Grant Agreement no. 777167

BOUNCE

Predicting Effective Adaptation to Breast Cancer to Help Women to BOUNCE Back

Research and Innovation Action
SC1-PM-17-2017: *Personalised computer models and in-silico systems for well-being*

Deliverable: D4.4a Accuracy Evaluation, Finalization and Quality Assurance of the BOUNCE In Silico Resilience Trajectory Predictor (Stage 1)

Due date of deliverable: (30-08-2021)
Actual submission date: (10-09-2021))

Start date of Project: 01 November 2017 Duration: 54 months (extended duration)

Responsible WP: ICCS

The research leading to these results has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 777167		
Dissemination level		
PU	Public	x
PP	Restricted to other programme participants (including the Commission Service	
RE	Restricted to a group specified by the consortium (including the Commission Services)	
CO	Confidential, only for members of the consortium (excluding the Commission Services)	

0. Document Info

0.1. Author

Author	Company	E-mail
Eleni Kolokotroni	ICCS	ekolok@mail.ntua.gr
Georgios Stamatakos	ICCS	gestam@central.ntua.gr
Georgios Manikis	FORTH	gmanikis@gmail.com
Konstantina Kourou	FORTH	konstadina.kourou@gmail.com
Panagiotis Simos	FORTH	Akis.simos@uoc.gr
Vangelis Karademas	FORTH	karademas@uoc.gr
Eugenia Mylona	FORTH	mylona.eugenia@gmail.com
Poikonen-Saksela Paula	HUS	paula.poikonen-saksela@hus.fi
Ira Haavisto	NHG	ira.haavisto@nhg.fi

0.2. Documents history

Document version #	Date	Change
V0.1	01/07/2021	Starting version, template
V0.2	10/07/2021	Definition of ToC
V0.3	07/09/2021	First complete draft
V0.4	09/09/2021	Integrated version (send to WP members)
V0.5	09/09/2021	Updated version (send PCP)
V0.6	09/09/2021	Updated version (send to project internal reviewers)
Sign off	10/09/2021	Signed off version (for approval to PMT members)
V1.0	10/092021	Approved Version to be submitted to EU

0.3. Document data

Keywords	breast cancer, resilience, predictive model, accuracy, evaluation, quality assurance
Editor Address data	Name: Georgios S. Stamatakos Partner: ICCS Address: Iroon Polytechniou 9, Zografos, ICCS, National Technical University of Athens Phone: +30 210 772 2287 Fax: - E-mail: gestam@central.ntua.gr
Delivery date	10/09/2021

1. Table of Contents

2. Introduction

This document provides an initial outline of the accuracy evaluation and quality assurance aspects of the BOUNCE In Silico Resilience Trajectory Predictor. It refers to the implementation status of the BOUNCE project at month 45 with the inclusion of data from the respective clinical pilot study from up to month 12. Therefore, this is an intermediate stage of the development of the BOUNCE predictive models. A full account of accuracy evaluation, finalization and quality assurance for the completed BOUNCE In Silico Resilience Trajectory Predictor is to be presented in deliverable D4.4b. The latter, which will include data from up to month 18, will be submitted at month 51.

The structure of the document is as follows. First the accuracy of the BOUNCE in silico resilience trajectory predictor is evaluated for models developed separately by FORTH and ICCS. It is noted that the models developed by the two organizations (presented in deliverable D4.3a) differ in the precise clinical questions addressed. Indicative approaches and results refer to the FORTH supervised models predicting mental health and quality of life at month 12 (M12) (section 3.1.1). Generalizability of the FORTH risk prediction models across sites is also addressed. An assessment of the concordance on the feature importance across sites through global interpretation analyses is also presented (section 3.1.2). Subsequently, a validation analysis

including external subset validation of indicative ICCS models is outlined. Several performance measures are provided including receiver operating characteristic (ROC) curves. The importance of the variables considered by the ICCS models is also visualized (section 3.2.1). Regarding quality assurance, at this early stage only a brief outline of the adopted approach is provided in chapter 5. Quantitative data are to be provided in deliverable D4.4b. Chapter 6 outlines an intervention pilot sub-study that has been designed (and is in progress) with the aim of testing the resilience predictor tools. Quantitative results stemming from the use of this study will be presented in D4.4b. Other planned studies are briefly outlined in chapter 7. Certain overall conclusions are provided in chapter 8.

3. Accuracy Evaluation of the BOUNCE In Silico Resilience Trajectory Predictor

3.1 FORTH models

The expected performance of the developed prediction models was evaluated on observed input – output data using statistical measures estimating the accuracy of forecast methods (e.g. root mean squared error, sensitivity, specificity) and approaches (e.g. sensitivity analysis).

3.1.1 Supervised Models predicting M12 Mental Health and QoL: Generalizability of risk prediction models across clinical sites (FORTH)

We assessed the generalizability of the supervised learning algorithms and associated pipeline across the four clinical sites. In this framework, for each one of the four models, outlined in detail in D4.3a, we considered each cohort separately as the testing/validation set of the training model. The training phase was implemented each time on three out of four datasets keeping the last cohort for testing purposes regarding the different combinations of the four available cohorts. This approach therefore validates the proposed ML-based pipeline when new unseen patient data are considered for predicting M12 outcomes given the patient characteristics at baseline (M0) and 3-months post diagnosis.

3.1.1.2. Model Description

Four sets of models have been considered and implemented according to the ML-based pipeline towards assessing the predictive power of aggregated M0 and M3 medical, psychological, socio-demographics and lifestyle variables for the overall mental health status and global QoL registered at 12 months post diagnosis and during recovery.

Model 1 entails prediction of overall mental health status at M12 based on all available M0 and M3 waves, with the exception of HADS Anxiety and Depression scores as well as QoL indices. This model sought to identify potentially modifiable behavioral and psychological characteristics that could inform clinicians regarding the needs of individual patients at the early phases of illness in order to maximize psychological resilience at later stages of recovery.
Model 2 entails prediction of M12 mental health risk prediction by including additionally data from the HADS and global QLQ30 questionnaires at M0 and M3. In view of the strong association

between M0/M3 and M12 scores on the same self-report scales, this model explored the optimal prediction capacity of the measurements used in the present study.
Model 3 involves prediction of global QoL at M12 based on all available M0 and M3 waves, with the exception of QoL indices, whereas **Model 4** involves prediction of Global QoL at M12 using all available variables at M0 and M3.

Tree-based classifiers were applied within the ensemble methodology, such as Random Forest (RF), Decision Trees (DTs) and Gradient Boosting Machines (GBM) estimators on a total of 532 (Models 1-2) and 528 patients with sufficient M12 data (Models 3-4).

3.1.1.3. Model performance metrics

To evaluate the performance of the classification models based on the ensemble-based methodology, 7 measures were calculated: accuracy, balanced accuracy (appropriate for balanced datasets), F1 score, precision, recall (sensitivity), specificity, and AUC. Additionally, the concordance of the external validation schemes on the most important predictors was assessed.

3.1.1.4. Results

3.1.1.4.1. External validation set: IEO

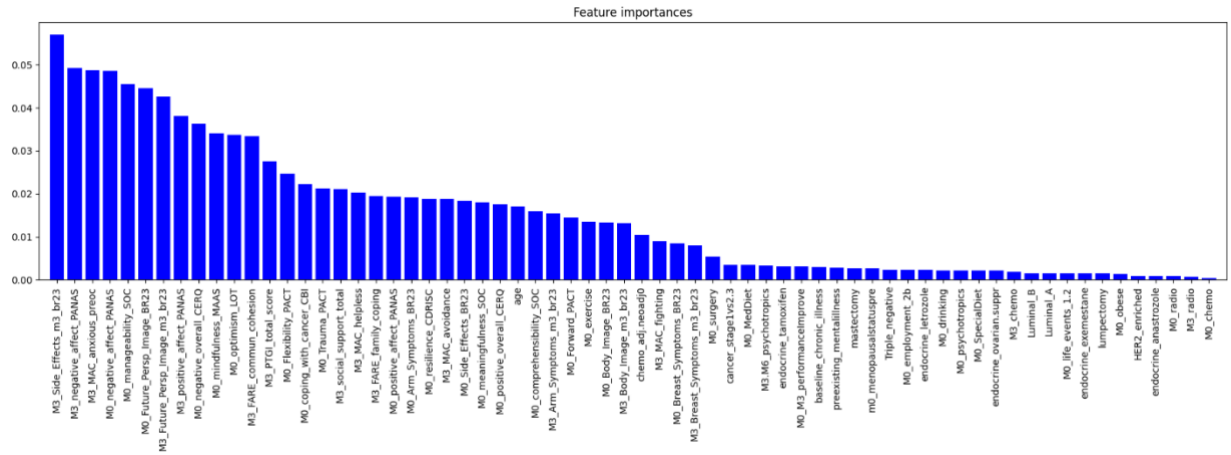
In Table 1 the results of the mental health and overall QoL risk prediction models are given for the IEO cohort that was considered for validation purposes according to the different modelling cases (i.e., inclusion or not of anxiety and depression symptoms and QoL predictors at M0 and M3).

Table 1. Performance of the ensemble methodology using the IEO data set for external validation.

	Model 1	Model 2	Model 3	Model 4
Accuracy	0.701	0.711	0.652	0.642
Balanced accuracy	0.694	0.766	0.678	0.719
AUC	0.800	0.850	0.765	0.793
Sensitivity	0.680	0.880	0.727	0.863
Specificity	0.708	0.652	0.630	0.575
F1	0.539	0.611	0.492	0.527

Note. Cross validation using grid search optimization against Area Under the ROC Curve (AUC) was applied following a 3-fold data division scheme for better and more generalizable results.

Feature importances in descending order in the validation scheme involving training on data from HUS, CHAMP, and HUJI and testing on data from IEO for Model 1 as example is presented in the figure below. Variable ranking denotes the significance of each individual predictor computed as the mean and standard deviation of accumulation of the impurity decrease within each tree.



3.1.1.4.2. External validation set: HUS

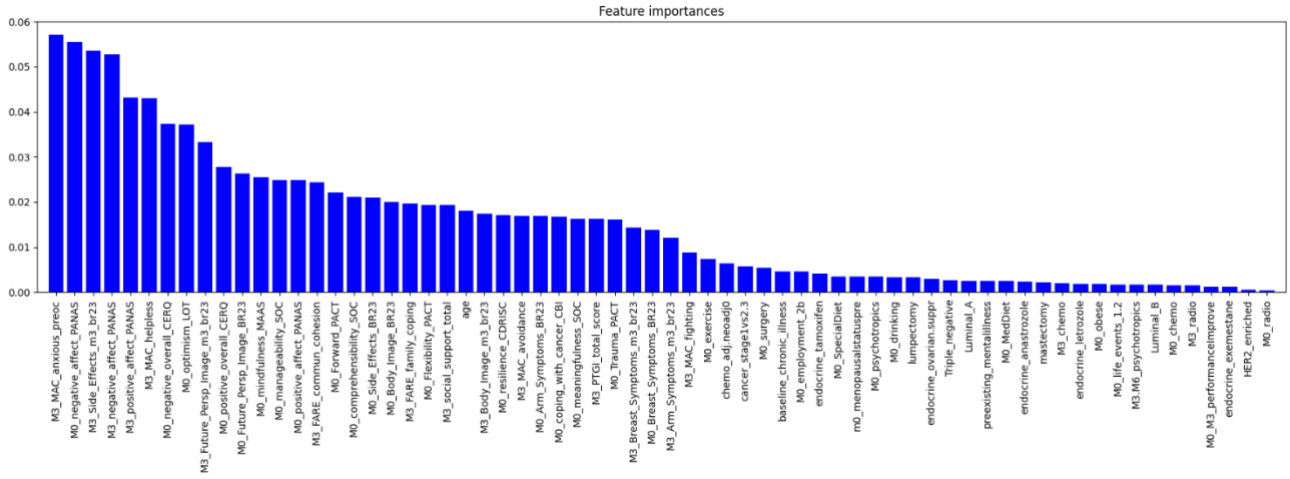
In Table 2 the results of the mental health and QoL risk prediction models are given when the HUS cohort was considered for validation purposes according to the different approaches (i.e., inclusion or not of anxiety and depression symptoms and QoL predictors at M0 and M3).

Table 2. Performance of the ensemble methodology using the HUS data set for external validation.

	Model 1	Model 2	Model 3	Model 4
Accuracy	0.783	0.788	0.599	0.633
Balanced accuracy	0.806	0.809	0.604	0.585
AUC	0.845	0.916	0.667	0.667
Sensitivity	0.833	0.833	0.612	0.5162
Specificity	0.780	0.785	0.596	0.654
F1	0.312	0.317	0.319	0.301

Note. Cross validation using grid search optimization against Area Under the ROC Curve (AUC) was applied following a 3-fold data division scheme for better and more generalizable results.

