Universitat Oberta de Catalunya

Functional trajectories of subjects with a first episode of psychosis: A two year follow up

Alejandro Keymer Gausset

Biostatistics
Máster in Bioinformátics and Bioestatistics

Daniel Fernández Martínez
(Carles Ventura Royo)

2 de Junio de 2022

FINAL WORK CARD

Title:	Functional trajectories of subjects with a first episode of psychosis: A two year follow up
Author:	Alejandro Keymer Gausset
Tutor:	Daniel Fernández Martínez
SRP:	Carles Ventura Royo
Date of delivery:	2 de Junio de 2022
Studies:	Máster in Bioinformátics and Bioestatistics
Area:	Biostatistics
Language:	English
Number of credits:	15
Keywords:	first episode psychosis, disease trajectories, remission, recovery, functional recovery, latent class analysis, mixed models

Abstract

First episode psychosis (FEP) is the first manifestation of some of the mean mental disorders; Mainly schizophrenia, and Bipolar disorder. Though longitudinal studies of FEP have focused mainly at symptom remission, currently there is a growing interest in functional outcome Likewise,, the relation between symptom remission and functional recovery in matter of current debate. Despite the fact of its pragmatic nature, functional assessment is not standardised, and different studies use different measures. On the other hand, longitudinal studies involve a handful of different methodologies, mainly derived from clustering methods, or an extension of mixed model theory.

This study explores different longitudinal latent growth models for the modelling of both functional and symptomatic outcomes for a sample of 335 FEP cases followed for a period of two years. The functional outcomes were measured by the Functioning Assessment Short Test (FAST) and symptoms outcome as scored by the positive, negative and general PANSS scale.

The results showed heterogeneous trajectories for both functional and symptom outcomes. There was an appreciable association between poor symptoms remission and impoverished functional recovery, specially within negative symptoms. Furthermore the study depicts a strong relation between some socio demographic characteristics, such as the educational levels, and some beeline variables such as the premorbid adjustment scale, IQ and days of untreated psychosis to both poor functional and negative symptoms outcomes. Moreover, more severe symptoms onset, especially in regards to negative and depressive symptoms, was related to both poor functional and symptomatic outcomes. Baseline functional assessment per the WHODAS, geopte and GAF scale also proves to be strongly related to both functional and symptomatic outcomes.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Context and justification

First Episode of Psychosis (FEP) is the first manifestation of some of the main severe mental disorders like Schizophrenia, Bipolar or Major Depressive Disorder, being their clinical characteristics often indistinguishable at illness onset. Traditionally, FEP studies were focused on schizophrenia and were mainly addressed to find effective treatments to achieve promptly remission of positive symptoms and prevention of relapses, aiming at better outcomes (Lieberman et al., 1993). In this line, longitudinal analysis have been noted as useful in FEP research as they can identify and model different outcome trajectories with most evidence suggesting that these trajectories are heterogeneous (Austin et al., 2015; Levine, Lurie, Kohn, & Levav, 2011; Schennach et al., 2012). Moreover, the analysis of longitudinal outcome variables proves to be challenging, especially in what regards mental health data; There is usually a rather large number of incomplete cases as missing data is frequently encountered, there are life events that should be accounted for, such as dropouts or death, and lastly, it usually is the case that outcome variables do not fit a Gaussian distribution.

In the last years, other important aspects of the disorder, such as cognitive impairment and functional outcome, have received greater attention (Miguel Bernardo & Bioque, 2014; Miguel Bernardo et al., 2017), mainly due to the fact that up to 60% of FEP patients have an unfavourable course of illness in spite of achieving high rates of symptomatic remission (Menezes, Arenovich, & Zipursky, 2006). The relationship between symptomatic and functional remission is still a matter of debate (Lambert, Karow, Leucht, Schimmelmann, & Naber, 2010), as it has been suggested by several groups that both outcomes are needed to achieve full recovery (Emsley, Chiliza, Asmal, & Lehloenya, 2011; Harvey & Bellack, 2009). While there is an agreement on clinical remission criteria (Andreasen et al., 2005), there is no consensus yet to define functional remission, as shown by a variety of criteria and scales that are used in the field (Liberman & Kopelowicz, 2005; Mausbach, Moore, Bowie, Cardenas, & Patterson, 2009; Wunderink, Sytema, Nienhuis, & Wiersma, 2009). Despite that impairment in functioning is a core feature in schizophrenia, it seems less evident in other disorders that may arise out of a

FEP.

Studies tend to use different tool to asses functioning. Recent works have reported that first-admission patients with a psychotic disorder had different social functioning trajectories in a 20-year follow-up, where schizophrenia spectrum disorders showed more severe and persistent social impairment compared to other diagnoses such as depression or bipolar disorder with psychotic symptoms (Velthorst et al., 2017). Although these authors suggested a stable trajectory after illness onset, this study only examined social functioning, missing other relevant domains of functioning as vocational or self-maintenance activities (Law & Morrison, 2014). Additionally, there was no assessment of affective symptoms (which may be present in other disorders out of schizophrenia-spectrum) or cognitive performance that might influence remission, relapse and functioning (White et al., 2009). Other findings regarding functional prognosis have also found heterogeneous trajectories using different approaches to measure functional outcome, such a s the GAF scale (Abdin et al., 2017) or a measure of social cognition (Hodgekins et al., 2015)

The heterogeneity of the functionality measurement tools used in different study proves to be problematic. On the one hand some of the tools can assess a restricted domain of functionality (e.g. social functioning), while others may be too general for the task (e.g. GAF). In this regard, the Functional Assessment Short Test (FAST), may provide a good measure of functioning problems experienced by psychiatric patients. The FAST scale is comprised of 24 items that can assess functional impairment id 6 different domains; autonomy, occupational functioning, cognitive functioning, financial issues, interpersonal relationships and leisure time. (González-Ortega et al., 2010)

With regard to predictor variables, there are studies that provide evidence of how different variables relate to different FEP trajectories. Among the most replicated factors, baseline negative symptoms (Verma, Subramaniam, Abdin, Poon, & Chong, 2012), female gender (Álvarez-Jiménez et al., 2012; Wing Chung Chang et al., 2016), duration of untreated psychosis (DUP) (W. C. Chang et al., 2012; Keshavan et al., 2003; Pelayo-Terán et al., 2014), and social, academic or occupational adjustment at illness onset A. K. Malla, Norman, Manchanda, & Townsend (2002) were related with functional outcomes. Cognitive impairment has also been extensively studied, as it is present in the majority of individuals with psychosis (Cuesta et al., 2015).

Therefore, we believe there is a need for more outcome trajectory studies, specifically for studies that may identify the different functional trajectories. The FAST scale may provide a coherent approach to assess functioning on different domains. Demographic, symptomatic and other baseline variables can then be related to the different trajectories in order to analyse the relation between them.

1.2 Objetives

1.2.1 General objetives

The intent of the present work is to determine functional trajectories of a cohort of FEP subjects over a two year follow-up, using a trajectory analysis methodology. Further analysis aims to characterize and differentiate the established classes with regard to the baseline and demographic variables.

1.2.2 Specific objectives

1. Establish a *state of the art* knowledge regarding functional trajectories in FEP.
2. Cleaning and transformation of raw data for further processing.
3. Missing data analysis and imputation.
4. Exploratory data analysis.
5. Identification of functional trajectories and their analysis of their relation to other relevant variables.

1.3 Planning

In order to properly fulfil the propose objectives, the following tasks are proposed:

1. Bibliographical research to establish background and *state of the art* knowledge about FEP prognosis and trajectories.
    a. Introduction
2. First examination and cleaning of *raw data*.
    b. Data import
    c. Error correction
    d. Data selection
3. Descriptive information of data.
    a. Descriptive statistics
    b. Exploratory data analysis
4. Analysis of *missing information* and assessment of possible data imputation.
    a. Descriptive statistic of NA values
    b. Multiple imputation model creation
5. Multiple imputation and assessment of imputed datasets.
    a. Multiple imputation.
    b. imputed data validation.

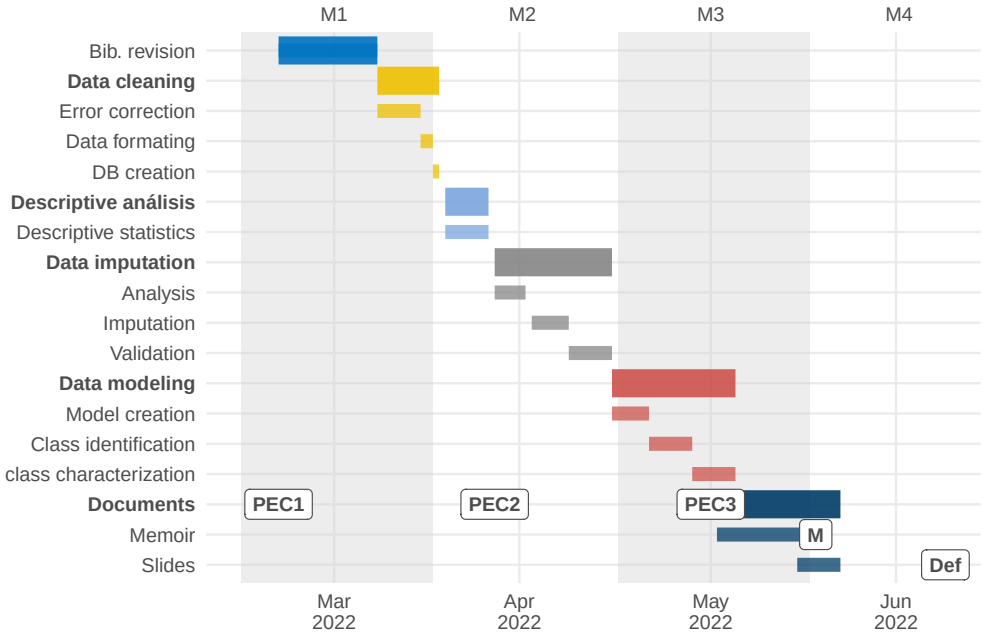


Figure 1.1: Project Gantt chart

6. Trajectories modelling.
    a. simple models
    b. complex models
    c. Identification of trajectory classes.
7. Trajectory analysis
    a. Relation of classes with other relevant variables.
8. Writing

The bibliographical research will be performed using traditional search engines such as Pubmed and Google Scholar, using key terms related to the current project (*first episode psychosis, recovery, functional recovery, remission, trajectories*) Data reading, exploratory analysis, descriptive statistic, data imputation and further data analysis will be performed using **R** programming language (R Core Team, 2021) The *raw* data has been obtained with the online application of the digital data records used by the research group. *Raw* data is in Microsoft Excel *xlsx* format, and is composed of up to a total of 30 files, which must be read, examined, and converted for further handling.

The proposed planning is pictured as a Gantt chart in fig 1.1.

Chapter 2

State of the Art

2.1 Trajectory Analysis

Population variables that change along time may elicit a challenge when analysed. Currently there is a handful of approaches that have been used in the fields of neurological and mental health diseases. More specifically, FEP trajectory studies have usually used some kind a *person centred* analysis, in contrast with *variable centred* analysis such as factor analysis (FA) or principal component analysis (PCA). Person centred analysis aims to provide ways to determine the trajectory of a variable, addressing both changes *between* as well as *within* the subjects themselves (Jung & Wickrama, 2007).

While some of the methodologies encountered in FEP trajectory analysis studies come from the family of cluster methods, such as **longitudinal k-means analysis** (Genolini & Falissard, 2009), most of the algorithms applied on current studies are related to **mixed models** and are globally references to as **Longitudinal Latent Growth Models (LGM)**.

2.1.1 Longitudinal latent growth models

LLGM comprise a family of model strategies for modelling longitudinal data. Presently there is some overlap in the terms that refer to similar models, as there is a confluence of similar algorithms that come from diverse disciplines and are termed differently.

A clear overview of the different approaches can be is described by Nest, Lima Passos, Candel, & Breukelen (2020) and is visually synthesised in Figure 2.1.

A first categorical divergence between models is depicted by a population assumption accepted by the model. Firstly, **single class models** consider that the population is *homogeneous* with regard of the outcome trajectory, hence differences in individual trajectories will only denote intra-subject differences, plus an error term. Secondly, **many class models**, presuppose that the population is *hetereogenic* in nature, with n different subgroups, where predictors tend to evolve similarly over time.

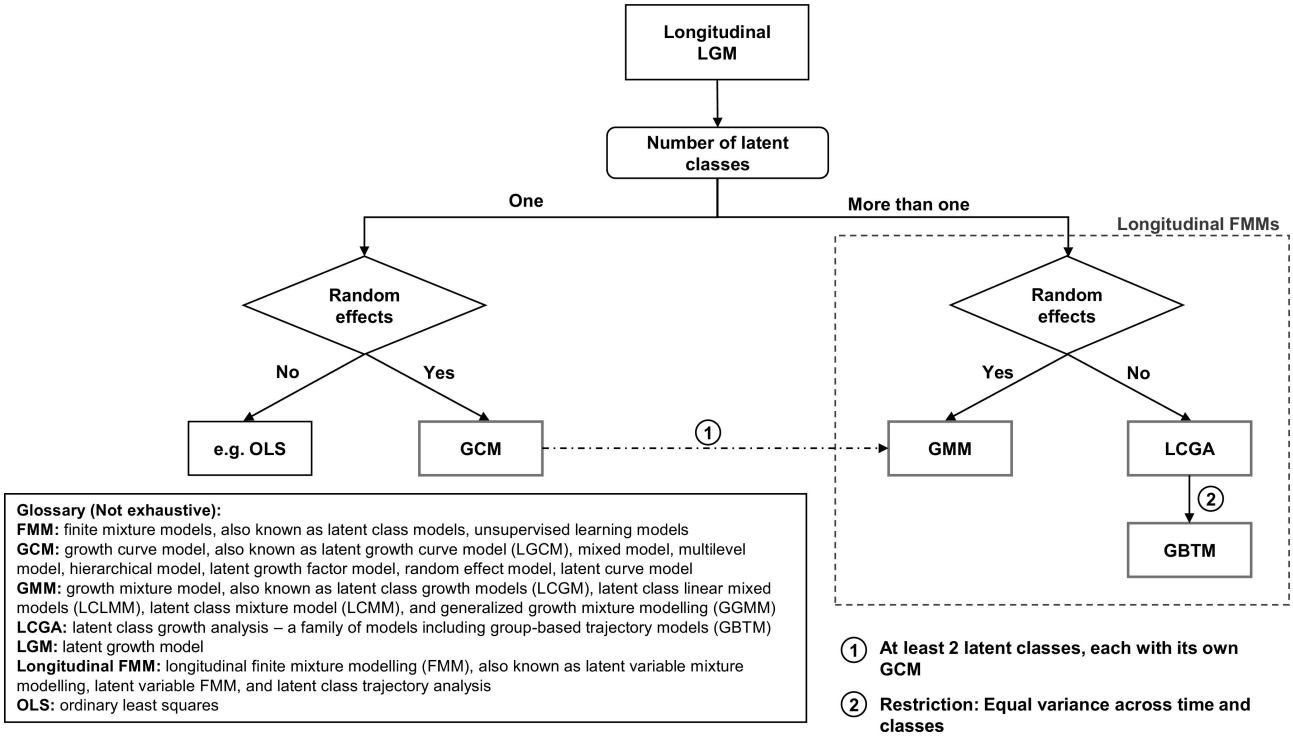


Figure 2.1: Interrelatedness of longitudinal LGM models (Nest et al., 2020).

2.1.1.1 Single class models

Models that presuppose an homogeneous populations are sometimes designated as; Growth Curve Models (GCM), Latent Growth Models (LGM), or just Mixed Models (MM).

In general these models aim to model the relation between predictor variables and a repeatedly measured outcome. In the case of GCM they are designed to model both fixed and random effects on the outcome, aiming to express both differences between and within individuals. The matrix equation can be seen in (2.1)[1]

$$y_i = X\beta + Zb_i + \epsilon_i, \tag{2.1}$$

where, $X : T \times p$ design matrix of constant plus $p-1$ predictors at time points $t = 1, \dots, T$, for fixed effects, $\beta : p \times 1$ vector of fixed effects coefficients, Z:design matrix for random effects. Usually a sub matrix of X obtained by leaving out those columns representing predictors with fixed slope, R: error covariance matrix, b_i: vector of random effects coefficients, $y_i \sim MVN(\mu, \Sigma)$, $b_i \sim N(0, D)$, where D is a covariance matrix, $\epsilon_i \sim N(0, R)$, $\mu = X\beta$, $\Sigma = R + ZDZ'$

[1]General Equation Assumptions $k = 1, \dots, K$ is the class. $t = 1, \dots, T$ is the time point. $i = 1, \dots, n$ is the subject.

Although single class model can only model one class trajectories, they can be extended to model a difference between two categorical groups.

2.1.1.2 Many class models

Models that aim to detect more than one trajectories of the outcome variable are globally known as finite mixture models (FMM). While the a priori number of classes is unknown, it can be established using a different criteria such as the *Bayesian information criteria* (BIC), the *Akaike information criteria* (AIC), and the post classification probabilities. Class assignment of each individual is then based on the similarity of de trajectories. Furthermore, post-hoc methods can be applied to assess the relations between classes and different variables.

Within FMM there, are mainly two different families that differ with regard of the random effect in the regression model. Firstly, latent class growth analysis (LCGA) do not include random effects (Jung & Wickrama, 2007). LCGA assumes that within a specific class, the repeated measures observed within individuals are *independent*. Therefore, as they do not account for between subject variability, few parameters need to be estimated, they usually find more classes and may be particularly useful in models that fail to converge, or have a smaller size (Equation (2.2))

$$y_i^k = X\beta^k + \epsilon_i^k, \tag{2.2}$$

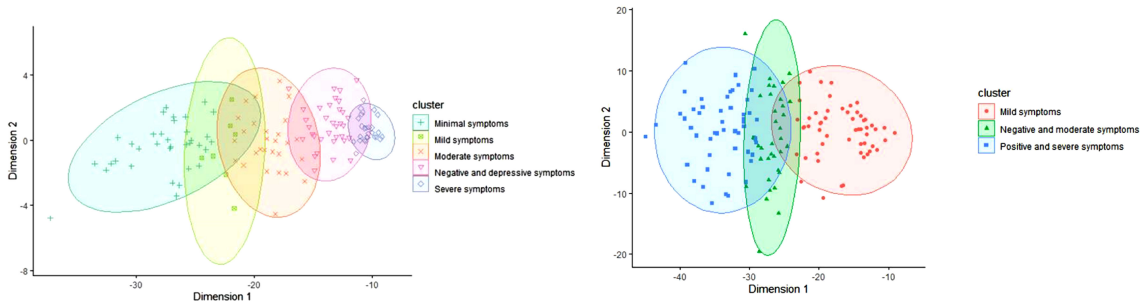
where, $y_i^k \sim MVN(\mu^k, \Sigma^k)$,
$\epsilon_i^k \sim N(0, R^k)$,
$\mu^k = X\beta^k$, $\Sigma^k = R^k$ is assumed diagonal.

In contrast, growth mixture models (GMM), and latent class mixed models (LCMM) both do account for between-subject difference. While, both GMM and LCGA are derived from the latent growth models usually developed in psychometrics, LCMM are derived from mixed model theory in biostatistics.

