rs6511720	CholestLDL	0,001431396
rs1173771	Compliance	0,003378696
rs2925979	CholestHDL	0,005806838
rs11708067	Ao-Res	0,005952394
rs10885122	HOMA-IR0	0,024077545
rs3184504	SBP-z	0,02657013
rs3184504	LVMassIndex	0,028335879
rs12940887	Compliance	0,029815141
rs386000	CholestHDL	0,041190159
rs4373814	LVMassIndex	0,054808144
rs7034200	Glucose0	0,06256259
rs11558471	HOMA-IR0	0,080788436
rs4607517	Glucose0	0,097969414
rs11558471	Insulin0	0,109242422

Figure 22 Interesting associations found within our data based on single SNP risk

B) Aggregated Risk Analysis

We performed a similar analysis after aggregating the risk of all SNP related to a particular domain. As in the previous analysis, some case might have low p-value but with an inverted relationship. Note that the p-values are here much higher. This can be explained by the fact that many SNP that were not found significant in our data have been merged together. Finally, the most significant association we found was using **diabetes2 SNPs to explain aortic resistance** with a **p-value of 0.09**.

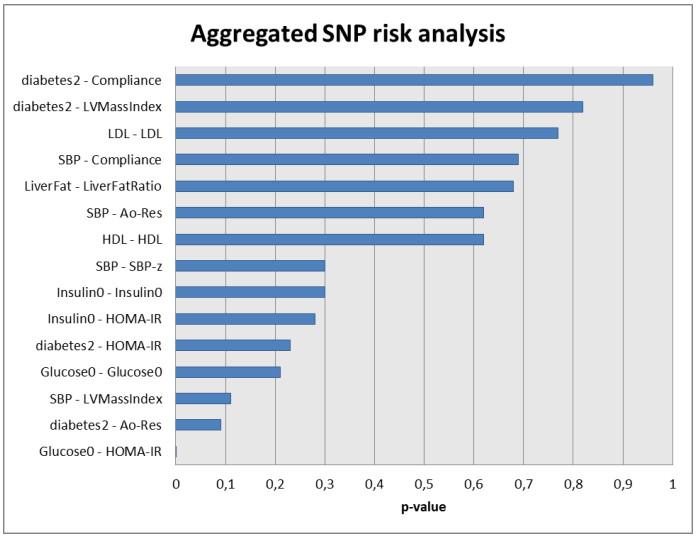


Figure 23 Analysis results for aggregated SNP risk. On the left the label of the row is encoding "SNP domain" - "target"

C) Multivariate Risk Analysis

In these experiments, we aim at predicting the different targets of interest by using multivariate SNP risk values from specific domains. Since risk values can influence the outcome only in the direction of increasing the probability of being at risk, an additional constraint was added to the model: the weights

as well as the bias are forced to be non-negative. Results depicted in Figures 24 and 25, showing the average sensitivity-specificity and the p-value compared to random for models using domain-specific SNPs and the complete set of SNPs respectively.

When using all SNPs, models get decreased performance and this, with a lower significance. Looking at the p-values, there might be one interesting result here: predicting compliance from SBP-specific SNPs. Our model trained using classification objective only reaches an **average sensitivity-specificity of 63%** with a **p-value of 0.04**. Clearly, results suggest that model suffered from the many SNPs that were not informative or even having an inverted association as seen in our previous analyses using single SNPs.

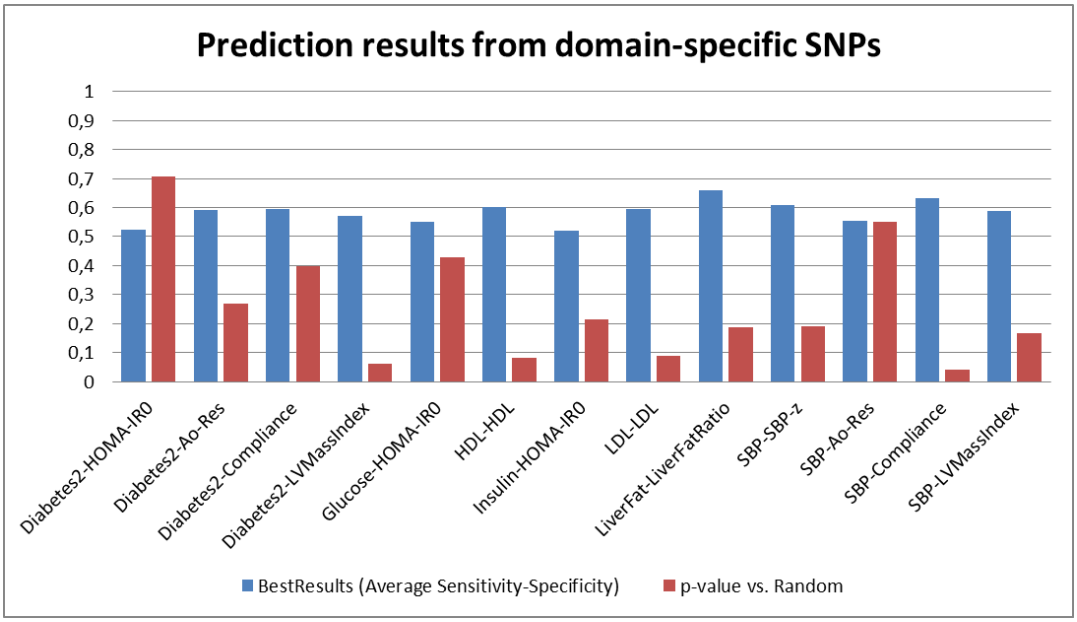


Figure 24 Prediction result for models using domain-specific SNPs. Labels are encoding "domain"-"target".

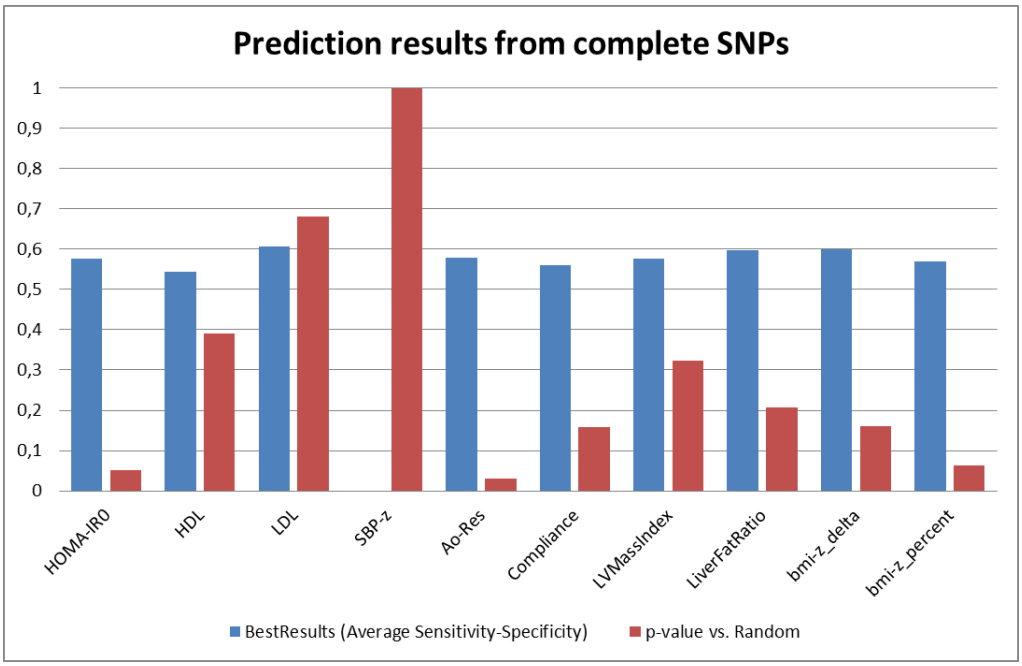


Figure 25 Prediction results for models using complete SNPs

3.7 Predicting intermediate outcome from clinical data (Level III)

In these experiments, we aim at predicting the different targets of interest by using clinical parameters such as fasting glucose, insulin, HOMA-IR, heart rate, diastolic and systolic blood pressure, HDL, LDL, triglycerides, BMI. Results in **Errore. L'origine riferimento non è stata trovata.** show the average sensitivity-specificity and the p-value compared to random for models using clinical parameters. Overall, it seems that clinical parameters alone provide only limited information for predicting intermediate outcome. However, the results obtained for predicting **compliance** could be considered as interesting: our multi-task model reaches an **average sensitivity-specificity of 67%** with a **p-value of 0.06**. Using blood pressure and BMI provides useful information for inferring compliance.

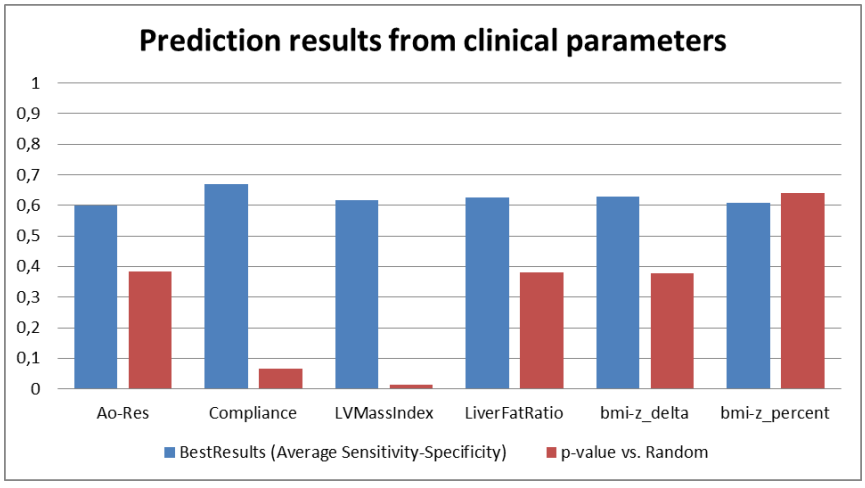


Figure 26 Prediction result for different target outcomes using clinical parameters

3.8 Predicting intermediate outcome from microbiome and clinical data (Level III)

In these experiments, we merge information from the microbiome with the clinical parameters and evaluate how well intermediate outcomes can be predicted. Similarly as when using microbiome and BMI, interesting associations are observed for inferring compliance and liver fat ratio, but also for change in BMI. Concerning compliance and liver fat ration, performances seem to have slightly decreased suggesting that clinical parameters might not bring information strengthening the predictive power of the model. However for predicting change in BMI, in particular bmi-z_delta, we obtain an **average sensitivity-specificity of 72%** with a **p-value of 0.004**. This provides an improvement compared to the model obtained when using microbiome L2 and BMI information only.

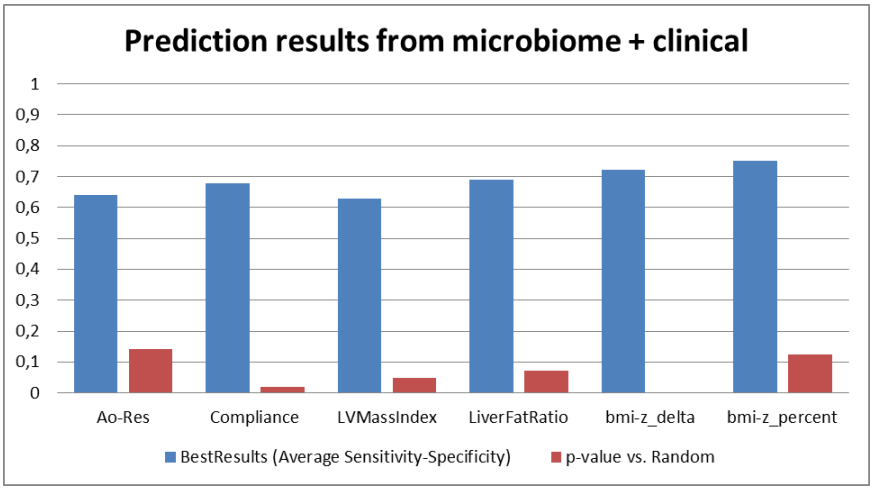


Figure 27 Prediction results for different target outcomes using microbiome and clinical data

Indeed, as shown below, our model reaches a **sensitivity of 67%** with a **specificity of 78%**.

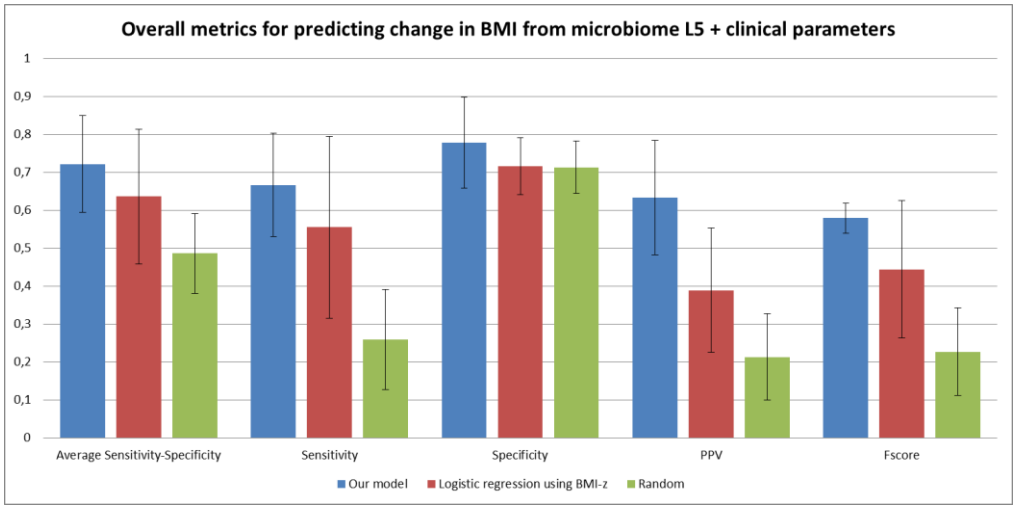


Figure 28 Metrics for predicting change in BMI based on microbiome and clinical data

3.9 Predicting change in BMI from parameters extracted from imaging (Level IV)

In these experiments, we evaluate the predictive power of models using parameters extracted from imaging, e.g. parameters characterizing the cardiac function or the fat distribution, for inferring the change in BMI at follow-up. Results are summarized by **Errore. L'origine riferimento non è stata trovata.** and **Errore. L'origine riferimento non è stata trovata.**. Those figures show the average sensitivity-specificity and p-value for bmi-z_delta and bmi-z_percent, i.e. change in BMI and change in BMI relative to baseline BMI in percentage. In the first case, experiments do not provide any results as there were not enough samples in both low and high class. In the second case, it was possible to conduct all experiments. According to **Errore. L'origine riferimento non è stata trovata.** 29 and 30, there is a very interesting result for predicting the change in BMI percentage given strain rate and BMI data, more precisely circumferential and radial strain rate during systole.

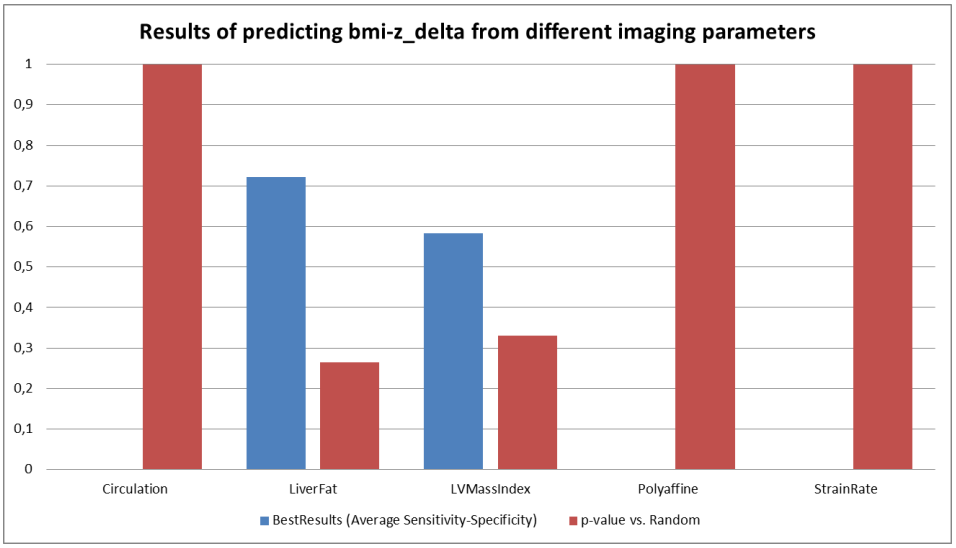


Figure 29 Results of predicting change in BMI from parameters extracted from imaging

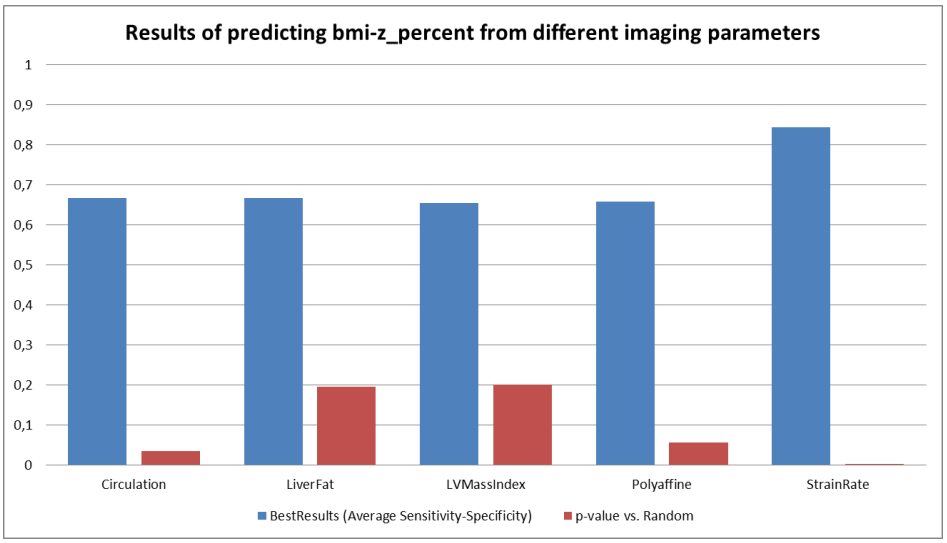


Figure 30 Results of predictive relative change in BMI (in percentage) from parameters exctracted from imaging

While this result is very intriguing, it needs to be considered carefully because of the small size of the dataset (intersection of patients having strain rate and follow up data). Indeed, it seems difficult to understand the association between strain rate and change in BMI. One could imagine that there is an association between strain rate and LV mass and that patients having a higher LV mass might be the one at higher risk where a diet won't help. But at this stage and with the amount of data at hand, this remains quite speculative. **Errore. L'origine riferimento non è stata trovata.** 31 provides detailed results for this association where our model reaches a **sensitivity of 89%** and a **specificity of 80%**, and this with a **p-value of 0.002**.

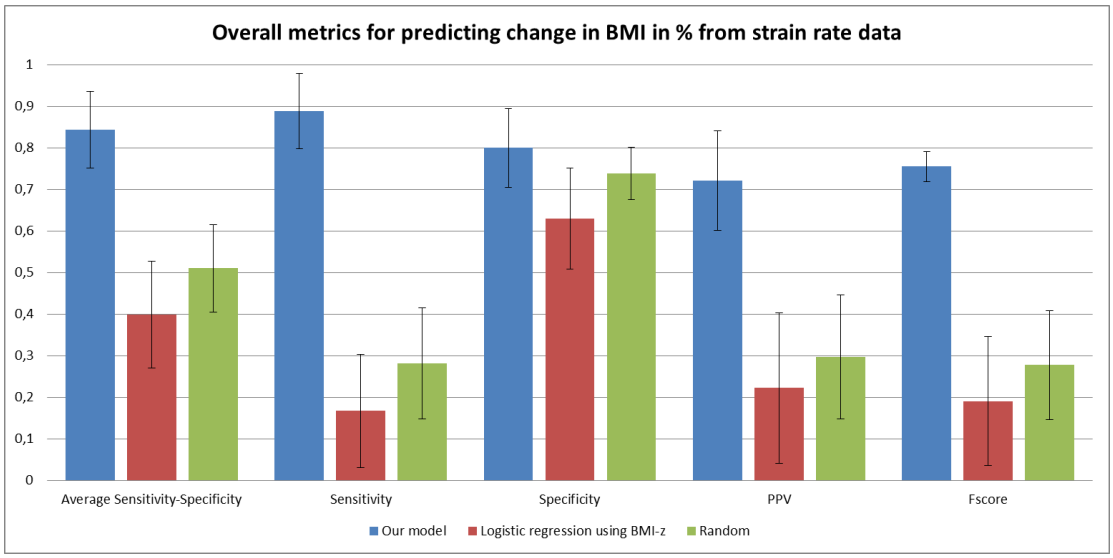


Figure 31 Metrics for predicting change in BMI in percentage from strain rate data

3.10 Predicting intermediate outcome from meal challenge data

For this range of experiments, the purpose was to check whether meal challenge blood test data acquired by our partners at UCL can be used to predict whether a patient is at risk or not based on heart model parameters extracted from MRI. An association would enable the use of the meal challenge protocol to decide whether further test such as MRI imaging and modelling are necessary. These experiments used a slightly different experimental setup which is described in section 4.7.1 of the deliverable D9.4. Results are detailed in sections 4.7.2 and 4.7.3 of the same deliverable.

Overall, the results for using just the first time point show a clearer picture of added value through using the meal challenge data over only BMI. This can be explained with the lower number of input dimensions in conjunction with the very limited amount of available data. In such a case it can be useful to use less dimensions for prediction in order to avoid overfitting of the model. This problem could be alleviated by running the developed algorithms with more data in the future. The good performance for some of the targets already gives an indication of the potential predictive value for these models.

4.0 Validation for the JIA models (WP5-WP10)

4.1 Clinical use cases -General clinical framing

An 8-year-old girl presents at the paediatric rheumatologist with pain in the left knee and ankle, difficulty walking and morning stiffness of about 1 hour. Furthermore, the parents noted swelling of the left knee for about two months. The girl has been treated with non-steroidal anti-inflammatory drugs for a month, without effect. She does not present any other symptoms, such as fever or cutaneous manifestations. She has never experienced a similar episode in her life. Her father has been diagnosed with rheumatoid arthritis a few years ago.

At the general physical examination, the rheumatologist does not find any abnormalities. The articular examination reveals swelling, pain and limitation of movement of the left knee and ankle. At the ankle joint, in particular the extension and flexion of the joint is limited, whereas the inversion and eversion of the joint appears to be normal. When asking the child to walk up and down the corridor, she is clearly limping.

The rheumatologist orders general blood tests, which reveal no pathologies, except for elevated inflammatory markers (ESR 32, CRP 0.83) and positive anti-nuclear antibodies (1:640, homogeneous pattern). Additionally, a plain X-ray of the left knee and ankle reveals no bone abnormalities. The articular ultrasound on the other hand shows a marked distension of the supra-patellar recess of the left knee due to joint effusion and synovial hypertrophy with signs of hyper-vascularization at power Doppler. At the level of the left ankle, a marked distension of the tibiotalar joint recess is seen, due to joint effusion and synovial hypertrophy with signs of hyper-vascularization at power Doppler. Additionally, the ultrasound shows a modest distension of the lateral recess of the subtalar joint, as well as modest distension of the tendon sheaths of the posterior tibial and the long flexor of the digits tendons.

Based on these findings, the rheumatologist makes a diagnosis of oligoarticular juvenile idiopathic arthritis. The following questions arise:

1. **What is the prognosis of this child? Does she have a high probability of achieving and maintaining disease remission?**

The answer to this question is pivotal to the treatment plan. Nowadays, no instruments exist to predict the prognosis of a child with juvenile idiopathic arthritis (JIA). Therefore, treatment is applied using a step-up approach: it is initiated using local treatment only (intra-articular steroids injections). In case the child does not respond well to treatment, systemic drugs are added to the treatment.

2. **Given the child is limping, is she at risk to develop joint damage?**

The changed intra-articular forces, due to different loading of the joint because of the limp, might cause joint damage. This is a serious, in general irreversible adverse condition.

3. **Does the monolateral joint involvement put the child at risk to develop arthritis at the contralateral side as well?**

Again, the altered gait pattern and joint loading forces at the contra-lateral side may be risk factors for the development of arthritis in the knee and ankle at that side. The biomechanical model will estimate probabilities and, if confirmed, the physician may use these probabilities to offer other treatment options to try to prevent disease extension.

The questions illustrated by the clinical case divide themselves naturally into two sections, corresponding to the two different models which will be developed within WP5 and WP10, respectively: the clinical prediction model and the biomechanical ankle model. The next two sections highlight the clinical use and validation of these two models.

4.2 Assessment and validation of the clinical prediction model

Clinical background

JIA is a heterogeneous group of diseases, of unknown cause. The clinical hallmark of all categories of JIA is joint inflammation (arthritis), sometimes accompanied by extra-articular manifestations. Affected children experience joint swelling, pain and limitation of motion of the inflamed joints. Chronic arthritis may cause permanent damage, such as bone and cartilage erosions or growth abnormalities. It is therefore crucial to treat patients well, in order to avoid these long-term sequelae.

Treatment of JIA is currently based on a step-up approach, in which a child is treated with first-line therapy, such as intra-articular joint injections with corticosteroids. In case of non-response, systemic therapy may be added, such as methotrexate. Only in case of an insufficient response to methotrexate, third-line treatment with biological agents (mostly anti-TNF alpha) is offered to the patient. However, in adult rheumatoid arthritis it has been demonstrated that early and aggressive treatment in the so-called window of opportunity leads to better outcomes. It is thought that a similar window of opportunity exists in paediatric rheumatic diseases, such as JIA. On the other hand, it is also known that not all children require aggressive therapy to induce disease remission. Therefore, a solution to avoid administering aggressive and costly drugs to children not needing them, while yet prescribing third-line therapy early to those in need of it is a clear need. The clinical prediction model that will be here developed will allow to stratify patients according to the risk of persistence of active disease and adapt treatment plans accordingly.

This model will use baseline (i.e. available at the time of diagnosis) predictors to predict the disease course in the first one or two years of disease. This timeframe was chosen since a) is the one in which we aim at inducing disease remission in all children; and b) short-term disease remission has been shown to be a predictor of longer-term disease outcome.

Validation

Validation of the clinical prediction model has been split up in three stages:

1. Clinical Protocol validation
2. Internal model validation
3. External model validation

Only stage 1 and 2 will be performed within the framework of MD-Paedigree. For stage 3, recruitment of additional patients, not used for the model development, is necessary. This needs therefore to be done in a later stage.