The figure below ranks predictor variables entered into Model 1 by the measure of impurity. The impurity decrease for each feature was averaged across trees of the RF classifier to determine the final importance of each individual variable.



3.1.1.4.3. External validation set: HUJI

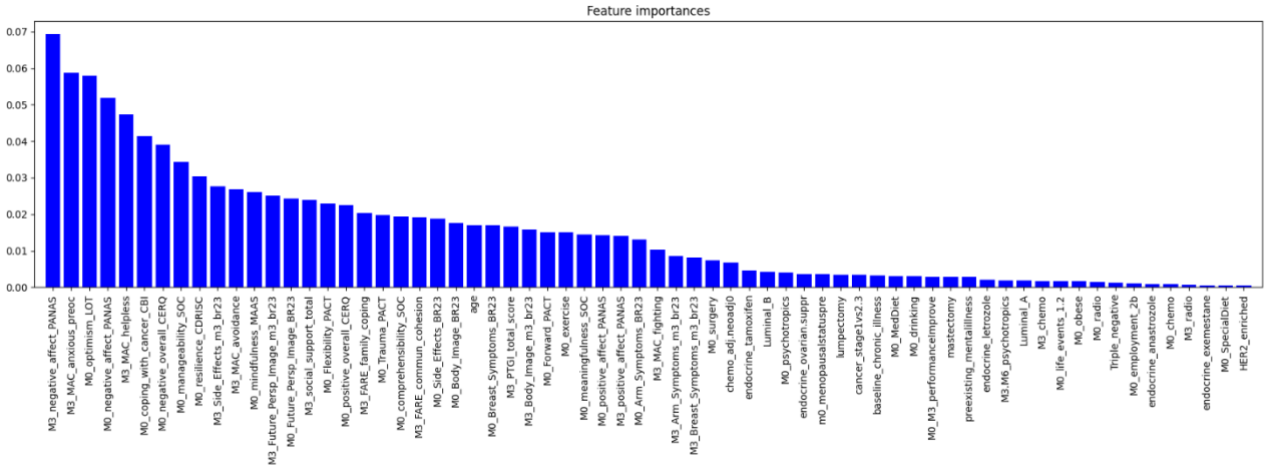
In Table 3 the results of the mental health and QoL risk prediction models are given when the HUJI cohort was considered for validation purposes according to the different approaches (i.e., inclusion or not of anxiety and depression symptoms and QoL predictors at M0 and M3).

Table 3. Performance of the ensemble methodology using the HUJI data set for external validation.

	Model 1	**Model 2**	**Model 3**	**Model** 4
Accuracy	0.798	0.815	0.661	0.635
Balanced accuracy	0.699	0.769	0.616	0.641
AUC	0.796	0.847	0.727	0.727
Sensitivity	0.550	0.700	0.550	0.650
Specificity	0.848	0.838	0.683	0.632
F1	0.478	0.560	0.354	0.376

Note. Cross validation using grid search optimization against Area Under the ROC Curve (AUC) was applied following a 3-fold data division scheme for better and more generalizable results.

The figure below ranks predictor variables entered into Model 1 according to the computed impurity measure.



3.1.1.4.4. External validation set: CHAMP

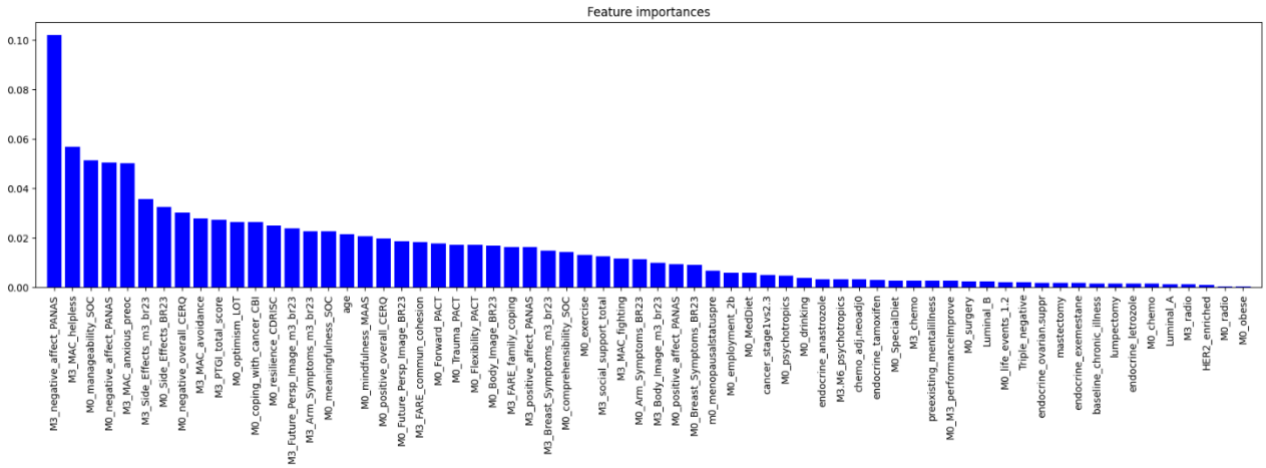
In Table 4 the results of the mental health and QoL risk prediction models are given when the CHAMP cohort was considered for validation purposes according to the different approaches.

Table 4. Performance of the ensemble methodology using the CHAMP data set for external validation.

	Model 1	Model 2	Model 3	Model 4
Accuracy	0.761	0.787	0.640	0.663
Balanced accuracy	0.801	0.834	0.739	0.687
AUC	0.879	0.913	0.807	0.783
Sensitivity	0.869	0.913	0.923	0.730
Specificity	0.733	0.755	0.556	0.643
F1	0.597	0.636	0.539	0.500

Note. Cross validation using grid search optimization against Area Under the ROC Curve (AUC) was applied following a 3-fold data division scheme for better and more generalizable results.

The figure below ranks predictor variables entered into Model 1 by feature importances computed based on the mean decrease of the impurity metric.



3.1.1.4.5. Concordance of predictors across sites

Table 5 lists the predictors of M12 overall Mental Health (Models 1-2) and M12 global QoL that were common across four cross-validation schemes (each tested with data from a single site that was not used in the training phase). The number of common features ranged from 6/10 (Model 3) to 8/10 (Models 2 and 4). The variables which emerged as significant predictors of M12 mental health and quality of life correspond to the same “clusters” of factors identified in similar analyses conducted on the entire patient cohort. These include (i) sense of coherence and self-efficacy to cope with cancer, (ii) emotion regulation strategies and habitual modes of thinking (e.g., anxious preoccupation), (iii) the experience of side effects, and (iv) previous levels of well-being and affect at and soon after diagnosis. As in the case of the supervised models conducted on the entire patient cohort (see D4.3a), besides side-effects (as perceived by the patient), the medical factors did not emerge as significant predictors of the M12 overall mental health and global quality of life.

Similarities and differences across sites in the relative importance of specific predictor variables was further explored within a global interpretation framework outlined in more detail in D4.3a (section 4.1) and presented in the subsequent Section (3.1.2).

Table 5. Common selected features for M12 mental health (Models 1-2) and QoL prediction (Models 3-4) in external cross-validation schemes.

MODEL 1	**MODEL 2**
M0_manageability_SOC	M0_manageability_SOC
M0_negative_overall_CERQ	M0_negative_affect_PANAS
M0_negative_affect_PANAS	M3_MAC_anxious_preoc
M3_MAC_helpless	M3_negative_affect_PANAS
M3_MAC_anxious_preoc	M0_anxiety_HADS
M3_negative_affect_PANAS	M0_depression_HADS
M3_Side_Effects_m3_br23	M3_anxiety_HADS
	M3_depression_HADS

MODEL 3	**MODEL 4**
M0_manageability_SOC	M0_manageability_SOC
M0_coping_with_cancer_CBI	M0_anxiety_HADS
M3_positive_affect_PANAS	M3_anxiety_HADS
M3_anxiety_HADS	M3_depression_HADS
M3_depression_HADS	M3_Side_Effects_m3_br23
M3_Side_Effects_m3_br23	M0_Global_QLQ30
	M3_Global_QLQ30
	M0_coping_with_cancer_CBI

3.1.2. Assessing concordance on feature importance across sites: Global interpretation analyses

Generalizability of classification results across clinical sites can be explored in more detail through global and local interpretation analyses. Here we demonstrate this concordance on results from Models 1 and 3 and illustrated in Figures 1-8.

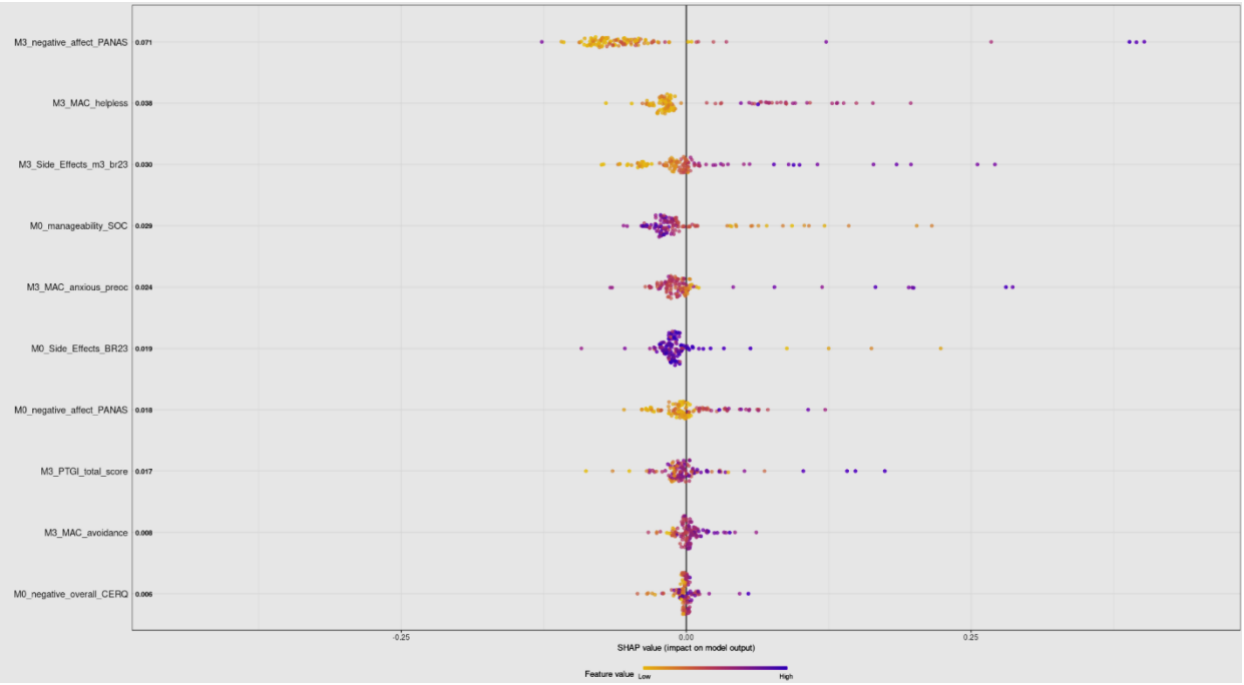


Figure 1 Shapley additive explanations (SHAP) summary plot generated using data from CHAMP in the testing phase of Model 1 (predicting overall Mental Health at M12 from M0 and M3 variables excluding QoL and mental health indices).

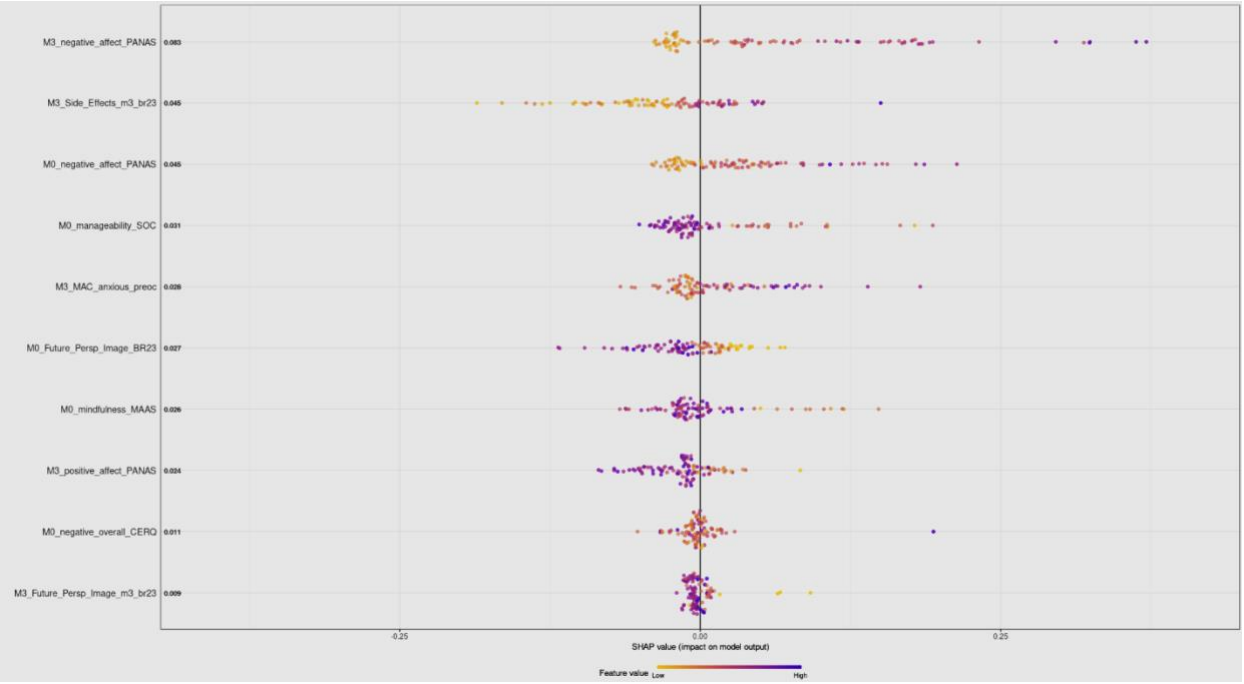


Figure 2 Shapley additive explanations (SHAP) summary plot generated using data from IEO in the testing phase of Model 1 (predicting overall Mental Health at M12 from M0 and M3 variables excluding QoL and mental health indices).