Thus, LCMM/GMM can be seen as an amalgamations of the former two approaches. It aims to detect many class trajectories as LCGA, including in its regression modelling the use of random effects, as GCM (Equation (2.3))

$$y_i^k = X\beta^k + Zb_i^k + \epsilon_i^k, \tag{2.3}$$

where
$b_i^k \sim N(0, D^k)$
$\Sigma^k = R^k + ZD^kZ'$



(a) Cluster distribution at the baseline. (b) Cluster distribution at the 2-year follow-up

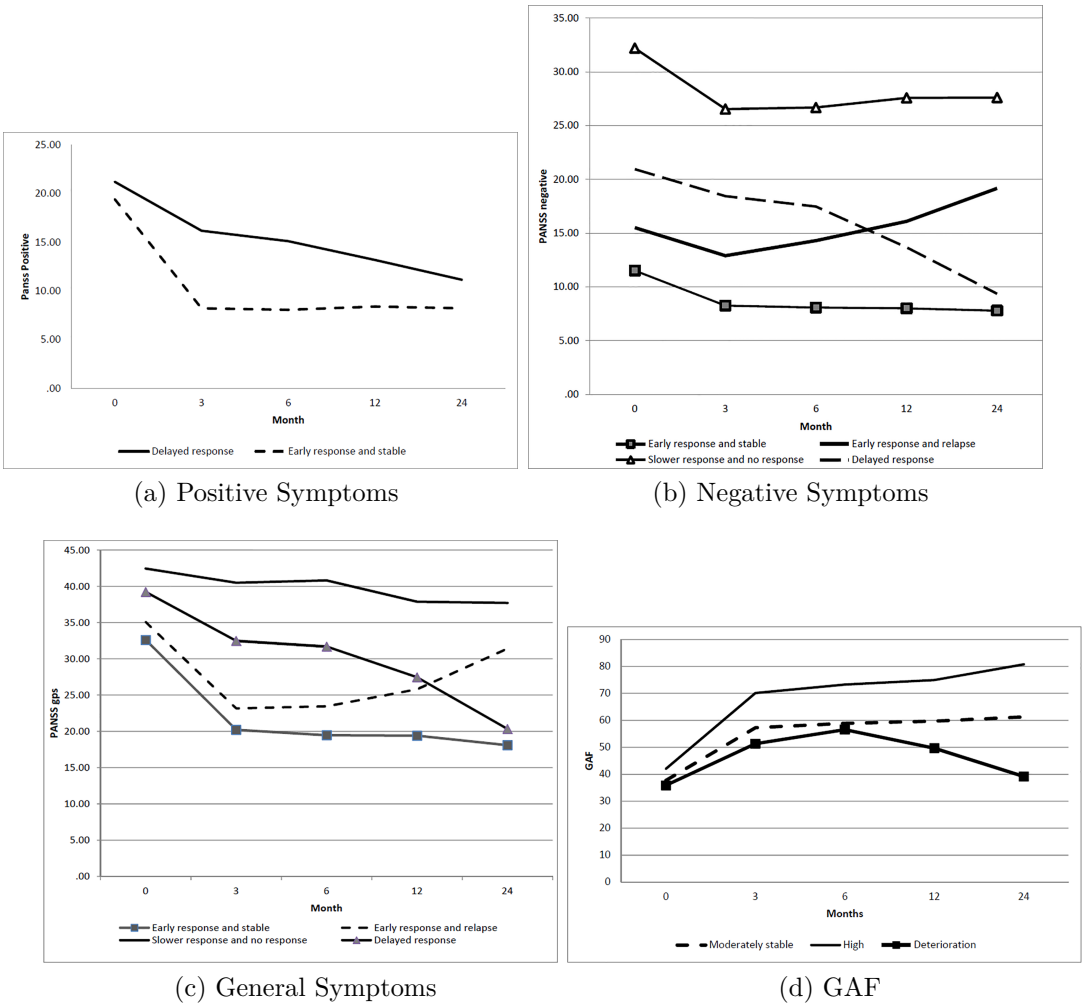
Figure 2.2: Distribution of clinical subgroups of subjects with first-episode psychosis at baseline and at 2-years using fuzzy clustering algorithm (Amoretti et al., 2021).

2.2 Current FEP trajectory analysis studies

Firstly, with regard of clustering methods, one study uses a specific k-means clustering method designed to work on longitudinal data, and is implemented in R packages *kml* and *kml3d.* One advantage of the *kml3d* approach is its ability to model the interaction of two or more variables over time, providing an efficient alternative to existing parametric algorithms (Genolini, Alacoque, Sentenac, & Arnaud, 2015). The study included a total of 369 FEP patients that were followed for one year. K-means cluster modelling for longitudinal data was used to model functional outcome trajectories using the GAF scale as a proxy of functional outcome, using psychotic symptoms scales as predictors. The results showed four distinct functional trajectories ("poor," "intermediate," "high," and "catchup") . The study found a correlation between low functioning and male gender, ethnic minority status, a low premorbid adjustment a low executive function or IQ, previous personality disorder or substance abuse (Hall, Holton, Öngür, Montrose, & Keshavan, 2019)

A similar approach that uses clustering methods, is developed in by (Amoretti et al., 2021). In this study 149 FEP patients where followed up to 2 years. A Fuzzy clustering method was utilized using information form the baseline and the 2 years follow up assessments, using R package *cluster*(Maechler, Rousseeuw, Struyf, & Hubert, 2021). This data was used to build a PCA model which identified dimension. The PCA information was then fed to the clustering methods to identify different subgroups of patients. At baseline, there were 3 subgroups identified; (mild; negative and moderate; and positive and severe symptom), whereas in the 2 year follow up, the best fit was achieved by a 5 cluster solution (minimal; mild; moderate; negative and depressive; and severe symptoms) (Figure 2.2). The main limitation of this approach is that the variability within subject is not accounted for.

With regard on proper trajectory modelling, there are more examples of current studies that use some kind of longitudinal growth model. Abdin et al. (2017) used the Singapore Early Psychosis Intervention Programme clinical database, and applied a a latent class growth anal-



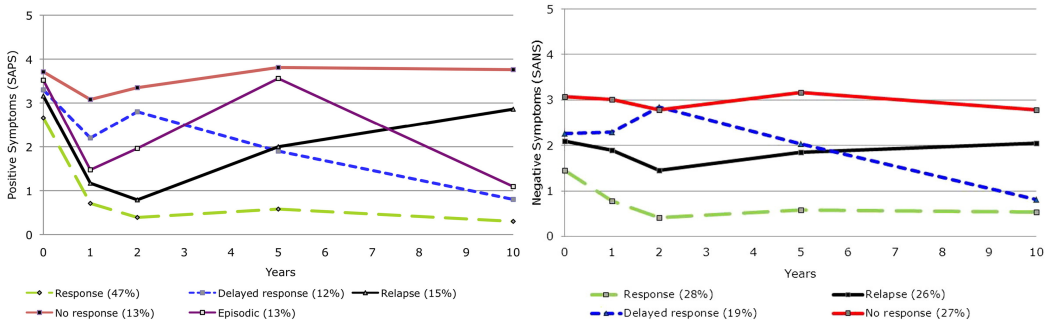
(a) Positive Symptoms (b) Negative Symptoms

(c) General Symptoms (d) GAF

Figure 2.3: Latent class trajectory of PANSS symptoms (Abdin et al., 2017).

ysis to identify distinct trajectories for positive, negative, general psychopathology symptom severity (PANSS positive, negative and general symptom scale) and functioning (GAF) over the 2-year follow-up period. In this approach, missing data was assumed to be missing at random (MAR) and was handled using full maximum likelihood ratio. BIC and AIC criteria were used to select the model with the best fit. In this case, posterior analysis were made using multinomial logistic regression, to explore the relation with other variables, as Age, gender, education level, employment status, marital status, ethnicity and DUP. The results identified two positive symptoms trajectories (*early response and stable trajectory* and *delayed response trajectory*) , four negative and general symptoms trajectories (*early response and stable trajectory*, *early response and relapse trajectory*, *slower response and no response trajectory* and *delayed response trajectory*) and three functional trajectories (*high functioning trajectory*, *moderately stable functioning trajectory* and *deterioration in functioning trajectory*, Figure 2.3).

A similar approach can be seen in Austin et al. (2015), where a large cohort of 496 patients



(a) Positive Symptoms (b) Negative Symptoms

Figure 2.4: Symptom trajectories for 10 year period. (SAPS scoring: 0 - absent, 1 - questionable, 2 - mild, 3 - moderate, 4 - marked, 5 - severe). (Austin et al., 2015)

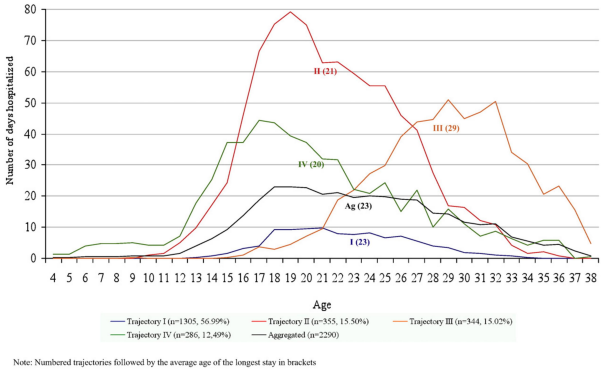


Figure 2.5: Trajectories of the course of illness (Levine et al., 2011).

where followed up to 10 years. The study used latent class analysis to model trajectories of positive end negative symptoms using the SPAS and SANS scales, in order find homogeneous classes of symptoms trajectories. As previous works, optimal number of classes where identified using BIC. With regard on the missing information, it was assumed that missing values where missing at random and, as LCA is tolerant to missing information, no imputation was made. Baseline variables where firstly modelled in univariate analysis with the obtained classes as outcomes to assess the individual variables contribution. Secondly, multivariate logistic models were fitted to see which of the variables still predicted the outcome with the shared variance. The results showed a total of five different positive symptoms classes and four classes for negative symptoms (Figure 2.4). The regression analysis showed that longer DUP and substance abuse, where associated with worst positive symptoms progression, while poor social functioning ans disorganization symptoms were correlated with worst negative symptoms trajectories.

On the same line, Levine et al. (2011) uses a mixed-model latent class regression modelling approach. The optimal number of classes was identified using BIC criteria. In this approach the models where age - adjusted, and all available data at each age was used. A main limitation of this study is that the study did not have access to a formal estimates of illness severity,

so disease severity could only be assumed using de length of the hospitalizations as proxy. ON the other hand, the longitudinal variable used was age, which may account for a more naturalistic approach, and may be less biased than other studies with a more rigid follow-up course. Differences between demographic variables where calculated for the identified classes. The study identified 4 class groups; a typical course with a first hospitalization at age 20, with mild deterioration until 23 and then amelioration. A second group with an earlier onset, a third group with a later onset and a longer deterioration period and finally a fourth group with an early onset and a refractory illness course (Figure 2.5)

Chapter 3

Methodology

3.1 Implementation of LCMM in R: the LCMM package

There are several implementation of the longitudinal model framework available. In the present work we will be using the approach of LGC as described in Cécile Proust-Lima, Philipps, & Liquet (2017) and implemented in the R package *lcmm* Cecile Proust-Lima, Philipps, Diakite, & Liquet (2022). The *lcmm* package includes both function to implement single-class and multi-class LCMM models. As noted by the package author, *lcmm* can also implement LCGA models, by specifying no random effect in the model design.[1] Moreover, it provides a straightforward strategy to deal with non-Gaussian outcome variables.

3.1.1 Linear mixed model

The linear mixed model (LMM) depicts the most basic approach for a longitudinal model, with a one class framework that aims to model the relation between predictor variables and outcomes measured along time, assuming a homogeneous population, that is, without trying to asses differential patterns of change. Furthermore, the outcome variable must fit a Gaussian distribution. Within this model, for each subject i in a sample of N we can consider a vector of n_i repeated measures of an outcome $Y_i = (Y_{i1}, \dots, Y_{ij}, \dots, Y_{in})^\top$ having Y_{ij} a certain outcome value at occasion j that is measured at time t_{ij}. The difference between occasion j and time allows the model to correctly deal with the fact that the number of measurements may vary from between subjects, therefore allowing for subjects to have missing data (Equation (3.1)).

The model represents an outcome variable across time. The type of relation can be modelled as linear or polynomial. Fixed effects denote the average growth overtime of the outcome variable among individuals. Then, individual differences from average time trend are expressed by the sum of the random effects (inter-individual) plus the error.

[1] https://github.com/CecileProust-Lima/lcmm/issues/44#issuecomment-600601121

$$Y_{ij} = \mathbf{X}_{Li}(t_{ij})^\top \beta + \mathbf{Z}_i(t_{ij})^\top u_i + w_i(t_{ij}) + \epsilon_{ij}, \tag{3.1}$$

where, X points to the $t \times p$ matrix for *fixed effects* of $p-1$ predictors, at times t with a β vector of *fixed effect* coefficients. Z is the *random effect* matrix with a vector of b_t random effects which have a D random effect covariance matrix, representing the inter individual random variability (between subjects). And R is the error covariance matrix which represent intra individual random variability.

The main limitations of the linear model is that outcomes have to be continuous and have Gaussian random deviations, and covariate effects B are constant across time, which is not usually the norm in psychological studies.

3.1.2 Latent process

To overcome LMM limitations, a first extension of the model uses a *latent variable* framework, that allows for the separation of the structural model itself from a measurements model, which links the quantities of interest with the actual observations. In this regard, the latent process is defined in the same way as standard linear mixed model, but without the error measurements (Equation (3.2))

$$\Lambda i(t) = X_{Li}(t)^\top \beta + Z_i(t)^\top u_i + w(t) \ , \ \forall t \in \mathbb{R} \tag{3.2}$$

Then, an Y_{tj} measurement model can be defined as a linking model between the the actual measurement Y_{tj} at time t_{ij} and the latent process Λ (Equation (3.3))

$$Y_{ij} = H(\tilde{Y}_{ij}; \eta) = H(\Lambda_i(t_{ij}) + \epsilon_{ij}; \eta) \tag{3.3}$$

The term ϵ corresponds to the standard error with variance of σ_e^2. The term H correspond to the link function utilized and the term $\tilde{Y}_{ij}$ is the latent process with noise at time t_{ij}

When the markers are quantitative there are different link function that can be used; a linear transformation, a re scaled cumulative distribution function (CDF) of a Beta distribution, and several quadratic I-splines with m knots.

3.1.3 Latent classes

A further extension of the latent process mixed model as described in section 2.1.1.2, aims at the proper treatment of *heterogeneous* populations. While the latent process assumes a homogeneous populations with a profile determined by $X_{Li}(t)^\top \beta$, latent *class* mixed models can model an heterogeneous population comprised of G number of different classes which have a G determined profile (Equation (3.4))

$$Y_{ij}|_{c_i=g} = X_{L1i}(t_{ij})^\top \beta + X_{L2i}(t_{ij})^\top v_g + Z_i(t_{ij})^\top u_{ig} + w_i(t_{ij}) + \epsilon_{ij} \tag{3.4}$$

Each subject belong to only one class, and the class memberships can be characterized by a random variable c_i which is a latent variable. Then c_i is equal to g for subject i belonging to class g. The probability of c_i is described by (3.5)

$$\pi_{ig} = P(c_i = g|X_{ci}) = \frac{e^{\xi_{0g}} + X_{ci}^\top \xi_{1g}}{\sum_{l=1}^{G} e^{\xi 0l} + X_{ci}^\top \xi_{1l}} \tag{3.5}$$

Just as for the one class models, the many class approach can use a latent process to outweigh the Gaussian assumption limitation regarding the outcome variable. For G classes the latent process equation is (3.6):

$$\Lambda_i(t)|_{c_i=g} = X_{L1i}(t)^\top \beta + X_{L2i}(t)^\top v_g + Z_i(t)^\top u_{ig} + w_i(t_{ij}) \tag{3.6}$$

3.1.4 Posterior classification

Moreover, in models involving latent classes, a posterior classification of the subjects in each latent class can be made. It is based on the posterior calculation of the class-membership probabilities and is used to characterize the classification of the subjects as well as to evaluate the goodness-of-fit of the model.

Consequently, class memberships probabilities van be calculated using the Bayes theorem. Thus, with the information assessed in the model the probability of class g membership can be determined for each individual i (3.7)

$$\hat{\pi}_{ig}^{(Y,T)} = P(c_i = g|X_{Li}, X_{Ci}, Y_i, T_i, E_i, \hat{\theta}_G) = \frac{\pi_{ig}\phi_{ig}(Y_i|c_i = g; \hat{\theta}_G)}{\sum_{l=1}^{G} \pi_{il}\phi_{il}(Y_i|c_i = l; \hat{\theta}_G)}, \tag{3.7}$$

where,
$\hat{\theta}_G$ is the vector of parameters estimated in the G latent class model.

3.2 Missing data handling

In the fields of health studies it is not rare to encounter missing information. This missing information has to be dealt with in some manner as it can reveal some kind of selection bias of the sample. Furthermore, most statistical algorithms need *complete* data. A first simple approach to deal with missing data is to eliminate all cases with one or more missing values. This approach provides just a few advantages; While it is straightforward and fast, the total number of cases can be reduced dramatically loosing statistical power. Moreover, if missing

data is associated with some characteristic of the sample, this strategy could inject a selection bias to the study.

A second approach is to use only complete *pairwise* data. This may provide a solution for certain statistical approaches which can properly deal with incomplete data. As examined in section 3.1, this is the case with LCMM modelling.

A third approach is to *impute* the missing information with the objective to try to use all of the available data. There are several imputation strategies, from the most straightforward ones that just replace missing values with the mean or the median of the variable, to more sophisticated strategies, that assume that the missing data has a certain underlying mechanism.

3.2.1 Missing data mechanisms

The question with regards of missing data has to do with the necessity of an imputation procedure. To solve this issue a first approach must define the pattern of missing data (Rubin, 1976). A first scenario of missing data is where missingness is *completely at random* (MCAR), which implies that the missing data does not depend neither on the observed nor on unobserved data. In this cases the imputation of the missing data would not provide much information gain. On the one hand imputation may be beneficial with a gain on statistical power. On the other hand, some noise in the form of uncertainty is added with the imputation process. In real life scenarios, specially in health sciences, MCAR cases are not really common, as most of the missing data found in them follow some kind of pattern. Thus, missing data is usually related to other observed variables.

When the missingness of a variable is related to the variable itself, the pattern is considered to be Missing Not at Random (MNAR). However, if the missingness does not depend on unobserved values, but is related to other observed variable, the data is treated as Missing at Random (MAR) (Schafer & Graham, 2002). Furthermore a MAR pattern implies that after controlling for all available data, the remaining missingness is completely random (Figure 3.1 adapted from (Schafer & Graham, 2002))

Both of the later patterns (MNAR & MAR) would be benefited from an imputation approach; While an incomplete dataset that follows a MAR can be imputed using other variable to model the missing values, a MNAR dataset proves to be much more problematic to impute.

3.2.2 Single v/s Multiple imputation

The most simple approach to impute is to use one of many *single imputation* strategies, using the *mean*, *median* or the *mode* for categorical, variables as a replacement for missing data. It may prove some benefit when dealing with MAR an MCAR datasets, but the main disadvantage is that *single imputation* methodologies can not distinguish between imputed and true values, giving no account for the uncertainty of the imputation methods. Furthermore, the use of central measure statistic should be avoided.

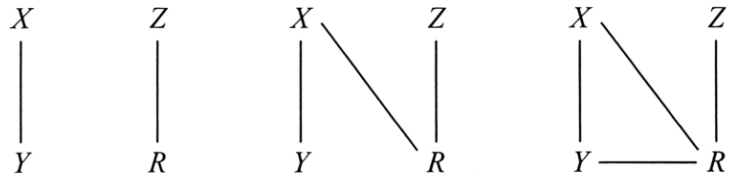


Figure 3.1: Graphical representations of (a) missing completely at random (MCAR), (b) missing at random (MAR), and (c) missing not at random (MNAR) in a univariate missing-data pattern. X represents variables that are completely observed, Y represents a variable that is partly missing, Z represents the component of the causes of missingness unrelated to X and Y, and R represents the missingness.

Multiple imputation strategies are based on the idea to of creating multiple "complete" datasets., thus creating a straightforward measure of uncertainty given by the standard error of the imputed variables. Multiple imputation by chained equations (MICE) describes a procedure in which the imputed values are modelled using n observed variables.

The unobserved variables are calculated iterating through regression models using all (or just some, if specified) observed variables as predictors. The number of iterations should be high enough to stabilize the regression parameters. This procedure then is repeated for n specified number of times to obtain n complete datasets.