Clinical Protocol validation

For the clinical prediction model, different predictors are taken into account, which can be divided into five broad

categories: clinical, laboratory, microbiological, immunological and imaging.
As far as the clinical and laboratory parameters are concerned, these are routine parameters collected during all visits patients with (suspected) paediatric rheumatic disorders perform. The standardized collection and use of these parameters have been described in a myriad of studies and can therefore be considered validated.
Concerning the microbiological (faeces samples) and immunological (Luminex analysis) parameters, similar remarks need to be made about both. Samples for these analyses have been collected in three different clinical centres using slightly different approaches (e.g. collect faeces samples at home and send them within 24 hours by express courier, versus storing them in a domestic freezer and bring them at the subsequent visit). Protocol validation for these samples will therefore consider comparability of the samples across the different centres. In particular, the quality of the samples will be assessed by evaluating the quantity and viability of the analysed material. Furthermore, any occurrence of clustering per centre will be verified using principal components analysis and clustering techniques, such as k-means clustering.
Finally, the use of imaging parameters will be validated, taking into account the ease of acquiring these data (with respect to equipment, needed training and time to acquire the images). Validation and standardization of imaging protocols is currently being discussed and studied by dedicated working parties (e.g. the OMERACT ultrasound working group) and falls outside the scope of this project.

Internal model validation

Internal validation of a model consists in re-using the data that were used to develop a model, to get an estimate of the performance of the model in independent populations. It is the first step in validating a prediction model. We will use standard statistical methods to perform internal validation. First of all, the performance of the model in classifying patients with a binary outcome (active disease versus inactive disease) will be evaluated using the area under the receiver-operating characteristics curve (AUROC). Furthermore, the sensitivity of the model to the data, and in particular its optimism in the predictions (the phenomenon that a model always performs best in the data set it was fitted to) will be tested using 10-fold cross-validation, a technique aimed at reducing the optimism of the model, thus yielding an estimate of its real performance in independent data sets.

External model validation

External validation of the model consists in collecting data of a new group of patients, not used to develop the model, to estimate the performance of the model in an independent population. This falls outside the scope of MD-Paedigree.

4.3 Assessment and validation of the subject-specific biomechanical models

With respect to the Juvenile Idiopathic Arthritis (JIA), the subject-specific modelling pipeline developed in the MD-Paedigree project will generate a large set of quantitative anatomical, functional, and biomechanical biomarkers, obtained through direct measurement (via medical imaging and biomedical instrumentation), data processing, and mechanistic data modelling. All these biomarkers have now been validated from a technical point of view, in term of accuracy and reproducibility / sensitivity.
The clinical validation aims to establish:

a) Robustness of the modelling protocol, i.e. how reproducible is the subject-specific modelling protocol in real-world clinical setting;

b) Clinical accuracy, intended as the ability of one or more biomarkers provided by the subject-specific modelling protocol to accurately answer a clinically relevant question;

c) Clinical efficacy, intended as the improvement (typically measured in term of cost-benefit and/or risk-benefit of care the new clinical pathway based on such biomarkers provides if compared to the previous standard of care.

In the context of MD-Paedigree we can aim to achieve a), and to some extent b); we would need a much larger cohort and a much longer follow-up to answer c). Assuming the clinical accuracy is good, future experimental medicine projects will have to assess the clinical efficacy.

Detailed clinical use case for validation

In order to define a protocol for the assessment of the clinical efficacy, the first step is to derive from the general clinical framing above, specific clinical use cases where to assess if any of the biomarkers provided is effective in answering a clinically relevant question. Four predictive medicine use cases were identified:

1. Identify from the biomarkers which side is affected, in mono-lateral cases: while this clinical question does not directly translate into a clinical use case, it is somehow preliminary to all the others. The determinants of the clinical progression of JIA are unknown; we suspect both systemic and local mechanisms. The ability of anatomically located biomarkers to correctly discriminate which side is affected, would confirm that local mechanisms play a major role, and justify the following use cases.
2. Identify from biomarkers single-joint from multi-joint cases: similarly, if some of the biomarkers can correctly separate those patients who have multiple joints involved, from those who only have one joint, this could cast some light on the mechanisms related to this significant difference in the clinical manifestation of the disease.
3. Treatment stratification: JIA patients can be stratified, at the 6-month control, into those who responded to the first-line treatment and went in remission, and those who did not: responders and non-responders. Prolonged disease activity in the non-responder patients, who at the second control start a more aggressive treatment, increases the risk of irreversible joint damage, which could be reduced or avoided altogether if we could from the outset stratify which patients would not respond to the first-line therapy. If these biomarkers can accurately stratify, on the basis of the data collected in the first visit, these two groups, we could modify the clinical pathway, so that those patients who are considered, on the basis of these biomarkers, to be at risk of non-response, could be immediately treated with the more aggressive therapies.
4. JIA progression: this is characterised by two major clinical signs, local inflammation and cartilage damage or bone erosions. The first is responsible for the acute symptoms (pain, swelling, stiffness, etc.), but the second has a much more profound long-term impact, as articular damage is largely irreversible. While all patients who have prolonged inflammatory flare will eventually develop some cartilage damage, the delay between the beginning of the flare and that of the cartilage damage, and the rate with which such damage progress varies considerably between patients. The most fundamental question is: do biomechanical alterations affect structural damage progression? If this is the case, then life style limitations, physical therapy, and possibly the use of assistive devices during the flare could reduce the extent of the cartilage damage; if this is not the case, then none of these interventions would make much difference. Assuming the answer to the question is yes, then the we could imagine a clinical pathway

where the subject-specific model is used to personalise the life style recommendations or the use of assistive devices until the inflammatory flare recedes.

Robustness

The robustness of the data acquisition, processing, and modelling will be quantified through the results of the quality assurance protocol applied over the entire workflow:

a) Data availability: for each dataset to be collected (for each patient-control) confirm if it has been collected and properly uploaded to the project repository;

b) Data processability: for each dataset and for each data processing step, confirm that the data can be processed into an output of quality sufficient for its subsequent use;

c) Data modellability: for each dataset and for each data processing step, confirm that the data processing outputs are of quality sufficient to inform the subject-specific model, without significant degradation of the expected predictive accuracy.

Clinical accuracy – laterality discrimination

All patient-control cases will be separated in left, right, and bilateral.
Initially we will compare the average value of each biomarker in the left and right groups, using a Student's t-test, or a Mann-Whitney test when variables are not normally distributed.
Given the large number of biomarkers we will use exploratory analysis techniques to explore the independence of the biomarkers, and their difference in relation to laterality.
If the exploratory analysis suggest that combinations of biomarkers might necessary for an effective discrimination, we will then use logistic regression models to test the discriminative power of independent biomarkers in combination. Depending on the completeness of the biomarkers matrix some case might have to be dropped from this analysis, as sparseness degrades the performance of logistic regressions.
Finally, if one or more biomarkers are confirmed to discriminate laterality, we use the Area Under the Receiver-Operating-Characteristic Curve (ROC-AUC) as a measure of the predictive accuracy of that biomarker for laterality; in case of combination of biomarkers we will calculate the ROC-AUC of the model provided by the logistic regression.

Clinical accuracy – Single-multi joint discrimination

The protocol is identical to the previous one, except in this case we will cluster biomarkers values for patient-controls in single-joint and multiple-joint groups.

Clinical accuracy – Treatment stratification

All patients for which it was possible to obtain a given biomarker at zero and six-month controls will be clustered in respondent and non-respondent groups, based on the clinical assessment at the second control. The two groups will then be analysed with the same approach described in the previous points.

Clinical accuracy – Cartilage damage prediction

This is the most complex use case, and it might not be possible to address it in full with the duration of the project. The first challenge, which will be possible to address only when most data are collected and processed,

is how exactly we can quantify the cartilage damage progression.
Assuming for simplicity that we can reduce this complex process to a binary condition (for example patients who do have cartilage damage at 6 months, opposed to those who have observable damage only at 12 months), the assessment protocol would be similar to those above.
If on the contrary we could achieve a quantitative representation of cartilage damage (area, volume, etc.) we could calculate the growth rate between 0-6-12 and where available also at 24 months. The average or peak growth rate could then be correlated to our set of biomarkers, to see if any of them could candidate as a predictor. If a robust correlation is found, we could use the 0-6 or 0-12 rate to identify the predictor, and then use the 24-month control to check its predictive accuracy.

4.4 Luminex analysis

As reported and discussed previously, Luminex analysis was performed of all baseline samples. A panel of 78 analytes was selected, however 10 of these produced out-of-range (OOR) results in more than 20% of cases and were subsequently removed from the analysis. Two samples showed a high false-positive response to mouse cytokines and were removed from the analysis. Five other samples were excluded, because they were withdrawn after methotrexate (MTX) start. This procedure left 68 analytes and and 129 cases to be analysed.
Principal components analysis (PCA) was performed of the log-transformed, centred and scaled data of the remaining analytes and a plot of the first and second principal component was made (Figure 32). Samples were coloured by centre of origin and an ellipse showing the 95% confidence region was drawn, indicating clearly that OPBG samples clustered separately from IGG and UMCU samples.

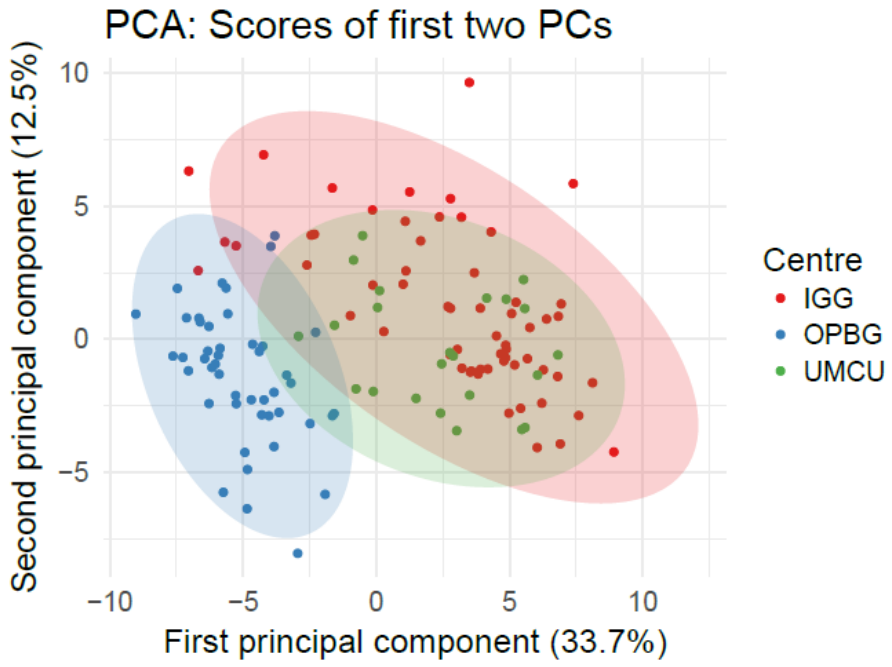


Figure 1: OPBG clearly separates from the other centres in a plot of the first two PCs.

Figure 32 OPBG clearly separates from the other centres in a plot of the first two PCs

The visual clustering was confirmed by k-means clustering (Figure 33). Four clusters were chosen, as this seemed to be the lowest number of clusters best explaining the variability in the data (determined using a plot of the total within sum of squares, not shown). It was confirmed that OPBG samples were collected in EDTA tubes, instead of sodium heparin tubes, explaining the marked differences between the centres. Therefore, it was decided that OPBG samples will be analysed separately.

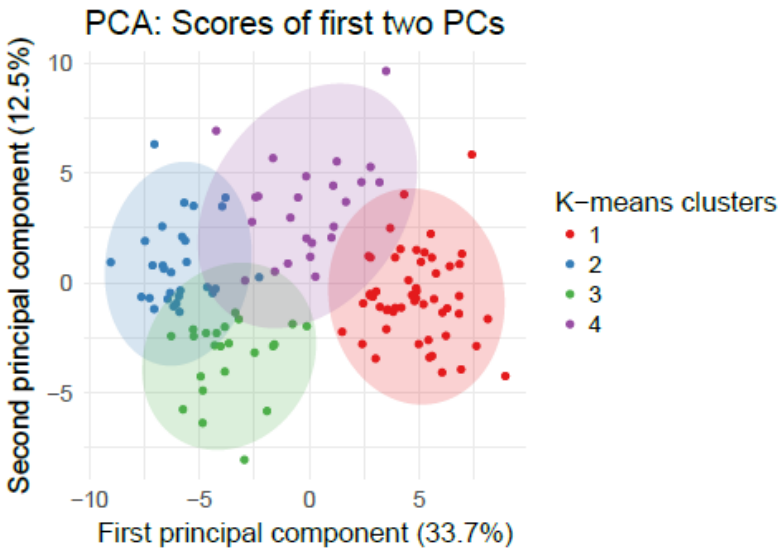


Figure 2: K-means clusters correspond to visually identified clusters.

Figure 33 K-meand clusters correspond to visually identified clusters

Analysis of IGG/UMCU samples

Data of IGG and UMCU samples were analysed using PCA. A plot the first and second principal component confirmed that, after exclusion of OPBG patients, IGG and UMCU patient still clustered together (Figure 34).

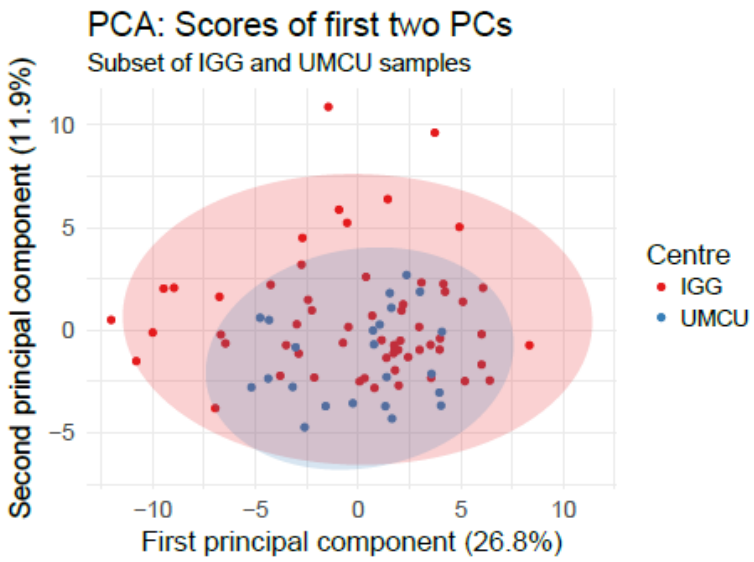


Figure 3: IGG and UMCU samples still cluster together.

Figure 34 IGG and UMCU samples still cluster together

Apart from the centre distribution, there still appears to be a clustering in two or three clusters in the data. This was analysed using k-means clustering, which identified the clusters as shown in Figure 35.

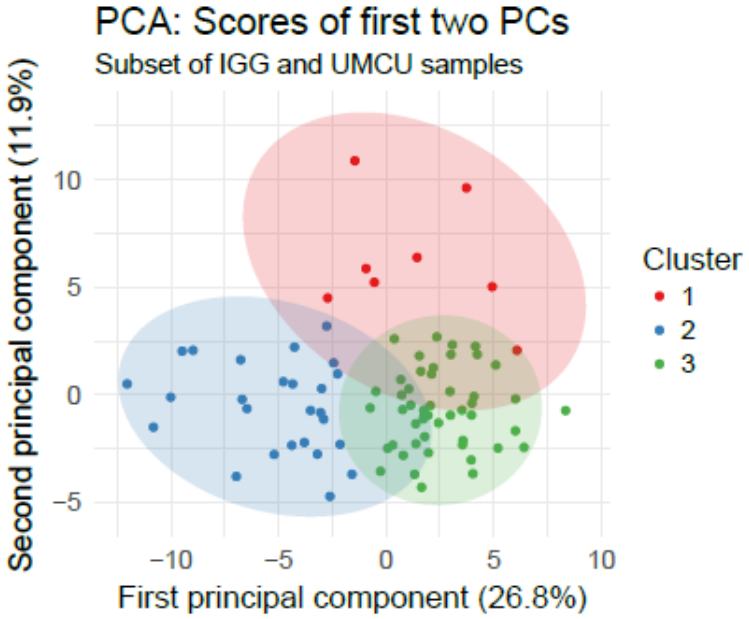


Figure 4: Three clusters identified by k-means clustering.

Figure 35 K-meand clusters correspond to visually identified clusters

The question is what are the Luminex analytes that define the separation into these clusters? This can be analysed by a plot of the variable loadings. These are the coefficients with which the variables are multiplied to obtain the principal component scores for each samples. So, if PC1 is the first principal component and $x_1, x_2, \ldots x_i$ are the variables, the variable loadings are the in the following equation:

$$PC_1 = 1 \times x_1 + 2 \times x_2 + \ldots + i \times x_i$$

In words, the variable loadings indicate the influence a variable has on a principal component. Since the first principal component captures most variability in the data, if a variable has a high loading on the first principal component, it means it has an important influence on the variability in the dataset. Moreover, if, as in our case, two clusters in the data are visible along the first principal component, variables having a high loading on this principal component are the ones that determine these clusters. The variable loading for the first and second principal component is shown in Figure 36.

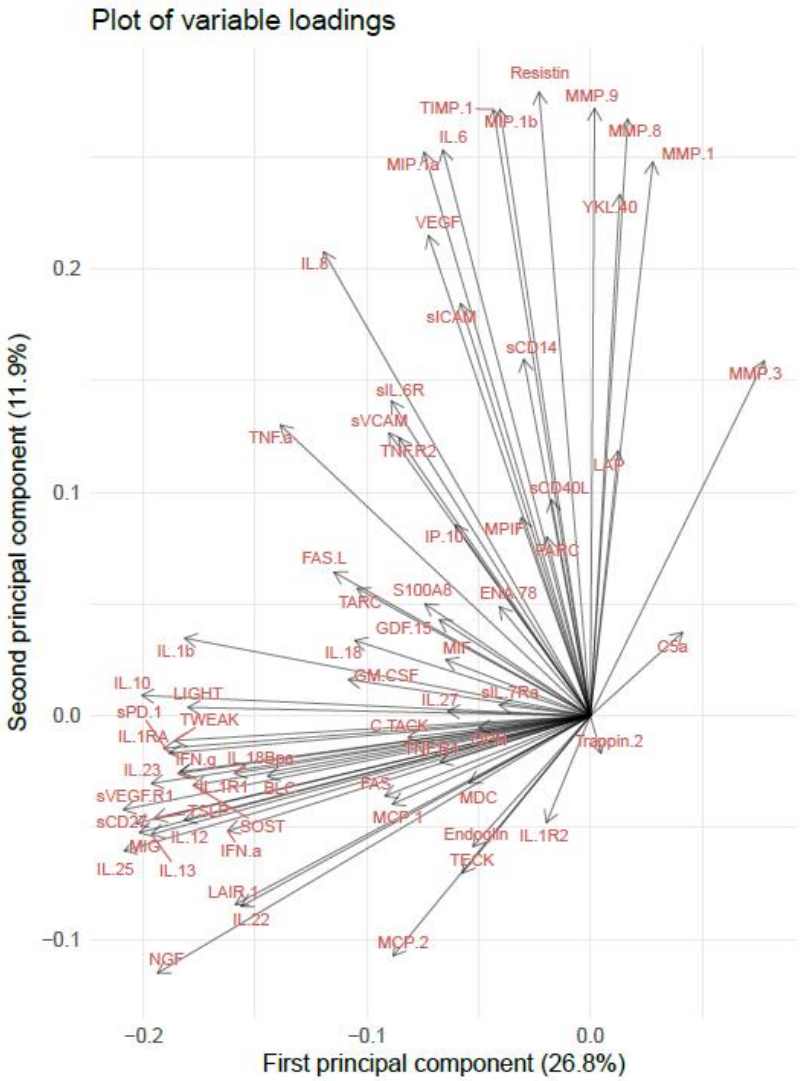


Figure 5: Variable loadings demonstrate two clusters of analytes.

Figure 36 Variable loadings demonstrate two clusters of analytes

The loadings are displayed as rays from the origin, as is usual. The farther the extension of the ray horizontally, the more influence the variable has on the first principal component and the farther the vertical influence, the more influence on the second principal component. This way of displaying the loadings helps in visualizing correlations among the variables: variables with rays parallel to each other are correlated, whereas those perpendicular to each other have no correlation at all.

This means that in our case a group of variables, containing most interleukins, but also interferon (IFN)- and IFN- are correlated and influence mainly the first principal component, but not the second. These are therefore the analytes responsible for the distinction between clusters 2 and 3 in Figure 35. Another group of variables, mainly the matrix metalloproteinases (MMP) 8 and 9 and YKL-40 influence the second principal component, but not the first. These analytes are mainly responsible for the separation between cluster 1 on the one hand and cluster 2 and 3 on the other in Figure 4. A third group of variables influence both the first and the second principal component, such as TNF-_, IL 8, S100A8 and VEGF. Finally, some analytes, like Trappin 2, IL-1R2 and sIL-7RA hardly influence the first or second principal component at all.

To sum up, performing this analysis, we have discovered that the samples cluster into three clusters (or maybe two clusters plus some outliers) and we have seen what are the analytes responsible for this clusterization. We don't know yet if the different expression of these analytes is becomes clinically apparent in some way, for example in causing a different subtype of the disease or in being associated with female or male sex. Therefore, Spearman correlations were calculated between all clinical variables and the principal components. They are displayed in Figure 37 as a dot plot. The bigger the dot and the intenser its colour, the higher the correlation coefficient. Blue colour indicates positive correlation and red a negative correlation.

From the plot it is appreciable that principal component 2 correlates with such disease activity measures as ESR, CRP and PGA, and, consequently, JADAS-71. The correlation with the number of active joints and onset subtype (basically, oligoarticular onset vs. polyarticular onset, i.e. presentation with more than 4 active joints or not) is practically absent. Moreover, there are no appreciable correlations with the first principal component, except for HLA-b27, but this was measured in only 23 subjects and positive in four of those.

This correlation should therefore be interpreted with caution. Of note, furthermore, is that correlations with variables defining clinically interesting subgroups of patients, such as subtype, uveitis (at baseline) and ANA are low, except for some minor exceptions (e.g. uveitis and the 8th principal component). On the other hand, an interesting observation is that knee involvement is correlated with the third principal component.

Overall, these observations suggest that there are no observable patterns in cytokine expression for different subgroups of patients. Furthermore, we did not succeed in identifying clinical patterns corresponding to the clusters that were visualized in Figure 35.

As a next step, we could correlate the analytes themselves (instead of the principal components) with the clinical variables. Since the principal components are not correlated among themselves and the analytes are, we expect to see more correlations in this plot. More in particular, if two analytes are correlated and both influence the second principal component (see Figure 36), a clinical variable that correlates with the second principal component can be expected to correlate with both these analytes. The correlations are shown in Figure 38. It is immediately apparent that analytes correlating with disease activity measures are MMP-1, MMP-3 and YKL-40. These associations are known from literature. These analytes were indeed the ones influencing the second principal component (Figure 36) and this was the principal component disease activity was correlated with,

making the circle round again. Figure 39 shows that correlations between disease activity and YKL-40 are indeed linear. Similar pictures were observed for MMP-1 and MMP-3 (not shown).

Other important correlations are found between IL-6, inducer of the acute phase response, and acute phase reactants/measurements such as CRP and ESR (Figure 38). Furthermore, age (age at onset, age at diagnosis, height and weight) correlates with various analytes, indicating that age-related variability is important in our data set. This visualization does, again, not aid us in the identifcation of clinical variables related to the clusterization of data. However, age seems to be an important source of variability, potentially obscuring other patterns in the data. Since no age-dependent normal values are available, we need to get rid of this source of variability by performing subgroup analyses.

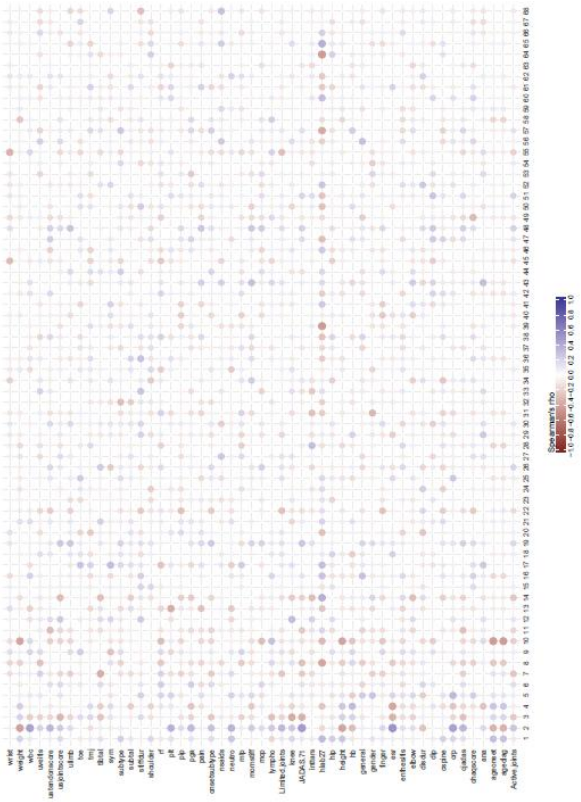


Figure 6: Correlations between clinical variables and principal components.