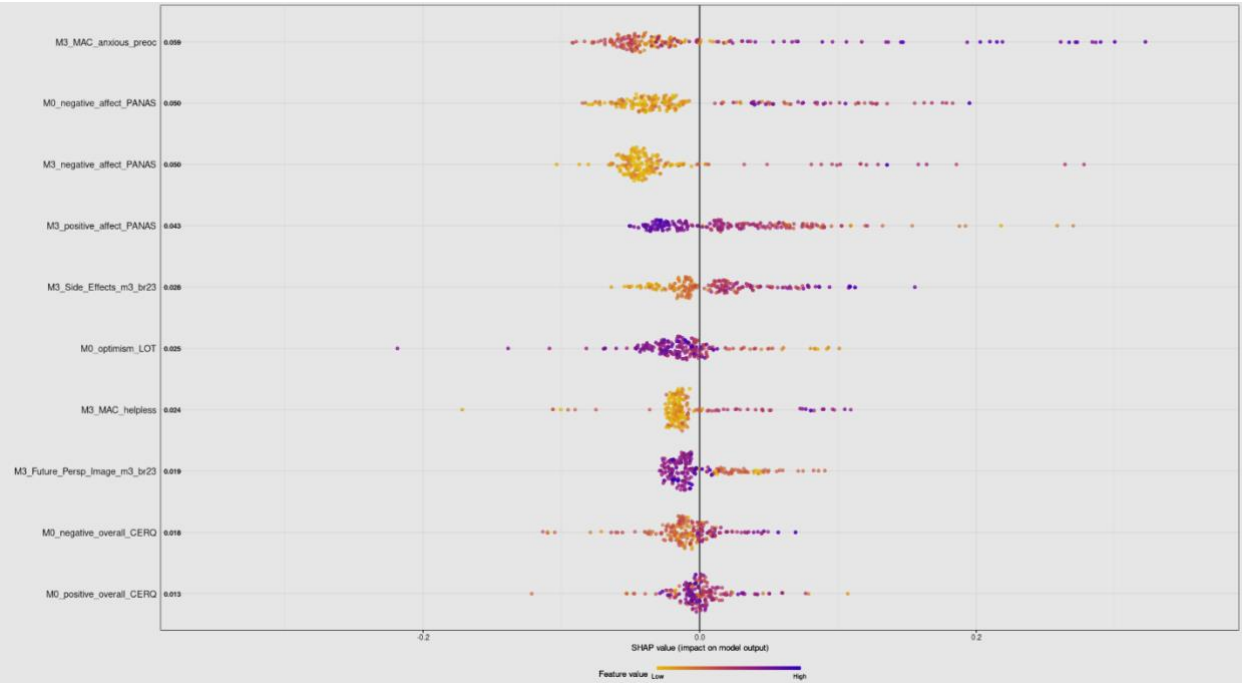


Figure 3 Shapley additive explanations (SHAP) summary plot generated using data from HUS in the testing phase of Model 1 (predicting overall Mental Health at M12 from M0 and M3 variables excluding QoL and mental health indices).

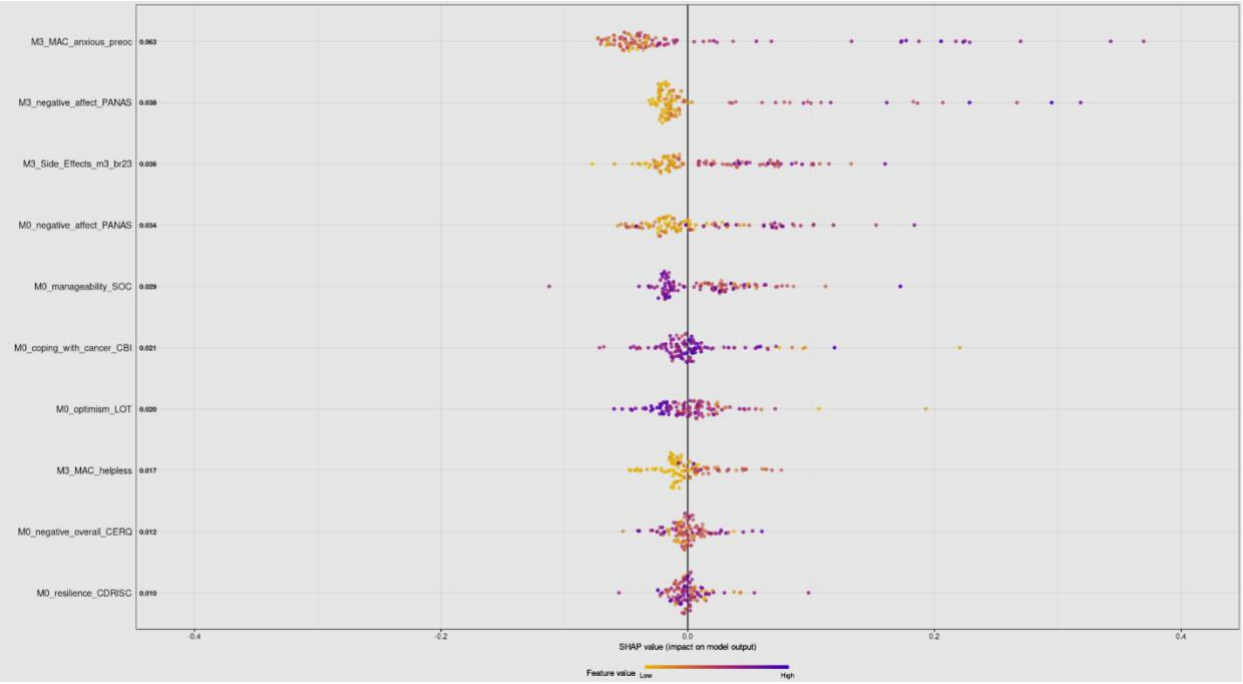


Figure 4 Shapley additive explanations (SHAP) summary plot generated using data from HUJI in the testing phase of Model 1 (predicting overall Mental Health at M12 from M0 and M3 variables excluding QoL and mental health indices).

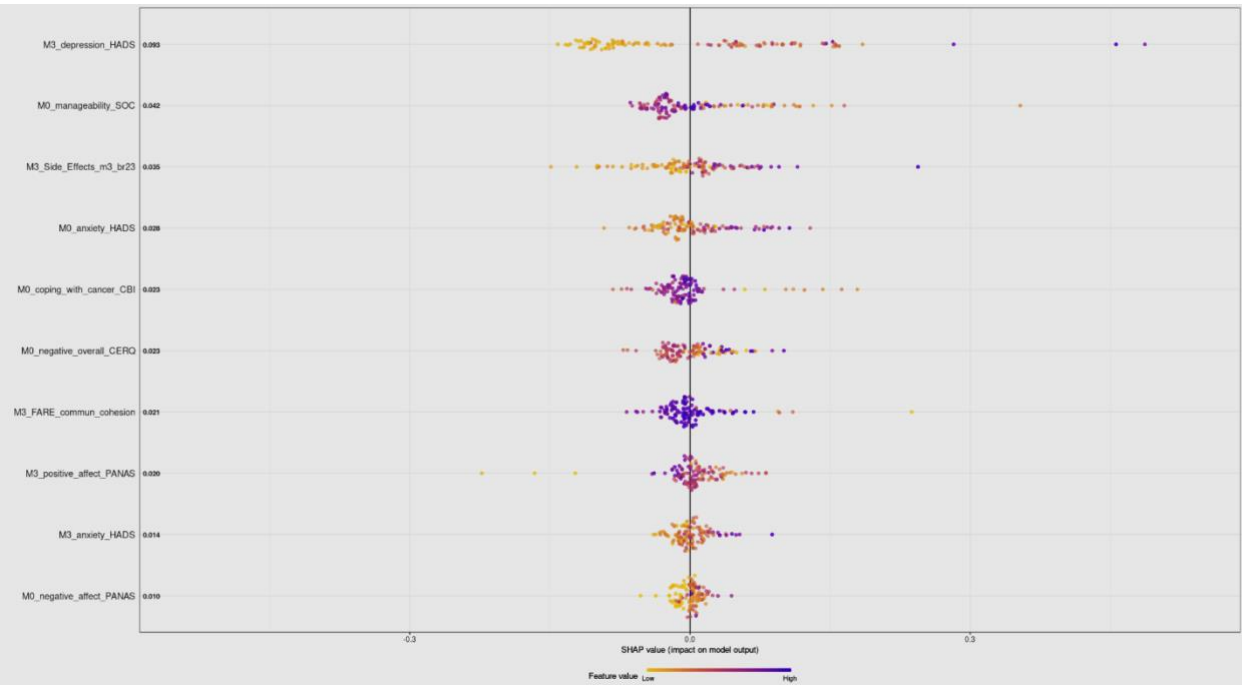


Figure 5 Shapley additive explanations (SHAP) summary plot generated using CHAMP dataset to externally validate Model 3 (predicting global QoL at M12 from M0 and M3 variables excluding QoL indices).

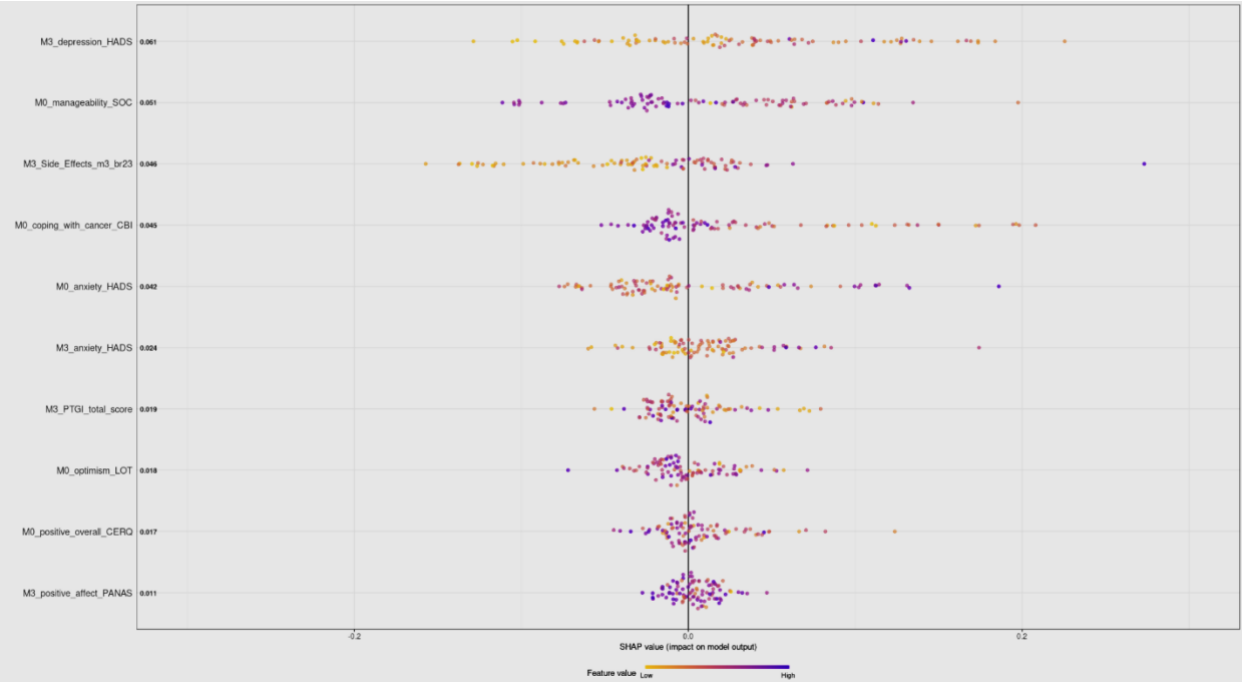


Figure 6 Shapley additive explanations (SHAP) summary plot generated using IEO dataset to externally validate Model 3 (predicting global QoL at M12 from M0 and M3 variables excluding QoL indices).