3.2.3 Longitudinal imputation

When dealing with a repeated reassure dataset, multiple imputation of missing values can follow different approaches (Buuren, 2018). A first method would be to impute the data in the **long** format, with a separate column taking into account de *time* variable. In this format each *row* correspond to one case in time t . This approach is straightforward, but may not provide the best means, as it would not take into account properly the intra-subject variability of the same measures along time. A second approach uses a **wide** format, spreading the *time* measure into different columns. In this way each *row* corresponds to one case, and the different measures across time are repeated in different columns.

Then, to reduce the number of predictors, for each imputed variable only variables *of the same time point* are included. Furthermore, *same variable* at *different time points* are also admitted to properly account for the repeated measures variability. This strategy can accurately portray predictions for longitudinal data as it provides a feasible imputation for a repeated data model, leaving out degrees of freedom for residual variation.

Following this later approach we can develop an example. Let MADRS be Y_a^p, the YMRS Y_b^p and the GEOPTE Y_b^p, the conditional prediction would be defined by (3.8)

$$P(Y_{a,1}^p | Y_{b,1}^p, Y_{c,1}^p, Y_{a,2}^p, Y_{a,3}^p, Y_{a,4}^p) \tag{3.8}$$

In this way, the prediction matrix for the former data would be (3.9):

$$\begin{bmatrix} & Y^p_{a1} & Y^p_{b1} & Y^p_{c1} & Y^p_{a2} & Y^p_{b2} & Y^p_{c2} & Y^p_{a3} & Y^p_{b3} & Y^p_{c3} \\ Y^p_{a1} & 0 & 1 & 1 & 1 & 0 & 0 & 1 & 0 & 0 \\ Y^p_{b1} & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 1 & 0 \\ Y^p_{c1} & 1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 \\ Y^p_{a2} & 1 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 \\ Y^p_{b2} & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \\ Y^p_{c2} & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 1 \end{bmatrix} \tag{3.9}$$

In summary, the proposed strategy to select the predictors for each variable is:

1. Keep all baseline data as predictors.
2. Keep all other variables *of the same time point.*
3. Of each variable, keep *the different time points*

3.2.4 Imputation and longitudinal latent modelling

Assuming the missing data follows a MAR pattern, there is less of a need to impute for class membership and a more strong need to do it for proper modelling of covariates and post-hoc modelling. What is more important is that both MICE imputation and Raw-maximun likelihood imputation that could also be used, make rather strong assumptions that the data to be imputed come *from a single population.* In contrast, as developed in the 2.1.1.2 section, many class LCMM is based on the hypothesis that data comes from *multiple* latent sub-populations.

The literature is heterogeneous with regards to the best approach of this problem. In the one hand, there is a strong case respecting the use of only complete data for latent class modelling (Colder et al., 2001). There are few specific approaches that are being developed that can deal with missing information, but they are still in an early stage. with this in mind, currently the best approach seems to be to first build the longitudinal models with the available data, and then, with the class information obtained by the models, impute the missing values in order to gain information for the post-hoc analyses.

3.3 Data cleaning

All data was obtained from the data managers of the FEP Project (Miquel Bernardo, Bioque, Parellada, Ruiz, et al., 2013). The raw data was distributed in a series of Excel files, which had one of two general schemes.

1. A *baseline* structure was used for data collected once at the *baseline assessment* (BA), such as demographic information (birth age, education…), baseline scales (PAS, LEWIS, …), arranged with first column with the subjects *id* number and the rest of the data in different columns.

ID	*X*	*Y*	*Z*	···
id	x	y	z	···

2. A *follow-up* structure, for repeated data obtained data from further assessments in time, with a first column with the subjects *id* and the repeated measures distributed along different columns, in what is known as a *wide* data format.

ID	X_{t1}	Y_{t1}	Z_{t1}	X_{t2}	Y_{t2}	Z_{t2}	···
id	x	y	z	x	y	z	···

A first visual inspection of the raw files allowed the identification of transcription errors and inconsistencies in the repeated variables measured. On the one hand, many inconsistencies in the coding of missing values where detected (e.g. missing values where sometimes coded as "NA," "NP," "np"…). Furthermore, in some of the scales (CGI and WHODAS, for example) missing values where coded as "0," or "999."

The Excel files where imported using the **readxl** R package (Wickham & Bryan, 2022). All corrections detected in the visual inspection were explicitly stated in the attached script `clean_data.R`. The scripts manges to read the different Excel files and correct the visual inspection errors detected. The repeated measures data is transformed to a *long* format, that is, one row for each observation with an added column to asses the time of the assessment ("BA," "2m," "6m," "1yr," "2yr"), and finally the data is exported in a *csv* format.

At the baseline assessment (BA), patients would have experienced different time periods from their illness onset. As the objective of the study is to study functionality along illness onset, it seems reasonable to account for this differences in the longitudinal variable. Therefore, the variable "Time from onset" (TFO) was calculated using the SOS scale, which recorded the illness onset date, for each patient at every assessment time.

Structured data sets were built using the clean data .csv files. all data was read into tidyverse tibbles (Wickham, 2021). *Baseline* data was entered into one tibble (`df_bs`), and the timely assessments data in two different tibbles, both in the *long* (`df_tb`) and *wide* formats (`df_wide`).

Table 3.3: Long data format

ID	*t*	*X*	*Y*	*Z*	···
id	1	x	y	z	···
id	2	x	y	z	···

3.4 Exploratory analysis methodlogy

Variable distribution for both baseline variables and primary & secondary outcomes were visually inspected and test for normality. Both outcome variables were visually examined in regards to the different predictors. Correlation between predictor and outcome variables was calculated using the Spearman correlation index. Outliers for the primary and secondary outcome variables where proposed as measures greater than 1.5 time the interquartile range (IQR).

3.5 Longitudnal modelling

Longitudinal models where built for both the primary (FAST) and the secondary outcomes (PANSS sub-scales), using a LCMM and a LCGA approaches. The general strategy followed for building the models was the one proposed in Nagin, Jones, Passos, & Tremblay (2016); The models were built using only the time variable. For the LCMM models both random and fixed effect where specified as the time variable. For LCGA, no random effect was specified. This strategy provides a simple approach which provides both a more comprehensible model and the possibility to further post-hoc analysis.

As both primary and secondary outcomes follow a non-normal distribution, the models were first built using a one-class approach with different link functions. The best model was selected using the BIC and AIC criteria.

Once the best link function was selected, models for n classes where built for all outcomes. A theoretical maximum number of 10 classes was favoured for all models. The optimal number of classes for each model was selected using the BIC and AIC criteria, as well as the post probabilistic classification of the n class model with a clinical criteria in mind.

3.6 Data imputation

The longitudinal imputation strategy followed the methodology exposed in section @ref(Longitudinal imputation). The prediction matrix can be seen in Figure 3.2.

The number of imputations needed was calculated using a two step approach; A first step with a “pilot” model with 5 to 10 imputed datasets, and a second step that uses the pilot results to calculate the imputations needed to establish a final analysis whose SE estimates will have a established level of replicability (von Hippel, 2018). The number of imputation obtained by this procedure is 136

Multiple imputation was made using *mice* function of the *mice* libary (van Buuren & Groothuis-Oudshoorn, 2021). The custom prediction matrix was used to predict 136 complete datasets, using the *Predictive mean matching* method (van Buuren & Groothuis-Oudshoorn, 2011).

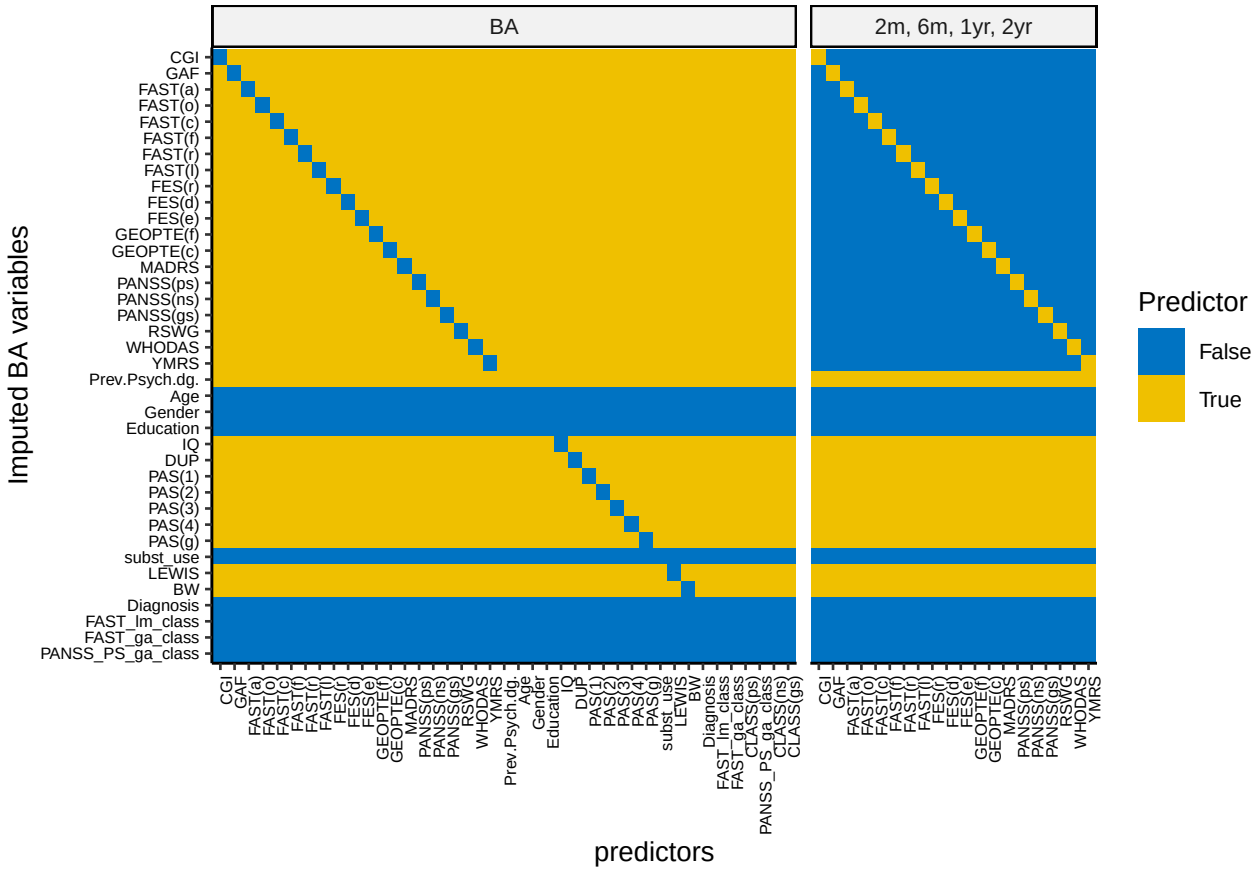


Figure 3.2: Prediction matrix for data imputation (fragment)

3.7 Statistical analysis

For both univariate and multivariate analysis the imputed datasets where used. The statistic for both univariate and multivariate models was calculated and then pooled using *Rubins* algorithm (Rubin, 1976).

A univariate analysis for both categorical and numerical variables was made between the different identified classes of the primary outcomes. For two level outcome classification, the *Wilcoxon signed-rank test* was used to assess mean differences of numerical variables. Where the outcome had more than two levels, the *Kruskal Wallis rank test* was used. For categorical variables, the *Chi-square test* was used. When categorical variables had levels with a low count (less than 5) *Fishers exact test* was preferred.

Furthermore, for outcomes with more than two levels, pairwise comparison were made for each pair of classes, using the *Dunn test* and correcting for multiple comparison using the *False discovery rate* methodology (Benjamini & Hochberg, 1995)

Chapter 4

Results

4.1 Sample description

The study sample is part of the "PEP National Study," which is a naturalistic, prospective, multi centre approach to assess variables from different domains (clinical, neuropsychological, neuroimaging, environmental, pharmacological) in a cohort of FEP patients. The full methodology of the study is described in Miquel Bernardo, Bioque, Parellada, Saiz Ruiz, et al. (2013).

The sample comprised a total of 335 subjects with a diagnosis of a FEP, who met the inclusion criteria for a first episode of psychosis, assessed by a semi-structured interview appropriate to the patient's age: the Spanish translation of the K-SADS-PL for children and adolescents and Spanish translation of the SCID-I and II for the adult population. Inclusion criteria were: a) age between 9 and 35 years at baseline assessment, b) presence of psychotic symptomatology of less than 1 year and c) fluent use of Spanish. Exclusion criteria included: a) DSM-IV-TR diagnosis of mental retardation (Intelligence Quotient (IQ) below 70 and/or impaired functioning), b) history of head trauma with loss of consciousness and c) systemic diseases with secondary mental symptomatology. All subjects provided written informed consent after receiving a complete description of the study. In case of children under 16 years of age, parents or legal guardians gave written informed consent before the beginning of their participation in the study.

At baseline, demographics such as gender, age and educational level were collected from all patients, as well as the retrospective assessment of the premorbid adjustment with the Pre-morbid Adjustment Scale (PAS). The scales used for the longitudinal assessments were: the Positive and Negative Symptoms Scale (PANSS), the Clinical Global Impression Severity Scale (CGI-S), the Global Assessment of Functioning scale (GAF), the Young Mania Rating Scale (YMRS), the Montgomery-Asberg Depression Rating Scale (MADRS), the Scale of Social Cognition for Psychosis (GEOPTE), the World Health Organization Disability Assessment Schedule scale (WHO-DAS) and the Functioning Assessment Short Test (FAST). The estimated premorbid IQ was measured with the Wechsler Adult Intelligence Scale-III (WAIS-III) Vocabulary Test. DUP was also measured as the time from the first onset of psychotic symptoms to first an-

Table 4.1: Socio-demographic categorical variables

		N	%
Gender	Men	225	67.2
	Woman	110	32.8
Prev.Psych.dg.	No	198	59.1
	Not assessable	19	5.7
	Yes	117	34.9
Education	Primary	79	23.6
	Secondary	202	60.3
	University studies	51	15.2
	Other	1	0.3
Substance abuse	FALSE	72	21.5
	TRUE	262	78.2
	NA	1	0.3

tipsychotic treatment. The longitudinal assessments were performed at baseline, 2, 6, 12 and 24 months after study entry.

At study entry the subjects had different times from illness onset. This was collected with the Symptom Onset in Schizophrenia (SOS) inventory, and the variable time from onset (TFO) was calculated for the different follow-up assessments.

Following the definition by the Remission Schizophrenia Working Group (RSWG) (Andreasen et al., 2005), we used the term 'clinical remission' for patients that fulfilled the proposed PANSS sub-items criteria for at least 1 year. This lag-time is more strict than the RSWG one, as the latter required sustained remission for only 6 months.

4.2 Data description

The cleaned data was composed of two data frames. A first "socio-demographic" data frame comprised of 12 variables, 5 of them categorical (1 of them, "Education" ordinal) and 6 numerical. The second was composed of repeated measures of 19 numerical variables in the *long format*, with a columns specifying "assessment time."

Socio-demographic categorical variables consisted in; Previous psychiatric diagnosis ('Prev.Psych.dg'), Gender, level of education ('Education'), Diagnosis, and previous substance abuse 4.1.

On the the hand, socio demographic numerical variables were consisted in; Age, IQ, Days of untreated psychosis (DUP), Pre-morbid Adjustment Scale sub-scales (PAS), Lewis pregnancy risk scale, an the Birth weight (BW). The full description of baseline numerical variables can be seen in table 4.2

In what regard to the "follow up" longitudinal data numerical data was composed of 19 variables; CGI, GAF, FAST 6 sub-scales, FES 3 sub-scales, GEOPTE 2 sub-scales, MADRS, PANSS 3 sub-scales, WHODAS and YMRS 4.3.

Table 4.2: Socio demographic numerical variables

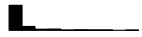
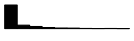
	Unique (#)	Missing (%)	Mean	SD	Min	Median	Max	
Age	26	0	23.6	6.0	9.0	23.0	36.0	
IQ	17	12	91.8	15.2	55.0	92.5	140.0	
DUP	149	14	103.6	119.5	0.0	59.0	807.0	
PAS (childhood)	22	6	6.1	4.1	0.0	6.0	21.0	
PAS (early adol.)	27	7	8.4	5.3	0.0	8.0	27.0	
PAS (late adol.)	29	12	9.5	5.9	0.0	9.0	28.0	
PAS (adult)	14	21	3.3	3.0	0.0	3.0	12.0	
PAS (general)	49	6	19.8	11.2	0.0	19.0	54.0	
Pregnany risk scale	11	2	0.8	1.5	0.0	0.0	10.0	
Birth Weight	51	30	3280.4	532.8	2000.0	3200.0	5000.0	

Table 4.3: Numerical variables of repetaed meassures at: study entry (BA), 2, 6, 12 and 24 months.

	Unique (#)	Missing (%)	Mean	SD	Min	Median	Max	
CGI	8	20	3.5	1.3	1.0	4.0	7.0	
GAF	66	20	64.3	17.9	10.0	65.0	100.0	
FAST (autonomy)	14	21	3.6	3.2	0.0	3.0	12.0	
FAST (occupational)	17	21	6.1	5.3	0.0	5.0	15.0	
FAST (cognitive)	17	21	4.8	3.8	0.0	5.0	15.0	
FAST (financial)	8	21	1.3	1.7	0.0	0.0	6.0	
FAST (interpersonal)	20	21	5.5	4.6	0.0	5.0	18.0	
FAST (leisure)	8	21	2.0	1.8	0.0	2.0	6.0	
FES (relationship)	21	31	15.0	3.1	4.0	16.0	27.0	
FES (development)	30	31	18.4	4.4	5.0	18.0	36.0	
FES (stability)	19	31	9.9	3.1	1.0	10.0	18.0	
GEOPTE (basic cognitive skills)	28	24	12.7	5.2	7.0	11.0	35.0	
GEOPTE (social cognition)	30	25	14.8	5.8	7.0	14.0	36.0	
MADRS	44	18	9.0	8.3	0.0	7.0	44.0	
PANSS (PS)	34	17	12.6	6.6	7.0	10.0	41.0	
PANSS (NS)	37	17	16.2	7.2	7.0	16.0	47.0	
PANSS (GS)	58	17	29.8	11.4	16.0	27.0	85.0	
WHODAS	23	20	7.4	4.6	1.0	7.0	23.0	
YMRS	41	18	4.1	7.1	0.0	0.0	49.0	

All numeric variable types are "double", aside from CGI which is an "integer"

Table 4.4: Baseline variables

Variable	**Overall**, N = 335	**Men**, N = 225	**Woman**, N = 110	**p-value**
Age	23.0 (19.0, 28.0)	22.0 (18.0, 28.0)	24.0 (20.0, 30.0)	**0.014**
Education				0.5
Primary	79 (24%)	53 (24%)	26 (24%)	
Secondary	202 (61%)	140 (62%)	62 (57%)	
University studies	51 (15%)	30 (13%)	21 (19%)	
Other	1 (0.3%)	1 (0.4%)	0 (0%)	
Substance abuse/dep	262 (78%)	186 (83%)	76 (69%)	**0.004**
LEWIS	0.00 (0.00, 1.00)	0.00 (0.00, 1.00)	0.00 (0.00, 1.00)	0.5
IQ	92 (80, 100)	95 (85, 105)	90 (80, 100)	**0.018**
DUP	59 (24, 138)	61 (22, 146)	49 (26, 135)	0.6
PAS	41 (27, 62)	42 (28, 63)	40 (26, 61)	0.4

[1] Median (IQR); n (%)