Figure 37 Correlation between clinical variables and principal components

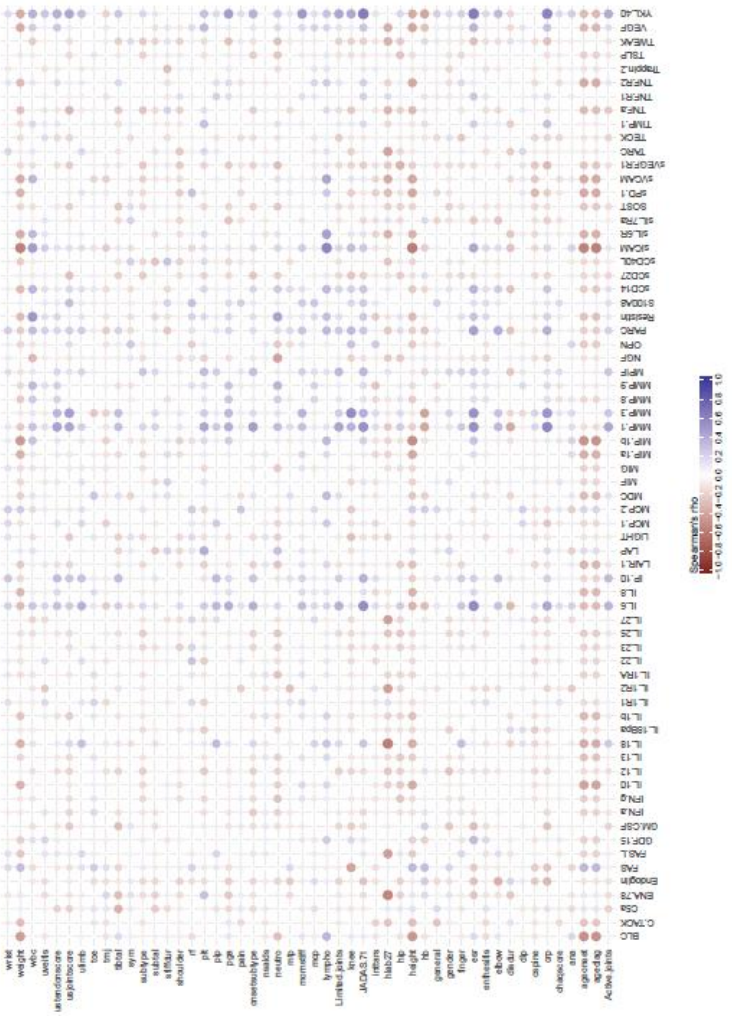


Figure 7: Correlations between clinical variables and analytes.

Figure 38 Correlation between clinical variables and analytes

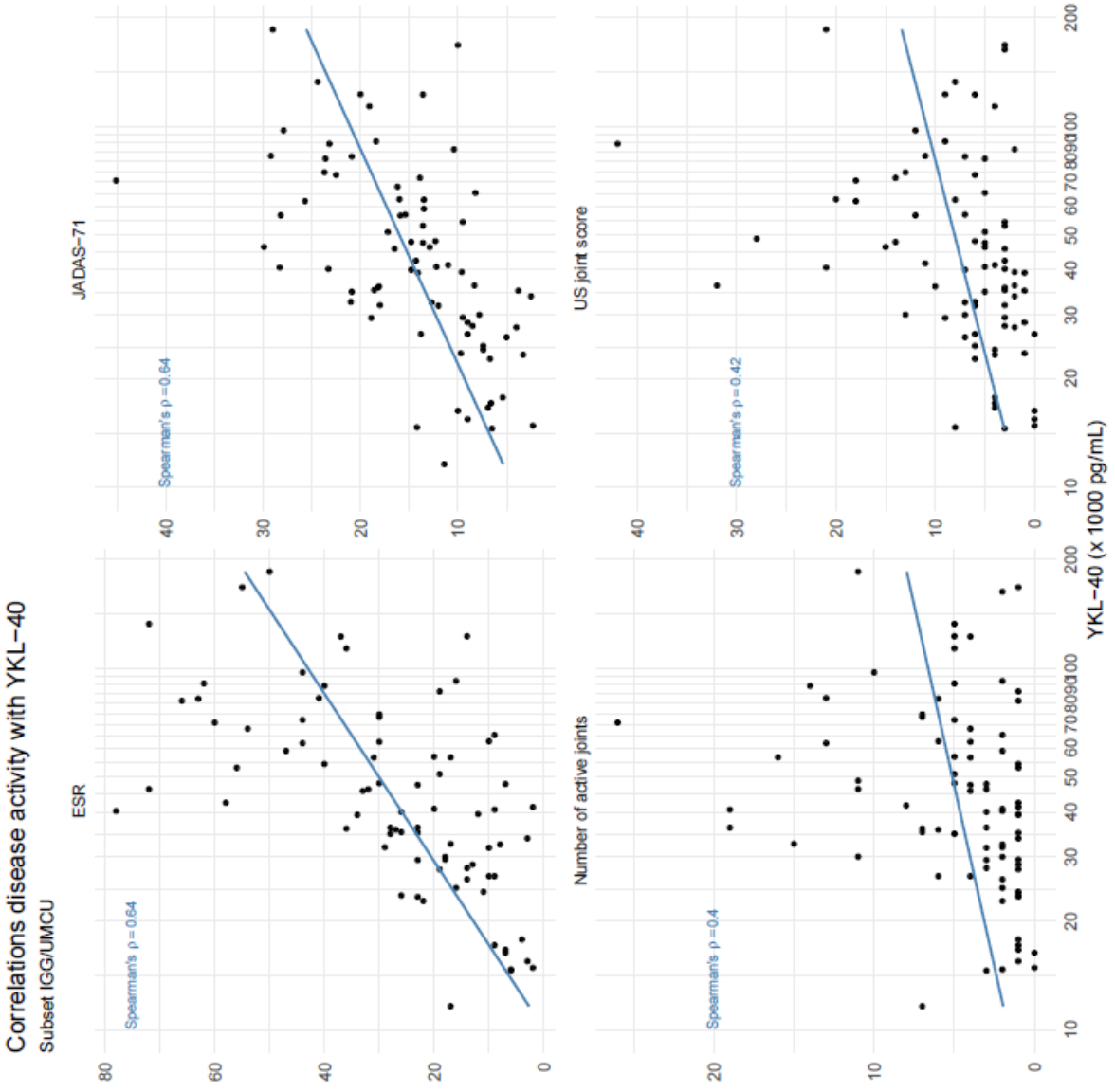


Figure 8: Linear correlations between YKL-40 and disease activity parameters.

Figure 39 Linear correlation between YKL-40 and disease activity paramenters

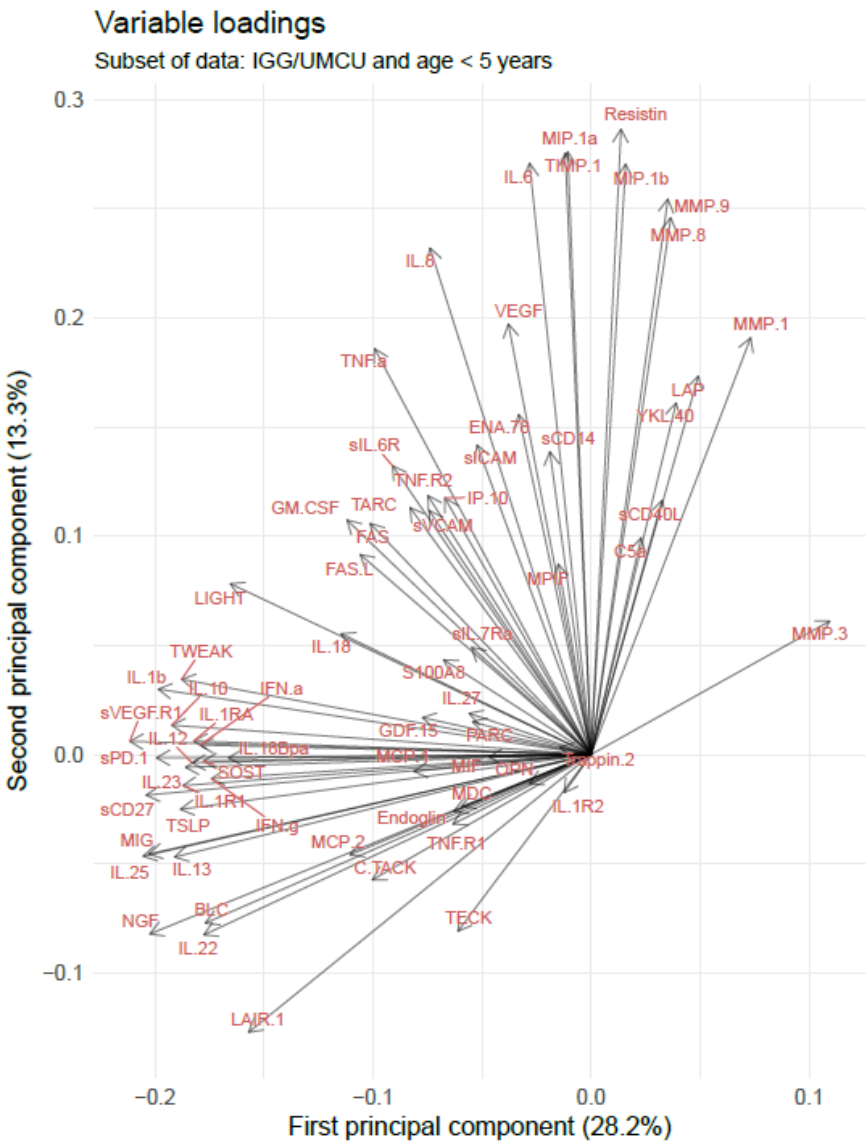


Figure 9: Variable loadings demonstrate two clusters of analytes.

Figure 40 Variable loadings demonstrate two clusters of analytes

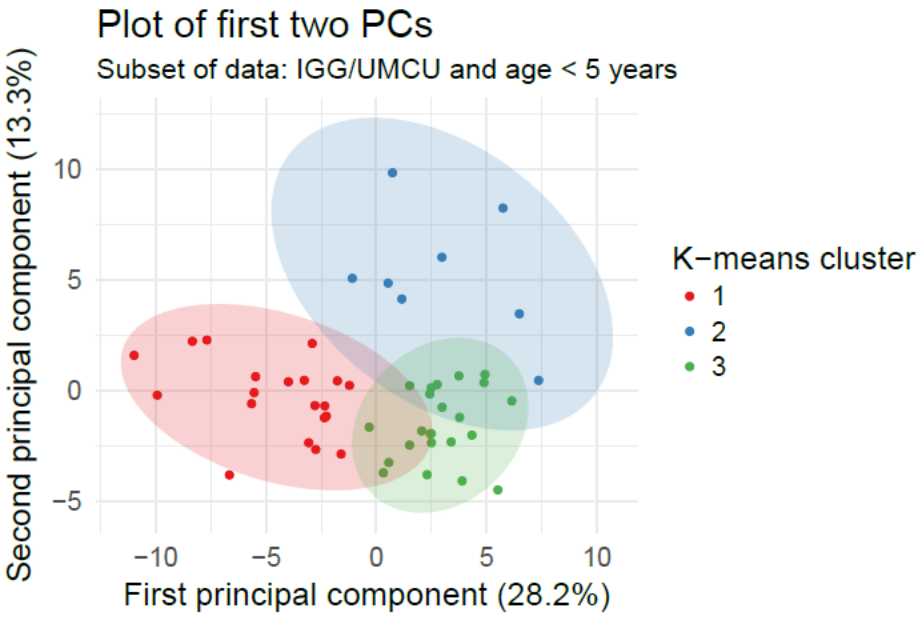


Figure 10: Similar clusters as previous.

Figure 41 Similar clusters as previous

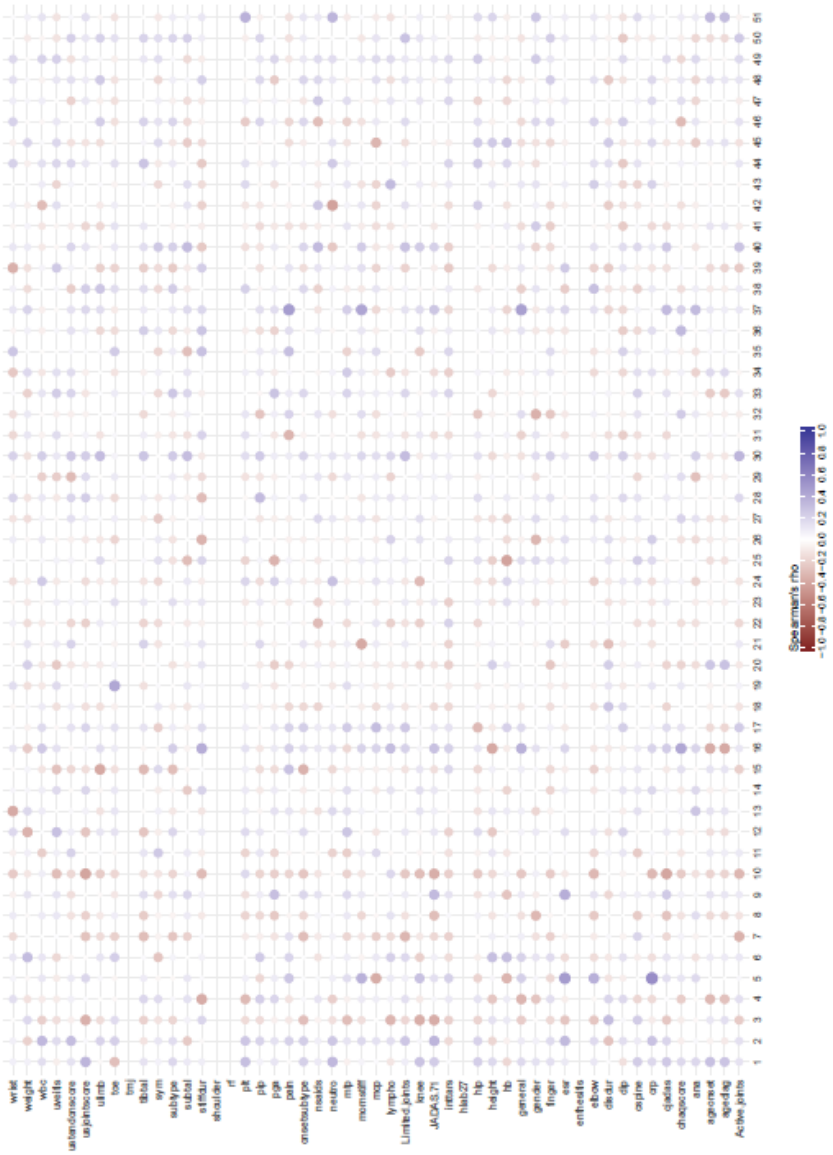


Figure 11: Correlations between clinical variables and principal components.

Figure 42 Correlation between clinical variables and principal components

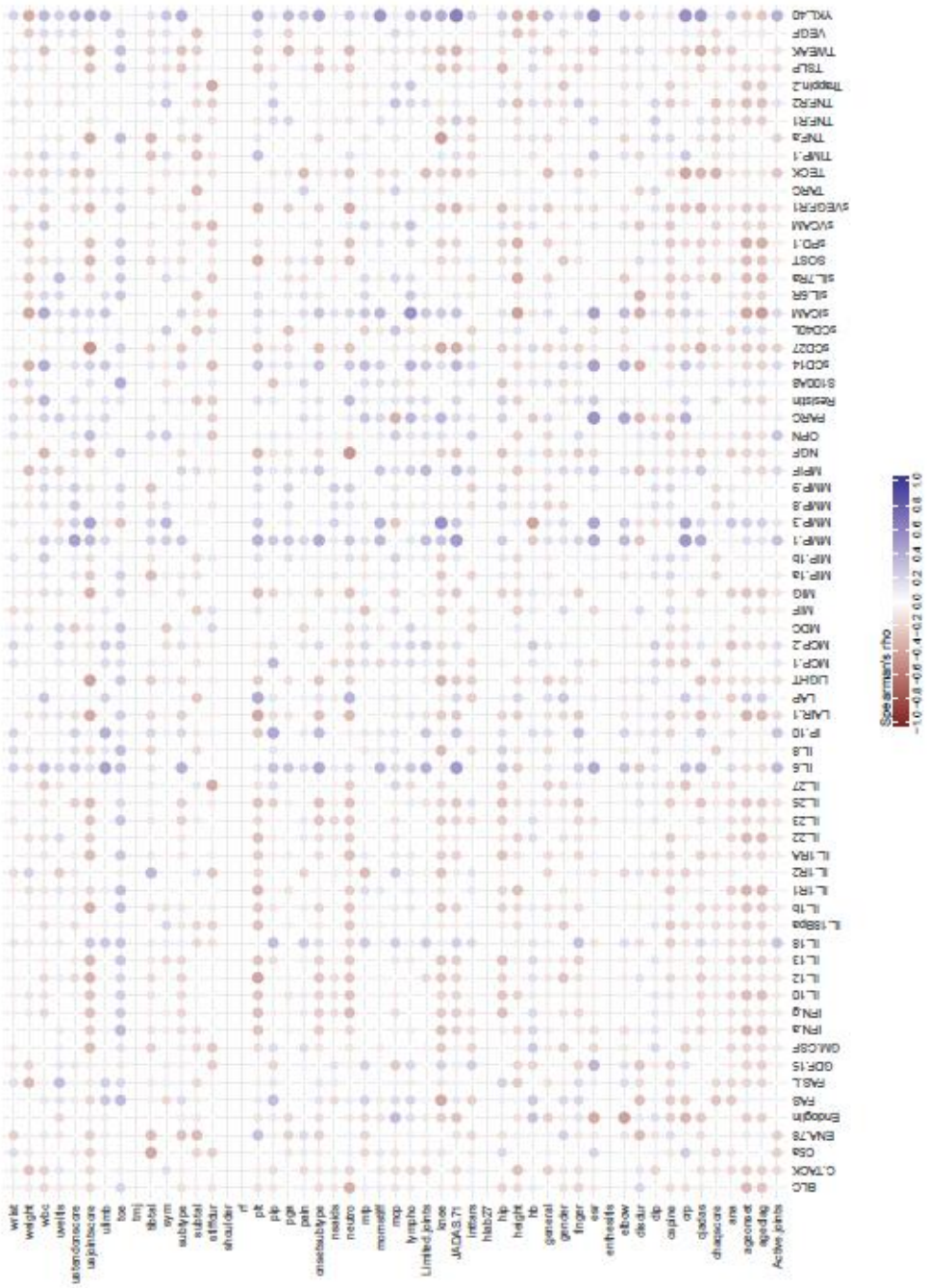


Figure 12: Correlations between clinical variables and analytes.

Figure 43 Correlation between clinical variables and analytes

5.0 Validation of metagenomics and metabolomics of gut microbiota studies

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.

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*** have been uploaded onto the ***Gnubila 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: *i*) ***OPBG repository database***, with household data processing and storage procedures; *ii*) NCBI BioProject submission, ***EBI repository database*** ; *iii*) Gnubila data submission, with the intent to generate a shared platform for model generation.

Detailed specific and comparative microbiota results:

a) metagenomics, metabolomics profiling of gut microbiota;

a) 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***. The distributions of OTUs were described in term of individual relative abundances and represented by using different criteria, *e.g*. ***age stratification***.
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 of JIA patients have been produced exploiting chemometrics models.

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. 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.
Additionally, ***dynamic models*** that link clinical phenomena across levels, have been designed and presented for JIA patients. 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 *chemometrics-based* computational pipelines.

The integration predictive model for JIA "disease" were generated as follows:

General notes. All raw data from NGS platform, from GC-MS, H-NMR generated from analyzed samples were reduced from multidimensional setting, harmonized amongst different platforms and integrated as described below. Furthermore the **Integration patterns and data validation** have been produced when possible, such as in the case of the extended panel of JIA patients samples analyzed fro both metagenomics and metabolomics.
Data Usage: Into Gnubila databases data from different platforms have been uploaded, first set of data, excel files, generation of NGS microbiota for obese and JIA patients, while integrated models have been generated by local duties, exploiting chemometrics procedures and pipelines only for JIA patients.
Data Integration: metagenomic data were integrated with metabolomic data from H-NMR and GC-MS origin. Integration patterns and data validation were done for JIA patients: please see computation below described.

Samples from Italian patients (more homogeneous data set, OPBG plus IGG patients) were analysed for metabolomics analysis by GC-MS/SPME and 1H-NMR for volatile and not volatile metabolite analyses (Figure 44).

HEALTHY (CTRLs)	BASELINE	INACTIVE DISEASE	PERSISTENT ACTIVITY
N-00-3	MDP-IGG-S45	MDP-IGG-S56	MD_A_101
N-00-4	MDP-IGG-S1	MDP-IGG-S33	MDP-IGG-S37
N-00-5	MDP-IGG-S9	MDP-IGG-S41	MDP-IGG-S23
N-01-1	MDP-IGG-S36	MDP-IGG-S55	MD_A_27
N-01-2	MDP-IGG-S32	MD_A_57	MD_A_41
N-02-6	MD_A_30	MD_A_102	MD_A_36
N-02-7	MD_A_81	MD_A_77	MD_A_19
N-03-3	MD_A_1	MD_A_97	MD_A_43
N-03-6	MD_A_32	MD_A_13	MD_A_86
N-03-7	MD_A_16	MD_A_70	MD_A_9
N-03-8	MDP-IGG-S6	MD_A_42	MD_A_51
N-04-1	MDP-IGG-S22	MD_A_24	MD_A_12
N-04-2	MDP-IGG-S19	MDP-IGG-S39	MD_A_34
N-04-8	MDP-IGG-S11	MD_A_35	MD_A_40
N-05-4	MDP-IGG-S8	MD_A_72	MD_A_76
N-05-5	MD_A_3	MD_A_28	MD_A_105
N-05-7	MD_A_21	MD_A_15	MD_A_74
N-05-9	MDP-IGG-S7	MD_A_60	MD_A_79
N-06-3	MDP-IGG-S2	MD_A_71	MD_A_66
N-06-8	MD_A_10	MD_A_49	MD_A_37
N-07-4	MD_A_14	MD_A_85	MD_A_58
N-07-5	MD_A_61	MD_A_109	MD_A_47
N-07-6	MDP-IGG-S28	MD_A_56	MD_A_26
N-08-5	MDP-IGG-S30	MD_A_96	MD_A_53
N-09-9	MDP-IGG-S27	MD_A_99	MD_A_67

Figure 44 100 samples analysed for metabolomics

Procedures of integration of metagenomics and metabolomics data

Metabolomic analyses for JIA patient samples

Amongst the 100 samples analysed from metabolomics, only 86 samples were also integrated with metagenomics data (Figure 45).

SAMPLES LIST of JIA and CTRL integrated for GC-MS/SPME and ^{1}H-NMR and metagenomics 86 (samples)

HEALTHY (CTRLs) (24 samples)		BASELINE (21 samples)	INACTIVE DISEASE (19 samples)	PERSISTENT ACTIVITY (22 samples)
N-00-3	N-07-5	MD_A_1	MD_A_13	MD_A_12
N-00-4	N-07-6	MD_A_14	MD_A_42	MD_A_34
N-00-5	N-08-5	MD_A_16	MD_A_56	MD_A_40
N-01-1	N-09-9	MD_A_30	MD_A_49	MD_A_41
N-01-2		MD_A_32	MD_A_57	MD_A_19
N-02-6		MD_A_81	MD_A_70	MD_A_26
N-02-7		MD_A_10	MD_A_96	MD_A_36
N-03-3		MD_A_21	MD_A_71	MD_A_37
N-03-6		MD_A_61	MD_A_77	MD_A_58
N-03-7		MD_A_3	MD_A_102	MD_A_74
N-03-8		MDP-IGG-S1	MD_A_97	MD_A_105
N-04-1		MDP-IGG-S2	MD_A_15	MD_A_86
N-04-2		MDP-IGG-S7	MD_A_28	MD_A_9
N-04-8		MDP-IGG-S32	MD_A_35	MD_A_43
N-05-4		MDP-IGG-S45	MD_A_72	MD_A_51
N-05-5		MDP-IGG-S9	MD_A_85	MD_A_53
N-05-7		MDP-IGG-S8	MD_A_109	MD_A_66
N-05-9		MDP-IGG-S19	MDP-IGG-S55	MD_A_67
N-06-3		MDP-IGG-S22	MDP-IGG-S56	MD_A_76
N-06-8		MDP-IGG-S27		MD_A_101
N-07-4		MDP-IGG-S28		MDP-IGG-S37

Figure 45 Samples list of JIA and CTRLs samples integrated for GC-MS/SPME , 1H-NMR and metagenomics (86 samples).

5.1 THE INTEGRATION MODEL: Strategy for the multivariate analysis of MG, GC/MS and NMR MB data and low data fusion (integrated platforms).

PLS-DA method using double check of validation. Partial least squares discriminant analysis (PLS-DA) is the most used classification methods in metabolomics. PLS-DA consists of a classical PLS regression where the dependent variable y is categorical and represents samples class membership e.g. y can be a vector with values of -1 and 1 where -1 represents each sample belonging to the class of controls and 1 represents each sample belonging to the class of cases. By making use of class information, PLS-DA tends to improve the separation between the (two) groups of samples. Two steps : 1) the selection of the optimal model complexity e.g. optimal number of latent variables (#LV) and 2) the assessment of the overall quality of the model. A double cross validation scheme consists of two nested loops CV1 and CV2, (see Smit et al. 2007). The aim of CV1 is to optimize complexity of the PLS-DA model and the aim of CV2 is to assess final model performance. In the outer loop (CV2) the complete dataset is split into a test set and a rest set: the test set is set aside and the rest set is used in a single cross validation (inner loop, CV1). In the CV1 the rest set is again split into a validation (sometimes called optimization) set and a training set. Statistical significance of each PLS-DA model is estimated by using the value of the diagnostic statistics (number of misclassification, NMC or Area under the Receiver Operating Characteristics, AUROC) to values of its null reference distribution H0 obtained by permutation tests. Also the Discriminant Q2 is showed. However DQ2 (in contrary to NMC and AUROC) prefers PLS-DA models with lower complexity. NMC and AUROC are more efficient and more reliable diagnostic statistics and have been recommended in two group discrimination metabolomic studies [Szymanska E et al. Metabolomics 2012 Jun; 8(Suppl 1):3-16]. Coomon samples analysed by NMR and GC were 86: 24 CTRLs, 21 baselines, 19 inactive, 22 persistent. Cuts for family and species were for presence of OTUs with at least >= 70 % of subjects.

Graphical illustration of use of diagnostic statistics: NMC, AUROC, and DQ2 in double cross validation procedure of PLSDA is below reported. a) Use of diagnostics statistics in selection of optimal number of latent variables in CV1; b) use of diagnostics statistics in assessment of overall PLS-DA model quality after double cross validation procedure (CV2) (Figure 46).