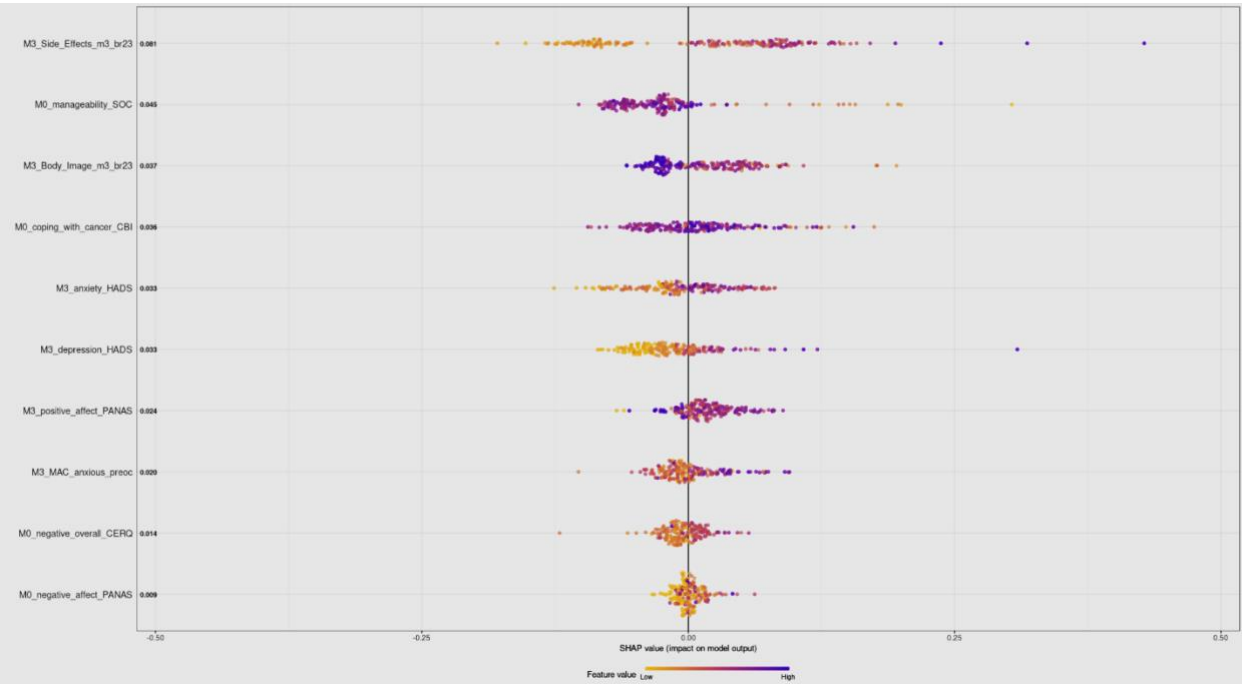


Figure 7 Shapley additive explanations (SHAP) summary plot generated using HUS dataset to externally validate Model 3 (predicting global QoL at M12 from M0 and M3 variables excluding QoL indices).

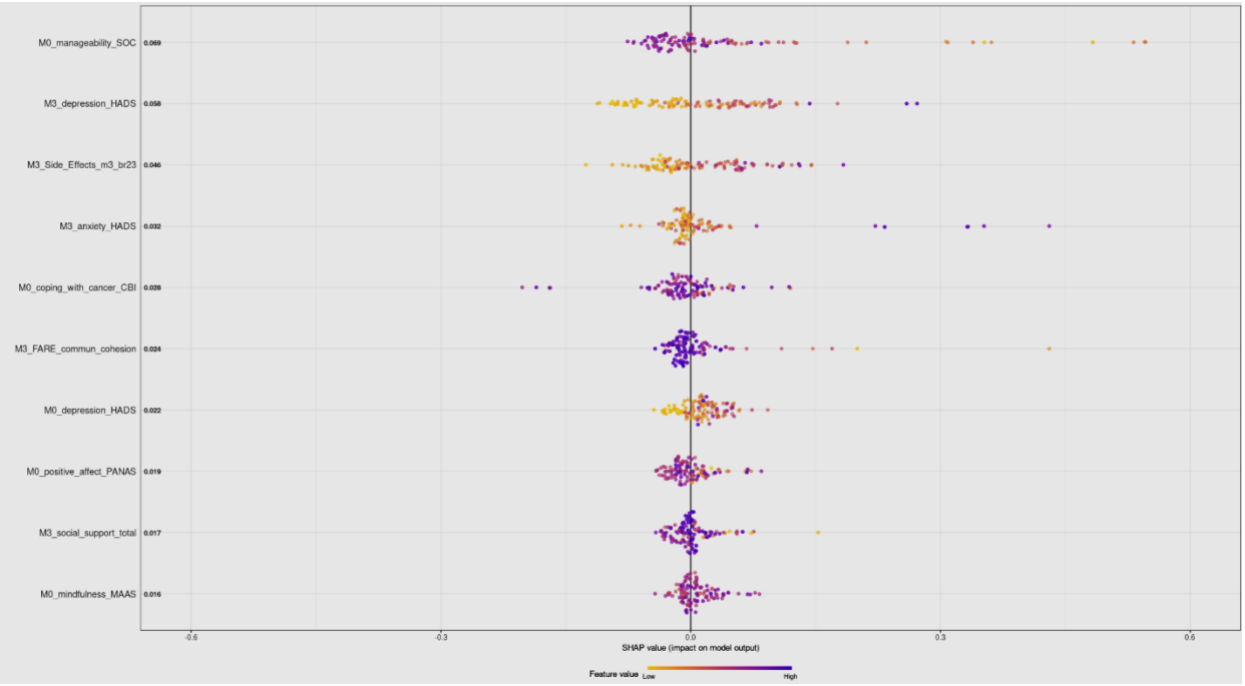


Figure 8 Shapley additive explanations (SHAP) summary plot generated using HUJI dataset to externally validate Model 3 (predicting global QoL at M12 from M0 and M3 variables excluding QoL indices).

To further explore global interpretability for Models 1 and 3, partial dependence plots (PDP) were generated to describe the marginal impact of Negative Affect at M3 on model mental health status (or global Qol, respectively) prediction, holding all other features in the model constant (model response fluctuations as a function of a single parameter keeping all other variables unchanged). As shown Figures 9-12 Negative Affect at M3 has a different impact on the prediction of M12 mental health within each external validation set.

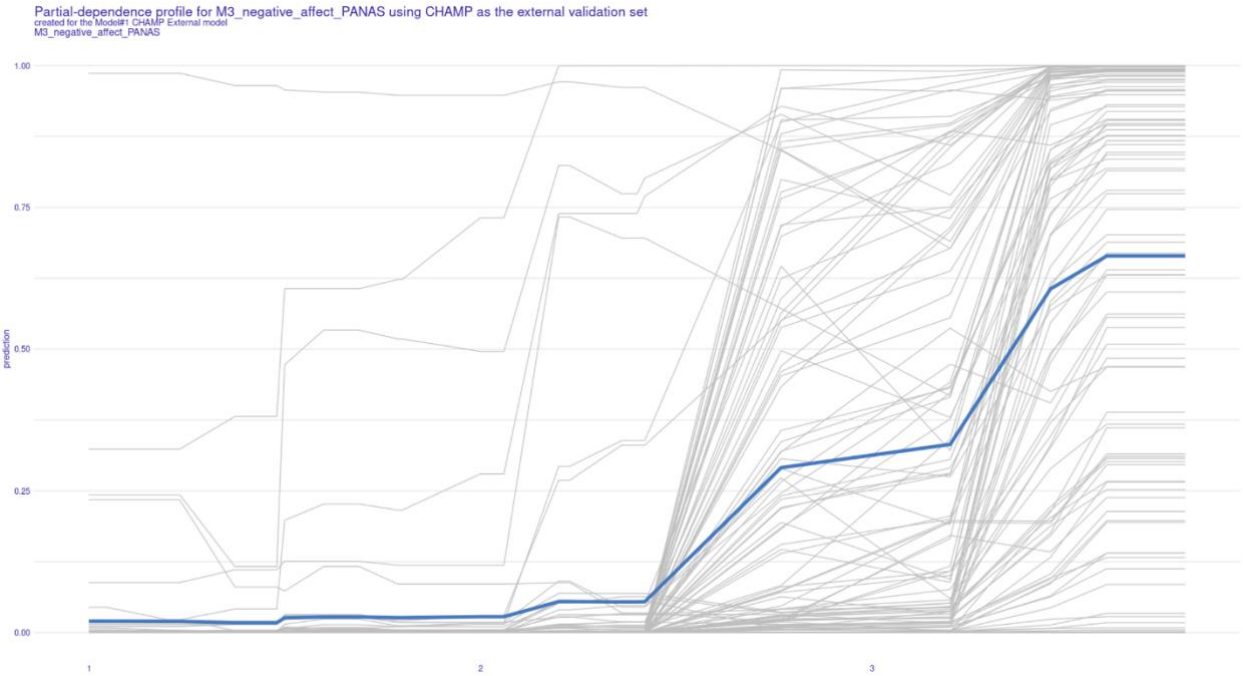


Figure 9 PDP of mental health status prediction by Negative Affect at M3 using CHAMP as the external validation set.

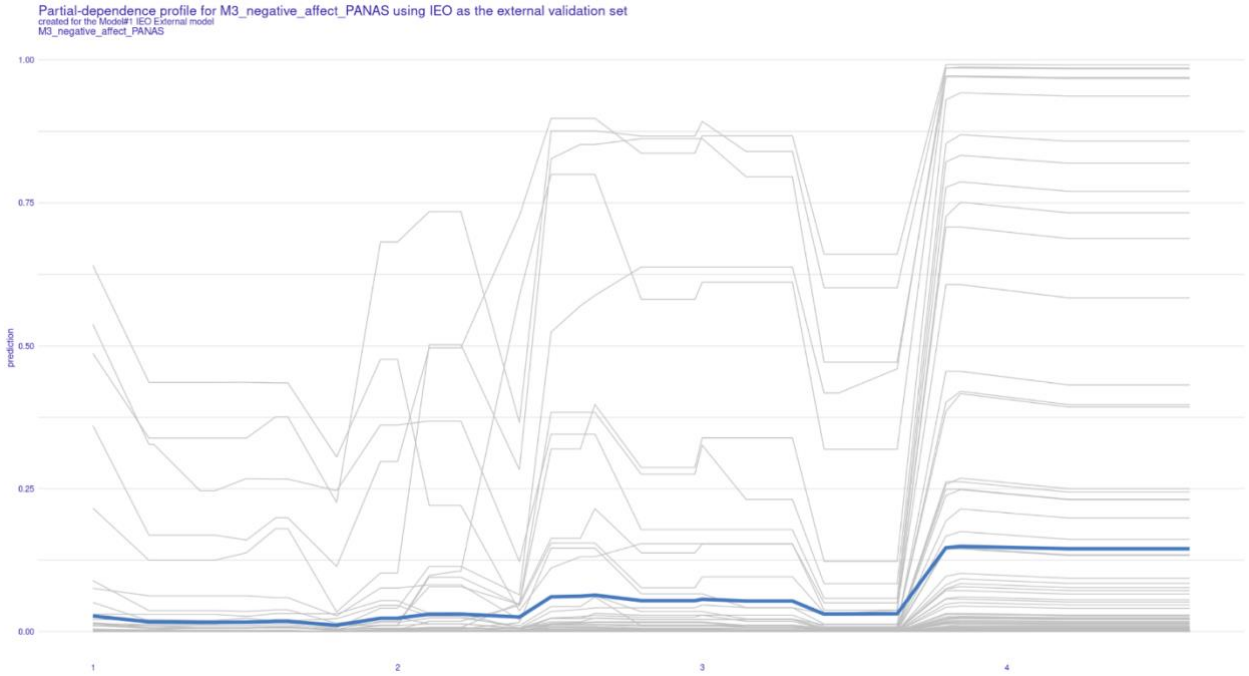


Figure 10 PDP of mental health status prediction by Negative Affect at M3 using IEO as the external validation set.