[2] Wilcoxon rank sum test; Fisher's exact test; Pearson's Chi-squared test

4.3 Exploratory analysis

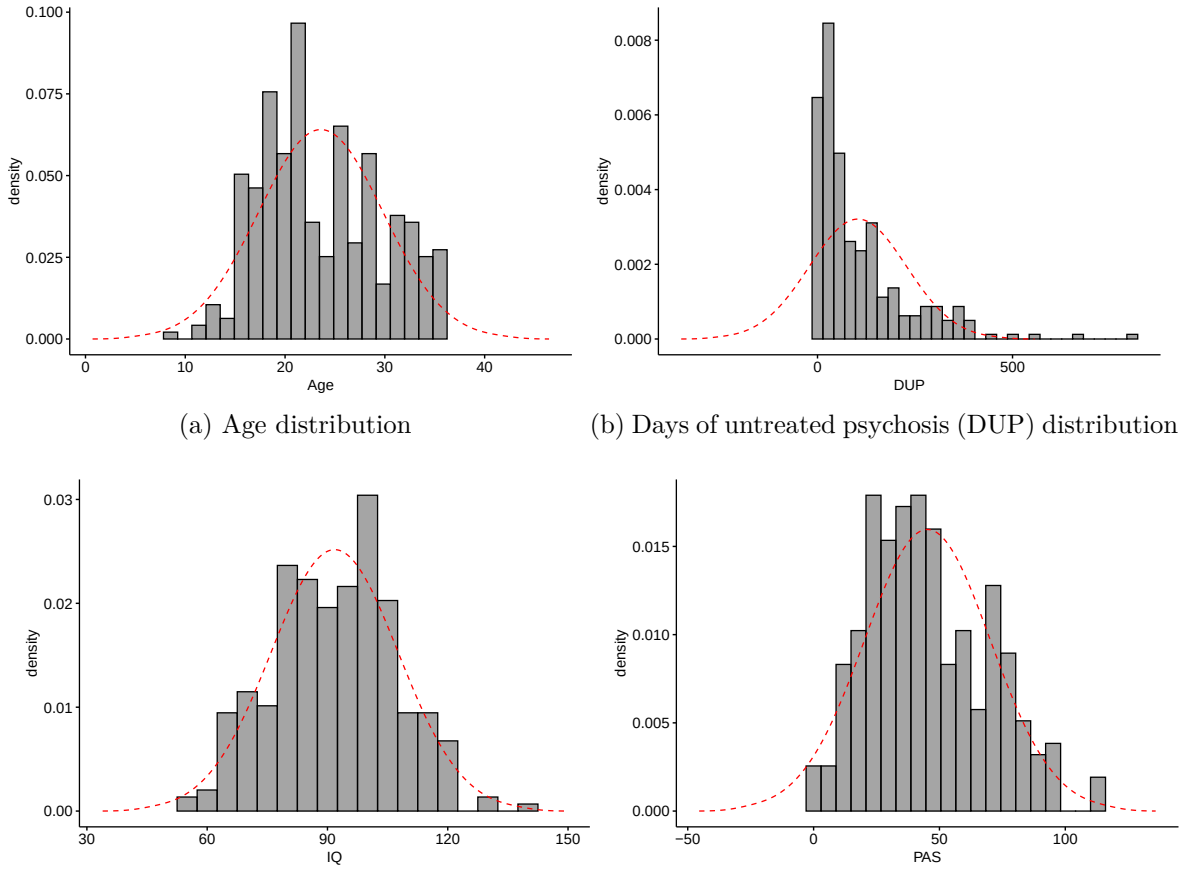
4.3.1 Socio-demographic variable univariate analysis

A total of 335 subject where initially included in the study, with most of them being males (67.2 %). The majority had at least finished secondary studies (99.4 %). Over half of the sample did not have a psychiatric previous diagnosis (59.3 %) while the 35% that did have one, was mainly diagnosed with anxiety disorders.

The mean DUP of the sample was 103.63 which is slightly over the mean DUP described in other FEP studies. Furthermore, the DUP distribution has a long right tail, as depicted by the fact that median is much higher than the mean.

There were signficative differences in age of onset along gender, which is coherent with the current evidence. Furthermore, gender differences can also be seen in the IQ median (95 for Men, 90 for Women). There is also a significative difference between Gender in what regards to substance abuse/dependence with 186 (83%) for men, and 76 (69%) for women (p=0.004) 4.4.

The univariate analysis of the socio-demographic variables show that most variables ('Age,' 'IQ,' 'PAS' and 'BW') differ slightly from the normal distribution. On the other hand, 'DUP' and 'LEWIS' follow a left skewed distribution, with a long right tail (Figure 4.1). Nevertheless, in all of the numerical socio -demographic variable the *Shapiro-Wilk* test discarded de null hypothesis, meaning that these variables do not fit a normal distribution A. These results suggest the use of non-parametric statistics when dealing with these variables.



(a) Age distribution

(b) Days of untreated psychosis (DUP) distribution

(c) IQ distribution

(d) PAS distribution

Figure 4.1: Baseline continuous variables distribution

Table 4.5: Follow-up longitudinal variables

Variable	**Overall**, N = 1,675	**BA**, N = 335	**2m**, N = 335	**6m**, N = 335	**1yr**, N = 335	**2yr**, N = 335
CGI	4.00 (3.00, 4.00)	4.00 (4.00, 5.00)	4.00 (3.00, 4.00)	3.00 (2.50, 4.00)	3.00 (2.00, 4.00)	3.00 (2.00, 4.00)
GAF	65 (55, 80)	51 (35, 65)	60 (55, 70)	70 (60, 80)	75 (60, 85)	75 (65, 85)
MADRS	7 (2, 14)	12 (5, 19)	8 (4, 15)	6 (2, 11)	4 (1, 10)	4 (0, 9)
PANSS (PS)	10.0 (7.0, 16.0)	19.0 (12.0, 25.0)	10.0 (7.0, 15.0)	9.0 (7.0, 12.0)	8.0 (7.0, 11.0)	8.0 (7.0, 13.0)
PANSS (NS)	16 (10, 21)	18 (12, 24)	17 (11, 21)	14 (10, 20)	13 (9, 19)	12 (8, 19)
PANSS (GS)	27 (21, 37)	37 (28, 47)	28 (22, 37)	25 (20, 32)	23 (19, 30)	23 (19, 31)
SWR remission criteria	772 (56%)	82 (25%)	178 (58%)	177 (63%)	185 (73%)	150 (74%)
WHODAS	7.0 (4.0, 10.0)	8.0 (5.0, 12.0)	7.0 (4.0, 11.0)	5.0 (3.0, 10.0)	6.0 (3.0, 9.0)	6.0 (3.0, 9.0)
YMRS	0.0 (0.0, 6.0)	6.0 (0.0, 15.0)	1.0 (0.0, 4.0)	0.0 (0.0, 2.0)	0.0 (0.0, 2.0)	0.0 (0.0, 3.0)
FAST	22 (10, 35)	27 (16, 40)	24 (13, 35)	20 (9, 34)	19 (7, 30)	19 (6, 30)
PANSS	54 (40, 73)	75 (57, 91)	56 (42, 70)	50 (39, 64)	46 (37, 60)	44 (35, 62)
GEOPTE	25 (19, 33)	30 (22, 38)	26 (20, 34)	24 (19, 32)	22 (18, 30)	22 (18, 29)
FES	44 (39, 49)	43 (38, 48)	43 (38, 48)	44 (39, 49)	44 (39, 49)	44 (40, 50)

[1] Median (IQR); n (%)

4.3.2 Follow up clinical variables univariate analysis

A first glimpse into the structure of the follow-upo variables returns some insights which are further developed in this work. Firstly, most of the numerical variables tend to drop as the time of the assessments progresses, excluding the 'GAF' scale. While most of the clinical and functional scales used measure severity with a higher score, the 'GAF' index is reversed; the lower the score the more severe impairment. Secondly, most measurements have left skewed distribution, as their median is near the lower IQR score 4.5.

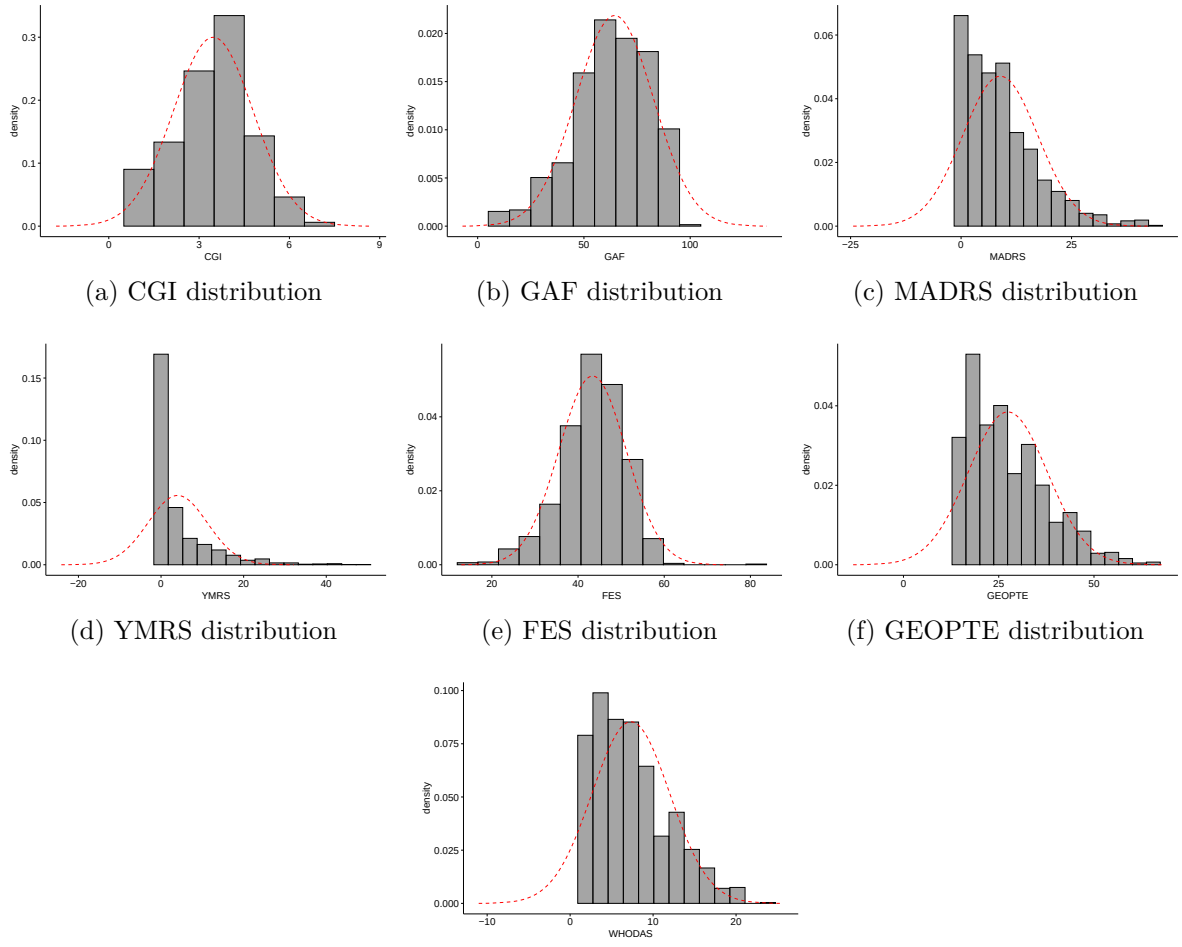
Further analysis of the follow-up variables reveal that most deviate from normal distribution. While CGI, GAF and FES have a distribution that nearly fits a Gaussian, the MADRS, YMRS, all PANSS scores, GEOPTE and WHODAS have a strongly left skewed distribution as depicted in figure 4.2. Regardless of the visual fit, the Shapiro-Wilk test rejects de null Hypothesis of normality A.

4.3.3 Variable correlations

Overall, while there are some variables that exhibit correlation, such as CGI and GAF or the PANSS negative and general sub scales, there does not seem to exist a pattern of excessive interrelationship. The highest degree of correlation is between the PANSS general and negative sub scales, with a Spearman correlation index of 0.71, that on the one hand may indicate a moderate to strong correlation, but on the other, dos not take into account the longitudinal distribution of the variables (Figure 4.3).

4.3.4 Outcomes and longitudinal time variable analysis

A special attention has to be put on the outcome variables, and the longitudinal time variable. On the first hand, the distribution of the primary outcome (FAST scale) and the secondary outcomes (PANSS sub-scales) markedly differ from the Gaussian distribution, at all time points where information was gathered. Moreover, as the time from study entry lengthens, all outcome



(a) CGI distribution (b) GAF distribution (c) MADRS distribution

(d) YMRS distribution (e) FES distribution (f) GEOPTE distribution

(g) WHODAS distribution

Figure 4.2: Longitudinal continuous variable distribution

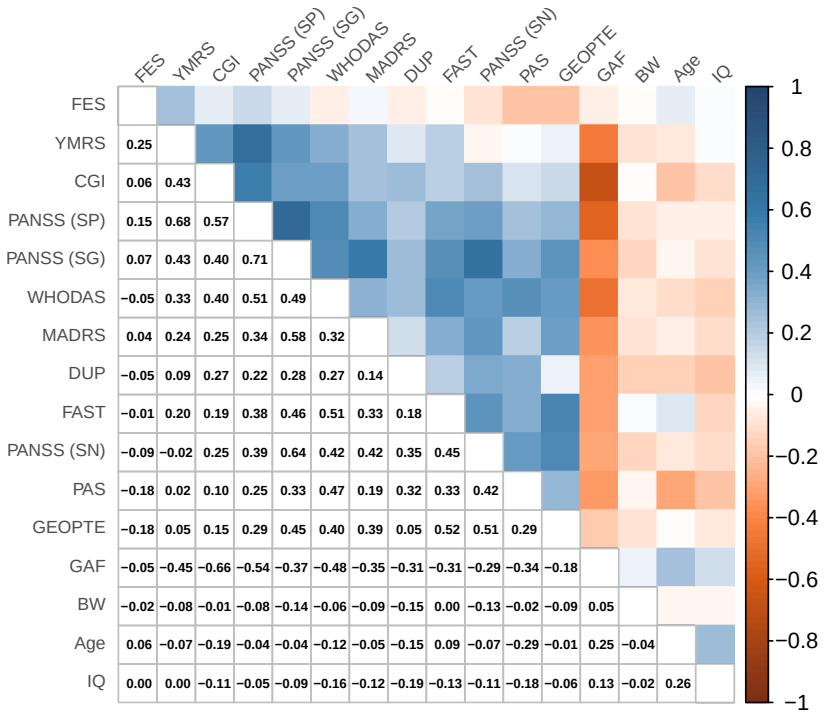


Figure 4.3: Variable correlations

variables tends to skew to the left. This has a clinical explanations, as both FAST and PANSS scores measure disease severity, and the general trend of PEP cases is towards the amelioration. Nonetheless, the pattern by which the different outcomes evolve in time has visible differences. While the PANSS(PS) shows a marked left skewness which translates as symptom remission, both PANSS(NS) and PANSS(GS) show a slower deviation. Likewise, the FAST scale, tends to skew to the left with time, but much slower than all of the PANSS sub scales. Besides, the distribution of the last assessment times has a bivariate shape, which is not as conspicuous as in the PANSS sub scales (Figure 4.4).

In what regards to the dependant time variable (TFO) from which the outcomes are measured, its distribution has a long right tail. A square root transformation of the TFO reveals a much better fit to a Gaussian distribution and will be preferred for further modelling (Figure 4.5).

4.3.5 Bivariate outcome analysis

The relation between the primary (FAST) and secondary (PANSS) outcomes is noticeable in the bivariate scatter plot (4.6). While at the study entry (BA) the relation between FAST and all PANSS sub scales is almost direct, the association weakens along time, specially in the positive and, slightly less, in the general symptom scales. Although, the negative symptom scale, seems to hold the correlation onward in time.

Furthermore, the analysis of the distribution along the time from onset (TFO) variable does

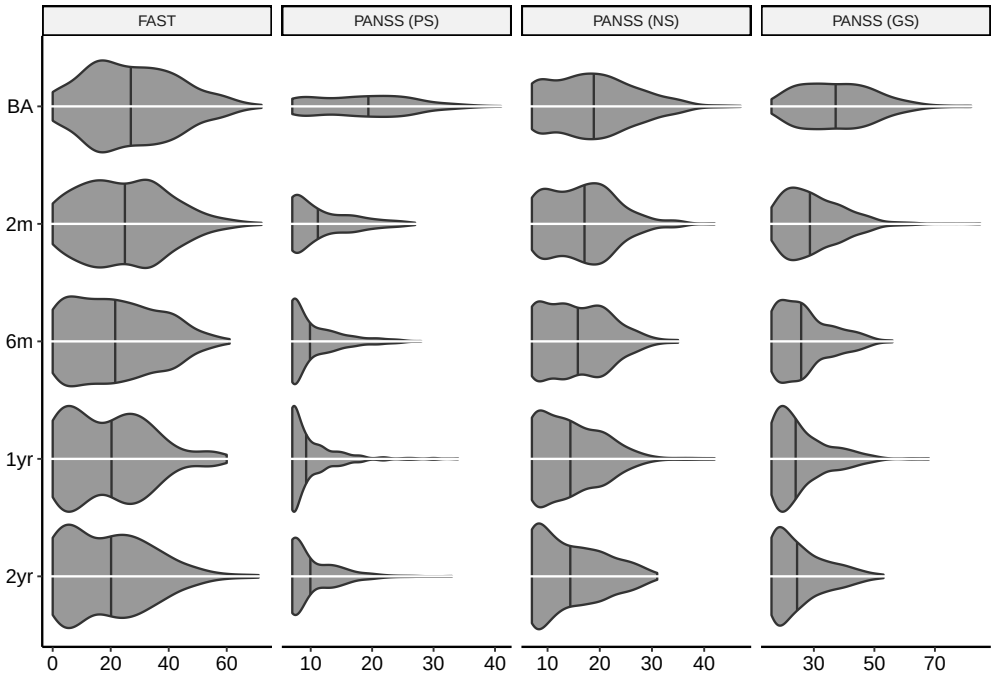
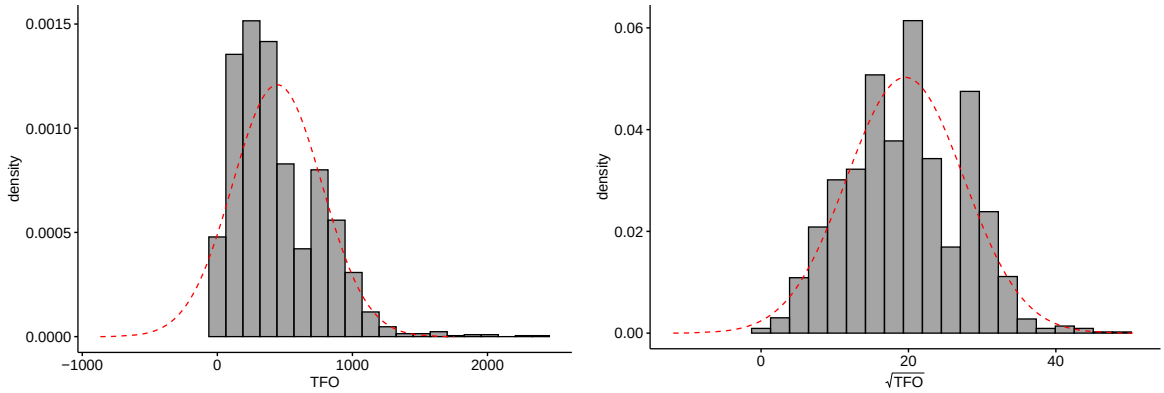


Figure 4.4: Primary and secondary outcome distribution along assessment time, from study entry ("BA"), 2 months ("2m"), 6 months ("6m"), 1 year ("1yr") and 2 years follow-up ("2yr")