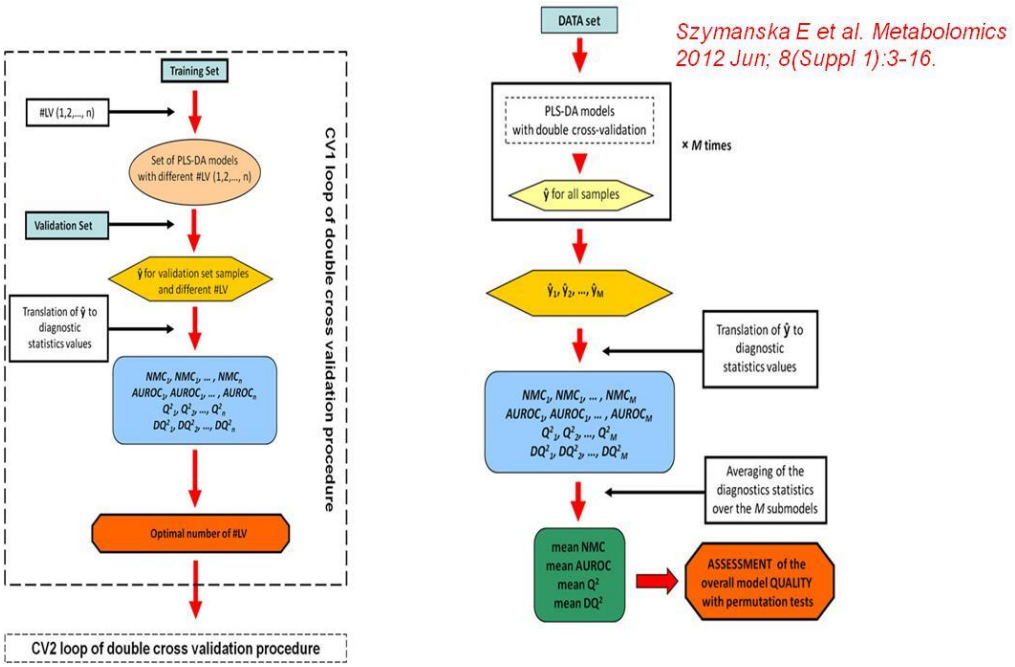
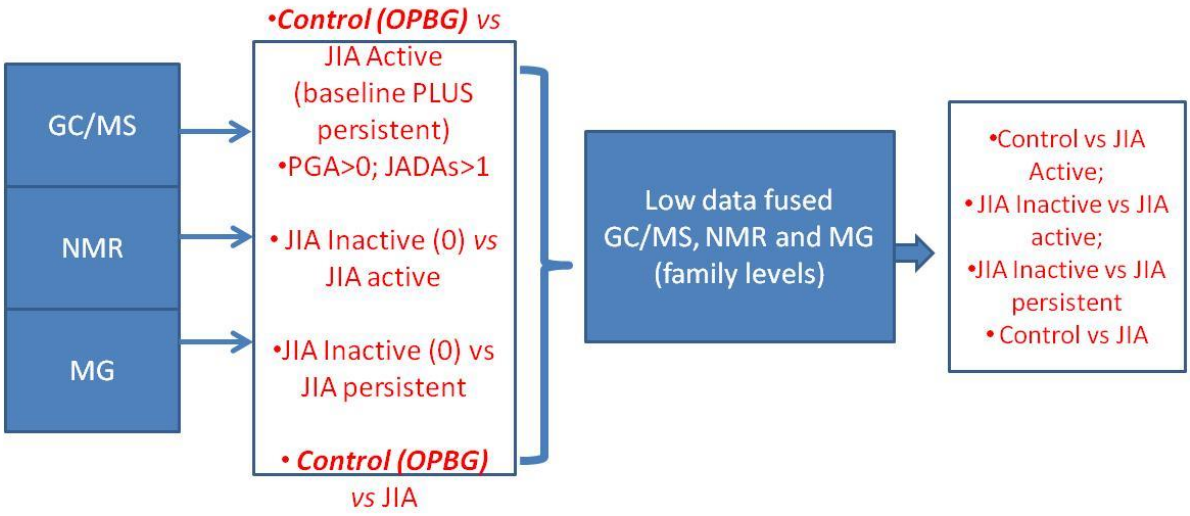


Figure 46 Principles of integration model.

Strategy for the multivariate analysis of metagenomics, GC/MS and NMR based metabolomics data and low data fused (integrated platforms)



Data Integration: metagenomic data were integrated with metabolomic data (H-NMR, GC-MS). Data Usage: Gnubila uploading, first set of data, excel files, generation of microbiota model. Obtained outcomes were compared with former. Integration patterns and data validation done: see ppt on MG and MB omics integration (**Figure 47-55**).

	Inner Loop (CV1)	Outer Loop (CV2: pred)
Control	99.20±1.94	99.73±1.01
Baseline+ Persistent (JIA active)	54.21±4.23	49.36±4.51
Total	70.99±2.44	68.16 ± 2.97

Correct classification rate

PLS-DA model is significant for NMC and AUROC

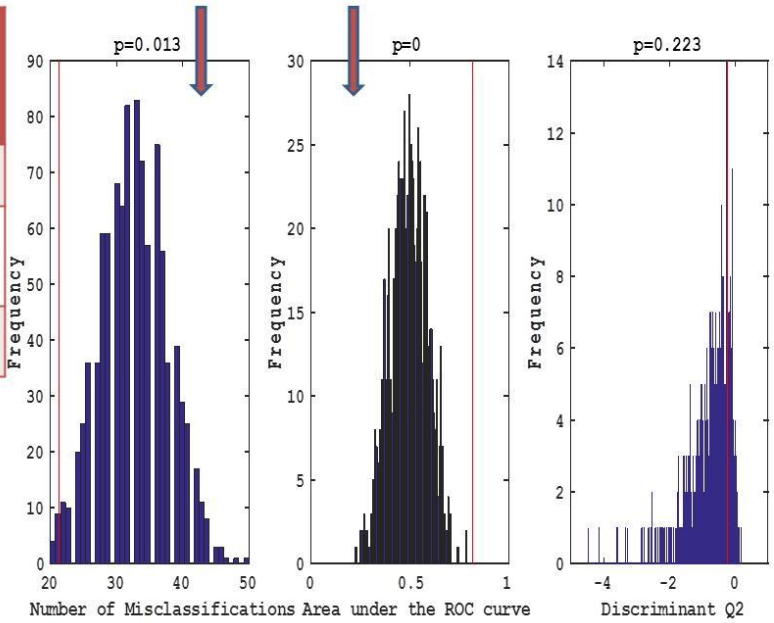
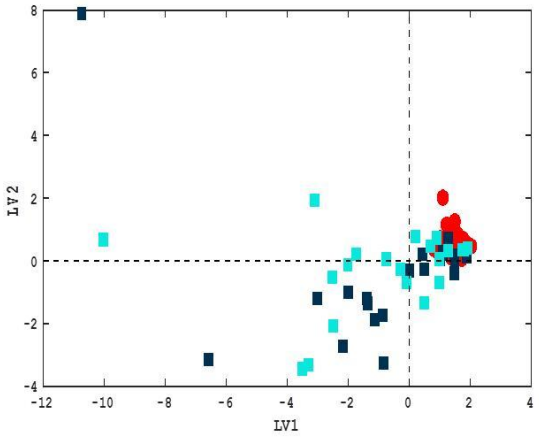
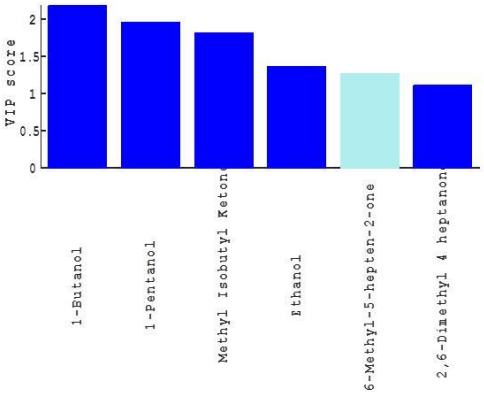


Figure 47 GC-MS Controls vs JIA active (baseline + persistent). Panel A.





Variables Important for Projection (VIP)
Dark blue: higher in JIA active
Pale blue: lower in JIA active

RED control
Black baseline
Pale blue persistent

Figure 48 GC-MS Controls vs JIA active (baseline + persistent). Panel B.

	Inner Loop (CV1)	Outer Loop (CV2: pred)
Inactive	41.05±8.11	39.65±4.53
Baseline+ Permanent	80.71±5.28	79.37±3.56
Total	68.36±4.58	66.99±2.94

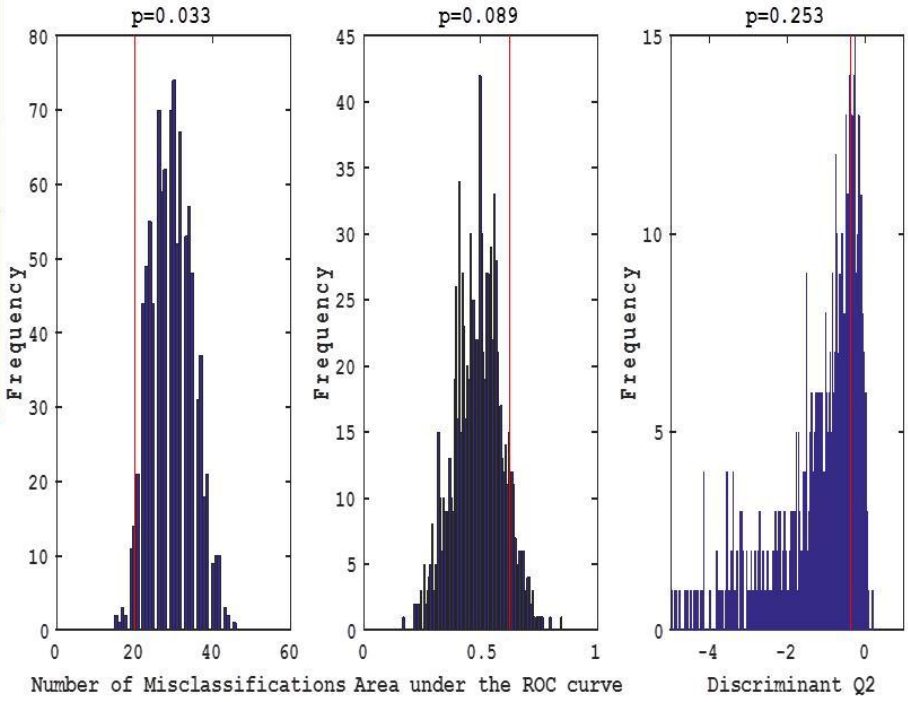


Figure 49 GC-MS JIA Inactive vs JIA active (baseline + persistent). Panel A.

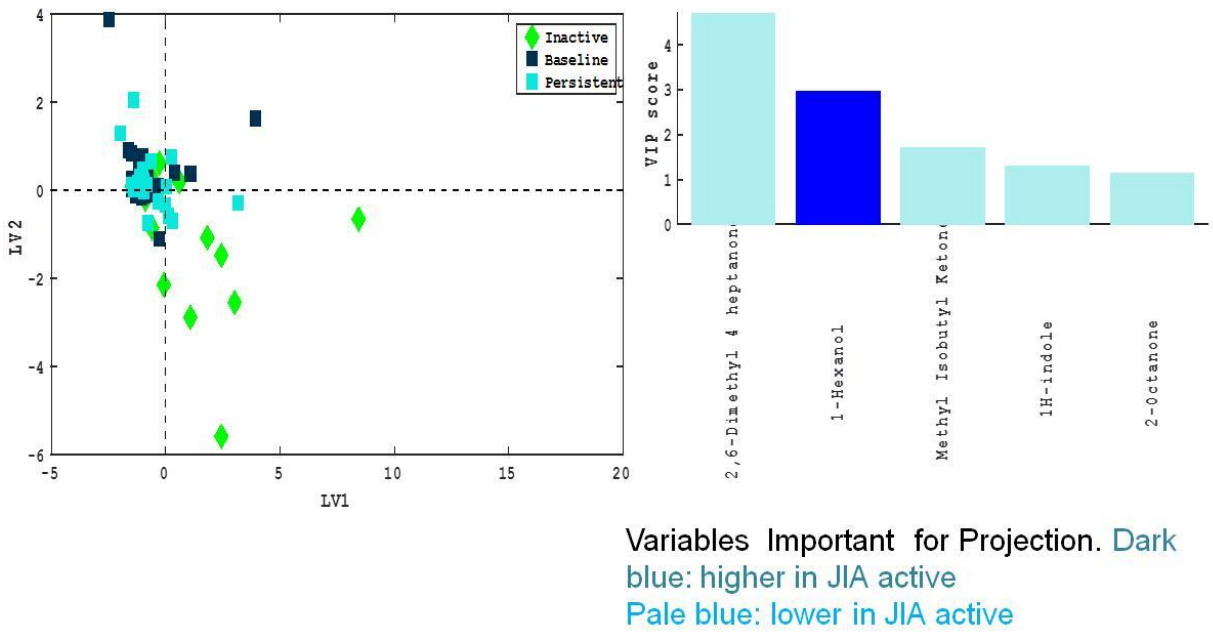
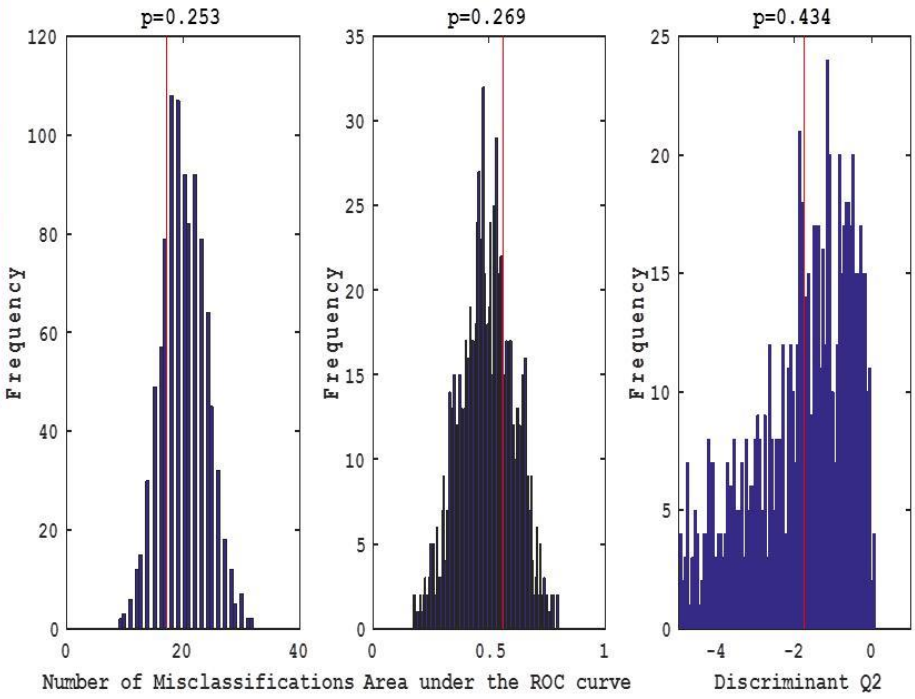


Figure 50 GC-MS JIA Inactive vs JIA active (baseline + persistent). Panel B.

	Inner Loop (CV1)	Outer Loop (CV2: pred)
Inactive	38.42±6.20	34.21±6.45
Permanent	82.06±7.68	78.41±5.56
Total	61.33±4.86	57.42±3.80

PLS-DA model is N.S



Leuven– 12-13 September 2016

Figure 51 GC-MS JIA Inactive vs JIA persistent active)

	Inner Loop (CV1)	Outer Loop (CV2: pred)
Control	99.20±1.63	99.20±1.94
JIA	50.77±3.81	47.92±2.61
Total	64.84±2.88	62.83±2.04

PLS-DA is a significant model but the misclassification for JIA is high (poor prediction)

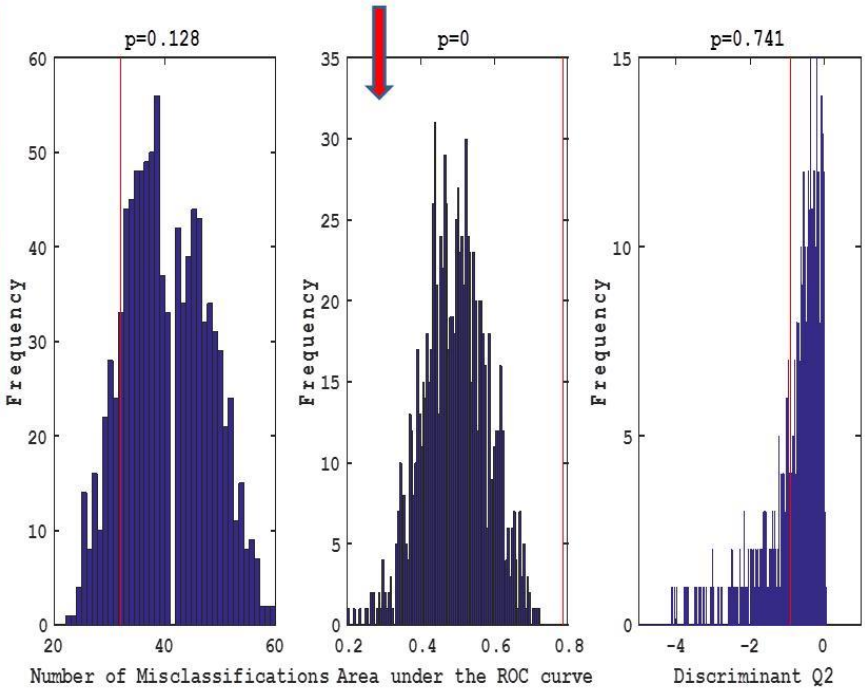
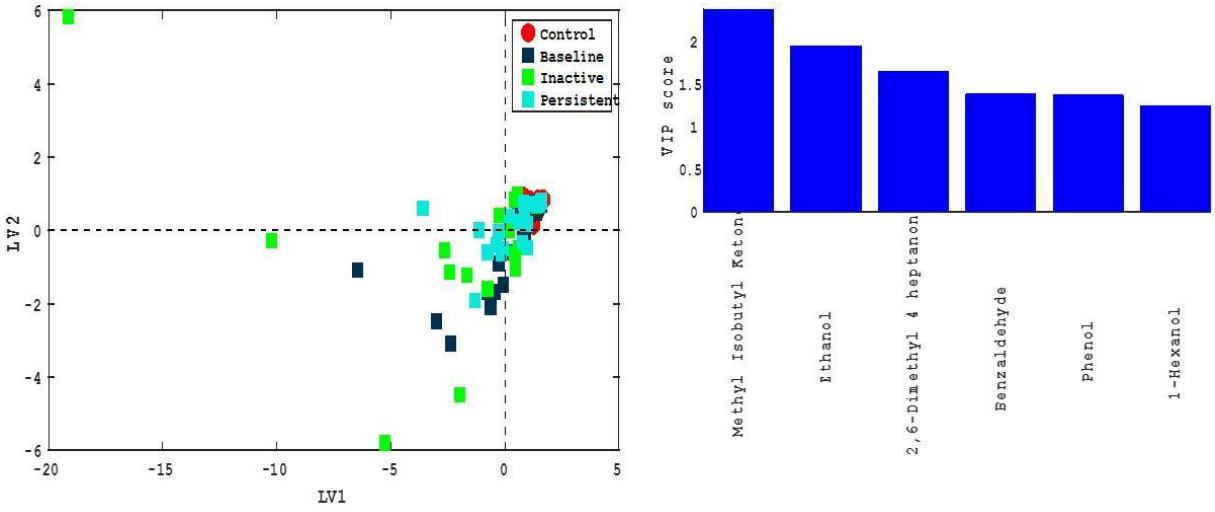


Figure 52 GC-MS Control vs JIA. Panel A.



Variables Important for Projection. Dark blue: higher in JIA

Figure 53 GC-MS Control vs JIA. Panel B.

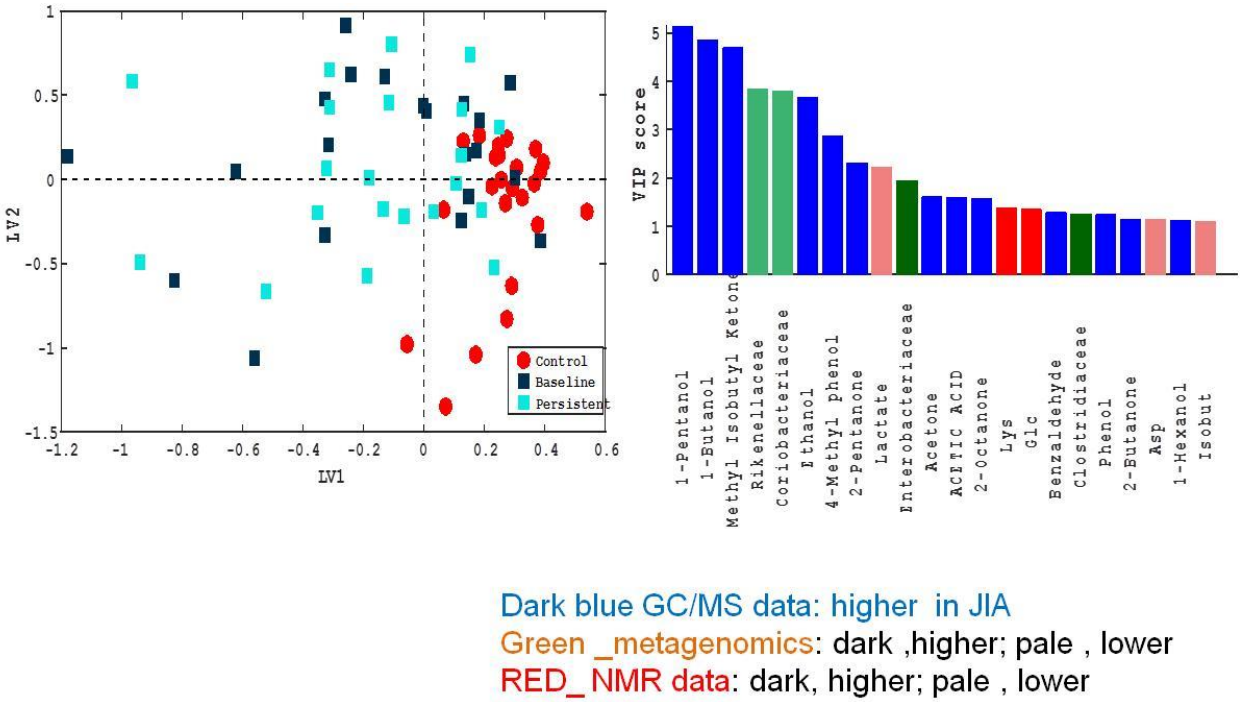


Figure 54 Low Level Data Fusion Controls vs JIA active (baseline + persistent)

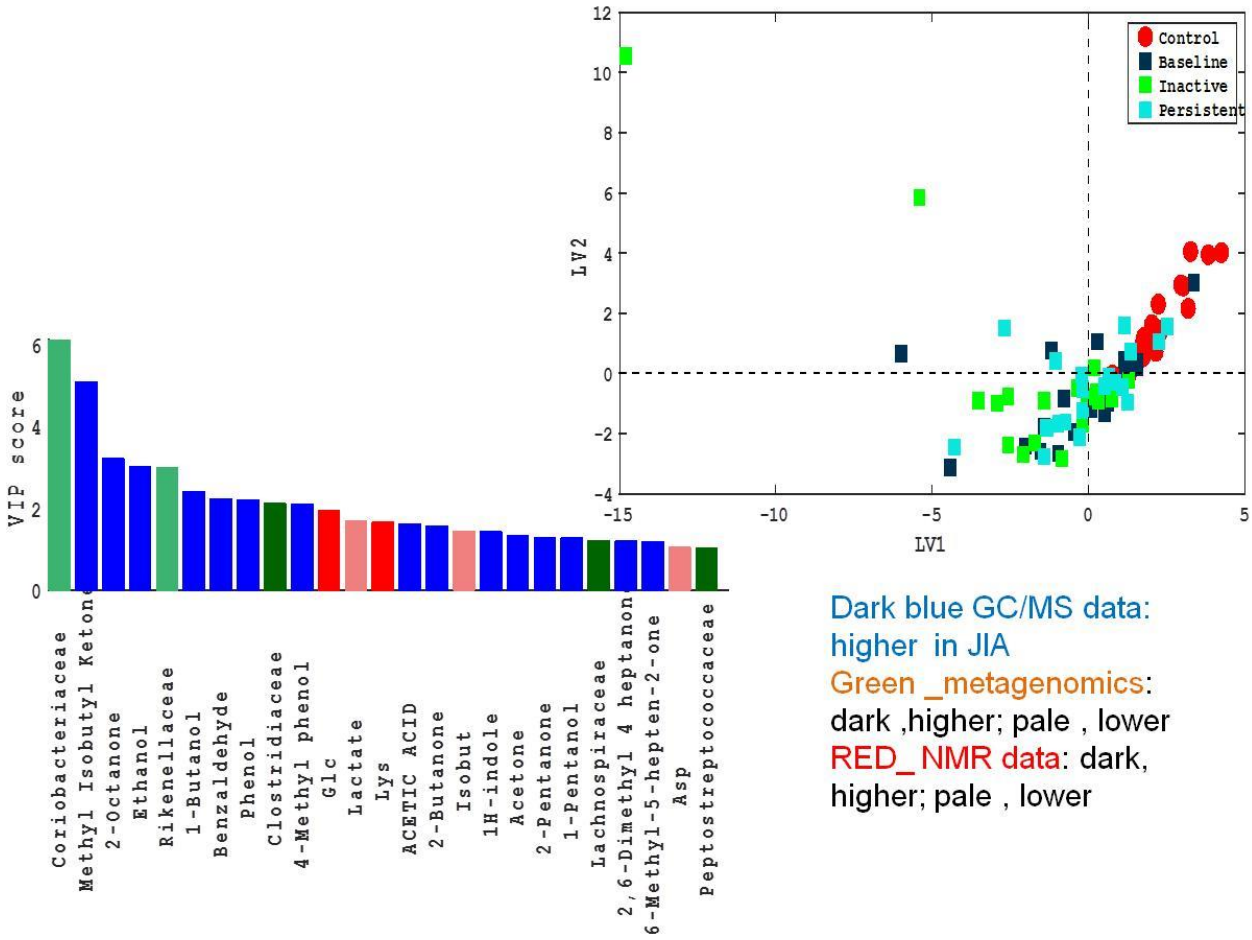


Figure 55 Low Level DF Control vs JIA.

The model produced the principal component describing the JIA disease model, compared to CTRLs. It resulted independent on disease activity (Figure 56).