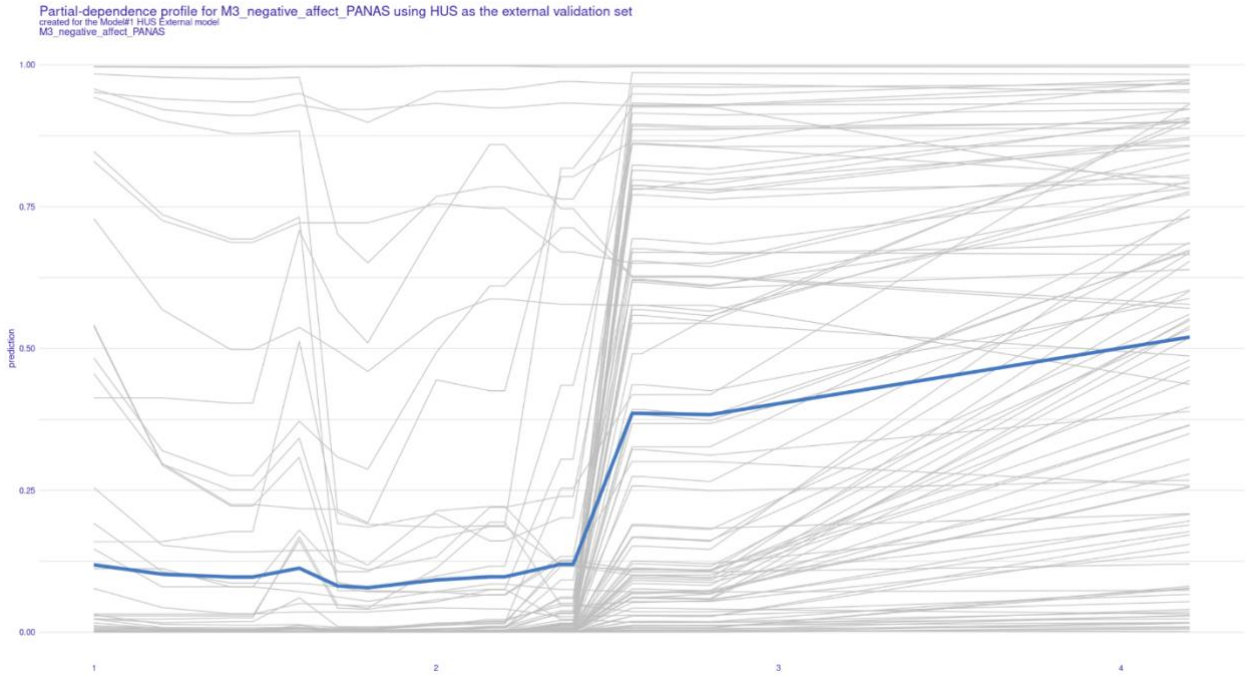


Figure 11 PDP of mental health status prediction by Negative Affect at M3 using HUS as the external validation set.

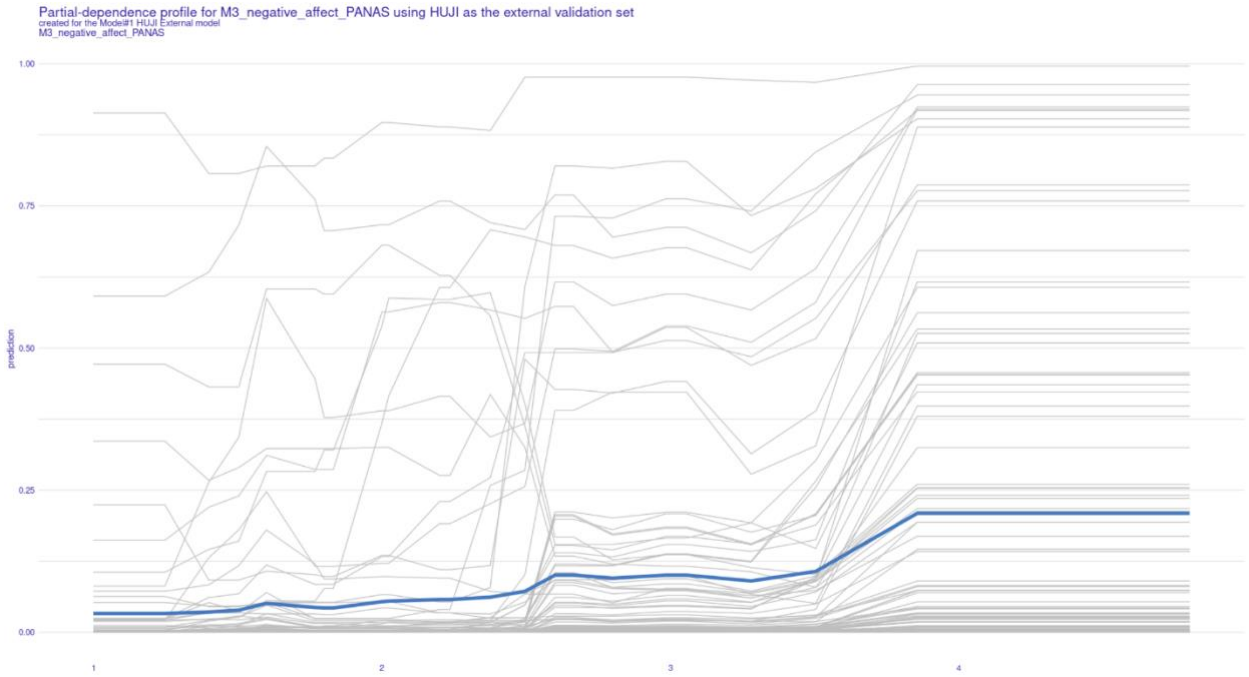


Figure 12 PDP of mental health status prediction by Negative Affect at M3 using HUJI as the external validation set.

Regarding Model 3, depression symptomatology at M3 was among the most important classification features for M12 global QoL. Variability across sites in the prediction value of this variable can be assessed in the conditional expectation plots that were superimposed on the partial dependence plot in Figures 13-16. The profile of dependence of global QoL upon earlier depression symptoms (individual conditional expectation plots) was more variable across participants, resulting in more shallow partial dependence plots at each site. The shape of the ICE plots raises the possibility that although earlier depression symptoms featured among the significant predictors of subsequent QoL, their actual import on the latter may not be critical is all other variables are taken into account at the same time.

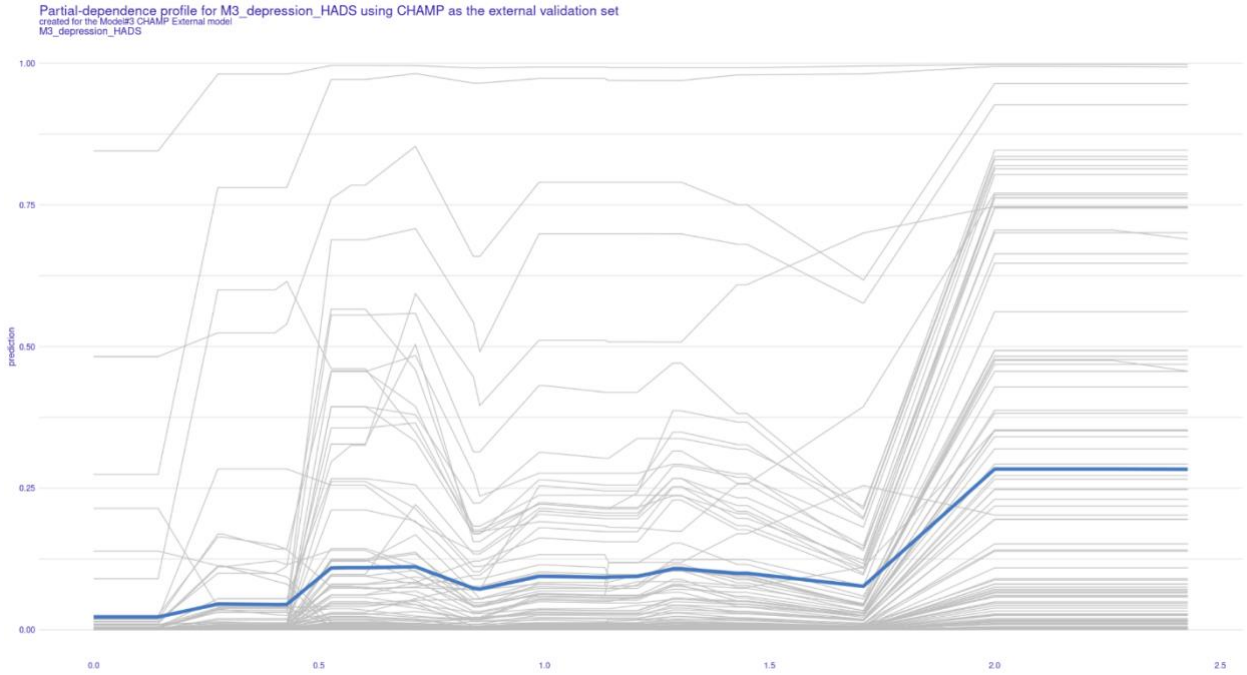


Figure 13 PDP of M12 global QoL prediction by Depression symptoms at M3 using CHAMP as the external validation set. Slight changes in the prediction value are evident when Depression symptoms at M3 values increase.

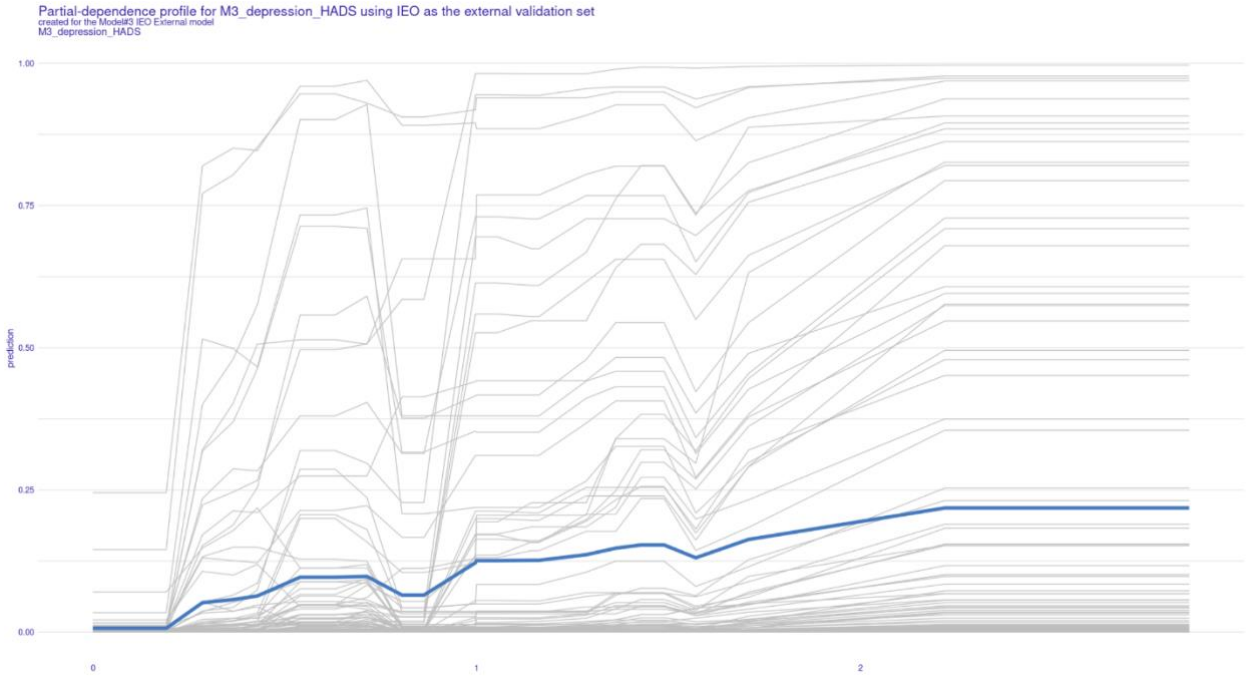


Figure 14 PDP of global QoL prediction by Depression symptoms at M3 using IEO as the external validation set.