(a) Original TFO (b) Square root Transformed TFO

Figure 4.5: Time from onset (TFO) distribution

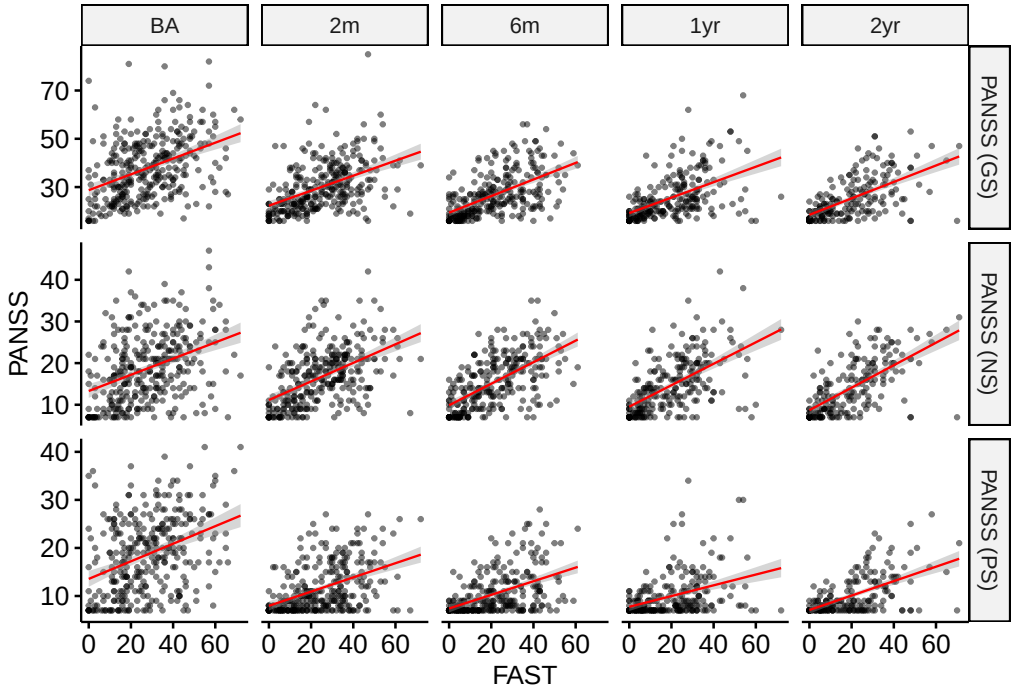


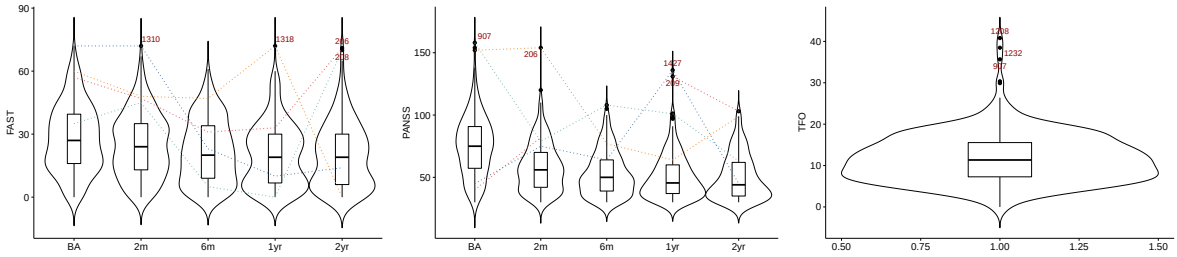
Figure 4.6: Relation between primary (FAST) and secondary (PANSS) outcomes along assessment from study entry

reveal the same trends seen on previous analysis. While both positive and general PANSS scores tend to fall along time, the FAST scale and to negative PANSS scores, decline in a slower fashion (Figures A).

4.4 Outliers

Outliers for both the primary and secondary outcomes were defined as measures greater than 1.5 times de inter quantile range (IQR) and are visually depicted in figure 4.7. Most of these cases were kept in the study, as thorough analysis revealed that the outlier measurement was coherent with other measures of illness severity. This was not the case for ID's "1318" and "208" at the "1yr" and "2yr" assessment respectively, which displayed an aberrant pattern which was not coherent with the rest of the clinical measures of severity and were thus deemed as probable errors, and accordingly, were discarded from further analysis.

On the other hand, TFO outliers where discarded as they where regarded as being at a clinically different stage of illness from rest of the sample, along the whole length of the study.



(a) FAST outliers (b) PANSS outliers (c) TFO outliers

Figure 4.7: Boxplot of variables along the different assessment visits, with possible outliers.

4.5 Missing data analysis

Firstly, all data was analysed in regard of missing values. On the one hand, a *drop-out* cases was characterised as a case for which all information was unavailable from a time point t and further time points. Moreover, if at a certain point there was only *some* missing values, or there was available data from further time points, cases where not considered as drop-outs, but as missing values cases. This contrast is important as imputation for the former was not considered as advantageous, but for the later was.

Furthermore, *early drop-out* cases, defined as having less than 2 time repeated measurements on all variables, were eliminated from the study.

Comparison between baseline variables (Age, Gender, BW, LEWIS, PAS, Diagnosis, Education, IQ, DUP, subst_use, CGI, GAF, MADRS, YMRS, PANSS, GEOPTE, FAST, FES) between drop-outs and kept cases was made, using the Wilcoxon rank-sum test to asses if there were significant differences between the two groups.

There where 13 cases of patients with only 1 assessment along time, that were dropped from the study. Table 4.6 shows that baselines demographic and baseline assessment data does not differ siggnficantly between the kept cases and dropouts.

The missing data patterns were visually depicted for inspection for baseline and the follow-up variables in the 5 time points. Furthermore, analysis regarding missing values in the main and secondary outcome variables (FAST, PANSS PS, PANSS NS and PANSS GS) was made. The relation between missing v/s not missing groups was examined using the Wilcoxon rank-sum test to asses significant differences. In addition, the association between the time variable (TFO) with the outcome primary and secondary variables was established visually and statistically with the Wilcoxon rank-sum test.

Furthermore missing data analysis concerning the baseline data, reveals no identifiable pattern; missing values seem to be distributed randomly and do not seem related (Figure 4.8)

In what respect the follow up information, missing values do display a an increase in their number further from the study entry. This was expected as information quality in longitudinal

Table 4.6: Descriptive table of basline variables of kept cases versus dropouts

Variable	**Kept**, N = 317	**Dropout**, N = 13	**p-value**
Age	22.0 (19.0, 28.0)	25.0 (19.0, 30.0)	0.3
Gender			0.4
Men	214 (68%)	7 (54%)	
Woman	103 (32%)	6 (46%)	
BW	3,200 (3,000, 3,600)	3,300 (3,000, 3,750)	0.9
LEWIS	0.00 (0.00, 1.00)	0.00 (0.00, 1.00)	0.7
PAS	40 (26, 62)	51 (27, 72)	0.6
Diagnosis			0.3
Toxic psychosis	14 (4.4%)	1 (7.7%)	
Schizophrenia disorders	260 (82%)	12 (92%)	
Affective disorders	43 (14%)	0 (0%)	
Education			>0.9
Primary	75 (24%)	3 (25%)	
Secondary	190 (60%)	8 (67%)	
University studies	50 (16%)	1 (8.3%)	
Other	1 (0.3%)	0 (0%)	
IQ	95 (80, 100)	72 (71, 74)	0.060
DUP	61 (25, 138)	9 (6, 171)	0.3
subst_use	247 (78%)	10 (77%)	>0.9
CGI	4.00 (4.00, 5.00)	4.00 (4.00, 5.50)	>0.9
GAF	51 (35, 65)	56 (41, 70)	0.5
MADRS	12 (5, 18)	12 (8, 20)	0.6
YMRS	6 (0, 15)	8 (2, 12)	0.8
PANSS	75 (58, 90)	72 (56, 92)	0.8
GEOPTE	30 (22, 38)	28 (26, 38)	0.7
FAST	26 (16, 39)	27 (17, 46)	0.7
FES	43 (38, 48)	45 (42, 50)	0.2

[1] Median (IQR); n (%)

[2] Wilcoxon rank sum test; Fisher's exact test

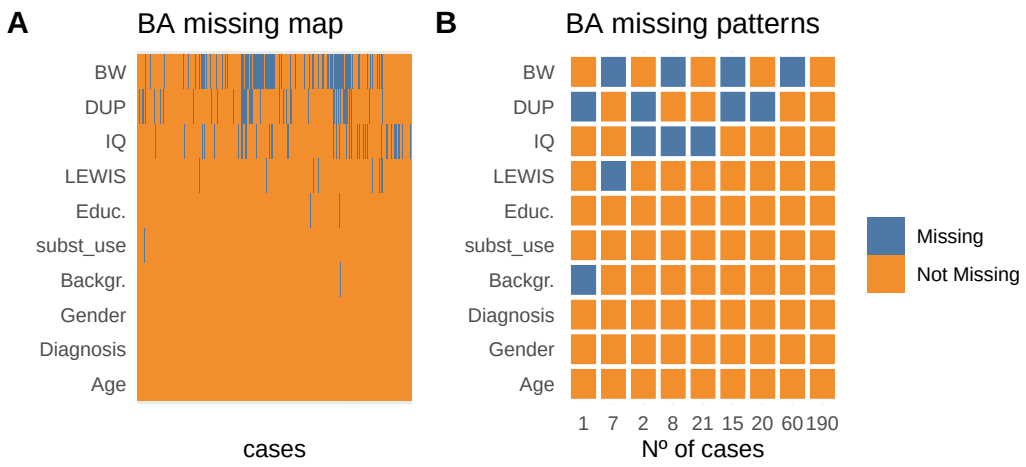


Figure 4.8: Missing values map and missing values pattern of baseline variables

Table 4.7: Comparison of missing versus not missing FAST

Variable	**Missing**, N = 290	**Not missing**, N = 1,295	**p-value**
CGI	4.00 (3.00, 5.00)	4.00 (3.00, 4.00)	0.6
GAF	60 (41, 75)	65 (55, 80)	**0.026**
PANSS	58 (38, 76)	54 (40, 73)	0.3
MADRS	6 (1, 12)	7 (2, 14)	0.3
YMRS	4.0 (0.0, 10.0)	0.0 (0.0, 5.0)	**<0.001**
WHODAS	5.5 (3.0, 9.2)	7.0 (4.0, 10.0)	0.3
FES	44 (39, 48)	44 (39, 49)	>0.9

[1] Median (IQR)
[2] Wilcoxon rank sum test

studies tends to decay along the length of the study 4.9).

4.5.0.1 Missing outcome variables and predictors

Missing data pattern from the primary outcome variable FAST reveals there is little difference between cases with or without missing FAST values. While there is a significative difference in the YMRS scale, further analysis proves that the difference is probable due to a high number of YMRS missing values. Moreover, there is a significant difference in the GAF scale (60 (41, 75) v/s 65 (55, 80), p=0.026) with the missing FAST group having a slightly lower GAF that the non-missing. This could point to the fact that the patient with better **functional** outcomes are the ones that have more missing information 4.7.

In the case of PANSS score the different in missing v/s not missing reveals there are differences in both the CGI score (2.50 (1.75, 3.00) vs 4.00 (3.00, 4.00), p=0.015) and the GAF (81 (81, 85) vs 65 (55, 80), p=0.041). Both differences show that cases with missing PANSS scores seem to have a milder symptoms 4.8, and on the other hand, better functional outcomes.

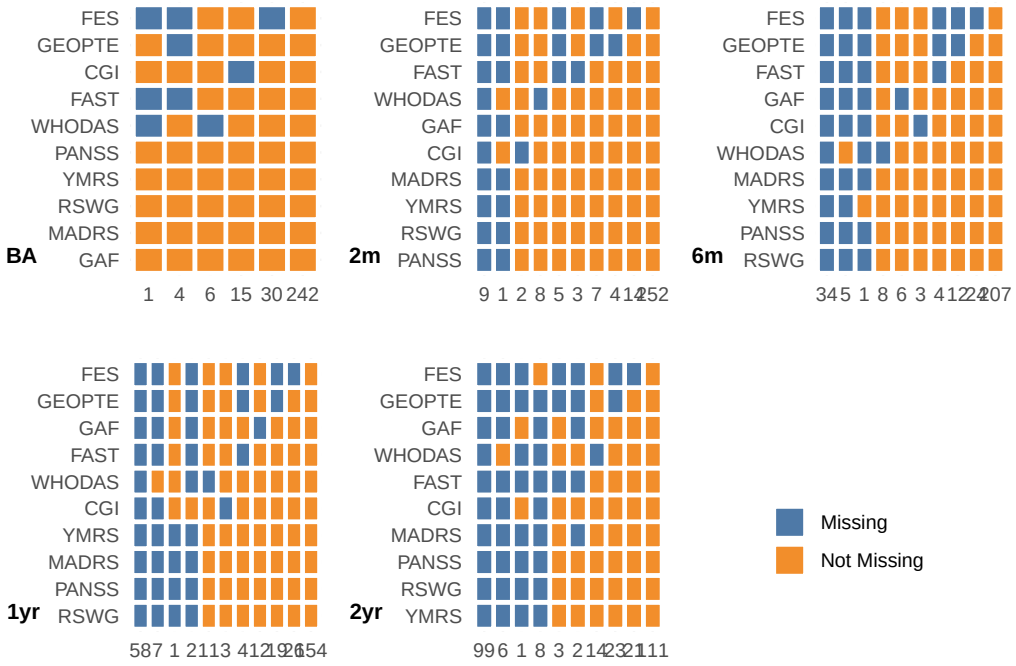


Figure 4.9: Missing values map and missing values pattern of follow up assessments

Table 4.8: Comparison of missing versus not missing PANSS

Variable	**Missing**, N = 235	**Not missing**, N = 1,350	**p-value**
CGI	2.50 (1.75, 3.00)	4.00 (3.00, 4.00)	**0.015**
GAF	81 (81, 85)	65 (55, 80)	**0.041**
FAST	18 (2, 38)	22 (10, 34)	0.8
MADRS	2 (1, 2)	7 (2, 14)	0.059
YMRS	1.5 (0.0, 4.2)	0.0 (0.0, 5.0)	>0.9
WHODAS	3.0 (3.0, 3.0)	7.0 (4.0, 10.0)	**<0.001**
FES	45 (44, 46)	44 (39, 49)	0.6

[1] Median (IQR)

[2] Wilcoxon rank sum test

Table 4.9: FAST models with different link functions.

Model	logLik	AIC	BIC
linear	-5136.17	10284.34	10306.89
beta	-4984.02	9984.03	10014.11
5q-splines	-5009.65	10041.30	10082.65
7q-splines	-4996.42	10018.83	10067.70
5e-splines	-5024.58	10071.17	10112.52
5m-splines	-4912.36	9846.73	9888.07

Table 4.10: FAST models with different number of latent classes.

G	logLik	AIC	BIC	%class1	%class2	%class3	%class4	%class5	%class6
LCMM									
2	-4907.91	9843.81	9896.44	42.27	57.73				
3	-4901.77	9837.54	9901.44	57.41	41.32	1.26			
4	-4898.45	9836.89	9912.07	48.90	45.11	4.73	1.26		
LCGA									
2	-4986.91	9995.82	10037.17	46.37	53.63				
3	-4952.37	9932.74	9985.36	44.79	12.30	42.90			
4	-4937.72	9909.44	9973.34	37.22	35.65	19.24	7.89		
5	-4916.39	9872.79	9947.96	32.49	3.47	23.66	33.44	6.94	
6	-4908.79	9863.59	9950.04	31.86	3.47	25.24	4.10	29.02	6.31

4.6 Longitudinal models

4.6.1 Primary outcome

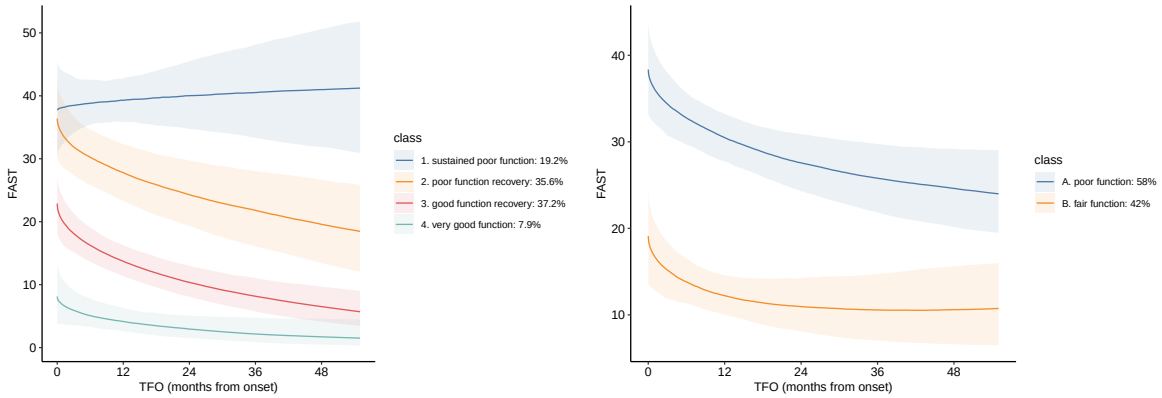
The parameters of the primary outcome, one class models built with their different functions can be seen in table 4.9. The model with the better fit uses a I-spline link function, with 5 manually set splines. The link function takes into account the non-linearity of the PANSS sub scores and allow the use of the longitudinal model algorithms. The QQ plots of both the “beta” and the “5-splines” solution show an overall fit, with a slightly better adjustment of the later (QQ plot are available in [Appendix A])

Both LCGA and LCMM many-class models where built using the optimal link function. Table 4.10 shows the different values for BIC and AIC. with regards to the LGCA model, the best solution was achieved by a 4-class model. While a 5 class models showed the lowest BIC values, some of the classes accounted for less than 5% of the cases. The LCMM best model was achieved by a 2-class model, based on the BIC and AIC criteria.

Posterior classification showed just a moderate classification power for the LCGA 4-class model. Two of the 4 classes classified less than 70% of cases with a probability of 0.7 or more. Moreover these percentages dropped considerably when the posterior probability of classification was raised 4.11. Nevertheless, the LCMM 2 class model achieved higher percentages for a posterior

Table 4.11: Posterior probability of FAST class assignment

Class prob.	class1	class2	class3	class4
LCGA				
prob>0.7	81.36	60.18	68.85	76
prob>0.8	71.19	49.56	42.62	68
prob>0.9	49.15	23.01	27.87	52
LCMM				
prob>0.7	77.61	82.51		
prob>0.8	68.66	71.58		
prob>0.9	48.51	55.19		



(a) LCGA model (b) LCMM model

Figure 4.10: FAST classes predictions along time.

probability of 0.7 for both classes.

The LCGA posterior prediction shows a four class solution with stable trajectories. Two classes show two levels of a maintained dysfunction (“1. sustained poor function” and “2. poor function recovery”) and two classes show two levels of recovery and good functional outcome (“4. very good function” and “NA”).

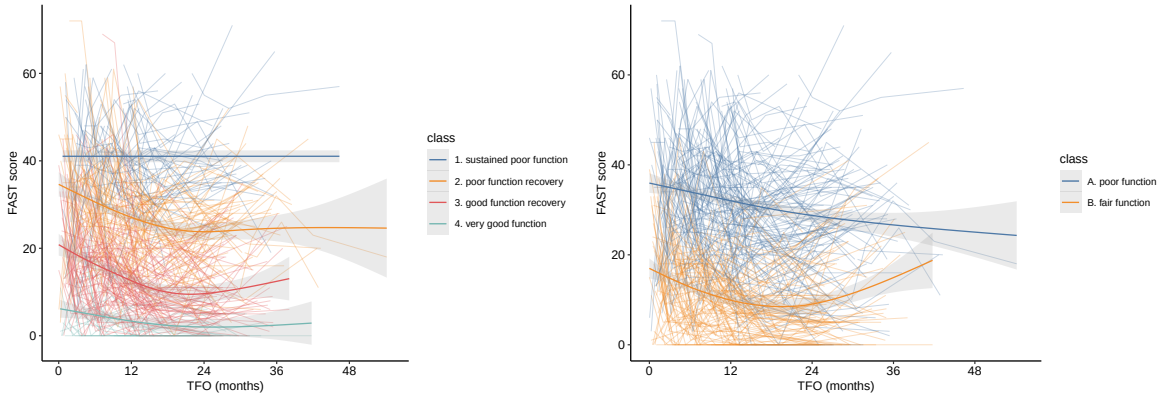
On the other hand the trajectory prediction for the LCMM FAST model, shows that the two classes have distinct trajectories; a first class called “B. fair function” that shows a moderate recovery and a sustained dysfunction group, and a second class named “A. poor function” with “58%” of cases, that shows an early recovery and overall good function (4.10).