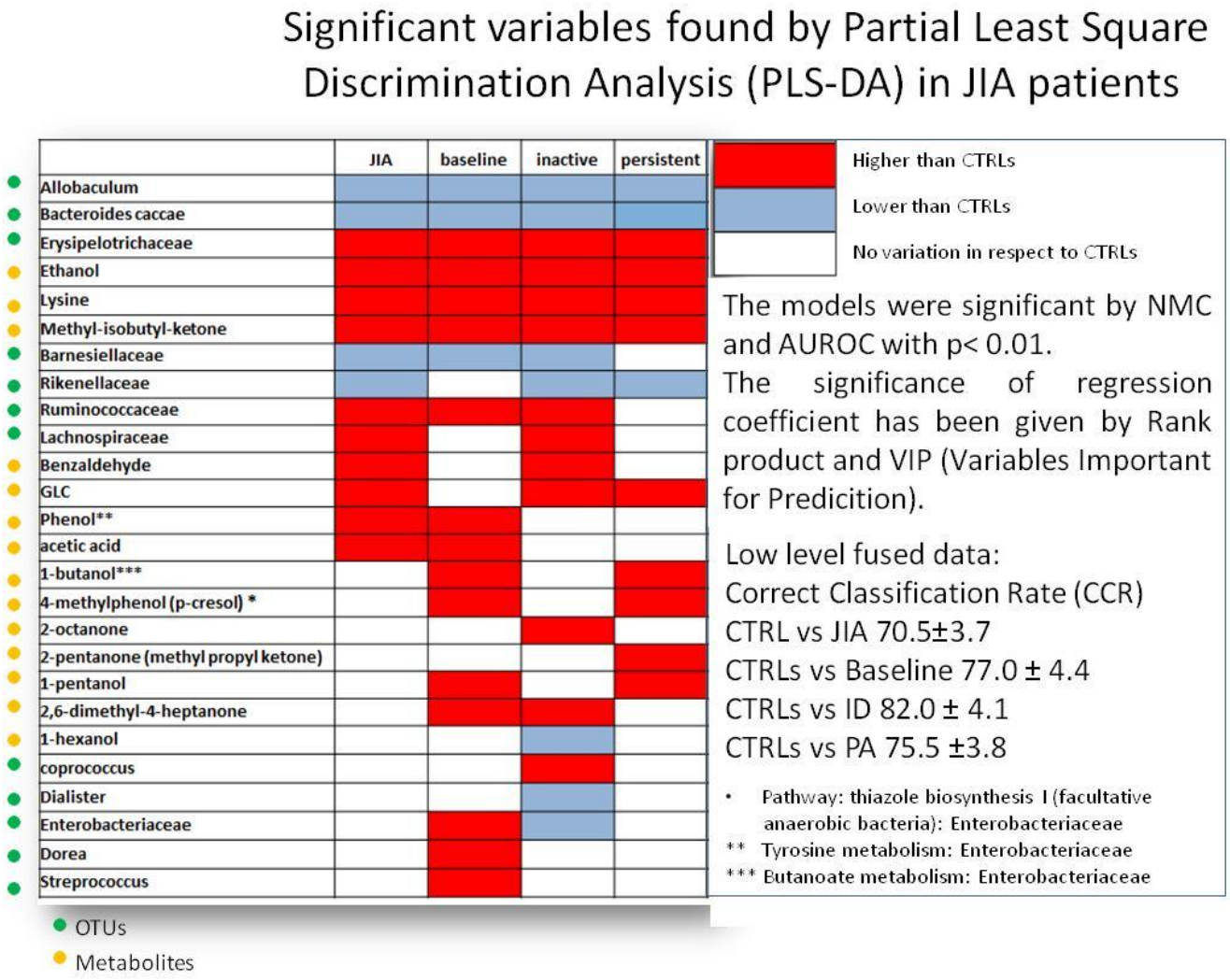


Figure 56 Predictive model of disease.

A goal of gut microbiota studies was to integrate the results into diagnostics and transferring them into clinical practice, providing model of microbiota enterophenotypes linked to disease, disease stages.

6.0 Validation for NND models (WP6-WP11)

6.1 Introduction

In this work package, the results of NND modelling efforts are presented. Three types of models were developed, of which the validation is described below :

- Clinically applicable Human Body Model (HBM) without use of MRI (paragraph 2.4.2)
- MRI-based modelling pipeline (paragraph 2.4.3 until 2.4.6)
- Forward dynamic simulation (paragraph 2.4.7 and 2.4.8)

Some overall considerations and conclusions are presented in paragraph 2.4.9, while scientific output (in progress) is listed in paragraph 2.4.10.

6.2 Validation of adapted Human Body Model

6.2.1 Comparison of HBM with two other conventional gait models (PIG and CAST)

VUmc performed a validation study on the new version of the Human Body Model (HBM), as developed in WP11, which can be run in real-time. In this study, the model was compared against two conventional gait models: the Plug-in-Gait (PiG) model and the Calibrated Anatomical System Technique (CAST), using ISB (International Society of Biomechanics) standards for anatomical frame definitions.

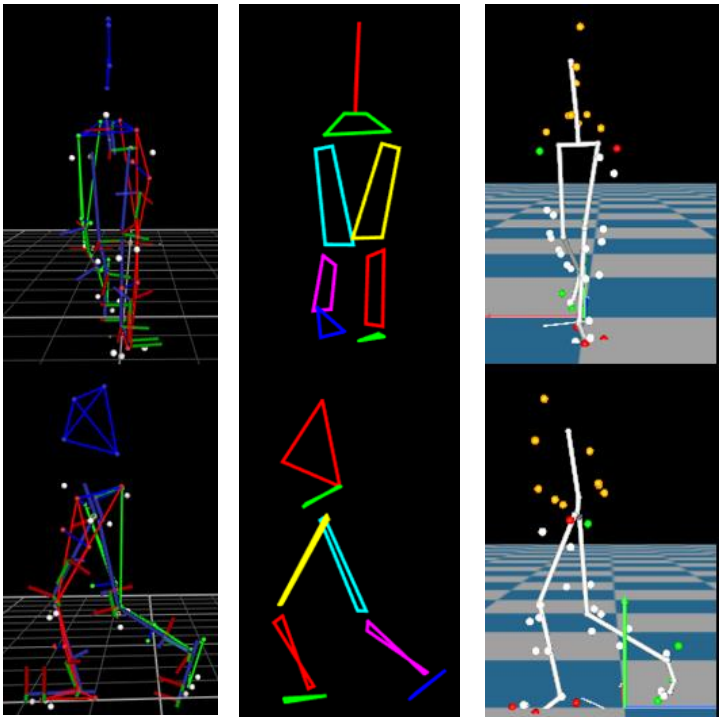


Figure 57 Schematic representation of PIG (left), CAST (middle) and HBM (right) models.

18 children with CP were included in the study, from all clinical centres: 6 from OPBG (overground); 6 from KUL (overground) and 6 from VUmc (GRAIL). 3 strides were selected for each subject. Statistical Parametric Mapping (SPM) statistics with ANOVA and post-hoc t-tests were used to quantify differences between gait graphs. Furthermore, root mean square errors (RMSE) between curves were quantified.

Results showed that differences could be seen between models, with similar variance between each of the three models. Most differences were within clinical relevant limits (5 degrees), except for differences in the transversal plane. The comparison between models in each of the three planes is seen as well as the RMSE values are shown in Figures below.

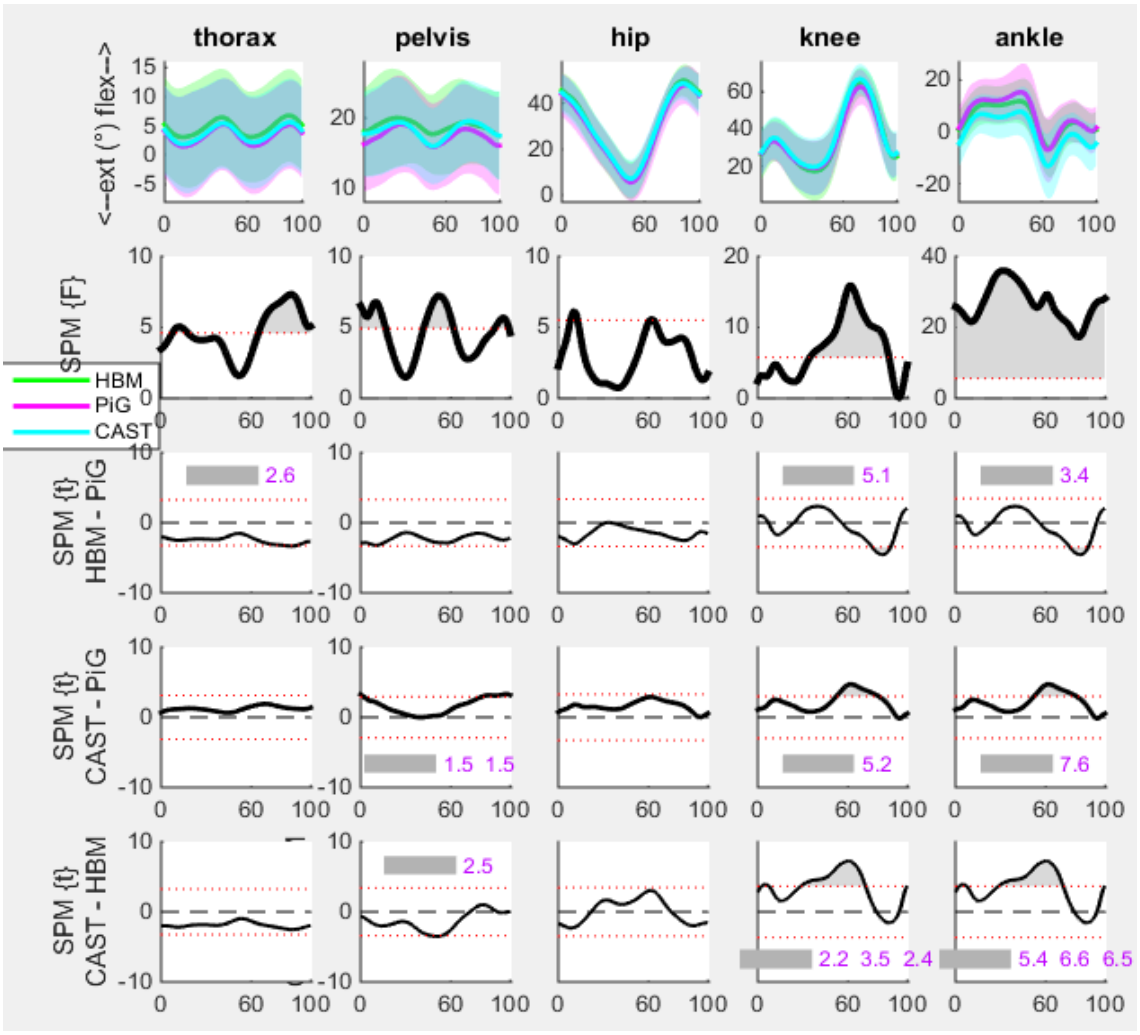


Figure 58 Sagittal plane angles for the three models (top row) and SPM outcomes of ANOVA (second row) and post-hoc analysis (bottom three rows), evaluating the differences between models. Grey areas and numbers indicate the size (RMSE value) of the significantly different areas.

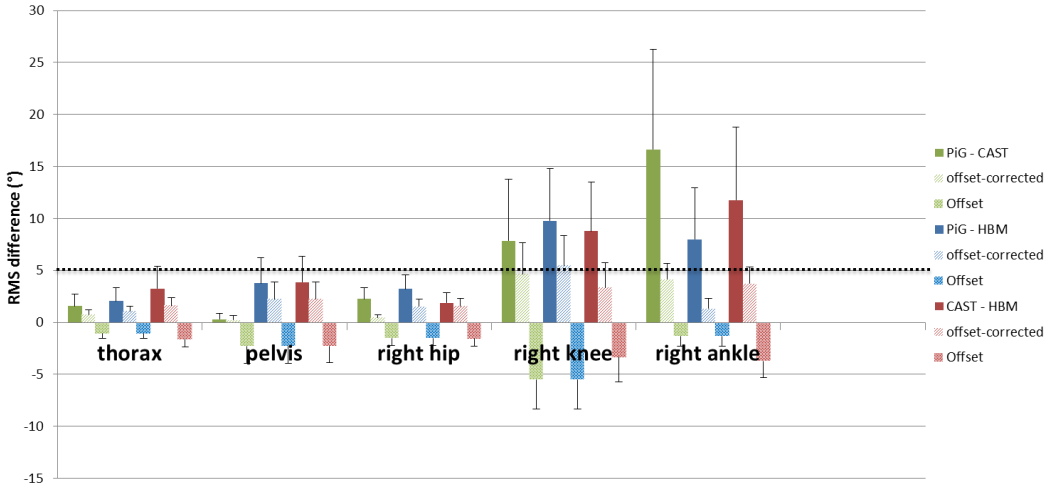


Figure 59 Sagittal plane RMSE differences between the three models; including a separation between the offset (systematic difference between curves) and offset-corrected differences (difference in shape of the curve). Black dotted line indicates a clinically relevant difference of 5 degrees.

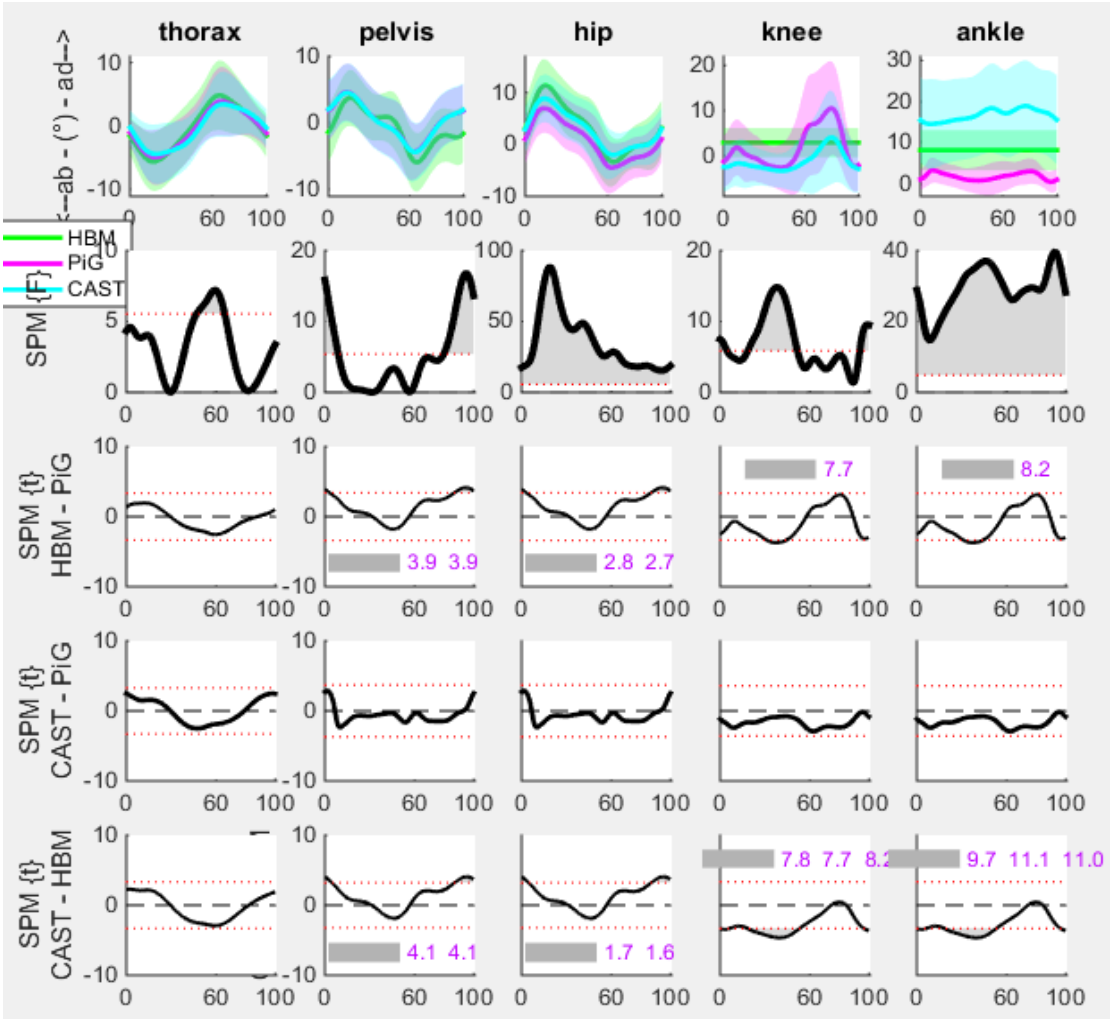


Figure 60 Frontal plane angles for the three models (top row) and SPM outcomes of ANOVA (second row) and post-hoc analysis (bottom three rows), evaluating the differences between models. Grey areas and numbers indicate the size (RMSE value) of the significantly different areas.

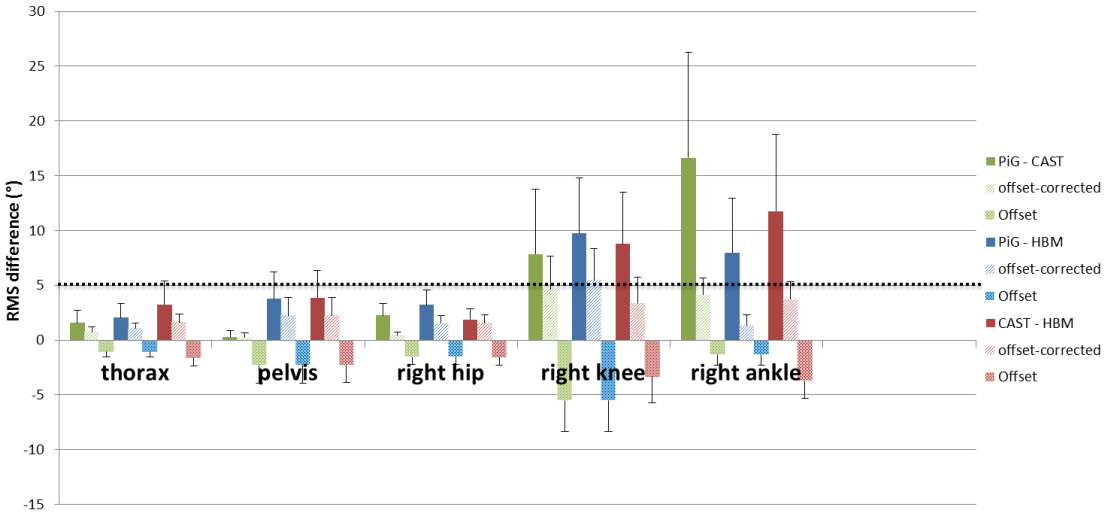


Figure 61 Frontal plane RMSE differences between the three models; including a separation between the offset (systematic difference between curves) and offset-corrected differences (difference in shape of the curve). Black dotted line indicates a clinically relevant difference of 5 degrees.

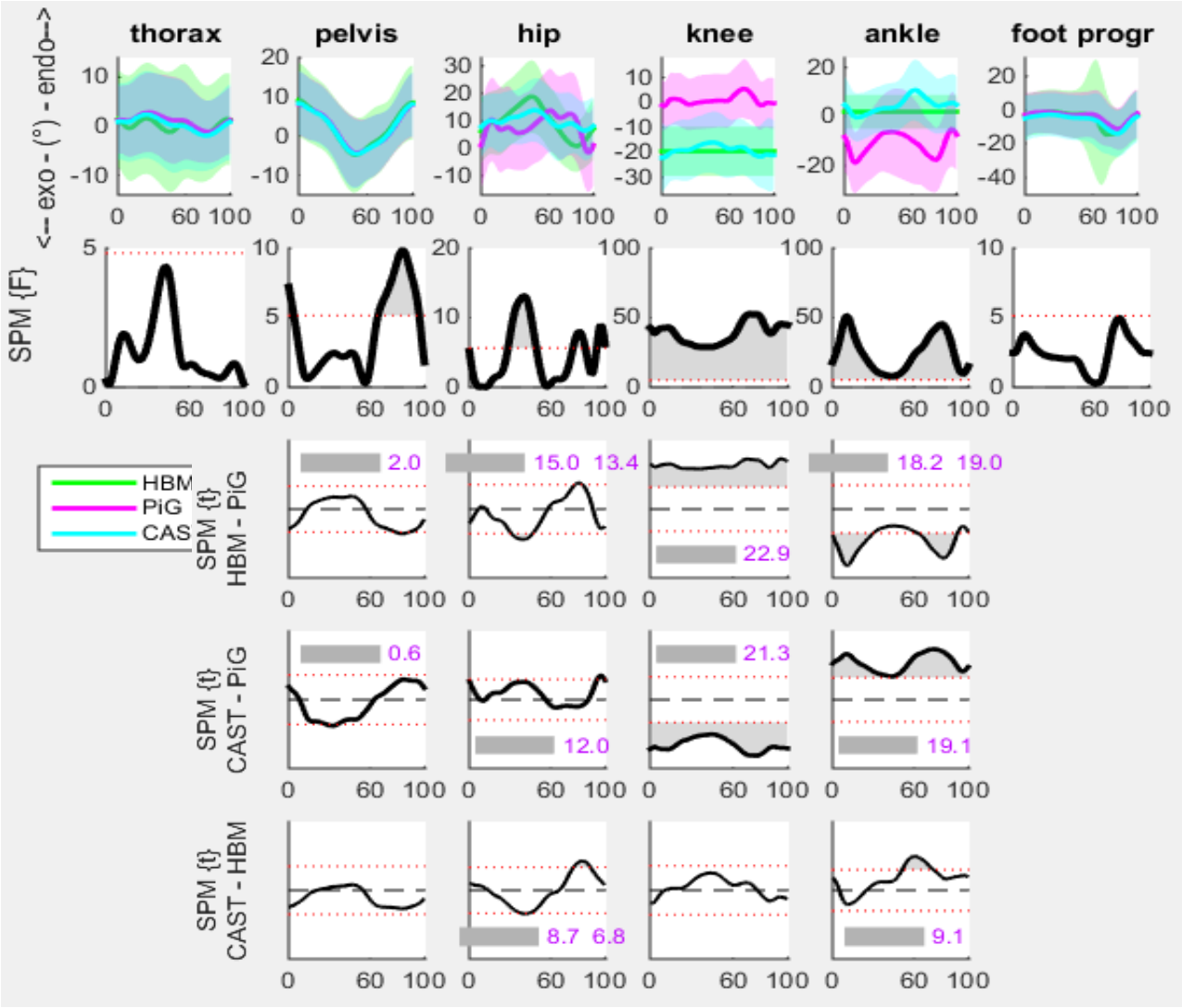


Figure 62 Transversal plane angles for the three models (top row) and SPM outcomes of ANOVA (second row) and post-hoc analysis (bottom three rows), evaluating the differences between models. Grey areas and numbers indicate the size (RMSE value) of the significantly different areas.

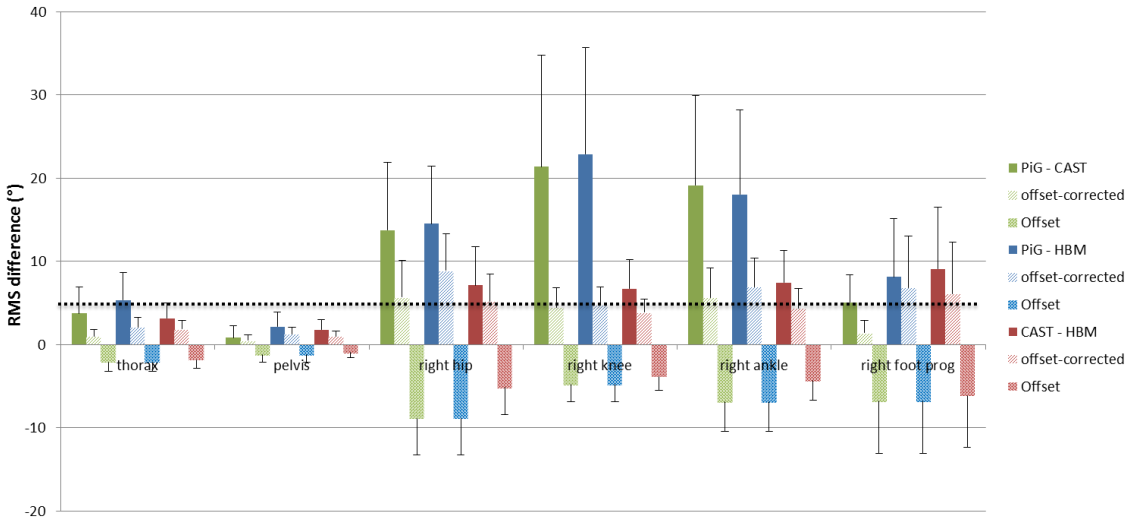


Figure 63 Transversal plane RMSE differences between the three models; including a separation between the offset (systematic difference between curves) and offset-corrected differences (difference in shape of the curve). Black dotted line indicates a clinically relevant difference of 5 degrees.

Based on the above described results, it was concluded that differences exist between models, but that none of these differences demonstrate the preferred use of CAST to PIG over HBM. Hence, the newly developed HVBM seems to be safe to be used for kinematic calculations in clinical settings. This will allow for advanced use of clinical gait analysis, as HBM can be run in real time, making it possible to for instance give real-time feedback to the patient, or apply perturbations based on the gait.

6.3.Comparison of HBM with commonly used modelling package

In order to evaluate the validity of the adapted HBM, the results of HBM were compared to those of OpenSim, a commonly used, but research-oriented modelling package.

The models in these systems rely on different definitions and assumptions (joint centre regressions), as well as other computational choices (objective functions). One goal of this work was to evaluate how these differences influence the biomechanical outcomes. The second goal was to evaluate the kinematic accuracy of both systems.

Walking trials and calibration poses were recorded for seven healthy adults and processed using the Gait Offline Analysis Tool 3.3 (using HBM) and OpenSim 3.3 (using the gait2392 model). The resulting joint kinematics, kinetics and muscle forces were compared using statistical parameter mapping. Different static muscle optimization options are available in both systems (HBM: muscle volume weighting enabled or disabled, OpenSim: force-length-velocity properties or ideal force generators) and all options are incorporated to evaluate their impact on the muscle force estimation.

The resulting joint kinematics from HBM and OpenSim showed similar patters but statistically differed for all degrees of freedom during large intervals. An offset in sagittal hip rotation is also observed (see Figure 1). Joint kinetics showed similar patterns but also differed statistically during several intervals for all degrees of freedom

(Figure 64). Muscle forces computed (for the same objective function) showed similar patterns but statistically differed during several intervals for most muscles (predominantly psoas, soleus and tibialis posterior).