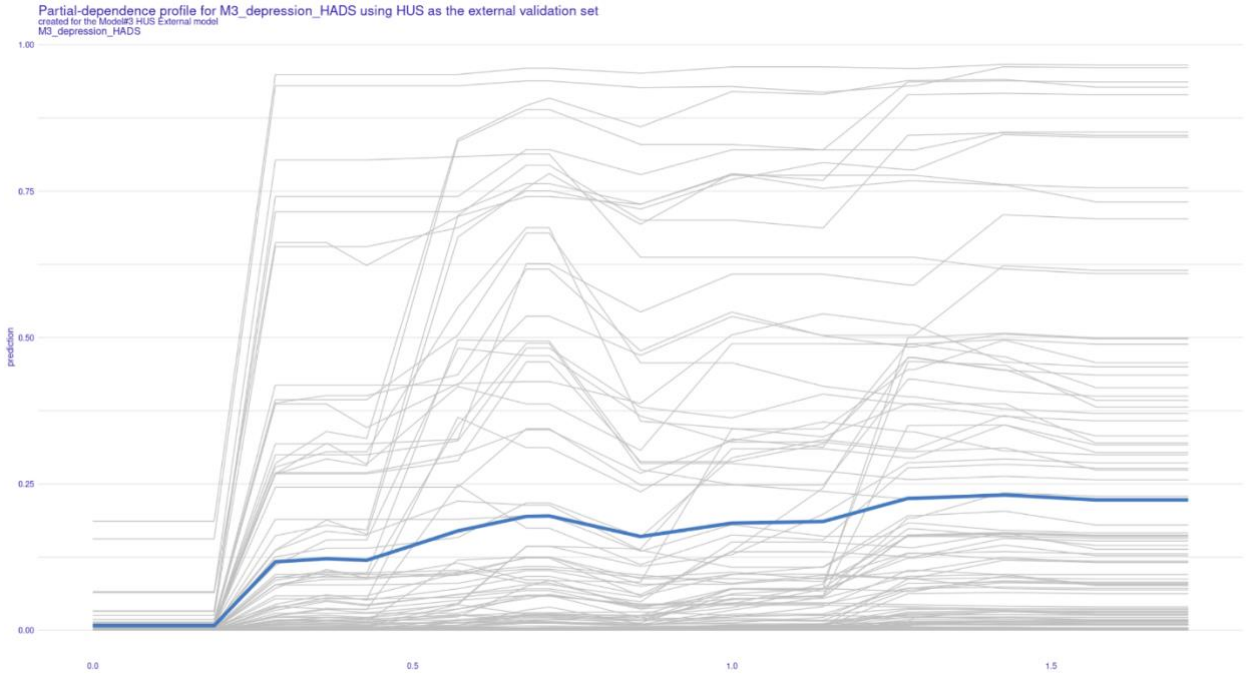


Figure 15 PDP of global QoL prediction by Depression symptoms at M3 using HUS as the external validation set.

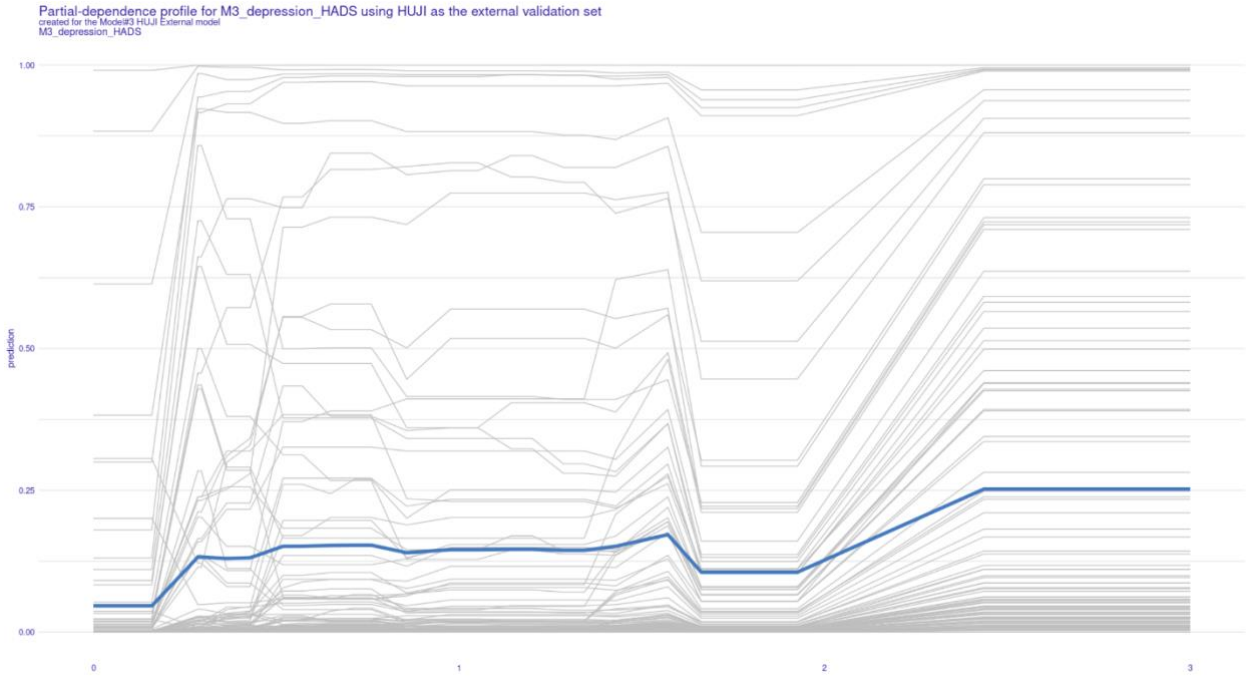


Figure 16 PDP of global QoL prediction by Depression symptoms at M3 using HUJI as the external validation set.

3.2 ICCS Models

3.2.1 Classification model for predicting the depression trajectories (ICCS)

By the time of the preparation of the present document, one model related to 'resilience as a process' is under implementation. The model refers to the cohort of patients with initial low HADS Depression (scores ≤1). Models addressing patients of moderate/high depression as well as other key psychological processes (trajectory of anxiety and/or quality of life) are to be examined.

3.2.1.1 Model Description

The model aims to predict which patients with initial low HADS Depression score (≤1) will have an increasing mean trajectory during the first year from baseline. Two classes are considered (Fig. 17) corresponding to the 'Low' and 'Increasing' trajectory groups identified using latent-class mixed effects model (D4.3a). Predictors considered are all sociodemographic, lifestyle, clinical, treatment and psychological variables at baseline and M3. The model is based on a random forest classifier. Currently all predictors are considered and no feature selection has been applied.

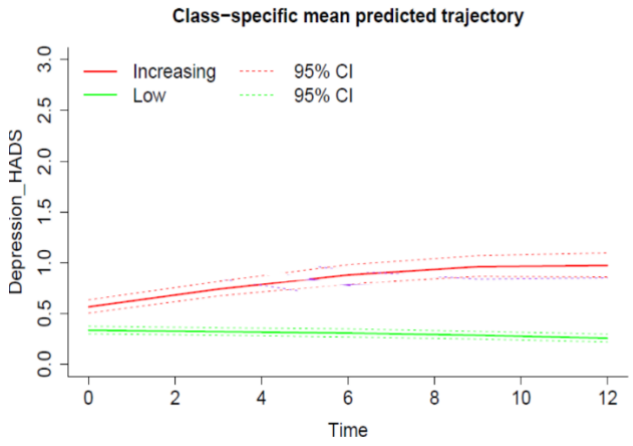


Figure 17: Mean Predicted trajectories over Time (in months) of the two classes considered by the classification model

3.2.1.2 Nested Cross Validation

The classifier performance is evaluated by means of nested cross validation (cv) and external dataset validation (Table 6). In the case of nested cv, the inner loop is responsible for model selection/hyperparameter tuning (validation set), while the outer loop is for error estimation (test set). A stratified four-fold resampling is used for the outer loop. Model parameters are tuned using grid search against AUC metric with 10-fold cross validation. For the results presented here, class imbalance was handled using smote subsampling, performed inside the inner cv resampling. Exclusion of well-being psychological variables (HADS depression and anxiety, C30 Emotional function, distress thermometer) is associated with poorer performance, however, the ability of the model to distinguish between the two classes is good in both cases (mean AUC>0.8) (Figures 18 & 19). Better tuning of the smote subsumbling is required to better balance sensitivity and specificity.

Table 6: Performance of Random Forest Classifier. Average values are given.

	All data	*No HADS, C30 Emotional function, distress thermometer (during training)*
Accuracy	0.83	0.80
Balanced Accuracy	0.80	0.72
Precision	0.69	0.65
Sensitivity	0.71	0.56
Specificity	0.88	0.89
AUC	0.88	0.85
F1	0.70	0.60

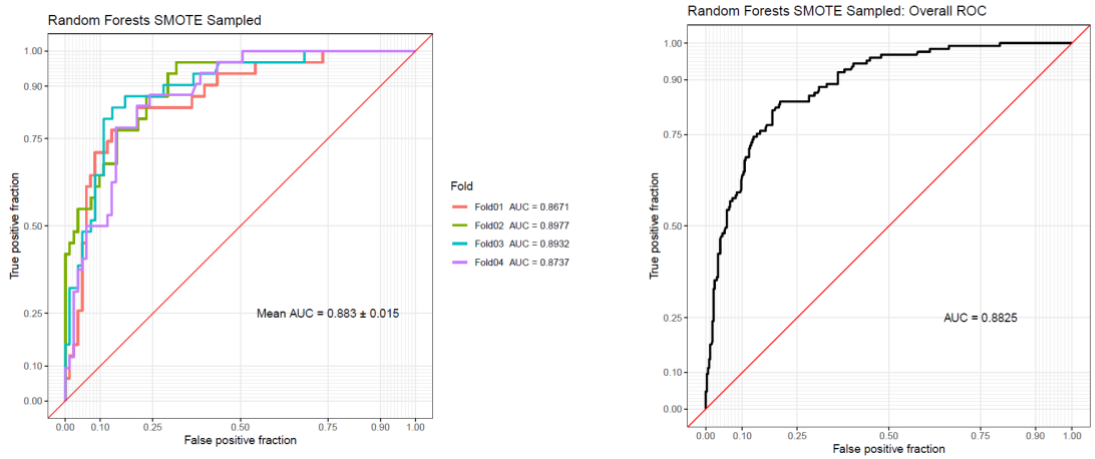


Figure 18: ROC curves for each outer fold of the nested cross validation. Model with all the predictors.

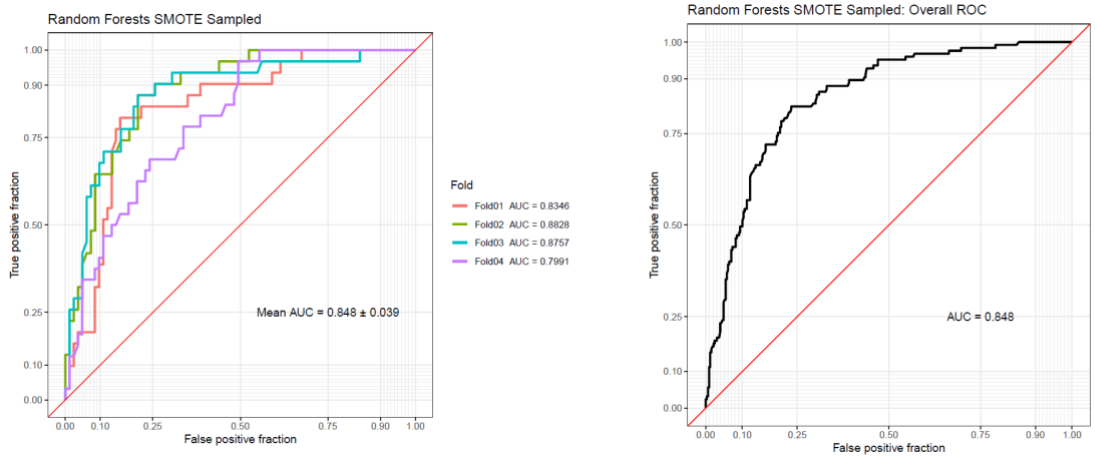


Figure 19: ROC curves for each outer fold of the nested cross validation. Model without HADS scales (depression, anxiety), C30 Emotional Function and distress thermometer.

3.2.1.3 External subset validation

The dataset is split into test and training set based on data origin. Each time, one of the four clinical subsets (i.e. CHAMP, IEO, HUS and HUJI) comprise the test set and the rest three the training set. Such a sampling is not stratified, meaning that the ratio between the target classes is not the same in each fold as it is in the whole dataset. Inner cross validation is as previously

described. The distribution of the trajectory classes between the clinical sites is reported in Table 7. The vast majority of patients in each clinical site is assigned to the 'Low' trajectory class, with the exception of IEO where the distribution between the two classes is more balanced. The distribution of 'Increasing' class between clinical sites is balanced. However, in the 'Low' trajectory class, IEO is under-represented, whereas the majority of patients comes from HUS.

Table 7: Class distribution (in percentage) in each clinical site

	'Low' trajectory class	*'Increasing' trajectory class*	*Both trajectory classes*
CHAMP	18%	7%	**24%**
IEO	6%	8%	**14%**
HUS	35%	6%	**42%**
HUJI	14%	6%	**20%**
Overall	**73%**	**27%**	

3.2.1.3.1 Performance metrics

Tables 8 and 9 report various performance evaluation metrics in the presence of all predictors and when psychological well-being scales are absent from the training set. Deviations between the clinical sites to be investigated. The best overall performance is observed when the test set comprise the CHAMP subset.

Table 8: Performance of Random Forest Classifier. Average values are given.