Cross validation between the two models further reveal their complimentary nature. While the LCMM model “A. poor function” class comprises both the “1. sustained poor function” and “2. poor function recovery” LCGA classes, the “A. poor function” is mostly composed of LCGA “3. good function recovery” and “3. good function recovery” classes 4.12.

Furthermore, the visual depiction of both solutions graphically depict their complimentary status. In this case while the 2 class LCMM model seem both feasible and parsimonious, the

Table 4.12: Cross validation between LCGA and LCMM FAST classification models

	FAST(LCMM)	
Classes	A. poor function	B. fair function
FAST(LCGA)		
1. sustained poor function	61 (100%)	0 (0%)
2. poor function recovery	109 (96%)	4 (3.5%)
3. good function recovery	13 (11%)	105 (89%)
4. very good function	0 (0%)	25 (100%)



(a) LCGA classes (b) LCMM classes

Figure 4.11: FAST trajectories along time

4 class model further broadens the identification of trajectories with clinical relevance. This is specially true for the " " class that is differentiated from the "A. poor function" group; it has a distinct trajectory and it can be differentiated from the "2. poor function recovery" class (Figure 4.11). Thus, both models were kept for further analysis as they were considered to be complimentary.

4.6.2 Secondary outcomes

4.6.2.1 PANSS positive symptoms [PANSS(PS)]

For the positive symptoms from the PANSS scale the best fit was also achieved using a manually set 5 spline solution 4.13, albeit it shows just a fair adjustment as seen on the qq-plot A.

In what concerns the many class solution, while the LCGA model finds 3 classes, the LCMM model settles for 2 4.14. Posterior classification of both models shows that, even thou the PANSS LCGA model has a better fit, as for the BIC criteria, the proportion of cases classified with a probability above .8 drops markedly, while the LCMM model maintains acceptable percentages at higher classifications probabilities 4.15.

Furthermore, the posterior trajectory prediction shows that the LCMM 2-class model has one

Table 4.13: PANSS positive models with different link functions.

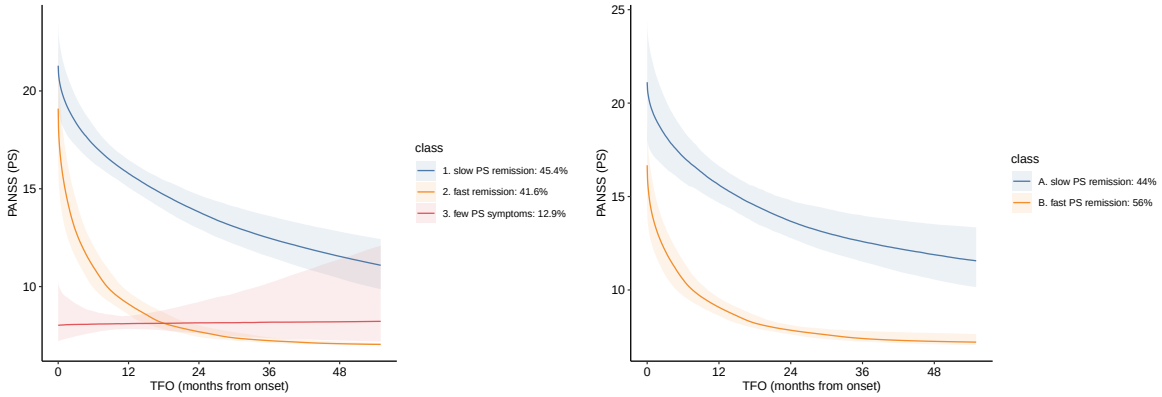
Model	logLik	AIC	BIC
linear	-4304.05	8620.10	8642.65
beta	-3494.81	7005.62	7035.70
4q-splines	-3403.53	6827.05	6864.64
5m-splines	-3386.27	6794.54	6835.89

Table 4.14: PANSS(PS) models with different number of latent classes.

G	logLik	AIC	BIC	%class1	%class2	%class3	%class4
LCGA							
2	-3390.23	6802.46	6843.81	50.79	49.21		
3	-3372.31	6772.62	6825.24	45.43	12.93	41.64	
4	-3364.11	6762.21	6826.11	31.55	33.75	23.97	10.73
LCMM							
2	-3373.59	6775.18	6827.80	55.84	44.16		
3	-3365.05	6764.10	6828.00	46.06	8.83	45.11	

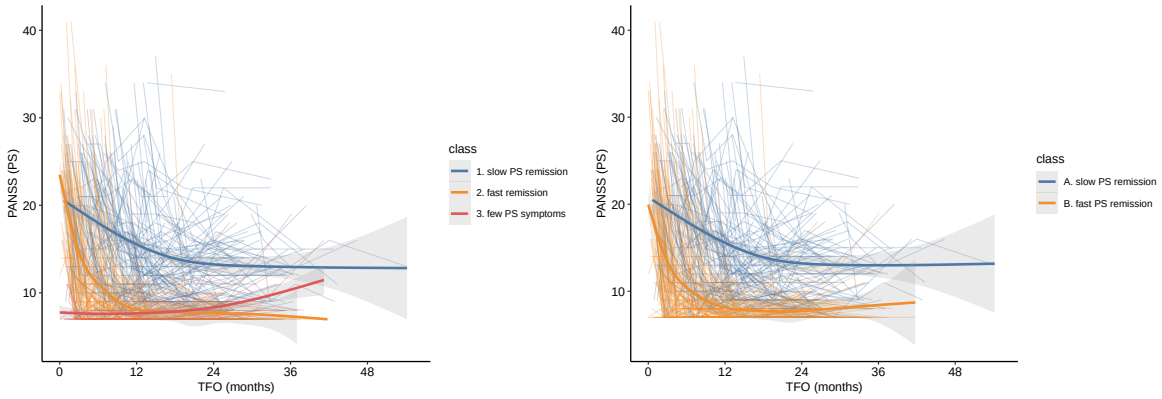
Table 4.15: Posterior probability of PANSS (PS) class assignment

Class prob.	class1	class2	class3
LCGA			
prob>0.7	86.11	78.05	72.73
prob>0.8	80.56	51.22	52.27
prob>0.9	72.22	29.27	35.61
LCMM			
prob>0.7	82.49	82.86	
prob>0.8	73.45	77.14	
prob>0.9	61.58	62.86	



(a) LCGA model (b) LCMM model

Figure 4.12: PANSS(PS) classes predictions along time form onset.



(a) LCGA classes (b) LCMM classes

Figure 4.13: PANSS positive symptom trajectories along time from onset

"B. fast PS remission" class with a rapid remission of symptoms versus a second class that falls gradually ("3. few PS symptoms"). The 3 class LCGA model constructs two analogous classes to the the LCMM model, though, this solutions identifies a third class, named "3. few PS symptoms" that has a lower starting PANSS (SP) scores, with maintained low PS values (Figure 4.12).

Moreover, the graphical depiction of the classes reveals that the "third" identified LCGA class does not appear to have a distinct pattern and seems difficult to differentiate from the others. Likewise, the clinical interpretation results cumbersome (Figure 4.13). Despite the fact that the 3 class LCGA model has a better fit according to the BIC criteria, the 2 class LCMM model has better probabilities capabilities. Moreover, the 2-class model is both clinically more feasible and parsimonious, whereas the three class model does not appear to add relevant information. Therefore, the 2-class model was the one used for further analysis.

Table 4.16: PANSS negative models with different link functions.

Model	logLik	AIC	BIC
linear	-4283.62	8579.24	8601.79
beta	-4044.62	8105.25	8135.32
5q-splines	-4021.37	8064.74	8106.09
5m-splines	-3933.63	7889.26	7930.61

Table 4.17: PANSS negative LCMM and LCGA many-class models.

G	logLik	AIC	BIC	%class1	%class2	%class3	%class4	%class5	%class6
LCGA									
2	-4031.58	8085.16	8126.51	43.22	56.78				
3	-3987.55	8003.09	8055.71	49.53	34.70	15.77			
4	-3959.05	7952.10	8016.01	38.49	48.58	4.73	8.20		
5	-3941.29	7922.58	7997.76	16.72	26.50	4.42	45.43	6.94	
6	-3919.85	7885.71	7972.16	22.71	4.42	36.28	6.94	11.04	18.61
LCMM									
2	-3929.26	7886.52	7939.14	65.62	34.38				
3	-3926.92	7887.85	7951.75	55.21	27.13	17.67			

4.6.2.2 PANSS negative symptoms [PANSS(NS)]

With regard on the negative symptoms and, just as is the case with the positive symptom scale, the best one-class model fit is achieved by a 5 manual spline solution 4.16, QQ plots in figure A.3, [Appendix A]).

In what regards the many class models, both the solutions given by the LCGA and LCMM models, provide a moderately good fit. The best LCGA many-class solution, returns a 3 class solution, as when more classes where identified, some comprise less than 5% of cases. On the other hand the LCMM model provides a 2 class solution which has a better fit (a lower BIC), is simpler and is clinically sensible 4.17.

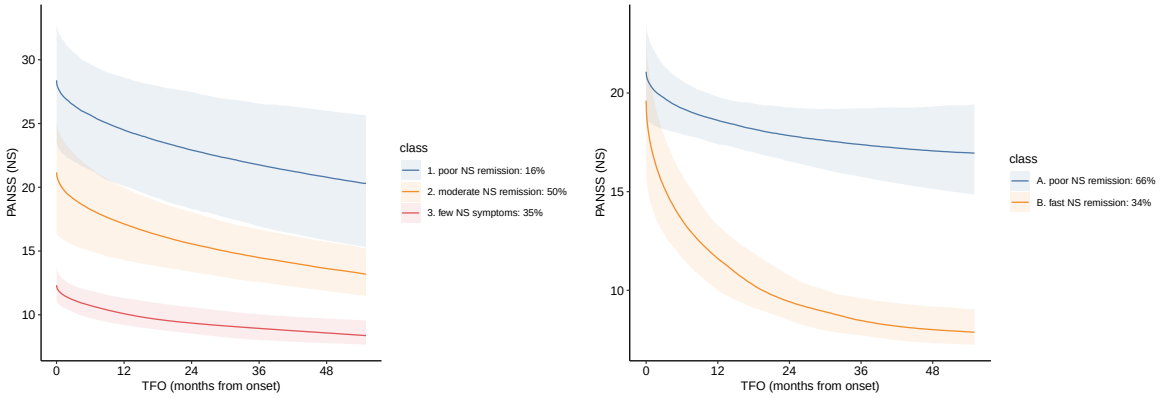
Moreover, both models show a good posterior classification ability. Nonetheless, even thou the LCGA model offers 3 classes, it has a better classification percentage at higher probability thresholds.

Class trajectory prediction further expands the results from both PANSS(NS) models. The 2 class LCMM solutions provides two somehow divergent classes, one with persistent negative symptoms and a second one with a symptom remission past the first year. On the other hand, the LCGA model, gives a complimentary solution, with 3 distinct classes (Figure 4.14).

Cross validation between the two classification models further expand their relation. While the LCGA “1. poor NS remission” class is just present in the LCMM “A. poor NS remission” LCMM class, the “2. moderate NS remission” is mainly construed by the former, with a lesser account of the “B. fast NS remission” class. In contrast, the “3. few NS symptoms” class is

Table 4.18: Posterior probability of PANSS(NS) class assignment

Class prob.	class1	class2	class3
LCGA			
prob>0.7	82.80	92.73	74
prob>0.8	71.34	86.36	62
prob>0.9	54.78	74.55	52
LCMM			
prob>0.7	80.29	71.56	
prob>0.8	67.79	58.72	
prob>0.9	49.04	37.61	

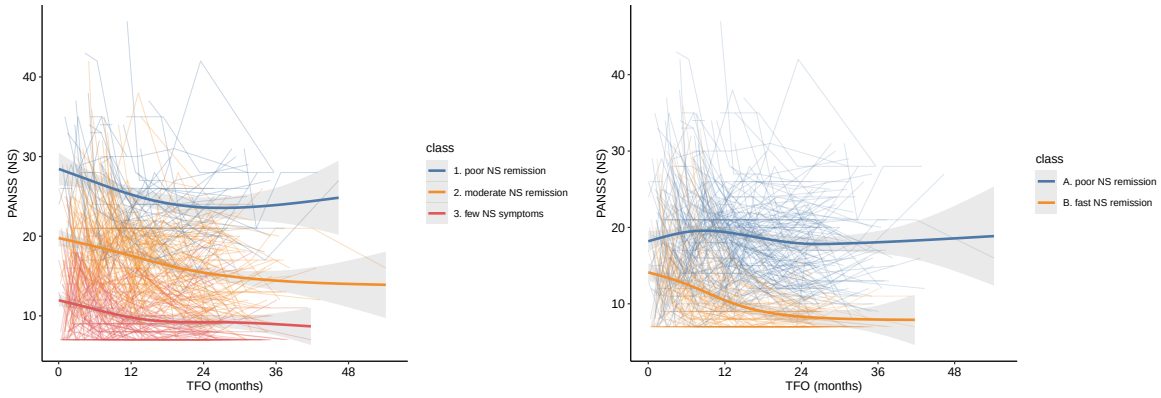


(a) LCGA model (b) LCMM model

Figure 4.14: PANSS(NS) classes predictions along time.

Table 4.19: Cross validation between LCMM and LCGA FAST models

	PANSS(NS) LCMM		
Classes	A. poor NS remission	B. fast NS remission	Total
PANSS(ns)			
1. poor NS remission	50 (100%)	0 (0%)	50 (100%)
2. moderate NS remission	134 (85%)	23 (15%)	157 (100%)
3. few NS symptoms	24 (22%)	86 (78%)	110 (100%)
Total	208 (66%)	109 (34%)	317 (100%)



(a) PANSS negative symptoms LCGA classes (b) PANSS negative symptoms LCMM classes

Figure 4.15: PANSS negative symptoms trajectories along time from onset

primary built with cases that fall into he LCMM "B. fast NS remission" class.

When the two model solutions are depicted in time, it can be seen the the 3 class LCGA model offers a good differentiation between classes (Figure 4.15). In contrast from what is seen in the PANSS(PS) models, the PANSS(NS) LCGA seems to provide a sensible and accurate depiction that is both a statistically sound and clinically feasible. Therefore, the LCGA 3 class model was kept and used for furether analysis.

4.6.2.3 PANSS general symptoms [PANSS(GS)]

In the same line as the previous one class analysis of the PANSS sub-scales, the best model is offered by a 5 manual spline model 4.20, QQ plots in figure A.

The many class solution offered a very similar pattern as the one seen for the PANSS(NS) case. Both the LCMM and LCGA serves a moderately good fit for a 2 and 3 class model respectively 4.21. The 3 class LCGA solution has a slightly lower BIC, and predicts trajectories that seem feasible.

Moreover, the LCGA 3-class model displays a better posterior probability classification that the one of the 2 class LCMM model, with higher percentage of classifications with a probability

Table 4.20: PANSS general models with different link functions.

Model	logLik	AIC	BIC
linear	-4953.88	9919.76	9942.31
beta	-4629.45	9274.91	9304.98
5q-splines	-4655.17	9332.33	9373.68
5m-splines	-4600.09	9222.18	9263.53

Table 4.21: PANSS General symptoms many-class models.

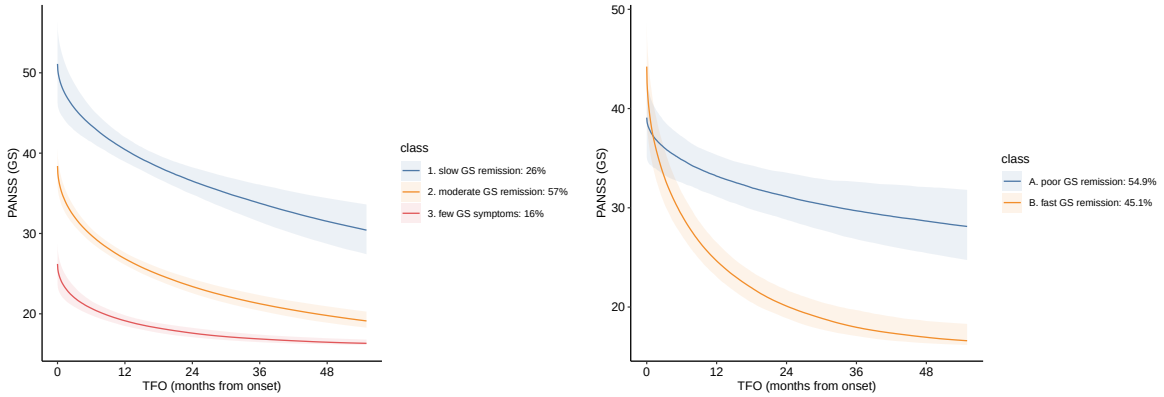
G	logLik	AIC	BIC	%class1	%class2	%class3	%class4
LCGA							
2	-4674.08	9370.16	9411.51	58.36	41.64		
3	-4614.21	9256.43	9309.05	57.41	16.09	26.50	
4	-4606.09	9246.18	9310.09	7.57	15.77	26.18	50.47
LCMM							
2	-4621.74	9271.48	9324.11	45.11	54.89		
3	-4621.74	9277.48	9341.38	0.00	56.15	43.85	

threshold of 0.7 or more.

In a similar fashion, the posterior prediction of the class trajectories for the 2 class model builds 2 divergent classes. A first class with a fast remission rate, and a second with a slower progression to remission. In contrast, the 3 class model distinguishes 3 different trajectories. A first class of maintained general symptoms, an intermediate class with moderate remission and a third group of less general symptoms and a faster remission rate.

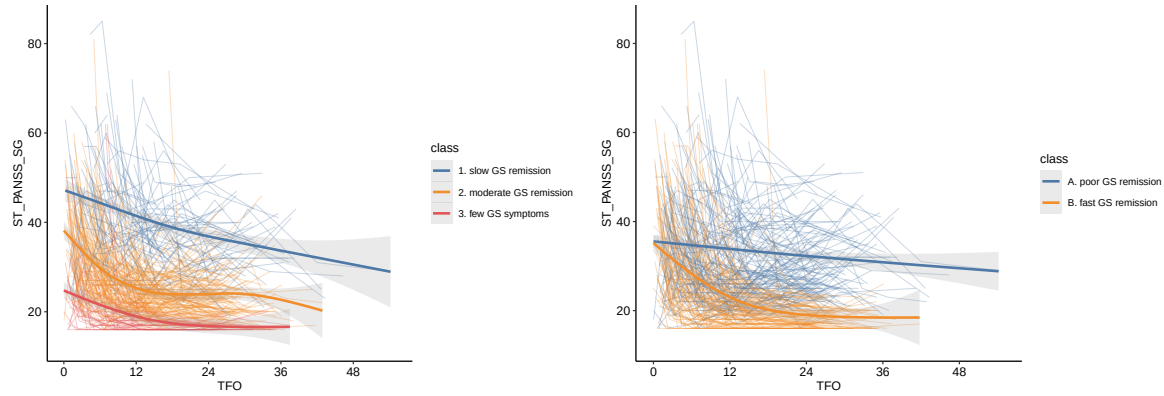
Table 4.22: Posterior probability of PANSS(GS) class assignment

Class prob.	class1	class2	class3
LCGA			
prob>0.7	91.76	90.20	86.90
prob>0.8	85.16	74.51	83.33
prob>0.9	68.13	64.71	71.43
LCMM			
prob>0.7	62.94	74.71	
prob>0.8	53.15	56.32	
prob>0.9	38.46	40.23	



(a) LCMM model (b) LCGA model (without confidence intervals)

Figure 4.16: PANSS(GS) classes predictions along time.