The resulting differences can be ascribed to differing joint definitions (for example, the pelvic reference frame had a different orientation, the hip joint centres are determined by different regressions and the subtalar axes have different inclinations). Since muscle-tendon parameters are defined similarly in both models, different joint definitions are most likely the cause of the muscle force differences. Nevertheless, OpenSim and HBM provide consistent estimation of joint kinematics, kinetics and muscle forces for this data set from healthy subjects.

Constraining the muscle optimization in OpenSim by force-length-velocity properties generally had a relatively small effect. Enabling muscle volume weighting (HBM) had a bigger effect: the contribution of smaller muscles increased at the expense of larger muscles.

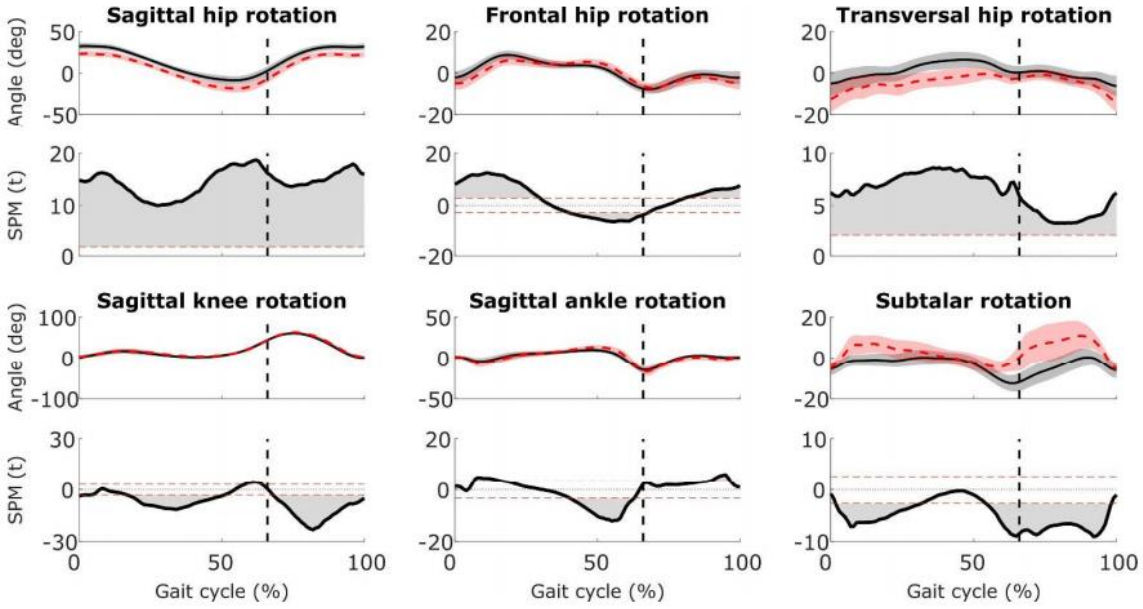


Figure 64 Comparison of joint kinematics calculated with OpenSim (dashed red) and HBM(black). Results from the statistical analysis (SPM) are shown underneath each kinematic plot (shaded areas indicate significant differences).

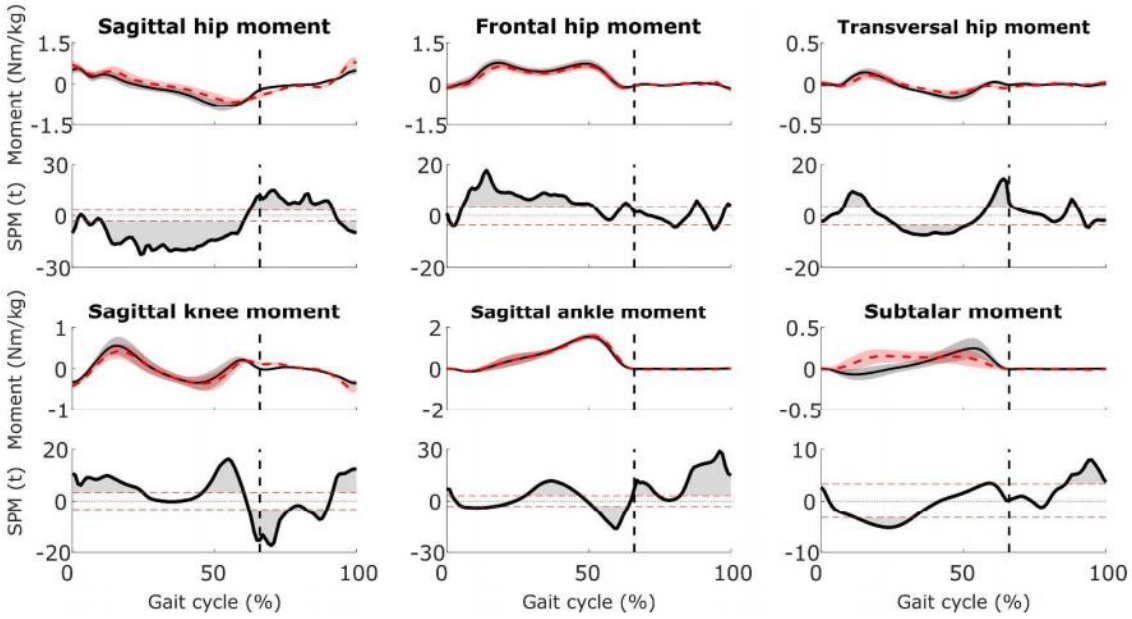


Figure 65 Comparison of joint kinetics calculated with OpenSim (dashed red) and HBM(black). Results from the statistical analysis (SPM) are shown underneath each kinematic plot (shaded areas indicate significant differences).

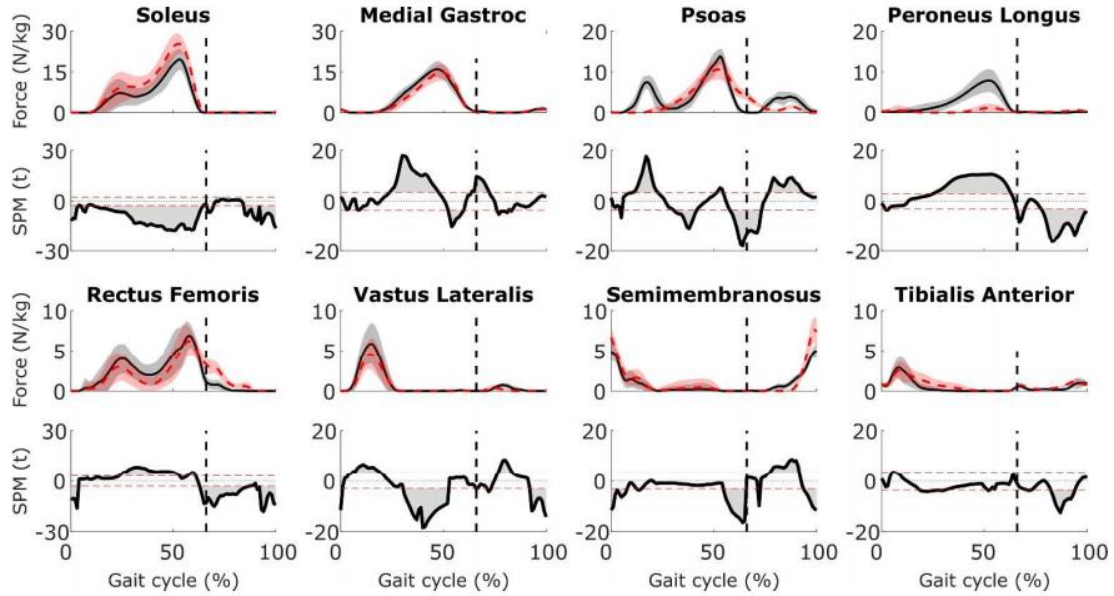


Figure 66 Comparison of muscle activations calculated with OpenSim (dashed red) and HBM(black). Results from the statistical analysis (SPM) are shown underneath each kinematic plot (shaded areas indicate significant differences).

As both OpenSim and HBM employ a least-square fit on the marker trajectories to fit the kinematic skeleton, it

is possible to compare the kinematic residuals (Euclidean distance between measured marker position and virtual (predicted) marker position). This measure provides a quantification of the kinematic accuracy of both systems. The RMS marker residuals were statistically smaller in HBM as compared to OpenSim (Wilcoxon signed rank statistical analysis, $p<.001$).

This study was done in close collaboration with the Department of Kinesiology of KU Leuven, Belgium. Paper has been submitted to the Journal of Applied Biomechanics.

6.3.1 Evaluating effect of functional calibration on kinematic residuals during gait

During this project, functional calibration algorithms (geometric sphere fit) have been integrated into Motek's commercial software package D-Flow (Figure 67). These algorithms estimate the hip joint centres and knee joint axes based on a trial were the subject performs a range of motion for either the hip (star-arc) or knee joint. The hip joint centre is determined by fitting spheres on the femur marker trajectories (Figure 68) and fitting discs on the tibia marker trajectories. (Figure 69).

In order to evaluate the effect of incorporating the results from functional calibration in the scaling process of HBM, we compared the kinematic residuals of trials processed with and without using functional calibration. In the first condition, the hip joint centres and knee joint centre and axes were determined by functional calibration. In the second condition, the hip joint centres and knee joint centre and axes were determined by predictive equations (Harrington). A pilot study on two healthy adults (single trial) showed that the kinematic residuals were lower in case of functional calibration (3.96 mm vs 4.40 mm, 4.23 mm vs 4.80 mm). The same test was done on the NND patient with the highest functional calibration range of motion, however, in this case the functional calibration did not perform better than predictive equations (7mm vs 6.5 mm). Adding extra technical markers slightly improved the result (6.9mm). Possible explanation might be an insufficient range of motion (not enough hip flexion) and that the subject may have been assisted (causing rapid movement).

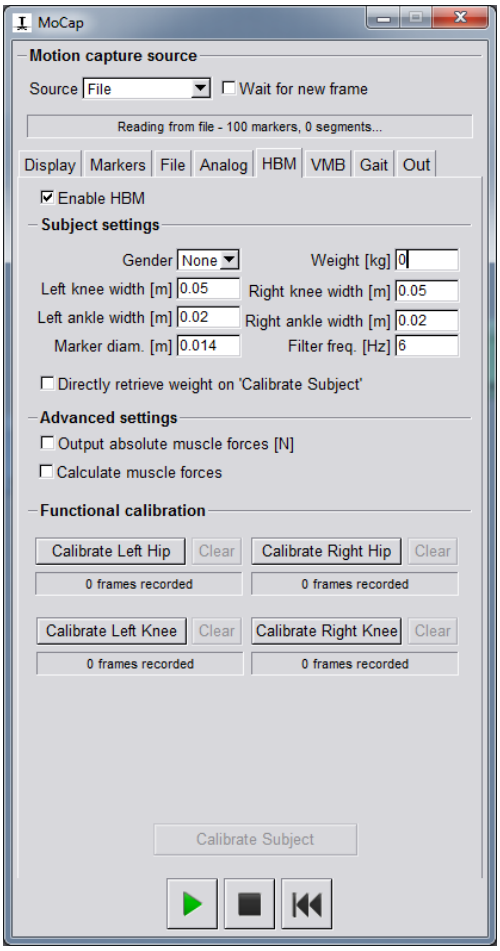


Figure 67 Functional calibration UI in D-Flow

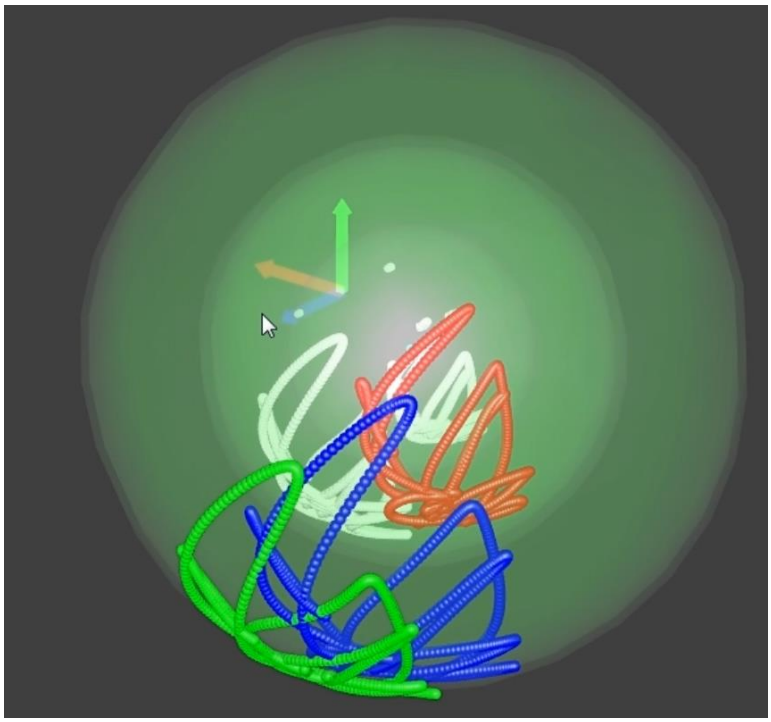


Figure 68 Spheres fit on the femur marker trajectories in pelvic coordinate system

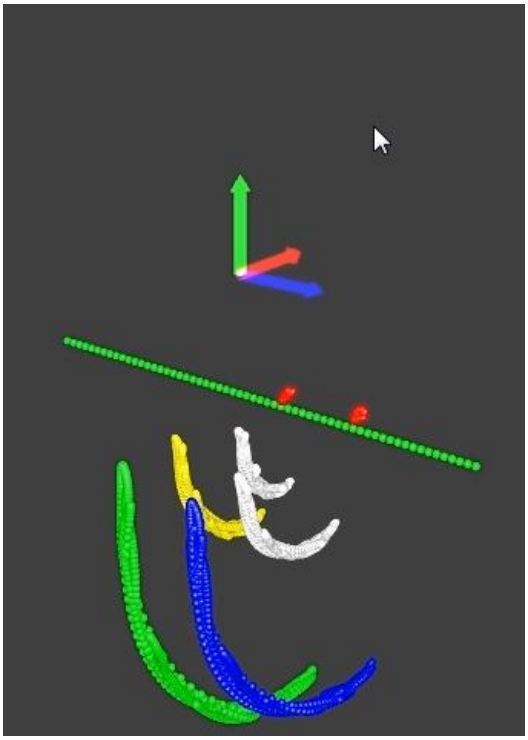


Figure 69 Knee axes fit on the tibia marker trajectories in femur coordinate system

6.4 Validation of the MRI segmentation

6.4.1 Basic Method

The method developed and used on healthy children's MRI images underwent special adaptations for each disease group. 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 Second, landmarks for the joints are detected in each target patient, and will be used to separate both legs into individual, non-articulated segments.
2. 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.
3. All affine transformations are then combined in a multi-affine approach, whose result provides both plausible and quite reliable pose information about the target 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.

6.4.2 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). We have created disease-specific atlas with detailed corresponding structures.
The original method's parameters were also extensively tuned for each group of patients. The evaluation results are shown in the Results section.

6.4.3 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 distances between the landmarks have been measured and the overall results on these landmarks, as well as the results per disease group, were measured.

6.4.4 Results

The overall landmark distance for the disease groups was comparable to the distances measured in healthy cases.

Overall, the method seems to perform quite well both quantitatively and qualitatively.

Figure 70 illustrates some qualitative results on different patients.

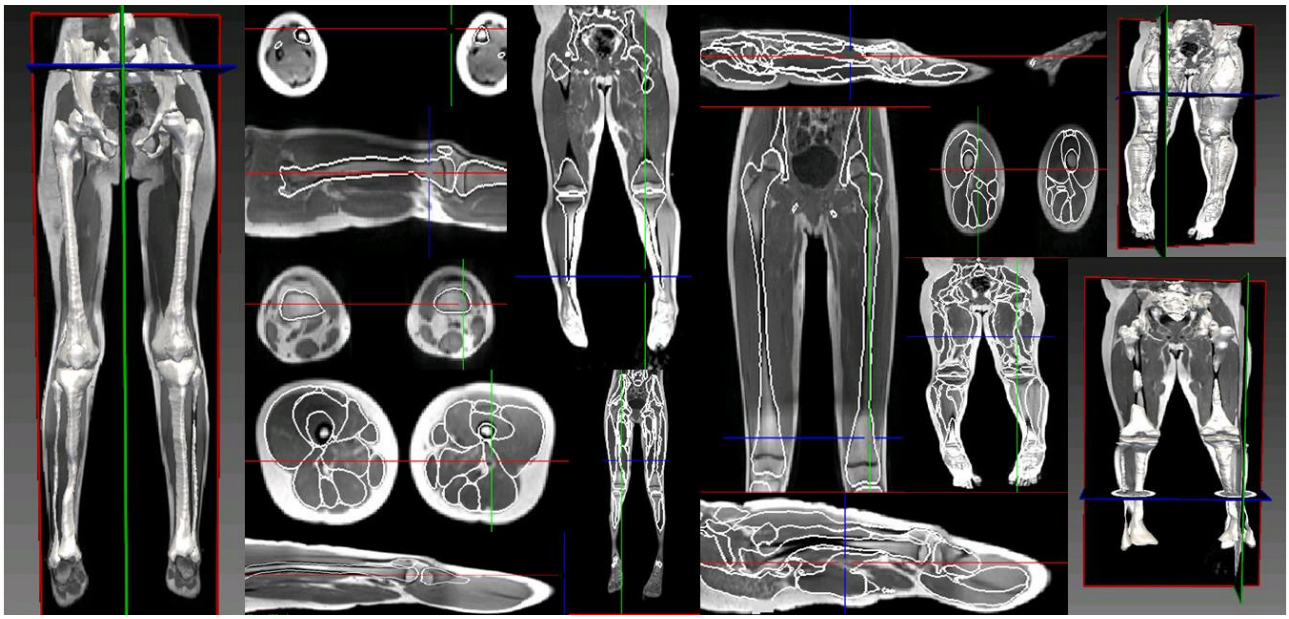


Figure 70 Qualitative results.

6.5 Validation of parameter extraction

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.

6.5.1 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.

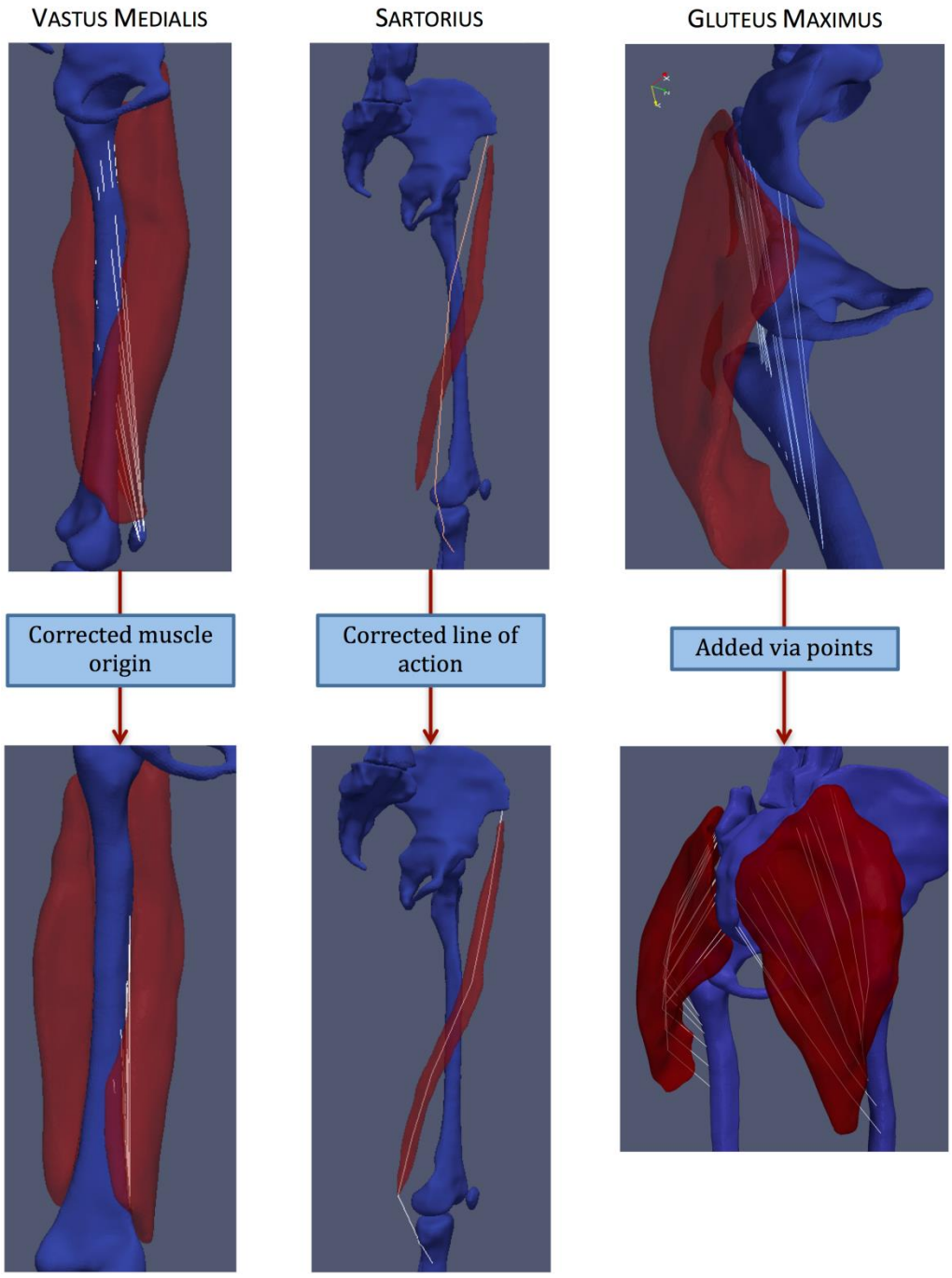


Figure 71 Three examples of the manual corrections of muscle-tendon lines of action performed during the evaluation of the anatomical accuracy of the complete MDPaedigree template for the lower limbs

6.5.2 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.

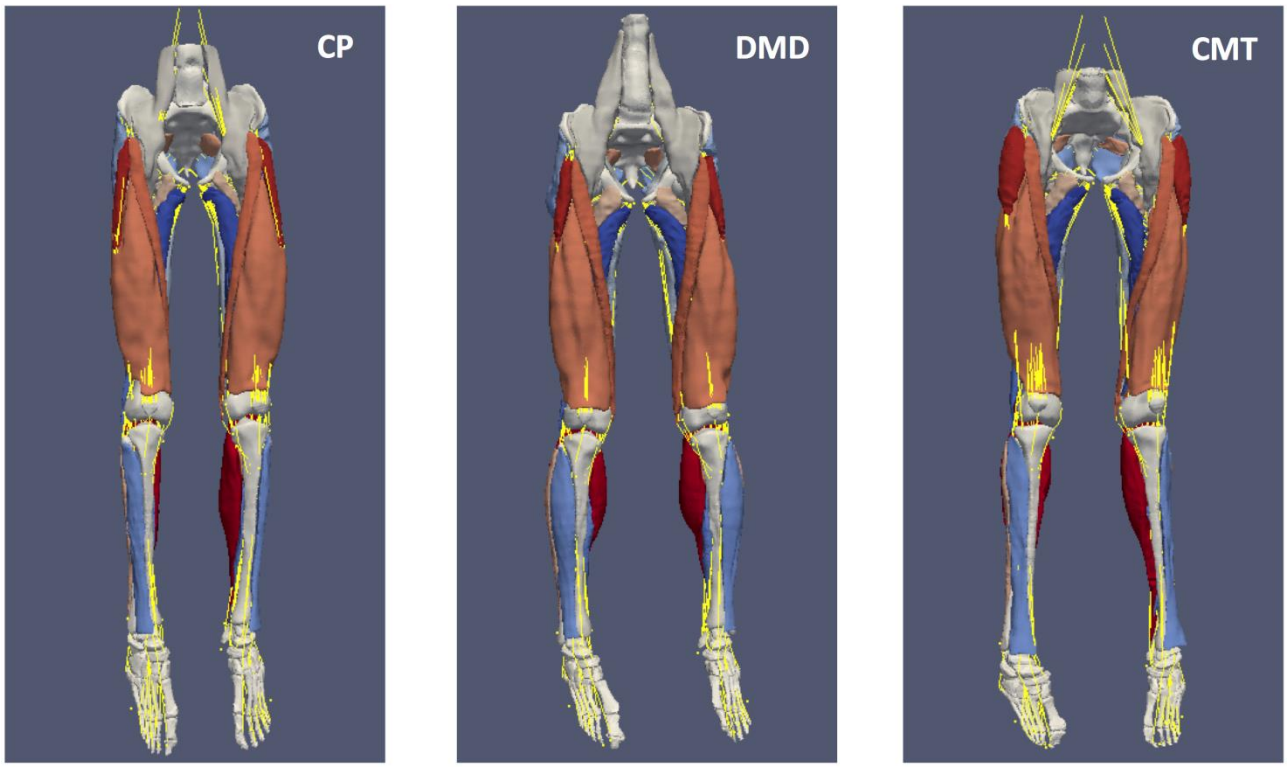


Figure 72 Three examples of patient-specific complete anatomical model, one for each of the diseases (CP, DMD, and CMT). They have been obtained by the morphing of the template anatomical model to the geometries segmented from MRI. Detailed visual assessment was performed for all the cases.

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.

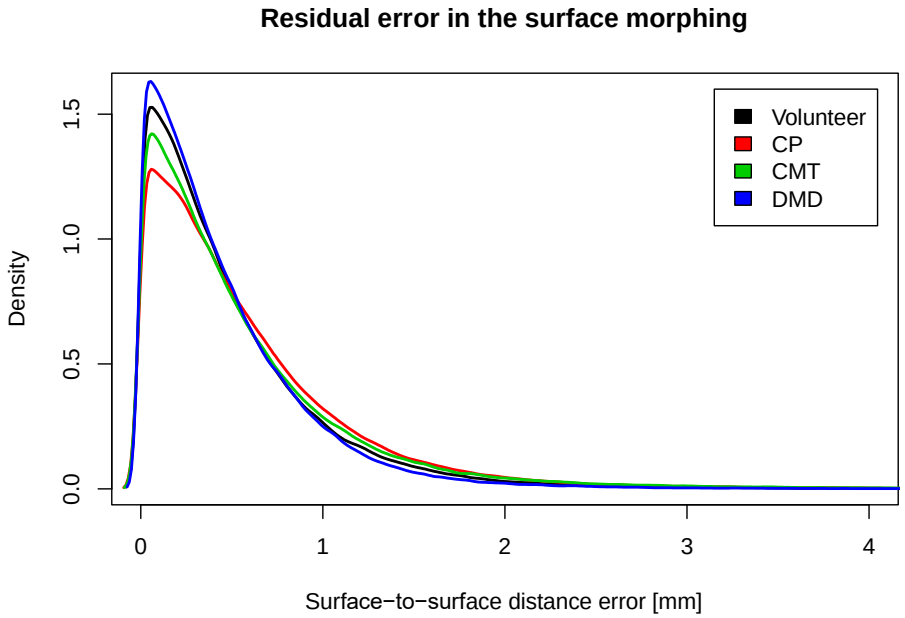


Figure 73 Distribution of the surface-to-surface distance error between the morphed model and the segmented geometries, across the complete model. The error has been stratified by disease group..