			All predictors	
Test set	**CHAMP**	**IEO**	**HUS**	**HUJI**
Accuracy	0.84	0.70	0.74	0.74
Balanced Accuracy	0.85	0.73	0.82	0.73
Precision	0.66	0.90	0.36	0.57
Sensitivity	0.87	0.53	0.93	0.72
Specificity	0.83	0.93	0.71	0.74
AUC	0.92	0.78	0.87	0.81
F1	0.75	0.67	0.52	0.64

Table 9: Performance of Random Forest Classifier. Average values are given.
No HADS, C30 Emotional function, distress thermometer (during training)

Test set	**CHAMP**	**IEO**	**HUS**	**HUJI**
Accuracy	0.80	0.55	0.73	0.74
Balanced Accuracy	0.76	0.59	0.73	0.76
Precision	0.59	0.76	0.33	0.55
Sensitivity	0.68	0.25	0.72	0.83
Specificity	0.85	0.93	0.73	0.69
AUC	0.89	0.76	0.81	0.82
F1	0.66	0.38	0.45	0.67

3.2.1.3.2 Variable Importance

Figure 20 depicts the importance of each predictor, in descending order, for each training set. Psychological well-being scales (HADS depression, HADS anxiety, C30 Emotional function, distress thermometer) are not included to better reveal the effect of the rest of the variables. These variables are ranked high in all training sets when included. Variable importance is assessed with a random forest specific metric (multivariate approach).
Differences exist between the order of variable importance, as anticipated.

The variables that are highly ranked in all cases (in the top-20) are: Negative_affect_PANAS.3, Fatigue_QLQ30.3, M3_MAC_helpless, Soc_Fun_QLQ30.3, Global_QLQ30.3, Negative_affect_PANAS.0.

Important variables are also: M0_coping_with_cancer_CBI, Pain_QLQ30.3, M3_MAC_anxious_preoc, M0_optimism_LOT, m3_domestic_help_days, Phys_Fun_QLQ30.3 Insomnia_QLQ30.3, bipq2.3

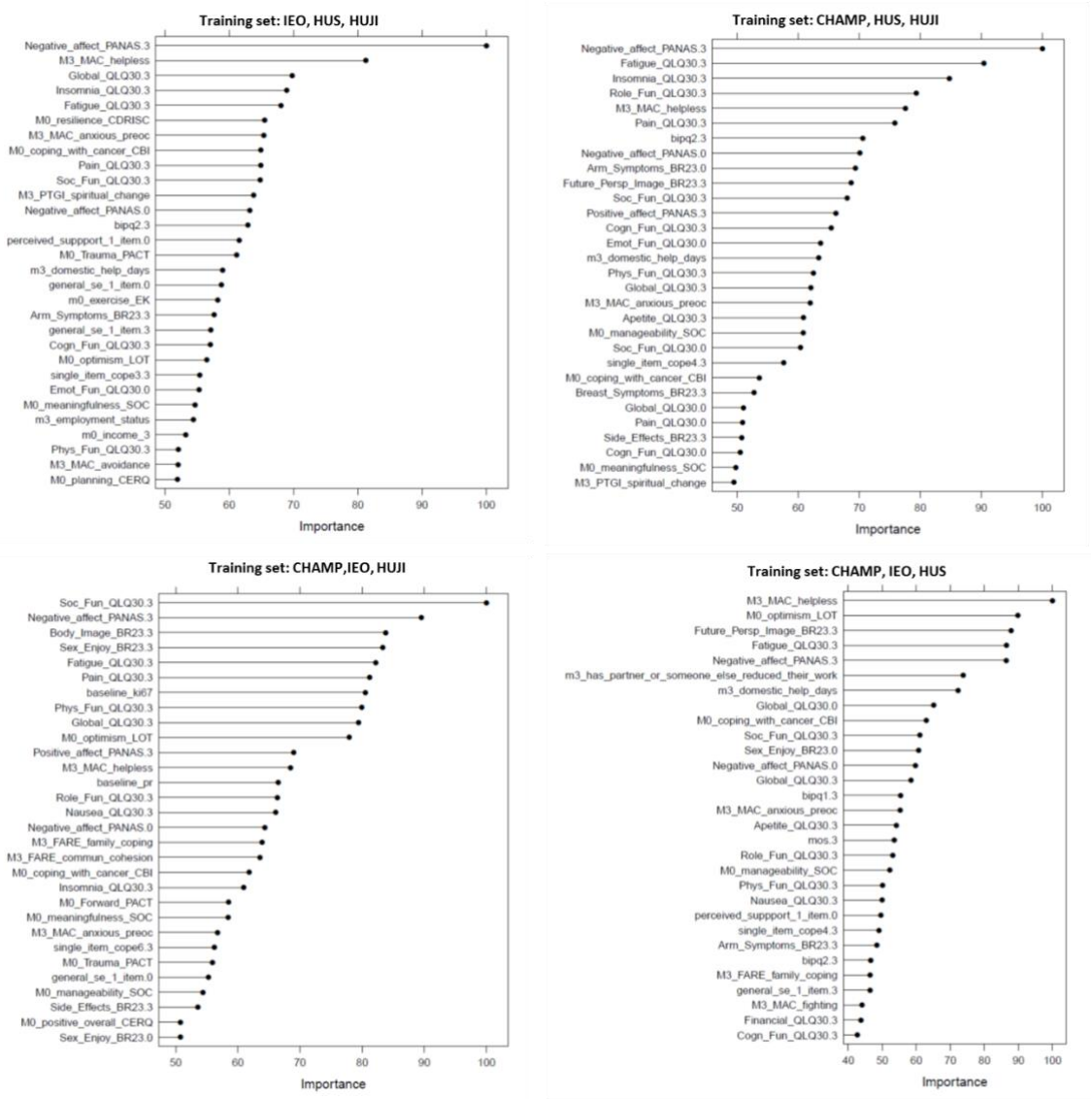


Figure 20: Variable importance chart in each cross-validation

4. Finalization of the BOUNCE In Silico Resilience Trajectory Predictor

This aspect is to be addressed in D4.4b to be submitted at month 51, since the finalization of the predictors will have been achieved by that time point.

5. Quality Assurance of the BOUNCE In Silico Resilience Trajectory Predictor

To ensure that the resilience prediction model is valuable in a clinical decision-making context, a cost-benefit model is developed in WP7. The cost-benefit model considers the entire care path of patients: the technical performance (e.g., sensitivity and specificity) of the resilience prediction model, the costs related to testing and alternative interventions, and health-related outcomes of interventions (e.g., quality of life, saved life years, reduction in sick leave days, reduction in need for services). The focus of the cost-benefit model is in the decision making of the clinician with and without the resilience prediction tools. The core model has been built and the final parameters will be obtained from the clinical pilot data and the prediction model development process.

As part of the model validation a user experiment is being conducted. The aim of the experiment is to measure whether the machine learning prediction incorporated in the clinical decision support system (CDSS) improves clinicians' performance to predict patients' QoL during treatment process. The experiment has bene piloted with a trained a machine learning model to predict a breast cancer patient's QoL after six months from the start of treatments. Until now, we have conducted an experimental setup in which six clinicians from HUS and 3 clinicians from CHAMP who used CDSS and predicted QoL for 60 breast cancer patients. Each patient was evaluated both with and without the aid of machine learning prediction.Our preliminary results proved that clinicians' performance to evaluate the patients' QoL was higher with the aid of machine learning predictions than without the aid. AUC of clinicians was 0.777 with the aid and 0.755 without the aid. When the machine learning model's prediction was correct, the average accuracy (ACC) of the clinicians was .788 with the aid and .717 without the aid. The experiment will be further developed based on month 12 data and conducted in HUS, CHAMP and IEO.

Furthermore, quality assurance tasks relate to work in WP8. Exploitation and Business Planning in WP8 consists of planning the possibilities for commercial exploitation of the BOUNCE In Silico Resilience Trajectory Predictor. As the commercial viability of the algorithm and the interest of potential partners in introducing it into their product portfolio depends on the final properties of the model in the BOUNCE project, the commercial planning work focuses on various possible scenarios. The potential business models are specified and visualized for these scenarios and their legal and technical feasibility is being evaluated. Planning work is done in close cooperation with Varian, as integrating the resilience predicting algorithm as an add-on product to Noona is currently considered to be the preferred path for commercialization of the algorithm. In addition, the views of potential customers in the key markets of interest and clinical end-users are being surveyed in order to evaluate the commercial viability and market interest towards the product. Initial views towards the product are positive, as clinicians have expressed the need

for tools to aid in the personalization of breast cancer care and gaining a holistic view of patients' well-being.

6. Intervention Pilot Study: Resilience predictor tool testing sub study

6.1 Introduction

Based on the BOUNCE prospective multicenter clinical pilot results, the first version of the resilience predictor tool has been tested in the clinical setting in a similar patient population as in the modelling cohort in HUS Comprehensive Cancer Center.

Online interventions are promising in addressing the psychosocial needs of breast cancer survivors. In the recent review by Ross et al 2020, websites were the most common platform for online intervention delivery and cognitive behavioural therapy was the most frequently used therapeutic approach (Ross 2020). Other means of intervention available are exercise and nutrition guidance to avoid obesity. Both exercise and weight control are known to have an effect on the quality of life and on prognosis of breast cancer patients (Holmes 2004, Dieli-Conwrigt 2018, Chlebowski 2020).

The COVID 19 pandemic further emphasizes the importance of digital supportive measures. During the pandemic patients have been unwilling to attend their planned hospital appointments. If we start to measure resilience of breast cancer patients systematically in the future and to offer interventions also for patients with moderately impaired resilience, more resources will be needed. In this testing pilot we aim to provide most of the interventions digitally. This approach would allow to start targeting supportive interventions to a large number of participants cost effectively. The planned interventions are digital psychosocial support, online exercise guidance and focused nutrition guidance. Digital interventions will be available in HUS digital service platform Digital Health Village My Path. This will be a feasible way to give participants support.

The interventions in this pilot are selected based on their availability and the prior knowledge of their effect on quality of life with breast cancer patients. Only the need for psychological support is predicted as the models where available at the correct time.

6.2 Objectives

The objective of this intervention pilot is to test the feasibility of our resilience predictor tool in clinical practise, from the start of oncological treatments and during the intervention period, and to describe the resilience level of the patients both before and after the intervention.

Within BOUNCE, a machine learning predictive analysis framework has been developed to address two of the four axes of the project each grounded on corresponding clinical scenarios.

Firstly, a cross-sectional resilience status prediction has been performed to identify those medical and psychological/behavioural factors that are related to and distinguish between distinct resilience categories as they will emerge in the first type of analysis. Secondly, The second axis addresses the clinical problem of long-term prediction of resilience based on patient trajectories at the early stages of cancer diagnosis and treatments. Within BOUNCE, the design of the predictive tools has been performed using the prospective multicenter clinical pilot of the project. Design of the models included model training, optimization and the selection of the most important parameters that contribute to resilience prediction (both cross-sectionally and longitudinally). Subsequently, BOUNCE models need to be technically validated in terms of their predictive performance. In machine learning aspects, validation of the trained models using external data sources (independent cohorts from those used for model training) is a significant phase of the analysis as it guarantees more accurate estimates of performance, generalizability, and robustness. To this end, this specific dataset will play a role in the validation phase towards the assessment of the model performance and the evaluation of the extracted knowledge regarding the significant patient parameters that were identified through the model design. When important parameters for patients resilience are available and the recruitment for testing pilot have been completed the usefulness of testing pilot cohort for model validation can be evaluated.

6.3 Primary Endpoint

To measure patient compliance towards the targeted interventions by using electronic log information, digital diaries and a specific questionary.

6.5 Secondary endpoints

To describe the resilience scores at baseline and during and after the targeted interventions and to compare the results to a control population of the former BOUNCE prospective pilot study without interventions.

The information about the usability of the tool and targeting the interventions will be collected from the trial nurse using the tool.

6.6 Methods and Study Design

Based on the first results from (baseline to months 3-6) of the BOUNCE prospective pilot study, a mental health status at the start of the oncological treatments in participants with early breast cancer will be evaluated. After oncologist´s appointment patients willing to participate will fill in baseline information about their psychosocial wellbeing and sociodemographic factors to Noona, similarly as described in this protocol earlier for the BOUNCE prospective pilot study. At baseline, the trial assistant will introduce the Noona platform face to face to the participants. The trial assistant will also collect data about the tumour characteristics and treatments received from the patient records of the hospital The information used will be a set of questionnaires and those parameters that predicted the resilience trajectory in the multicenter clinical pilot study

described in the study protocol. The collected and pseudonymized data of each participant will be exported from Noona platform to FORTH. This will provide input to a decision support system to predict mental well-being for the following months based on the four models of anxiety and depression. The platform will also provide a recommendation if amount of exercise is sufficient compared to recommendations based on the amount of exercise patient have recommended. This amount will be compared to the general recommendations (UKK Institute).