4.7 Primary and secondary class relationships

A first analysis of the relation between the primary and secondary classes reveals a close relation of the two domains. The cross tabulation of the functional and symptomatological classes affirms the relations between the two domains. As expected, there is a strong association between the poorer symptomatic trajectories and a poor functional outcome for all of the PANSS sub-scale, and specially in what regards the negative and general symptom scales. As table 4.23 shows, the "1. sustained poor function" class is mainly composed of "slow remission" symptons classes, especilly in what regards to the PANSS(NS) where up to 95% of the poor functional outcome belong to the "1. poor NS remission" or "1. poor NS remission" PANSS(NS) groups. In contrast, for the "4. very good function" class, both PANSS(PS) and PANSS(NS) models classify more than 90% of the cases as "B. fast PS remission" or "3. few NS symptoms"

In what regards the relation between the PANSS sub-scores classes identified, there is a strong relation between the negative and positive symptom classes, specially between the "3. few NS symptoms" class and the "A. slow PS remission" group.

Table 4.23: Cross validation between PANSS and FAST models

	FAST(LCGA)			
	1. sustained poor function	2. poor function recovery	3. good function recovery	4. very good function
PANSS(ps)				
A. slow PS remission	47 (34%)	59 (42%)	33 (24%)	1 (0.7%)
B. fast PS remission	14 (7.9%)	54 (31%)	85 (48%)	24 (14%)
PANSS(ns)				
1. poor NS remission	26 (52%)	18 (36%)	6 (12%)	0 (0%)
2. moderate NS remission	32 (20%)	78 (50%)	45 (29%)	2 (1.3%)
3. few NS symptoms	3 (2.7%)	17 (15%)	67 (61%)	23 (21%)
PANSS(gs)				
1. slow GS remission	34 (40%)	40 (48%)	10 (12%)	0 (0%)
2. moderate GS remission	26 (14%)	71 (39%)	77 (42%)	8 (4.4%)
3. few GS symptoms	1 (2.0%)	2 (3.9%)	31 (61%)	17 (33%)

[a] Row-wise perecentages

Table 4.24: Cross validation between PANSS negative symptoms and positive subscores

	PANSS(ps)		
	A. slow PS remission	B. fast PS remission	p-value
PANSS(ns)			<0.001
1. poor NS remission	38 (76%)	12 (24%)	
2. moderate NS remission	87 (55%)	70 (45%)	
3. few NS symptoms	15 (14%)	95 (86%)	

[1] Pearson's Chi-squared test
[a] Row-wise perecentages

The relation within PANSS sub scores can also seen within the negative and general classes, mostly within the negative "1. poor NS remission" class and both the "1. slow GS remission" "2. moderate GS remission" general symptoms classes

4.8 Baseline characterization of modelled classes

For the class characterization, the imputed data was used. Firstly, a visual check of the imputed datasets was made, in order to establish deviation of the imputed data from the original data dsitribution. Figure 4.17 shows a close fit of the imputed data to the original data.

Table 4.25: Cross validation between PANSS negative symptoms and general subscores

	PANSS(gs)			
	1. slow GS remission	2. moderate GS remission	3. few GS symptoms	p-value
PANSS(ns)				<0.001
1. poor NS remission	31 (62%)	19 (38%)	0 (0%)	
2. moderate NS remission	52 (33%)	101 (64%)	4 (2.5%)	
3. few NS symptoms	1 (0.9%)	62 (56%)	47 (43%)	

[1] Pearson's Chi-squared test
[a] Row-wise perecentages

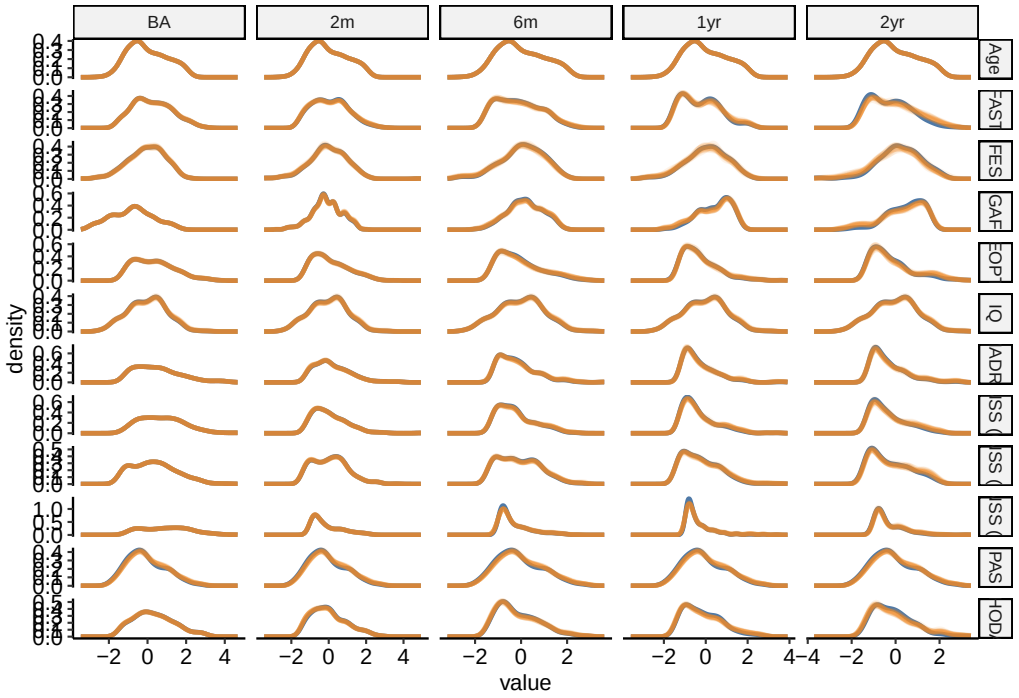


Figure 4.17: Density plot of original and imputed datasets

4.8.1 Primary outcome (FAST score)

The univariate analysis of the LCMM primary outcome classification acknowledges differences in many of the baseline variables along the two classes. Firstly, in what regards to prior adjustment, the "B. fair function" class has a lower level of education; the proportion of primary education is higher in this group, with regards to Secondary and higher education number on the "A. poor function" group (corr.p<0.001).

Furthermore, there is a mean 10 point difference between both classes IQ (90 (80, 100) v/s 100 (85, 105), corr.p=0.001.

The pre-morbid adjustment scale for the "B. fair function" also showed a significant difference, with a mean score of 52 (39, 72) for the poor outcome group, and 33 (22, 48) for the "A. poor function" class (corr.p=<0.001. All of WHODAS, GAF and the GEOPTE scale also displayed significant differences between both groups.

Regarding symptom scales, all PANSS sub scales where higher in the poor outcome group. as well as baseline CGI

Strikingly, the MADRS scale revealed a strong and significant difference between groups, with a value as high as 15 (8, 21) for the "B. fair function" class, while the A. poor function was much lower (7 (3, 13), corr.p<0.001). One other finding is that there is a significant difference between the days of untreated psychosis (DUP) between the two FAST classes (91 (31, 175) v/s 37 (13, 106), corr.p<0.001). Although there were more affective disorders as final diagnosis

Table 4.26: Univariate differences of socio-demographic and baseline functional and symptomatic assessment variabels within FAST LCMM classes

Variable	**A. poor function**, N = 183 (58%)	**B. fair function**, N = 134 (42%)	**pooled p-value**	**pooled sd**	**corrected p-value**
Age	23.0 (20.0, 28.0)	22.0 (18.0, 28.0)	0.5	0.00	0.5
Woman	67 (37%)	36 (27%)	0.070	0.00	0.089
Education			<0.001	0.00	**<0.001**
Primary	53 (29%)	22.213 (17%)			
Secondary	116 (63%)	74.61 (56%)			
University studies	14 (7.7%)	36.176 (27%)			
Other	0 (0%)	1 (0.7%)			
Diagnosis			0.12	0.01	0.14
Toxic psychosis	7 (3.8%)	7 (5.2%)			
Schizophrenia disorders	157 (86%)	103 (77%)			
Affective disorders	19 (10%)	24 (18%)			
IQ	90 (80, 100)	100 (85, 105)	<0.001	0.00	**0.001**
DUP	91 (31, 175)	37 (13, 106)	<0.001	0.00	**<0.001**
Subst. abuse	142.801 (78%)	105 (78%)	>0.9	0.04	>0.9
PAS	52 (39, 72)	33 (22, 48)	<0.001	0.00	**<0.001**
FES	43 (36, 48)	44 (39, 48)	0.2	0.08	0.2
CGI	5.00 (4.00, 5.00)	4.00 (4.00, 5.00)	0.002	0.00	**0.002**
PANSS (PS)	20 (14, 25)	16 (9, 23)	<0.001	0.00	**<0.001**
PANSS (NS)	21 (15, 26)	14 (9, 19)	<0.001	0.00	**<0.001**
PANSS (GS)	41 (32, 48)	30 (23, 41)	<0.001	0.00	**<0.001**
MADRS	15 (8, 21)	7 (3, 13)	<0.001	0.00	**<0.001**
YMRS	8 (0, 16)	4 (0, 14)	0.031	0.00	**0.042**
GAF	50 (31, 60)	60 (45, 70)	<0.001	0.00	**<0.001**
FAST	36 (26, 45)	16 (10, 22)	<0.001	0.00	**<0.001**
WHODAS	10.0 (7.0, 14.0)	5.0 (3.0, 8.0)	<0.001	0.00	**<0.001**
GEOPTE	33 (27, 42)	23 (19, 33)	<0.001	0.00	**<0.001**

[1] Median (IQR); n (%)
[2] Wilcoxon signed-rank test; Fisher's exact test; Pooled values from imputed datasets
[3] False discovery rate correction for multiple testing

in the "A. poor function" group, the differences were not significant. There were no differences between substance abuse, nor within sex or Age 4.26.

The univariate analysis using the LCGA FAST 4 classes maintain a similar pattern of differences between the four groups 4.27.

Further pairwise analysis gives a comprehensive view. Firstly the differences in "Education" is significant between the "4. very good function" and both the "2. poor function recovery" and "1. sustained poor function" classes (p=0.0972 & p=0.374, respectively, Figure 4.18)

Whereas significant differences between IQ where only seen between the "4. very good function" and "1. sustained poor function" classes, DUP was significantly lower for the "1. sustained poor function" class, in contrast with the "3. good function recovery" and "4. very good function" classes, and between "3. good function recovery" class and the "1. sustained poor function" class. In what regards the PAS score, significant differences where seen in all class comparisons; As classes expressed a worst functional outcome, the PAS score showed a poorer adjustment (Figure 4.19).

In what regards the symptoms scales, the study entry PANSS positive symptom score indicate significant higher figures for the "1. sustained poor function" class than both "3. good function recovery" and "4. very good function" classes. Similarly to what is seen with the PAS score, the PANSS negative symptom scale illustrate significant differences in all comparisons, with

Table 4.27: Univariate differences of socio-demographic and baseline functional and symptomatic assessment variabels within FAST LCGA classes

Variable	1. sustained poor function, N = 61 (19%)	2. poor function recovery, N = 113 (36%)	3. good function recovery, N = 118 (37%)	4. very good function, N = 25 (7.9%)	pooled p-value	pooled sd	corrected p-value
Age	23.0 (20.0, 30.0)	22.0 (19.0, 27.0)	21.5 (18.0, 28.0)	23.0 (20.0, 29.0)	0.2	0.00	0.2
Woman	22 (36%)	39 (35%)	31 (26%)	11 (44%)	0.2	0.01	0.2
Education					0.003	0.00	**0.006**
Primary	19 (31%)	30 (27%)	24.213 (21%)	2 (8.0%)			
Secondary	38 (62%)	72 (64%)	66.61 (56%)	14 (56%)			
University studies	4 (6.6%)	11 (9.7%)	26.176 (22%)	9 (36%)			
Other	0 (0%)	0 (0%)	1 (0.8%)	0 (0%)			
Diagnosis					0.2	0.01	0.2
Toxic psychosis	2 (3.3%)	4 (3.5%)	7 (5.9%)	1 (4.0%)			
Schizophrenia disorders	56 (92%)	94 (83%)	90 (76%)	20 (80%)			
Affective disorders	3 (4.9%)	15 (13%)	21 (18%)	4 (16%)			
IQ	90 (80, 100)	90 (80, 100)	95 (80, 105)	100 (90, 105)	0.014	0.01	**0.023**
DUP	92 (42, 264)	78 (30, 150)	50 (15, 127)	27 (7, 75)	<0.001	0.00	**0.001**
Subst. abuse	43 (70%)	94.801 (84%)	90 (76%)	20 (80%)	0.2	0.03	0.2
PAS	63 (45, 79)	48 (35, 66)	37 (25, 53)	19 (12, 36)	<0.001	0.00	**<0.001**
FES	42 (34, 49)	43 (38, 48)	43 (38, 48)	46 (41, 48)	0.3	0.07	0.3
CGI	5.00 (4.00, 5.00)	5.00 (4.00, 5.00)	4.00 (4.00, 5.00)	4.00 (3.00, 5.00)	0.019	0.01	**0.028**
PANSS (PS)	21 (17, 26)	19 (14, 25)	17 (10, 24)	10 (7, 20)	<0.001	0.00	**0.001**
PANSS (NS)	23 (18, 29)	20 (15, 25)	16 (10, 21)	8 (7, 16)	<0.001	0.00	**<0.001**
PANSS (GS)	43 (37, 51)	40 (32, 48)	32 (24, 43)	25 (20, 32)	<0.001	0.00	**<0.001**
MADRS	15 (9, 21)	15 (8, 23)	8 (4, 15)	2 (0, 8)	<0.001	0.00	**<0.001**
YMRS	7 (0, 14)	9 (0, 17)	4 (0, 14)	0 (0, 14)	0.2	0.00	0.2
GAF	50 (35, 55)	50 (30, 60)	55 (41, 70)	70 (50, 85)	<0.001	0.00	**<0.001**
FAST	40 (33, 50)	32 (23, 43)	17 (12, 26)	3 (0, 12)	<0.001	0.00	**<0.001**
WHODAS	13.0 (8.0, 15.0)	10.0 (7.0, 12.0)	6.0 (4.0, 9.0)	2.0 (1.0, 5.0)	<0.001	0.00	**<0.001**
GEOPTE	39 (30, 46)	32 (26, 39)	24 (20, 33)	20 (17, 28)	<0.001	0.00	**<0.001**

[1] Median (IQR); n (%)
[2] Kruskal-Wallis test; Fisher's exact test; Pooled values from imputed datasets
[3] False discovery rate correction for multiple testing

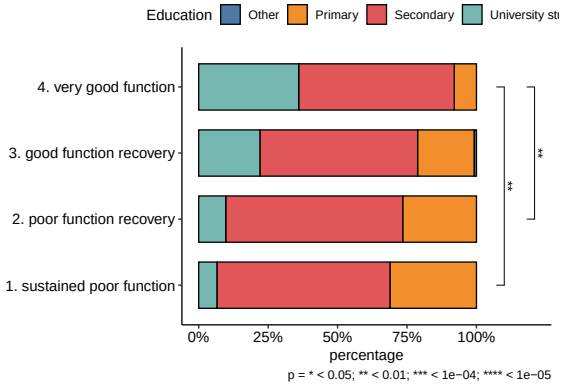


Figure 4.18: Pairwise differences of Educational level within the 3 class LCGA FAST model.

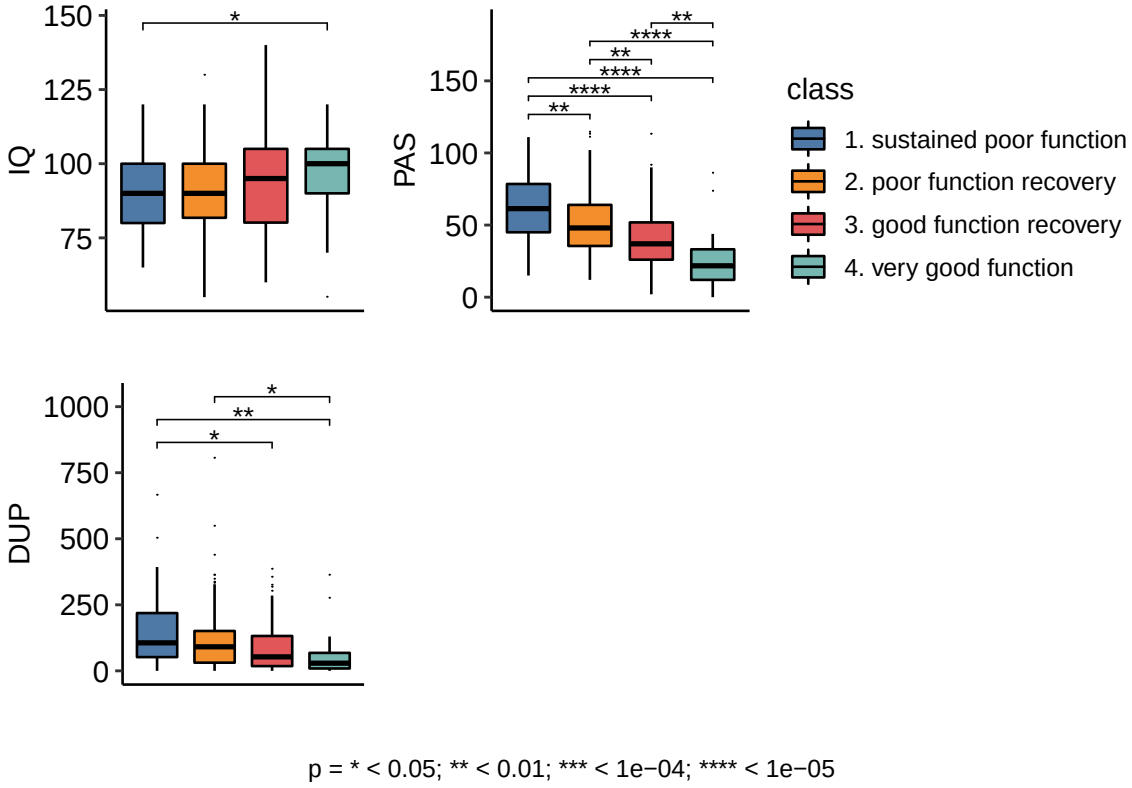


Figure 4.19: Pairwise differences between the 3 class LCGA FAST model, for the premorbid IQ, the days of untreated psychosis (DUP) and the premorbid adjustment scale (PAS)