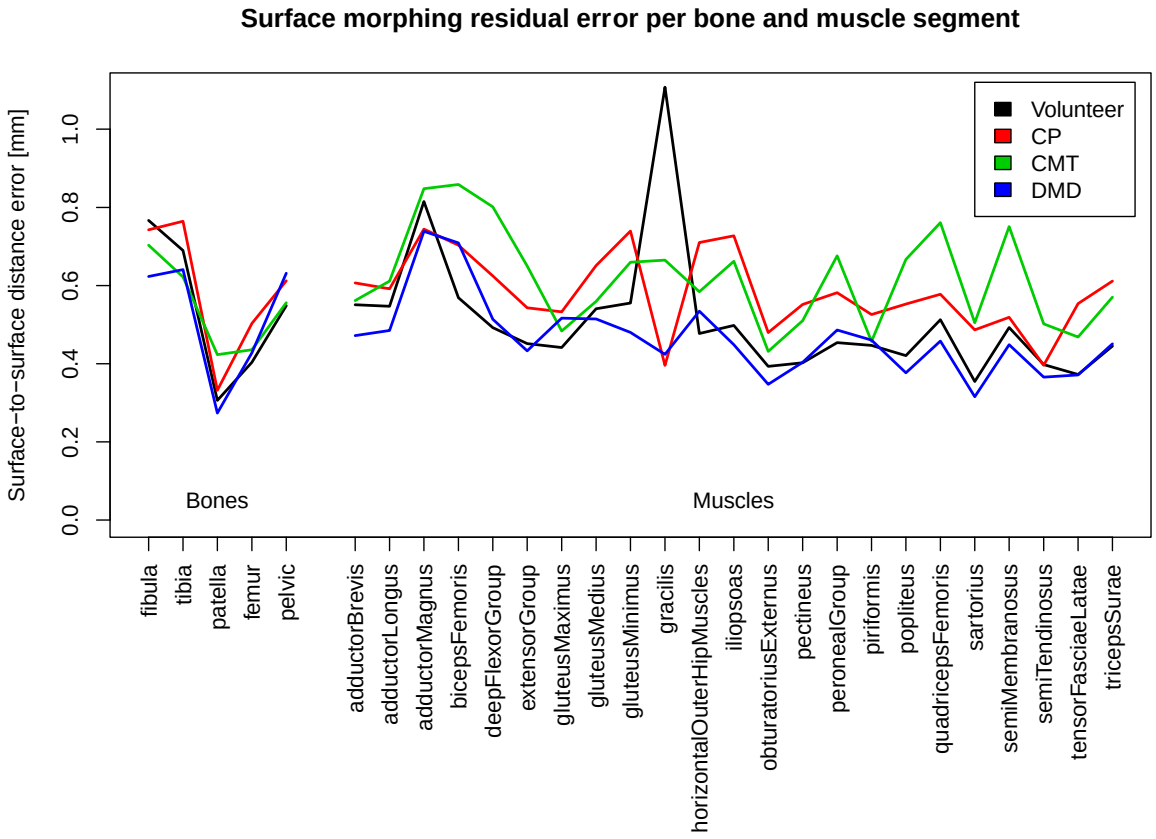


Figure 74 Mean surface-to-surface distance error for each individual bone and muscle, between the morphed model and the segmented geometries. The error has been stratified by disease group.

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.

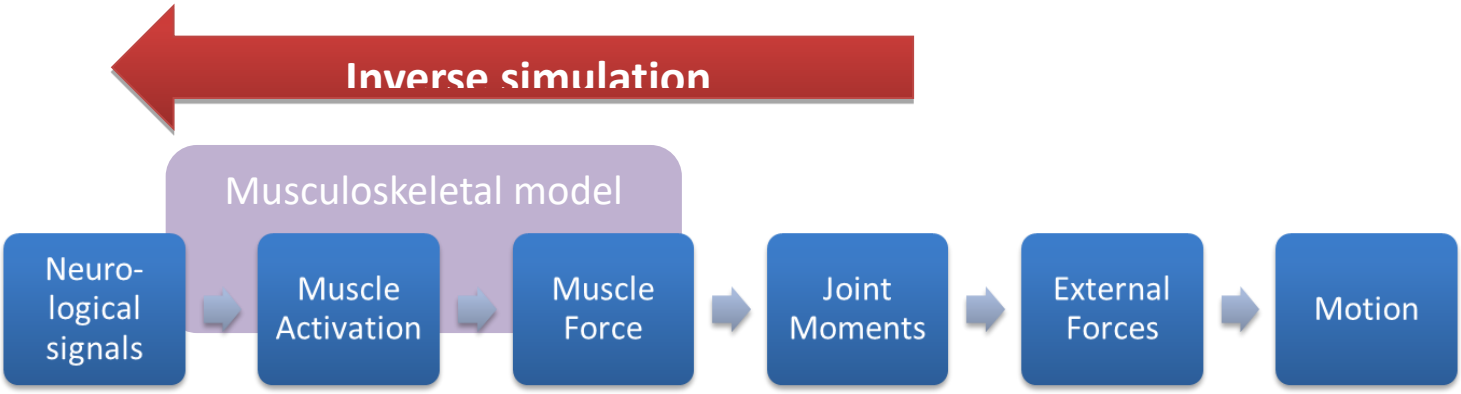
6.5.3 Number of patient-specific complete anatomical models generated from MRI data

A total of 43 valid MRI were collected by the 3 clinical partners (KUL, VUmc, and OPBG) and for the 3 diseases (9 CMT + 22 CP + 12 DMD). All of them where processed to extract the corresponding 43 patient-specific complete anatomical models. For that the MRI images where segmented by SIEMENS. Subsequently, USFD completed the anatomical model and extracted all the patient specific geometrical parameters.

6.6 Validation of inverse dynamics modelling pipeline

Inverse dynamics simulation

A common approach for validation of musculoskeletal models is the use of inverse dynamics simulation. In this approach, motion data and ground reaction forces are measured, and different quantities are derived from these measurements.

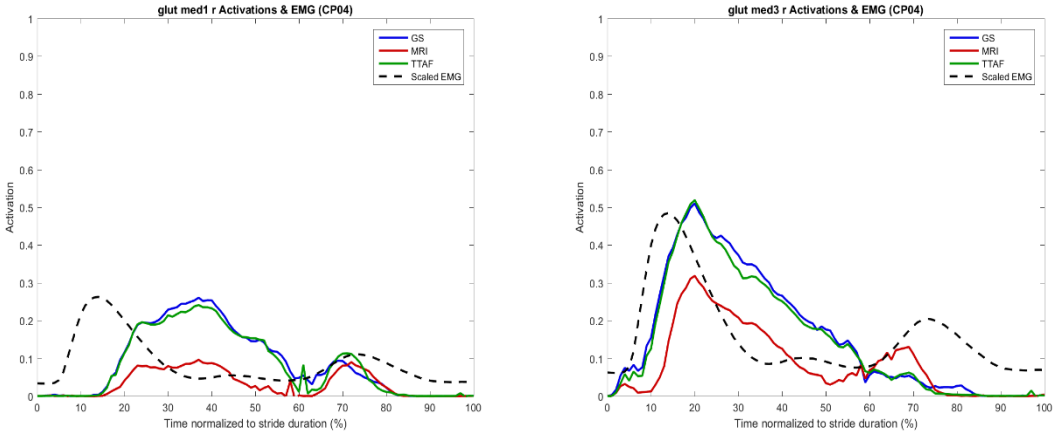


For the validation of the complete anatomical model, we have compared estimated muscle activation to EMG measurements of 43 available personalized models. In order to investigate the benefit of using subject specific models, we have made this comparison with three different modelling approaches:

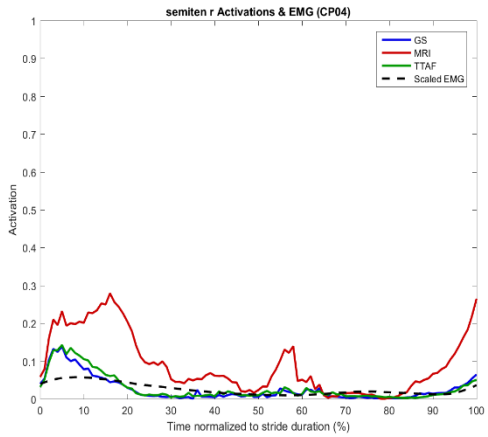
- Using a scaled generic model
- Using a generic model with additional morphing, based on tibial torsion and femoral anteversion measurements from the physical exam
- Morphing a generic model based on segmented MRI data

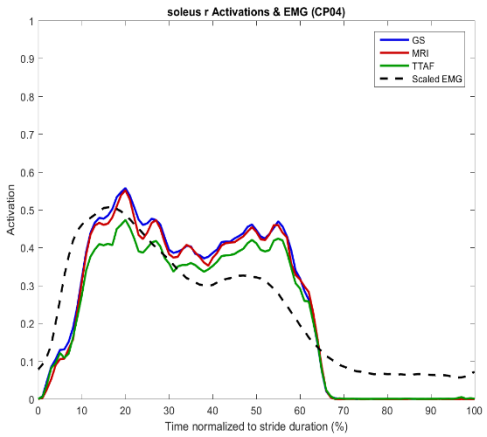
Results

We have performed inverse dynamics analysis for 43 subjects and 11 muscles, and compared these to normalized EMG patterns. A representative selection of the results is shown below:

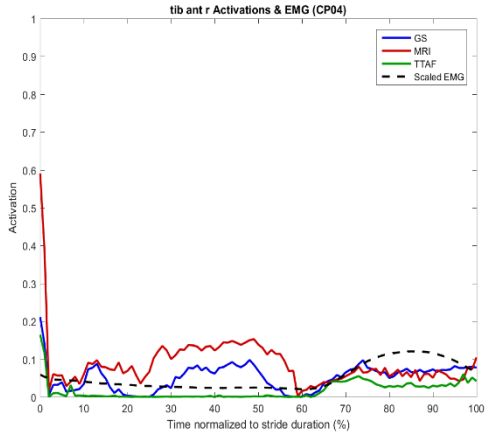


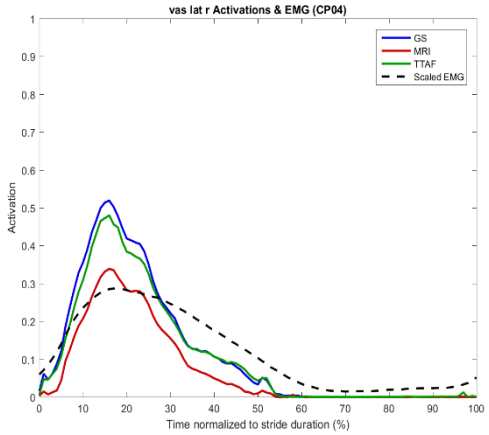
Gluteus Medius 1 + 2 (right leg)





Semitendinosus + Soleus (right leg)





Tibialis Anterior + Vastus Lateralis (right leg)

In analysing the results, we have drawn the following preliminary conclusions:

- The EMG data agrees reasonably well with the estimated muscle forces
- The difference between the estimated force of the three models is typically smaller than the difference between the estimated muscle force and the EMG

Discussion and Limitations

During this study we have come across a number of limitations:

- EMG does not provide us a “golden standard” to which we can compare the modelling results. This makes it difficult to evaluate different models using inverse dynamics optimization.
- Neurological constraints, such as spasticity or involuntary synergy, are not taken into account with inverse dynamics optimization. This is especially relevant for CP, the most prevalent patient group in NND.
- Key parameters are missing because they cannot be extracted from MRI, including muscle optimal fiber length and tendon slack length. Previous sensitivity studies by TUD have shown these parameters to be the most sensitive in inverse optimizations.

6.7 Clinical application and validation of modelling pipeline (VUmc)

The potential clinical relevance of the MRI-based modelling pipeline was illustrated using a clinical case example. The video still below shows a girl with CP, with a functional gait problem of frequent falling and tripping.

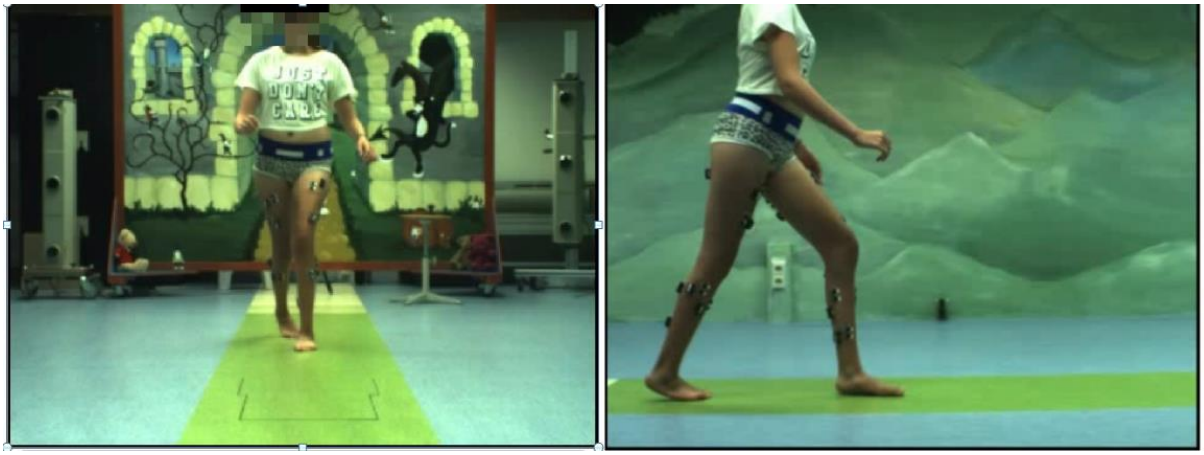


Figure 75 Video still of patient with a right leg foot tripping problem

The 3D gait analysis using the newly developed clinical HBM model showed excessive supination (a combination of ankle inversion and endorotation) especially during preswing:

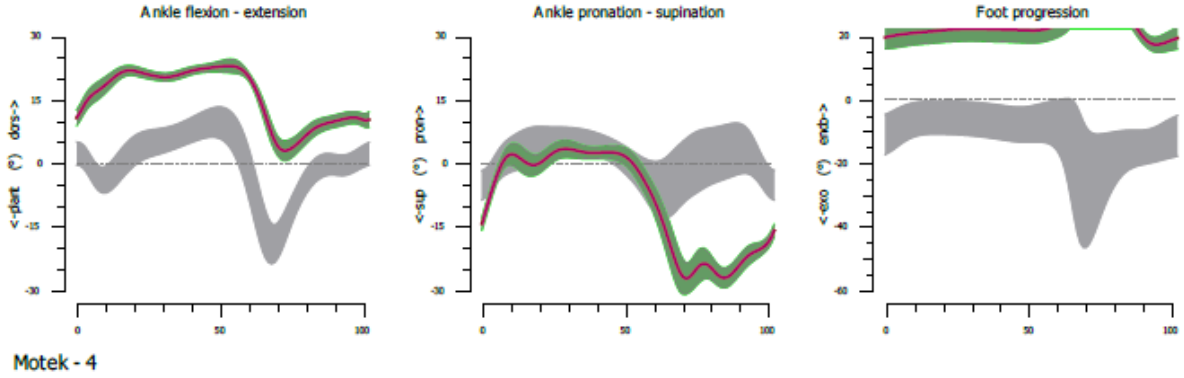


Figure 76 Ankle joint flexion/extension and pro/supination angles and foot progression angle for the patient shown in the previous figure.

This problem was hypothesized to be due (in part) to m. tibialis posterior (TP) over-activity or spasticity, but this is impossible to measure clinically, neither in physical examination, nor in EMG.

The MRI-based modelling pipeline, however, resulted in personalized estimates of the tibialis posterior muscle-tendon length as well as its activation during the gait, as shown in figure. These figures illustrate that, compared to the left leg, the right TP was not short during gait. According to the personalized model, it was even quite a bit longer than the left. However, the right tib post was overly active compared to the left throughout stance and even continuing into swing. This overactivity was even more pronounced in the personalized model. This shows

than indeed the TP muscle contributes to the excessive supination in swing, leading to the functional problem of tripping. It was decided to lengthen the TP surgically, resulting in less effect of the TP activity and reducing the functional problem.

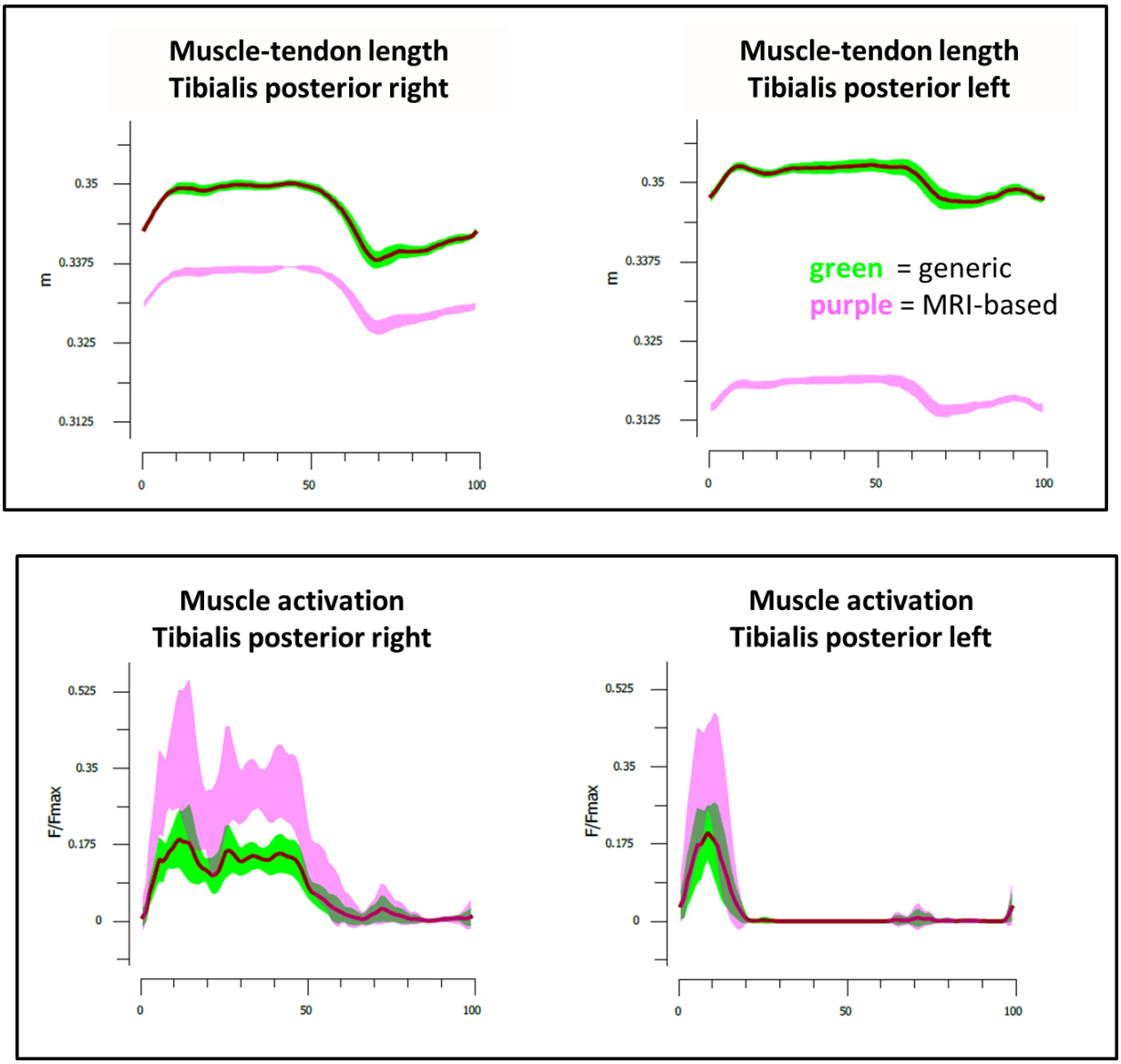
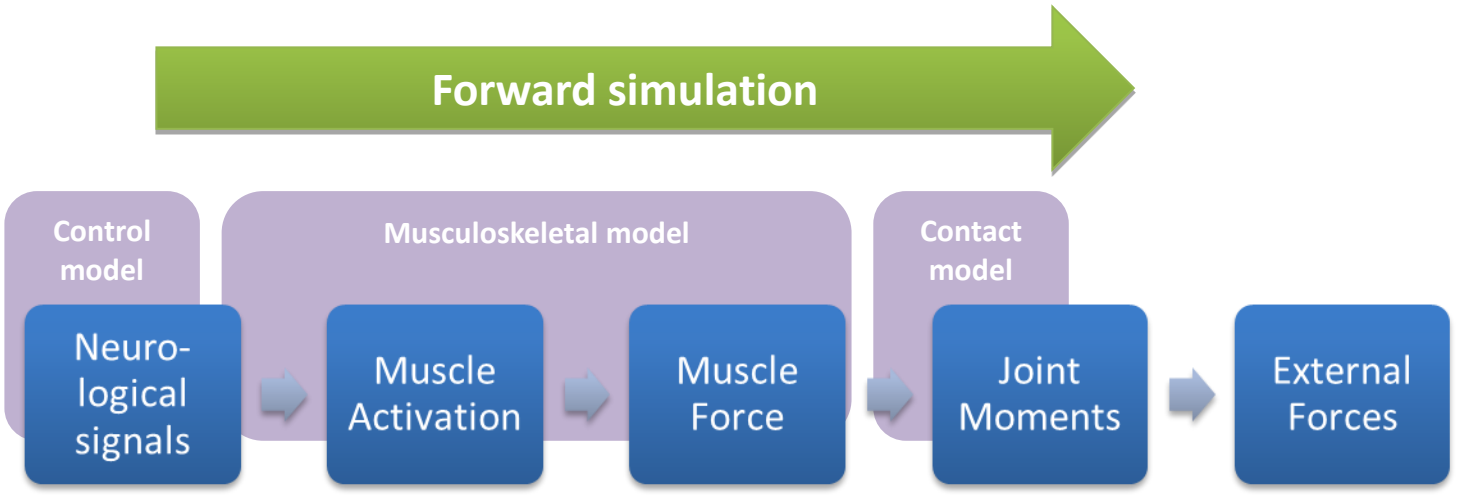


Figure 77 Tibialis posterior muscle-tendon lengths (top) and activation (bottom) during the gait cycle.

This example is just one of the many potential applications of personalized musculoskeletal models in clinical practice. A next step would be to perform a more extensive cost-benefit analysis, to identify whether the added value of MRI-based models outweighs the costs and burden to perform the MRI scans, analyses and modelling calculation in daily clinical practice.

6.8 Validation using forward dynamic modelling

Predictive forward simulations attempt the opposite from inverse simulations: they compute the motion and external reaction forces that optimally perform a given task, according to the principle that human motion is optimized for high-level objectives such as stability, metabolic energy expenditure and pain avoidance. Predicted motion patterns can take into account both musculoskeletal and neurological constraints, including neural delay and pathologies such as muscle weakness or spasticity. Musculoskeletal models can be augmented with electromechanical models of devices such as prostheses or exoskeletons, allowing researchers to predict human-machine interaction.



TU Delft have developed a method that allows forward dynamics validation through predictive simulation. For the analysis of personalized musculoskeletal models, the advantage of predictive simulation is that resulting motion patterns can take into account both musculoskeletal and neurological constraints, including neural delay and pathologies such as muscle weakness or spasticity.

Results

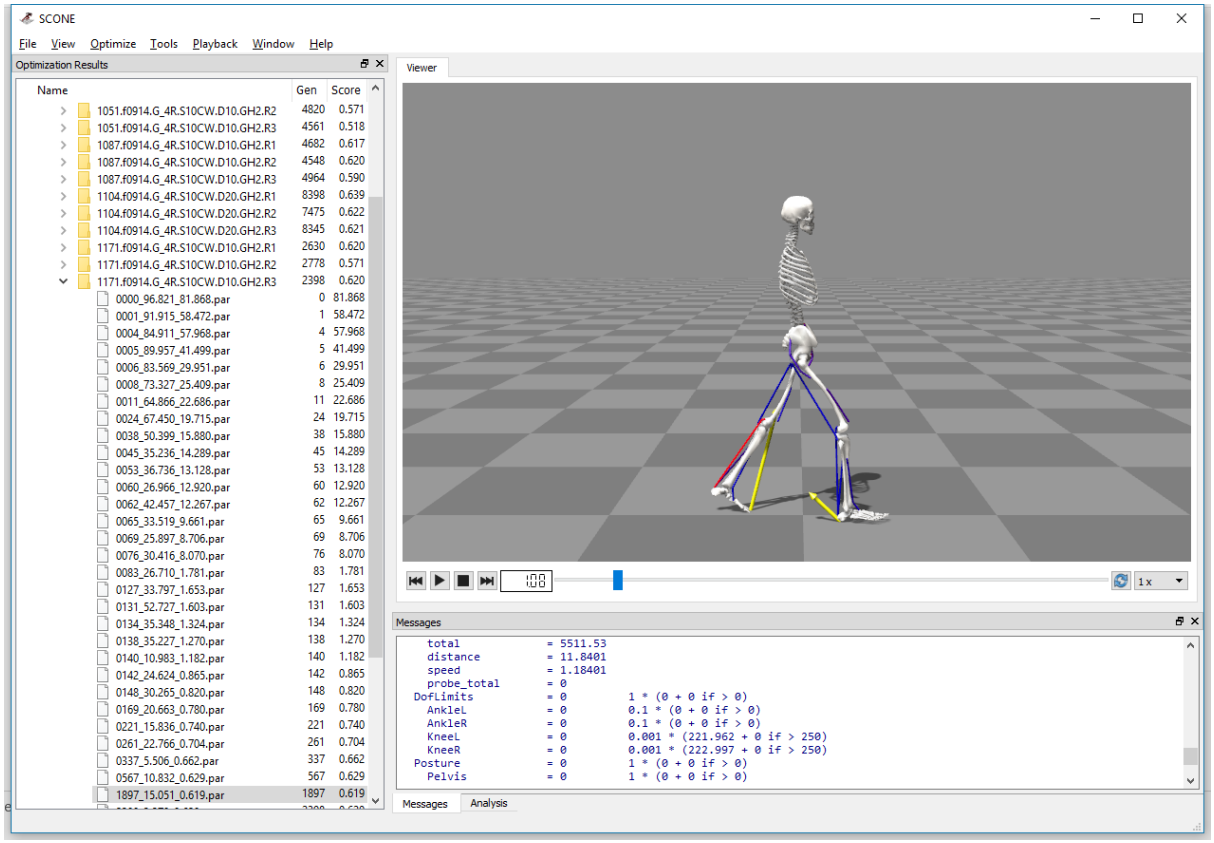
For our validation, we have modeled a set of NND-related pathologies, and compared the results with patients (face value). Pathologies include:

- Shortened hamstrings
- Shortened hamstrings + tendon elongation surgery
- Weakened vasti muscles
- Spastic gastrocnemius and soleus muscles
- Spastic gastrocnemius and soleus muscles + botulin toxin injection

The preliminary results of these studies are very promising, with a likeness to gaits found in subjects with similar deficiencies. We will perform more validation studies to further improve our model, and work towards a clinical test case in which we support clinicians in surgical decision making.
Videos of each of these results are available upon request **(t.geijtenbeek@tudelft.nl).**

SCONE – Software for predictive simulation of human gait

A major deliverable of TU Delft will be the release of the software that has been developed for performing the predictive simulations. The software is named SCONE (short for Simulation and Control OptimizatioN Environment) and is estimated to be released in fall 2017, under the open source LGPL license.