The participants will be classified to have a good, intermediate or poor resilience (Table 9). In this testing pilot good resilience means that all the mental health models shows a good mental health during the following months and the amount of exercise is sufficient, poor resilience means that all the mental health models show risk for problems during the following months and the amount of exercise is insufficient. Intermediate resilience means that some of the models show impairment during the following months and/ or the amount of exercise is insufficient. Participants with an impaired resilience (intermediate or poor score) at baseline or with a risk for impaired resilience score in the future, will be offered targeted interventions based on the factors that impaired their resilience. A trial nurse will select the interventions needed for each participant separately. The intervention methods will be a patient empowerment module in the Noona platform, physical exercise through an online group and online nutrition counselling. Nutrition counselling is selected independently based on the age of 40-55 years and risk for early menopause and weight gain.

Table 9 The resilience groups in the testing pilot

Intervention	Good resilience group	Intermediate resilience group	Poor resilience
Patient empowerment module	all the models predicting good mental health	some of the models shows need for support AND/ OR	all of the models shows need for support AND
Physical exercise	amount of exercise is in line with the recommendations*	amount of exercise is insufficient	amount of exercise is insufficient

*moderate aerobic exercise 150 minutes per week or 75 minutes hard aerobic exercise per week (or combination of these two where 1minute of moderate aerobic exercise=2 minutes of hard aerobic exercise) plus 2 times muscle work per week

Resilience of each participant will be followed at baseline and after 3 and 6 months. The active intervention period will be approximately 10 weeks. Participants will have access to the online materials also during the follow up period up to 1 year. Participants' opinions about using the resilience predictive tool and the use interventions will be collected with a questionnaire after 3 and 6 months. Electronic logs about using the digital interventions will be collected. All interventions will take place digitally in HUS Digital Health Village My Path platform. Participants will log in with their bank access codes. Participant information and consent will be signed digitally. The workflow and contents of the Health Village My Path in this study is described in the study protocol. Participant will found the contents described below in the My Path.

1.Patient empowerment module in HUS digital service platform

The participants with moderate or severe psychosocial problems will be guided to use the HUS patient empowerment platform in Digital Health Village My Path (Manuscript in Finnish). Participants with severe psychosocial problems at baseline will also be offered an appointment in the psychosocial unit according to clinical practise of the center. The contents of the module are planned by a psychologist with prior experience about digital support systems in HUS. In the empowerment module participants will have digital support. The module consists of information, detection and practices on different themes: anxiety, relaxation, coping with the everyday life, worrying, presence, thoughts and believes, self-compassion and strength. Anxiety levels will be measured repeatedly when resilience is measured. If the anxiety level indicates severe anxiety (score 13 or more), participants are told to contact the trial nurse. An appointment at the psychosocial unit is then organized after consulting the participant´s oncologist. Participants are asked to use the supportive module 2-3 times a week during the first 10 weeks and to continue according to their own preference after that. Logging information about the use of the module will be collected from the platform and participants will keep an electronic diary. The information about participants who are already referred to the psychosocial unit by the oncologist before using the resilience trajectory predictor will be collected.

2. Physical exercise

The participants with problems related to lack of exercise will be guided to group exercise intervention. Participants are asked to exercise at least once a week with a recorded online session in the My Path platform. These exercise session are planned by the physiotherapists of the HUS Comprehensive Cancer Center. Sessions have various themes like walking outside, relaxation, stretching and strength training. One session will take 30-45 minutes. Once a month during six months online group chat will be available to participants to receive more personal guidance and discuss their questions related to the exercises. Logging information about the use of the module will be collected from the video platform. From the My Path participants will found information about the importance of exercise for the cancer patients and they are encourages to exercise according to the guidelines during the 6 months intervention period.

3. Nutrition counselling

This interventions will be targeted to the participants between 40 to 55 years of age who are at risk for weight gain caused by early menopause caused by the breast cancer therapies. Patients will receive information about the effect on weight gain to the prognosis and quality of life and wellbeing with breast cancer. Nutrition counselling includes also digital practises during 10 weeks and monthly group chat during 6 months.

6.7 Patient Selection

Participants will be breast cancer patients with stage I-III histologically confirmed diagnosis.

6.7 Total number of participants

During 6 months from the start of this study all the consecutive patients fulfilling the inclusion criteria will be asked to participate. The estimated number of participants is 100, based on the multicenter clinical pilot with a similar patient population.

6.8 Inclusion criteria

To be eligible for inclusion in the study, each participant must fulfil the criteria below:

- Presence of a devoted informed consent signed by the participant and the physician
- Female participants, 40-70 years of age at the time of recruitment of diagnosis
- Histologically confirmed invasive early or locally advanced operable breast cancer
- Tumour stage I, II and III
- Patients receiving surgery as part of the local treatment
- Patients receiving any type of systemic treatment for breast cancer regardless of treatment type
- Patient have a computer or mobile device available and she is capable and willing to use it for this study.
- Finnish speaking patients as Digital Health Village My Path is available at the moment only in Finnish

6.9 Exclusion Criteria

Patients who meet any of the following criteria will be excluded:

- Refusal to sign informed consent
- Presence of distant metastases
- History of another malignancy or contralateral invasive breast cancer within the last five years except cured basal cell carcinoma of skin or carcinoma in situ of the uterine cervix
- History of early onset (i.e., before 40 years of age) mental disorder (i.e., schizophrenia, psychosis, bipolar disorder, diagnosis of major depression) or severe neurologic disorder (i.e., neurodegenerative disorder, dementia)
- Serious other diagnosed concomitant diseases such as clinically significant (i.e. active) cardiac disease (e.g. congestive heart failure, symptomatic coronary artery disease or cardiac arrhythmia not well controlled with medication) or myocardial infarction within the last 12 months.
- Major surgery for a severe disease or trauma which could affect patient's psychosocial wellbeing (for example, major heart or abdominal surgery) within 4 weeks prior to study entry or lack of complete recovery from the effects of surgery
- Treatment for invasive cancer

- Treatment for any major illness in the last half year
- Pregnancy or breastfeeding at time of recruitment
- Patient don´t have a computer or mobile device available and she is not capable and unwilling to use it for this study.

6.10 Statistical Analyses

Descriptive statistics will be collected about the compliance for the interventions. Resilience at different time points and between the testing group and controls will be analysed in terms of the normality of their distribution using Shapiro-Wilk test and corresponding tests will be applied to compare the different groups (e.g. a Mann-Whitney U test if the assumption of normality will no met). Resilience of the testing pilot cohort can be compared to a matched cohort from the original BOUNCE pilot.

Status of the pilot 26.8.2021:

85 participants have given their consent, 7 drop outs, 78 active patients
65 patients have the recommendation for some type of intervention or a combination of different interventions
6 patients didn´t need any intervention

Activity in the intervention paths:

Intervention n=78	Need for this intervention	Some activity in the path	Completed intervention	Inactive patients
Patient empowerment	30 (38%)	23 (77%)	1 (3%)	7 (23%)
Exercise	63 (81%)	63 (100%)	19 (30%)	0 (0%)
Nutrition	38 (49%)	34 (89%)	4 (11%)	4 (11%)

Results of the compliance questionnaire at M3 n=43

1. Responding to questionnaires in Noona to determine my resilience has been effortless mean 3,7 (scale 1-5)

2. Measuring my resilience has felt meaningful to me mean 3,3 (scale 1-5)

3. The support measures that I have received have felt helpful 3,3 (scale 1-5)

4. I recommend that breast cancer patients determine their resilience and use the customized online support services 4 (scale 1-5)

5. Would you have liked to receive a different type of support than digital support services, such as in-person meetings? 14/43 (33%) all these patient have used at least occasionally digital services

6. 7. Would you also have managed well without any support at all? 24/43 (56%)

At the beginning of this testing pilot in December 2020 only the first models predicting anxiety, depression and mental health based on the baseline parameters were available. The digital empowerment module will be offered to the patients when even one of the four models shows increased risk for problems. With this low threshold 27/78 (35%) of the patients are in need for psychological support. The most severe cases were referred to a psychologists appointment by the treating oncologist according to local practise.

51 patients are naive for psychological intervention this could bring an opportunity to use this cohort for validation. However the importance of possible other interventions exercise and nutrition should be checked based on the parameters with importance in the final models.

The use of exercise intervention in the testing pilot was based on the prior knowledge about positive effect of exercise for patients health in general as well as to the prognosis and quality of life. At the time this pilot was started we didn't have yet models with the effect of exercise available but an exercise calculator was added to the platform based on self-reporting by patient. Clearly there is a need for exercise intervention as the amount of exercise is insufficient at baseline in 63/78 (81%) patients. This means that 81% patients exercised less than recommended for general population as well as breast cancer patients and survivors.

Also weight gain during breast cancer treatments have effect on breast cancer patients prognosis and quality of life based on literature. We offered digital nutrition guidance for patients 55 years of age or less who are especially at risk for weight gain because of the biological menopause or treatment related early menopause. At the time this pilot was started we didn't have yet models with the effect of weight gain available. 36/78 (46%) patients received nutrition counselling.

Conclusions:

At the time of this report we are nearly finalizing the recruitment with 85 patients who have given their consent and 78 participant continuing in the study. The results of the endpoints are not yet available. 31% of the screened patients have given their consent so obviously there is need for measuring resilience and supportive interventions from patients' perspective.

The data export process and programming the interventions have been a step by step process and have needed work of the trial assistant and research doctor. This clearly needs to be automatized for the final tool. The schedule in the normal clinical practise is tight and the whole process must be very fluent and fast. The use of three different systems in testing one for data collection, one for prediction and one for interventions has neither been optimal and must be developed further. Based on the information received from the trial assistant some of the

patients need help for signing in to the Noona and the Health Village. Only one interface for the patient would be better.

At the end of BOUNCE project all patients will have approximately 6 moths follow up and part patients up to 12 moths follow up. We will collect and report the information about patients compliance and their view towards measuring resilience and targeting interventions as well as the resilience outcomes measured during the study. We will also collect information about the process in general from a single trial nurse.

Based on the experience from the testing pilot from December 2020 to August 2021 we can say that there is a need for this type tool. The technical process from collecting the information from patient to predicting resilience and programming the interventions should be technically improved and automatized if this tool will be implemented to standard clinical practise.

6.11 References

Chlebowski R. T., Aragaki A. K., Anderson G. L., Simon M. S., Manson J. E., Neuhouser M. L., Pan K., Stefanic M.L., Rohan T.E., Lane D., Qi L., Snetselaar L., Prentice R. L. (2018) Association of Low Fat Dietary Pattern With Breast CancerOverall Survival: A Secondary Analysis of the Women´s Health Initiative Randomized Clinical Trial *JAMA Oncol*, 4(10):e181212.

Dieli-Conwright C. M., Courneya K. S., Demark-Wahnefried W., Sami N. , Lee K. , Sweeney F. C., Stewart C, Buchanan T. A. , Spicer D., Tripathy D., Bernstein L., Mortimer J. E. (2018) Aerobic and resistance exercise improves physical fitness, bone health and quality of life in overweight and obese cancer survivors: a randomized controlled trial *Breast Cancer Res*. 20(1), 124.

Holmes M.D., Chen W.Y., Feskanich D., Kroenke C.H., Colditz G.A. (2015) Physical activity and survival after breat cancer diagnosis *JAMA* 293(20), 2479-86

Ross X.S., Gunn K. M., Olver I., Willems R.A., Lechner L., Mesters I., Bolman C. A. (2020) Online Psychosocial Interventions for Posttreatment Cancer Survivors: An International Evidence Review and Update *Curr Opin Support Palliat Care* 14(1), 40-50

7. Other Planned Studies

Descriptions in terms of Accuracy Evaluation, Finalization and Quality Assurance

Although the sets of predictors that emerged from the external (site-specific) validation sets and from the models conducted on the entire cohort present a rather uniform picture regarding the types of modifiable variables accounting for 12-month key resilience outcomes, certain discrepancies were noted with respect to specific patient-report scales that correspond to these general psychological characteristics. To address this issue we plan to examine both the global interpretation results for each scale (PDPs) in order to identify potential measurement-scaling issues, as well as identify subgroups of patients within each clinical site that (through inspection of breakdown profiles examples of which are displayed in Figures 7-14 in D4.3a).

Moreover, additional supervised model cross-validation is under way using M0-M12 prospective longitudinal data recently compiled on an independent cohort of breast cancer patients at IEO.

8. Conclusions

In this document a palette of initial steps, methods, actions and indicative results aiming at estimating the accuracy and assuring the quality of the set of in silico resilience trajectory predictors of the BOUNCE project has been presented. The results obtained so far have led to a generally positive assessment of the respective models developed, although certain discrepancies with respect to specific patient-report scales have been identified. The latter are currently being addressed. However, since the clinical pilot study and a special sub-study aiming at testing the resilience predictor tools are still in progress, a complete account of the aspects under consideration referring to the finalized version of the predictors will be presented in deliverable D4.4b. The latter is to be submitted at month 51. It will contain the exploitation of data collected up to month 18 of the BOUNCE clinical study.