For more information and news updates, see **https://simtk.org/projects/scone**.

6.9 Clinical validation of forward dynamic modelling

To evaluate the face validity of the forward modelling results, five walking patterns were generated by the SCONE software. Besides normal walking, four walking types were synthesized all based on inducing a specific impairment into the neuro-musculoskeletal system. The choice of these impairments are more or less regularly observed in the populations of interest to NND (CP, DMD and CMT1)

A panel of ten clinicians from our network, i.e all engaged in promoting gait in paediatric diseases, and experienced in observing gait deviations, were approached by email , and provided with hyper links to a video , showing the result of our simulations

The following information and question was put forward;

(…) Part of the goals of MD-Paedigree are to create forward simulation models that would enable clinicians to perform what-if analysis, to evaluate tentative interventions in silico . Acknowledging this is a challenging tasks,

we have set a first attempt, for which we need your opinion.

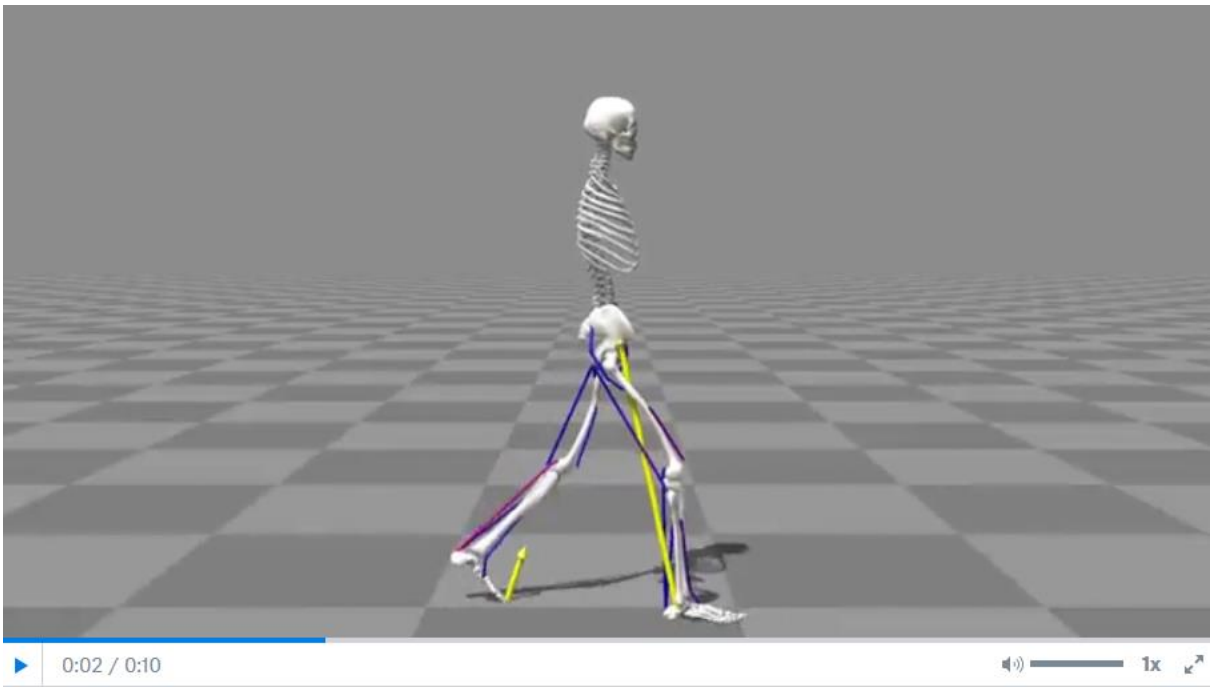
Five simulations have been performed, to let a human neuro-musculoskeletal system walk. All of these are based on a model that was optimized (using many interations) to get a synthesized movement. The model is 2D (sagittal plane only) although the visualizations are rendered into 3D, and has no arms. Four specific impairments (bilaterally implemented) were built in the model, affecting the gait pattern, to evaluate its face validity.

The long term goal is obviously to let these models predict interventions, and compare them to actual interventions, which would mean real clinical validation. This face validity is only a first step, that needs a trained eye to asses it .

I would like to ask your opinion to whether or not the 5 mimicked gait patterns, look as a realistic result to you, based on your clinical experience in this field. I know that isolated impairments are very rare in clinical practise, but I am convinced we picked sensible examples.
I would like to ask you whether you ***consider the animation realistic*** *based on your own professional, clinical judgement. Please rate as : Non/Poor/Fair/Reasonable/Full. In addition I would like a short comment, that reflect your consideration(s).*

The rates were assigned to a 5 point scale to allow for averaging. Essential remarks are summarized.

6.9.1 Result of the forwards dynamical simulation : healthy gait

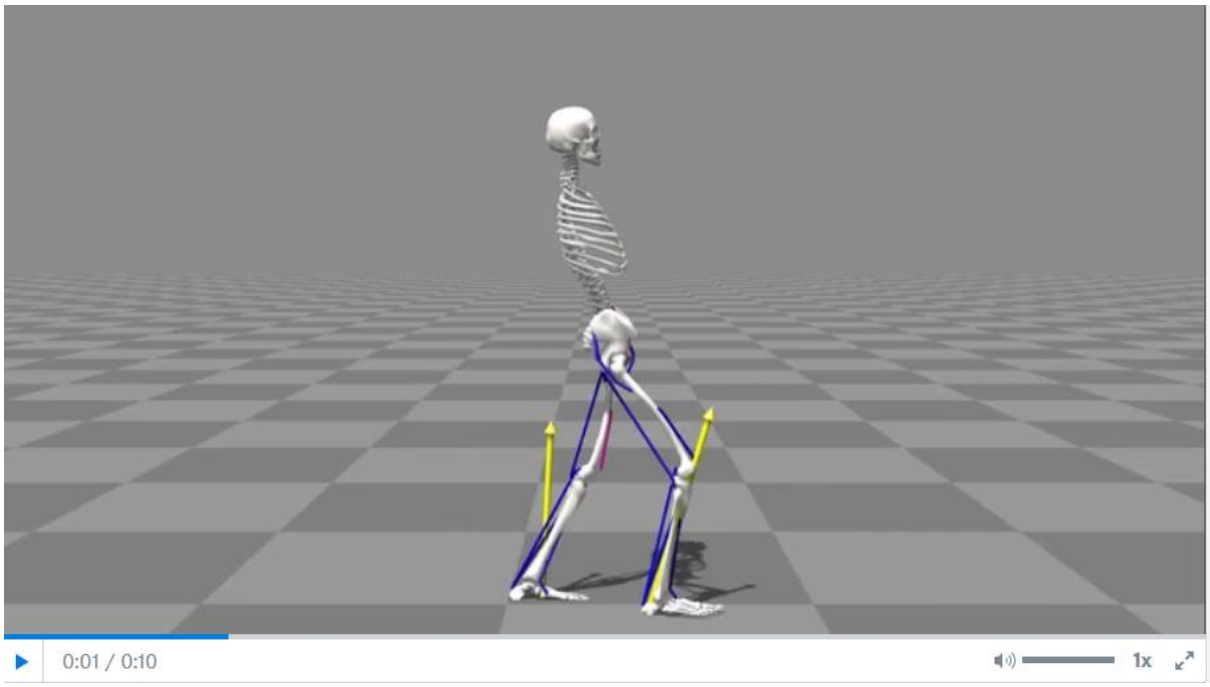


link: https://www.dropbox.com/s/cuu4qkquk0f4a8f/healthy.avi?dl=0

level of realism : 4.1 (/5) > quite realistic

Comments summary: Legs are OK, but the forward/back ward sway of the upper body does not seem realistic.

6.9.2 Result of the forwards dynamical simulation : Shortened Hamstrings

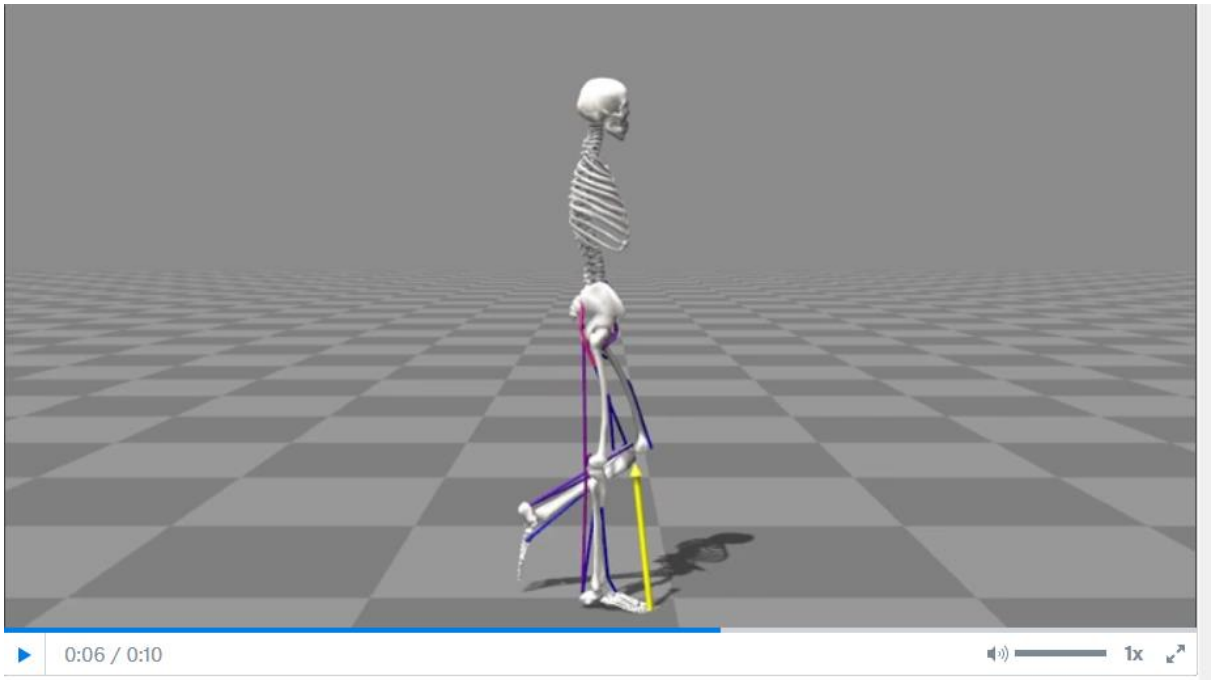


https://www.dropbox.com/s/wtl7bv918gilu9v/short_ham.avi?dl=0

Level of realism : 4.2 (/5) > quite realistic

Comments summary: Body too posterior, and (*one member*) a lack of posterior tilt movement in the pelvis at late swing.

6.9.3 Result of the forwards dynamical simulation : Weakened Vastii

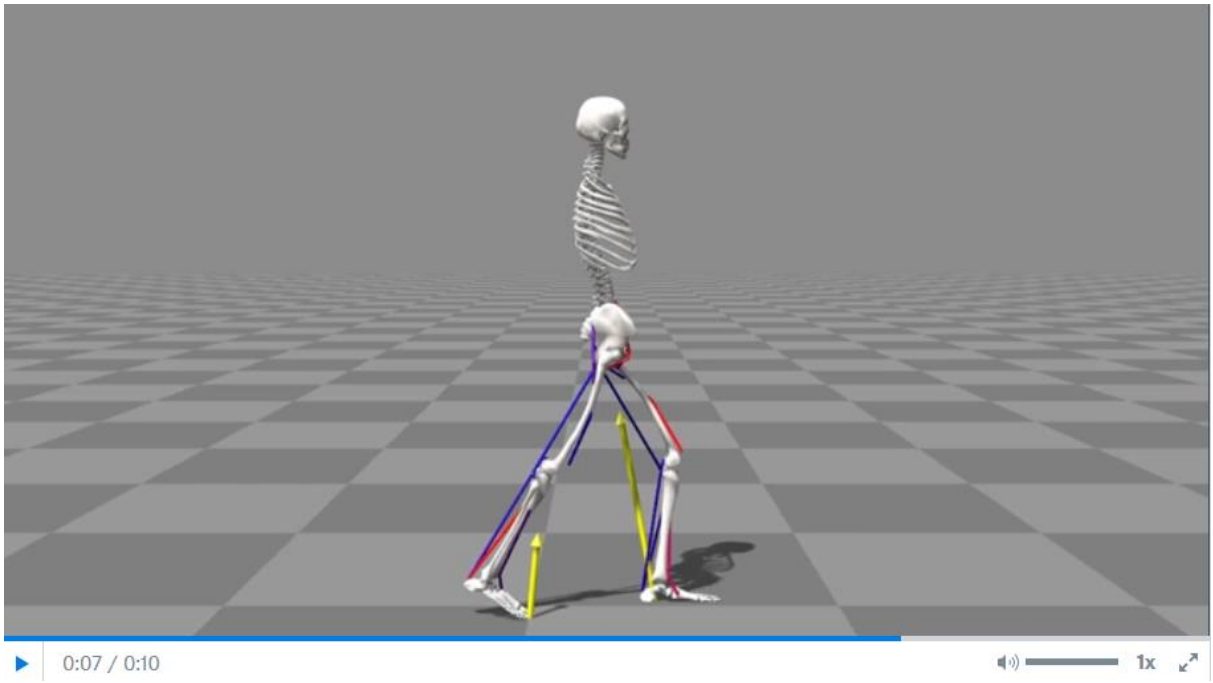


https://www.dropbox.com/s/llo0j1mjp8q2pqe/weak_vas.avi?dl=0

Level of realism : 3,0 (/5) > fair

Comments summary: [*divers reactions from the panel*] Somewhat more hyperextension in stance expected; bouncing seems unnatural, too much forefoot contact in later stance .

6.9.4 Result of the forwards dynamical simulation : General muscle weakness

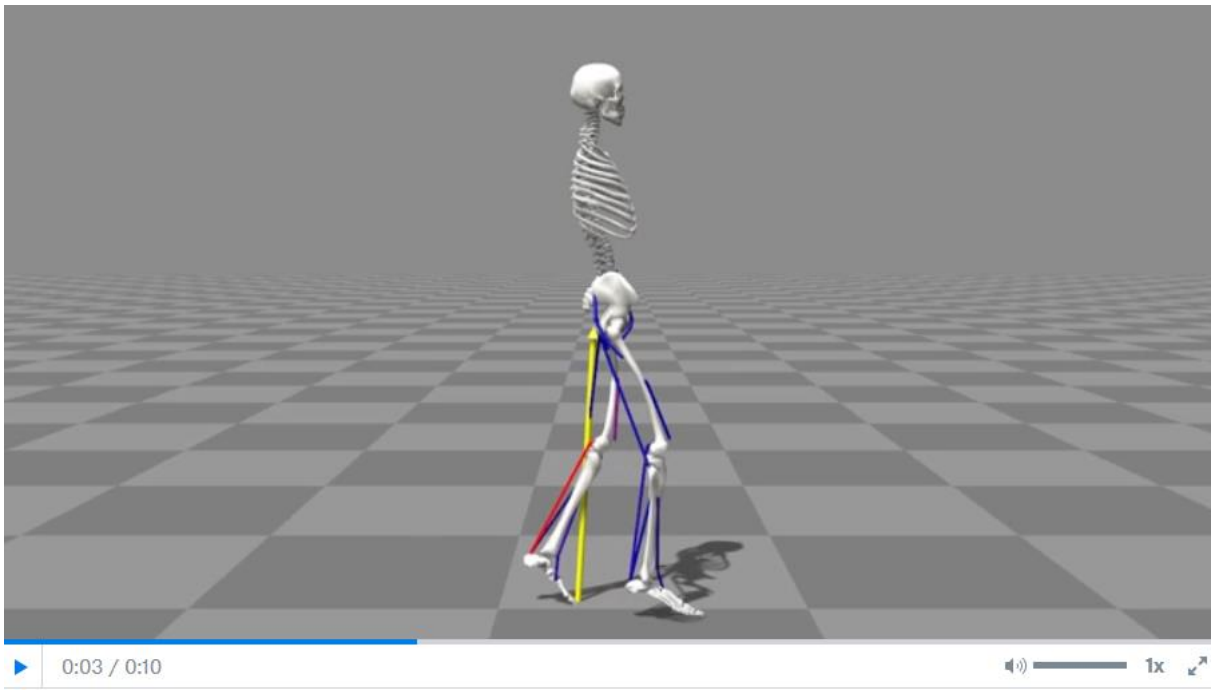


https://www.dropbox.com/s/vushpe9y9yv30l9/weak_all.avi?dl=0

Level of realism : 3,2 (/5) > fair

Comments summary: [*sparse comments from the panel*] more foot drop in sway was expected; net knee flexion moment (from force vector visualization) appears unrealistically high

6.9.5 Result of the forwards dynamical simulation : Spastic Gastrocnemius



https://www.dropbox.com/s/78kixmbsp6orudd/spastic_gas.avi?dl=0

Level of realism : 4,4 (/5) > very realistic

Comments summary: (*all (very) positive*) e.g. “the best of all five!”

6.9.6 Forwards dynamical simulation: Discussion & Conclusion

Overall the gait patterns generated by forward dynamics simulation were rated reasonably realistic. There is still room for improvement , but these experts in the field mentioned also their enthusiasm on the future results on these developments , and assessed the result sof this MD paedigree effort as a major step forward towards implementation in clinical practice .

Part of the critical remarks of the panel might be explained by the fact that no arms were (yet) part of the model , meaning that its dynamics must be compensated for using trunk movements, yielding unrealistic movement for some walking patterns.

It is striking that the two weakness scenarios scored only “fair” , meaning that these should be subjected to a more in depth evaluation, to improve these forwards simulations at follow-up.

The spastic gastroc simulation was assessed a the best forward dynamics simulation, meaning that this type of forward modeling is very promising especially in the case of neurological diseases like Cerebral Palsy, which is in fact the most prevalent chronic pediatric disease

In conclusion : these forwards dynamic simulations in gait were the first ever to be performed for clinical application scenarios. For typical cerebral palsy related applications the simulations were quite to very realistic, while for muscle weakness (DMD and CMT1) the results were fair. Given the potential for further improvements of forwards dynamics simulations (incorporating 3D, arms, advanced optimization criteria) the unpreceded forward modeling efforts in the NND area of MD Paedigree have demonstrated an impressive level of clinical validity .

6.10 Considerations and conclusions

- The newly developed HBM model (WP11) has the potential to be used for regular as well as more advanced, interactive (real-time) clinical gait analysis.
- Inverse dynamics musculoskeletal modelling has the potential to be used instead or as an extension to EMG during clinical gait analysis
- For patient specific modelling of muscle forces and activation, segmented MRI should be augmented by measurements of tendon slack length and optimal fiber length
- The muscle moment arms derived from segmented MRI require additional validation
- Predictive forward dynamics simulation is a promising approach that can circumvent many of the obstacles found in inverse analysis.

7.0 REFERENCES

1. J.B. Bell, P. Colella, H.M. Glaz: A second-order projection method for the incompressible navier-stokes equations. Journal of Computational Physics, 85(2):257-283, 1989.

2. R.I. Ionasec, I. Voigt, B. Georgescu, Y. Wang, H. Houle, F. Vega-Higuera, N. Navab, D. Comaniciu: Patient-Specific Modeling and Quantification of the Aortic and Mitral Valves From 4-D Cardiac CT and TEE. IEEE Transactions on Medical Imaging, 9(29):1636-1651, 2010.

3. L. Itu, P. Sharma, B. Georgescu, A. Kamen, C. Suciu, D. Comaniciu: Model based non-invasive estimation of PV loop from echocardiography. In: 36th Annual Inter. Conf. of the IEEE Engineering in Medicine & Biology Society - EMBC, pp. 6774-6777, 2014.

4. L.M. Itu, P. Sharma, T. Passerinin, A. Kamen, C. Suciu, D. Comaniciu: A Parameter Estimation Framework for Patient-specific Hemodynamic Computations. Journal of Computational Physics, 281:316-333, 2015.

5. R.C. Kerckhoffs, M.L. Neal, Q. Gu, J.B. Bassingthwaighte, J.H. Omens, A.D. McCulloch: Coupling of a 3D finite element model of cardiac ventricular mechanics to lumped systems models of the systemic and pulmonic circulation. Annals of biomedical engineering, 35(1):1-18, 2007.

6. P.J. Kilner, G.Z. Yang, A.J. Wilkes, R.H. Mohiaddin, D.N. Firmin, M.H. Yacoub: Asymmetric redirection of flow through the heart. Nature, 404(6779):759-761, 2000.

7. X. Lu, Y. Wang, B. Georgescu, A. Littman, D. Comaniciu: Automatic Delineation of Left and Right Ventricles in Cardiac MRI Sequences Using a Joint Ventricular Model. In: Proc Functional Imaging and Modeling of the Heart (FIMH), 2011.

8. T. Mansi: Image-Based Physiological and Statistical Models of the Heart, Application to Tetralogy of Fallot. Ph.D. thesis, 2010.

9. V. Mihalef, R. Ionasec, P. Sharma, B. Georgescu, M. Huber, D. Comaniciu: Patient-specific modeling of whole heart anatomy, dynamics and hemodynamics from 4D cardiac CT images. In: Proc VPH Conference, 2010.

10. V. Mihalef: Flow-aware intra-cardiac segmentation. Invention disclosure 2016E04263US, 2016.

11. J.P. Mynard, M.R. Davidson, D.J. Penny, J.J. Smolich: A Simple, Versatile Valve Model for Use in Lumped Parameter and One-Dimensional Cardiovascular Models. International Journal for Numerical Methods in Biomedical Engineering, 28:626–641, 2012.

12. C. Nita, L. Itu, V. Mihalef, P. Sharma, S. Rapaka: GPU-accelerated model for fast, three-dimensional fluid-structure interaction computations. In: Proc IEEE EMBC, 965-968, 2015.

13. M. Sermesant, P. Moireau, O. Camara, J. Sainte-Marie, R. Andriantsimiavona, R. Cimrman, D.L. Hill, D. Chapelle, R. Razavi: Cardiac function estimation from MRI using a heart model and data assimilation: advances and difficulties. Medical Image Analysis, 10(4):642-656, 2006.

14. N. Westerhof, G. Elzinga, P. Sipkema: An artificial arterial system for pumping hearts. Journal of Applied Physiology, 31(5):776, 1971.

15. Y. Zheng, A. Barbu, B. Georgescu, M. Scheuering, D. Comaniciu: Four-Chamber Heart Modeling and Automatic Segmentation for 3D Cardiac CT Volumes using Marginal Space Learning and Steerable Features, IEEE Trans. Medical Imaging, 27(11):1668-1681, 2008.

16. 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.

*17. 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**.*

18. Staemmler M. Service delivery for e-Health applications. Stud Health Technol Inform. 2011;169:537-41.

19. Torchio M, Molino F, Seidenari C, Molino G. Clinical database. Structure, development strategies and expected clinical applications. Minerva Med. 2003Apr;94(2):111-22.

20. Schout BM, Hendrikx AJ, Scheele F, Bemelmans BL, Scherpbier AJ. Validation and implementation of surgical simulators: a critical review of present, past, and future. Surg Endosc. 2010 Mar;24(3):536-46.

21. Ohmann C, Kuchinke W. Future developments of medical informatics from the viewpoint of networked clinical research. Interoperability and integration. Methods Inf Med. 2009;48(1):45-54.

22. Wong MC, Turner P, Yee KC. Involving clinicians in the development of an electronic clinical handover system - thinking systems not just technology. Stud Health Technol Inform. 2008;136:490-5.

23. Angelow A, Schmidt M, Weitmann K, Schwedler S, Vogt H, Havemann C, Hoffmann W; German Transregional Collaborative Research Centre on Molecular Pathogenesis and Therapy of Inflammatory Dilated Cardiomyopathy. Methods and implementation of a central biosample and data management in a three-centre clinical study. Comput Methods Programs Biomed. 2008 Jul;91(1):82-90.

24. Park J, MacRae H, Musselman LJ, Rossos P, Hamstra SJ, Wolman S, Reznick RK. Randomized controlled trial of virtual reality simulator training: transfer to live patients. Am J Surg. 2007 Aug;194(2):205-11.

25. 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.

*26. 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**.*

27. Coble JM, Maffitt JS, Orland MJ, Kahn MG. Contextual inquiry: discovering physicians' true needs. Proc Annu Symp Comput Appl Med Care. 1995:469-73.

28. 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.

29. 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.

30. 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.

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

32. 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.

33. 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.

34. 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.