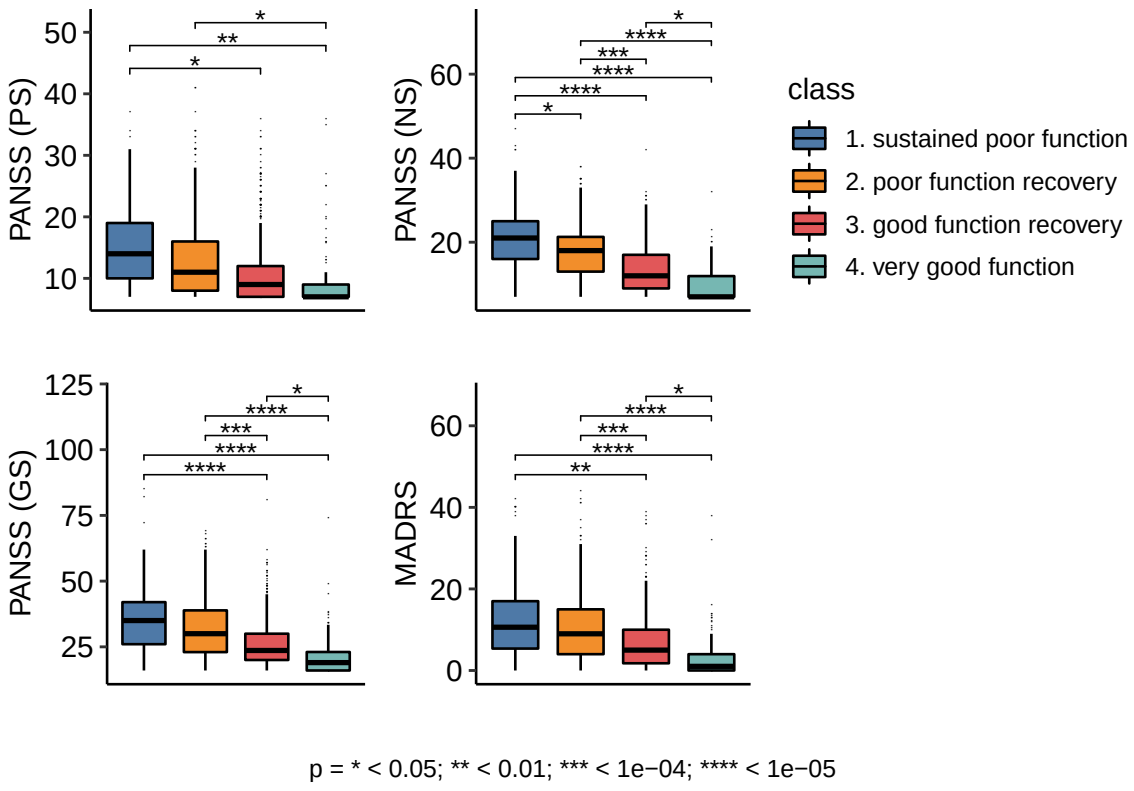
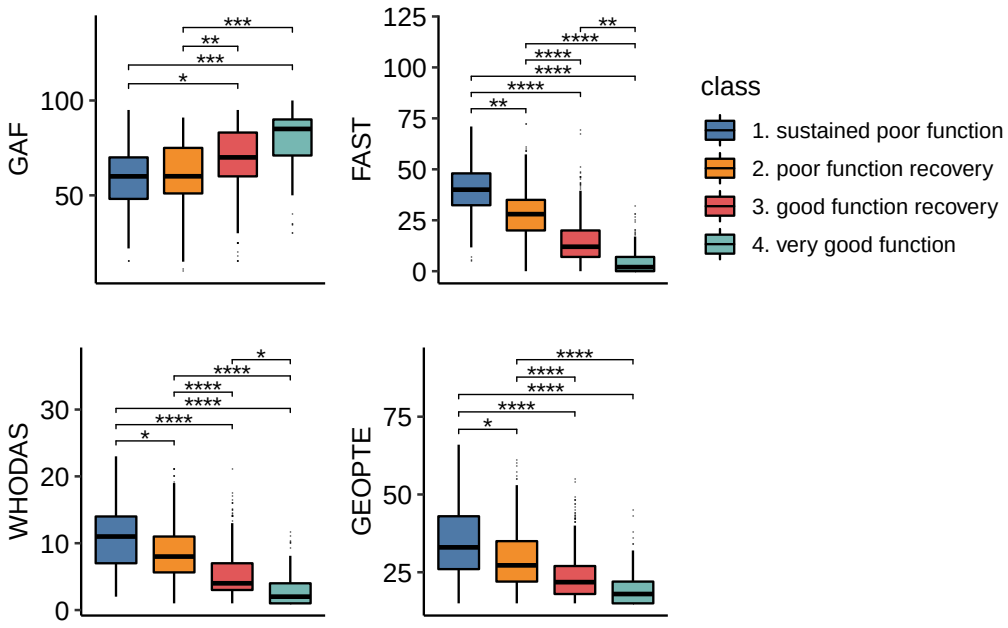


Figure 4.20: Pairwise differences of the 5 class LCGA FAST model, of baseline PANSS positive (PS), negative (NS), general (GS) , and the Montgomery–Åsberg Depression Rating Scale (MADRS).

higher symptom in the more impaired functional classes.

For the functional and social cognition scales, significant differences where found between the "4. very good function" class and all other classes for the FAST, WHODAS and GEOPTE. Moreover the FAST and GeOPTE scores was significantly different between all classes, as well as for the GEOPTE scale which only lacked a difference between the "4. very good function" and "3. good function recovery" class.



4.8.1.1 Secondary outcome (PANSS subscores)

In what regards to the PANSS(PS) classes, there where significant differences in the educational level, with a marked lower educational level for the "1. slow PS remission" class. Distinctly, there was a difference in what regard to diagnosticgroup, with a higher proportion of affective disorders in the "2. fast remission" group. Further differences were seen in all of IQ, DUP, the PAS score, the CGI at study entry, all three PANSS subscales, the MADRS and YMRS, with significantly more impaired values for the "1. slow PS remission" class. Additional, all functional scale were significantly different, with a better functional baseline assessment for the "2. fast remission" class, with lower scores for the FAST, GAF, WHODAS and in the social cognition scale (GEOPTE).

For the negative symptom PANSS score, the three classes displayed a similar pattern. For instance, a lower educational level was associated to the more impaired negative symptom classes. Additional, there was a slight but significant difference in the proportion of affective disordered within the "2. moderate NS remission" and specially the "1. poor NS remission" class Premorbid IQ levels showed a marked relation with the PANSS(NS) classes, as the "3. few NS symptoms" presented 10 more points in than the "2. moderate NS remission" which had an average of 90 (80, 100). Likewise, the value dropped 5 points for the "1. poor NS remission" group. Similarly, the DUP augmented and the PAS score was lower the more impaired the class. For all of the baseline PANSS sub-scores, the MADRS and CGI there was a declining trend the better level of remission of the class. The same was true for All of the funcional and social cognition score 4.29.

Pairwise analysis between the 3 PANSS(NS) classes for Educational level and Diagnostic group,

Table 4.28: Univariate differences of socio-demographic and baseline functional and symptomatic assessment variabels within PANSS positive symptoms LCMM classes

Variable	**A. slow PS remission**, N = 140 (44%)	**B. fast PS remission**, N = 177 (56%)	**pooled p-value**	**pooled sd**	**corrected p-value**
Age	22.0 (19.0, 27.0)	23.0 (19.0, 29.0)	0.12	0.00	0.14
Woman	39 (28%)	64 (36%)	0.15	0.00	0.2
Education			0.001	0.00	**0.003**
Primary	44.213 (32%)	31 (18%)			
Secondary	83.61 (60%)	107 (60%)			
University studies	12.176 (8.7%)	38 (21%)			
Other	0 (0%)	1 (0.6%)			
Diagnosis			0.027	0.00	**0.037**
Toxic psychosis	7 (5.0%)	7 (4.0%)			
Schizophrenia disorders	122 (87%)	138 (78%)			
Affective disorders	11 (7.9%)	32 (18%)			
IQ	90 (80, 100)	95 (85, 105)	0.003	0.00	**0.004**
DUP	123 (52, 243)	35 (14, 90)	<0.001	0.00	**<0.001**
Subst. abuse	103.801 (74%)	144 (81%)	0.13	0.01	0.14
PAS	55 (38, 74)	37 (24, 52)	<0.001	0.00	**<0.001**
FES	42 (37, 49)	44 (38, 48)	0.8	0.15	0.8
CGI	5.00 (4.00, 5.00)	4.00 (4.00, 5.00)	0.003	0.00	**0.005**
PANSS (PS)	21 (17, 26)	16 (9, 23)	<0.001	0.00	**<0.001**
PANSS (NS)	21 (15, 25)	16 (9, 22)	<0.001	0.00	**<0.001**
PANSS (GS)	41 (34, 50)	32 (25, 43)	<0.001	0.00	**<0.001**
MADRS	14 (8, 21)	10 (4, 17)	0.001	0.00	**0.003**
YMRS	9 (1, 15)	4 (0, 14)	0.023	0.00	**0.034**
GAF	50 (40, 60)	55 (35, 70)	0.033	0.00	**0.042**
FAST	32 (18, 42)	22 (13, 35)	<0.001	0.00	**<0.001**
WHODAS	10.0 (7.0, 13.0)	7.0 (4.0, 11.0)	<0.001	0.00	**<0.001**
GEOPTE	33 (26, 42)	26 (20, 34)	<0.001	0.00	**<0.001**

[1] Median (IQR); n (%)

[2] Wilcoxon signed-rank test; Fisher's exact test; Pooled values from imputed datasets

[3] False discovery rate correction for multiple testing

Table 4.29: Univariate differences of socio-demographic and baseline functional and symptomatic assessment variabels within PANSS negative symptoms LCGA classes

Variable	**1. poor NS remission**, N = 50 (16%)	**2. moderate NS remission**, N = 157 (50%)	**3. few NS symptoms**, N = 110 (35%)	pooled p-value	pooled sd	corrected p-value
Age	22.5 (17.0, 28.0)	23.0 (19.0, 28.0)	22.0 (19.0, 28.0)	>0.9	0.00	>0.9
Woman	14 (28%)	51 (32%)	38 (35%)	0.7	0.01	0.8
Education				<0.001	0.00	**0.001**
Primary	17 (34%)	39.213 (25%)	19 (17%)			
Secondary	30 (60%)	101.61 (65%)	59 (54%)			
University studies	3 (6.0%)	16.176 (10%)	31 (28%)			
Other	0 (0%)	0 (0%)	1 (0.9%)			
Diagnosis				0.004	0.00	**0.006**
Toxic psychosis	1 (2.0%)	11 (7.0%)	2 (1.8%)			
Schizophrenia disorders	47 (94%)	129 (82%)	84 (76%)			
Affective disorders	2 (4.0%)	17 (11%)	24 (22%)			
IQ	85 (75, 95)	90 (80, 100)	100 (85, 110)	<0.001	0.00	**<0.001**
DUP	125 (34, 246)	92 (39, 180)	30 (9, 55)	<0.001	0.00	**<0.001**
Subst. abuse	32.801 (66%)	127 (81%)	88 (80%)	0.077	0.02	0.10
PAS	69 (47, 87)	47 (34, 66)	33 (21, 44)	<0.001	0.00	**<0.001**
FES	42 (38, 48)	43 (37, 48)	44 (38, 48)	0.6	0.15	0.6
CGI	5.00 (4.00, 5.00)	5.00 (4.00, 5.00)	4.00 (4.00, 5.00)	0.017	0.01	**0.023**
PANSS (PS)	20 (14, 26)	20 (14, 25)	15 (9, 23)	0.001	0.00	**0.002**
PANSS (NS)	29 (24, 34)	20 (16, 24)	11 (8, 16)	<0.001	0.00	**<0.001**
PANSS (GS)	44 (37, 52)	39 (32, 47)	28 (22, 39)	<0.001	0.00	**<0.001**
MADRS	16 (10, 23)	13 (8, 19)	7 (2, 14)	<0.001	0.00	**<0.001**
YMRS	4 (0, 13)	7 (0, 15)	4 (0, 15)	0.4	0.00	0.5
GAF	41 (30, 55)	50 (35, 60)	60 (40, 70)	<0.001	0.00	**<0.001**
FAST	39 (20, 48)	32 (19, 40)	17 (11, 28)	<0.001	0.00	**<0.001**
WHODAS	12.0 (7.0, 15.0)	9.0 (6.0, 12.0)	6.0 (3.0, 10.0)	<0.001	0.00	**<0.001**
GEOPTE	38 (30, 45)	32 (26, 39)	22 (19, 33)	<0.001	0.00	**<0.001**

[1] Median (IQR); n (%)
[2] Kruskal-Wallis test; Fisher's exact test; Pooled values from imputed datasets
[3] False discovery rate correction for multiple testing

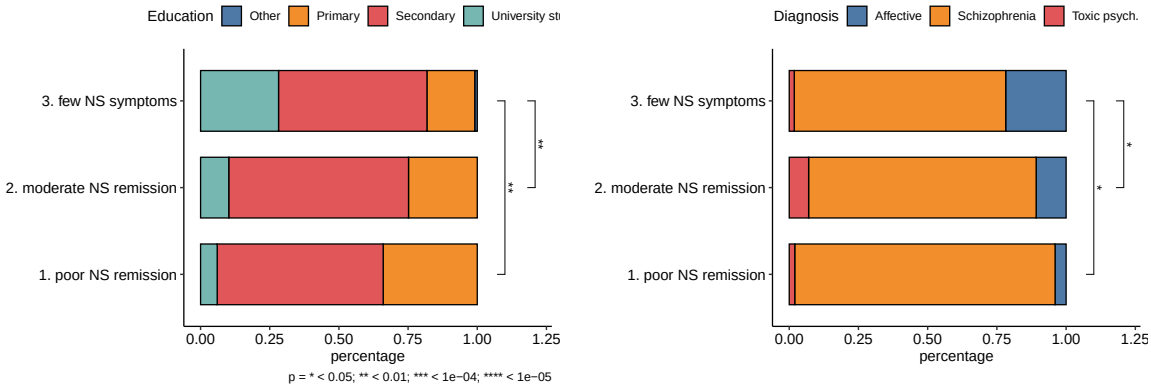
kept significant differences between the "3. few NS symptoms" class and both the "1. poor NS remission" and "2. moderate NS remission" classes (Figure 4.21).

In what regards with the IQ and the PAS score, the differences expressed in table 4.29 where significant for all class comparisons. However, for DUP differences were only significant between the "3. few NS symptoms" class and the other two.

In line with previous results, the symptomatic baseline scores reflected de differences within the negative symptoms classes. While both the negative and the general PANSS sub scores maintained significant differences along all comparisons between the three classes, the positive symptoms score differences were only significant between the "3. few NS symptoms" and the other two. Likewise, the baseline MADRS score was significantly higher in both the "1. poor NS remission" and "2. moderate NS remission" groups is contrast to the "3. few NS symptoms" class.

Finally, the baseline assessment of the CGI was only significantly different between the "1. poor NS remission" and "3. few NS symptoms" classes

The functional baseline profile also maintained significative differences within the class comparisons. Both the WHODAS and GEOPTE social cognition score displayed differences between all classes, while both the GAF and FAST saw differences between the "3. few NS symptoms" and both the "1. poor NS remission" and "2. moderate NS remission" groups.



(a) Education (b) Diagnosis

Figure 4.21: Pairwise differences between the 3 class LCGA PANSS negative symptoms model.

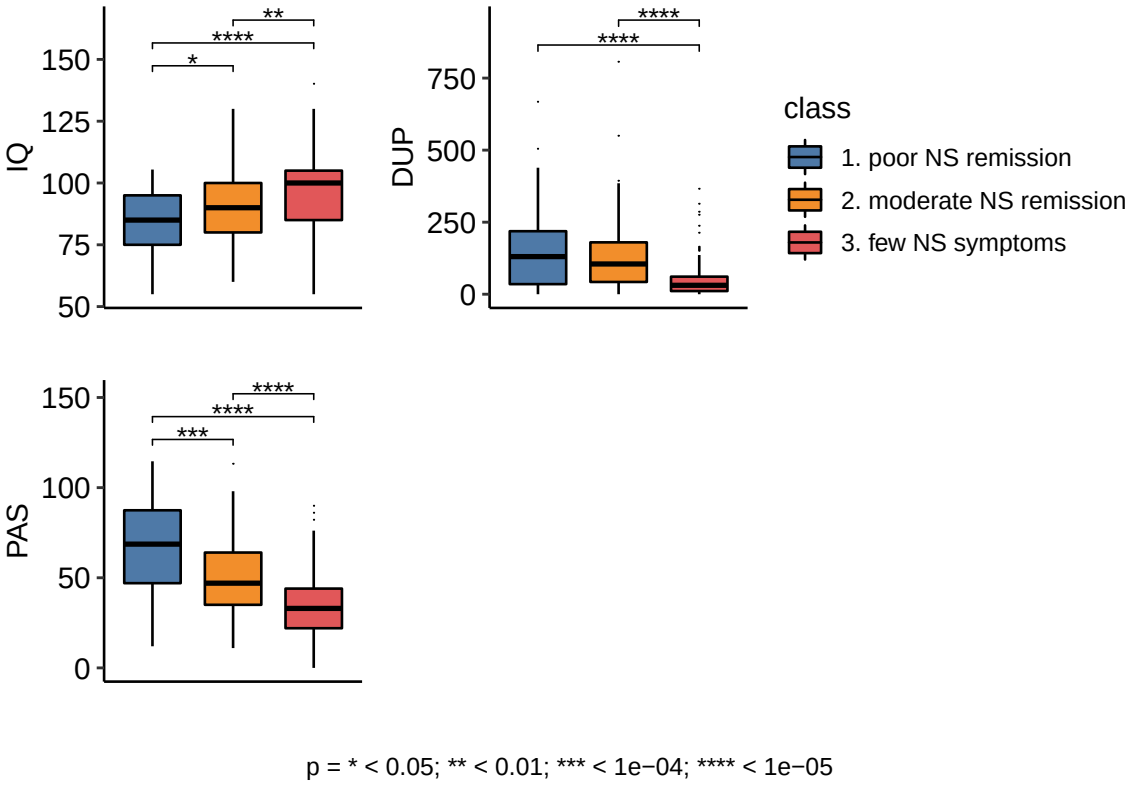


Figure 4.22: Pairwise differences between the 3 class LCGA PANSS negative symptoms model, for the premorbid IQ, the days of untreated psychosis (DUP) and the premorbid adjustment scale (PAS).

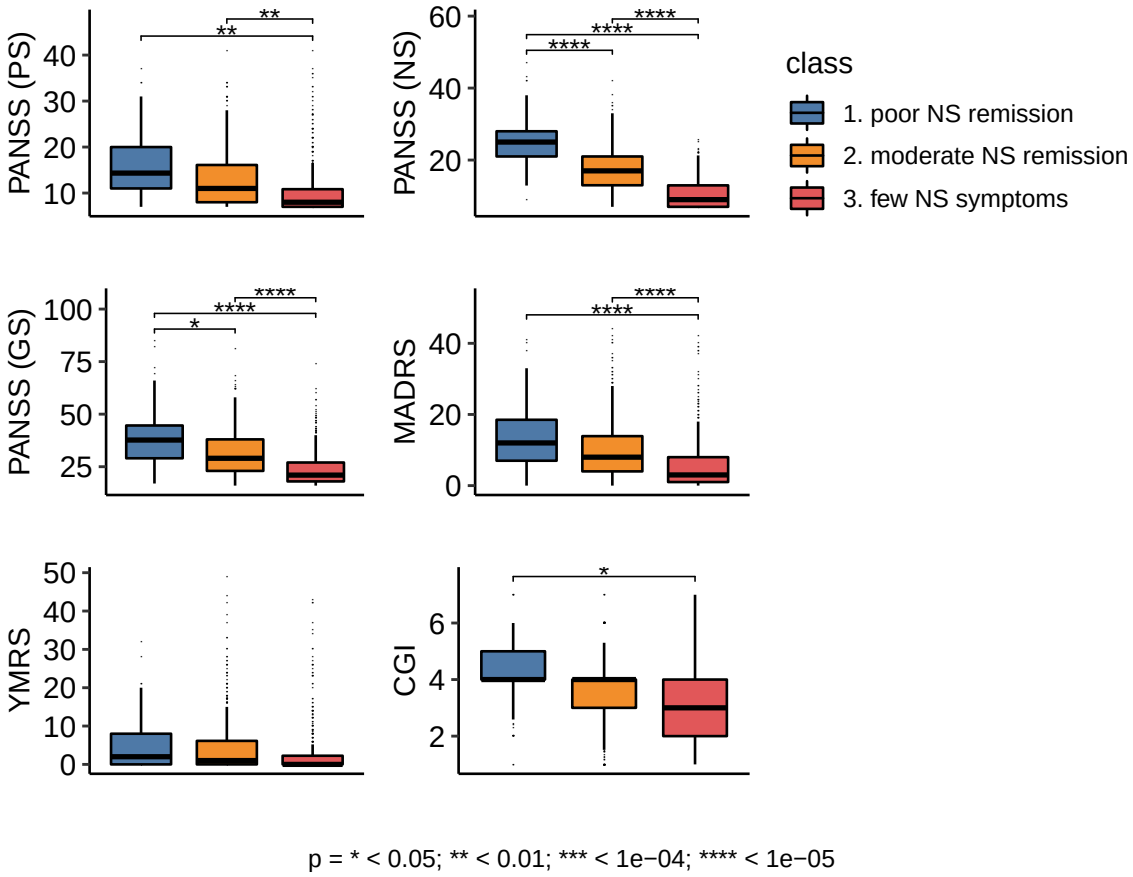


Figure 4.23: Pairwise differences between the 3 class LCGA PANSS negative symptoms model for the baseline PANSS positive (PS), negative (NS) and general (GS) symptom scale, and the Montgomery-Asberg Depression Rating scale (MADRS)

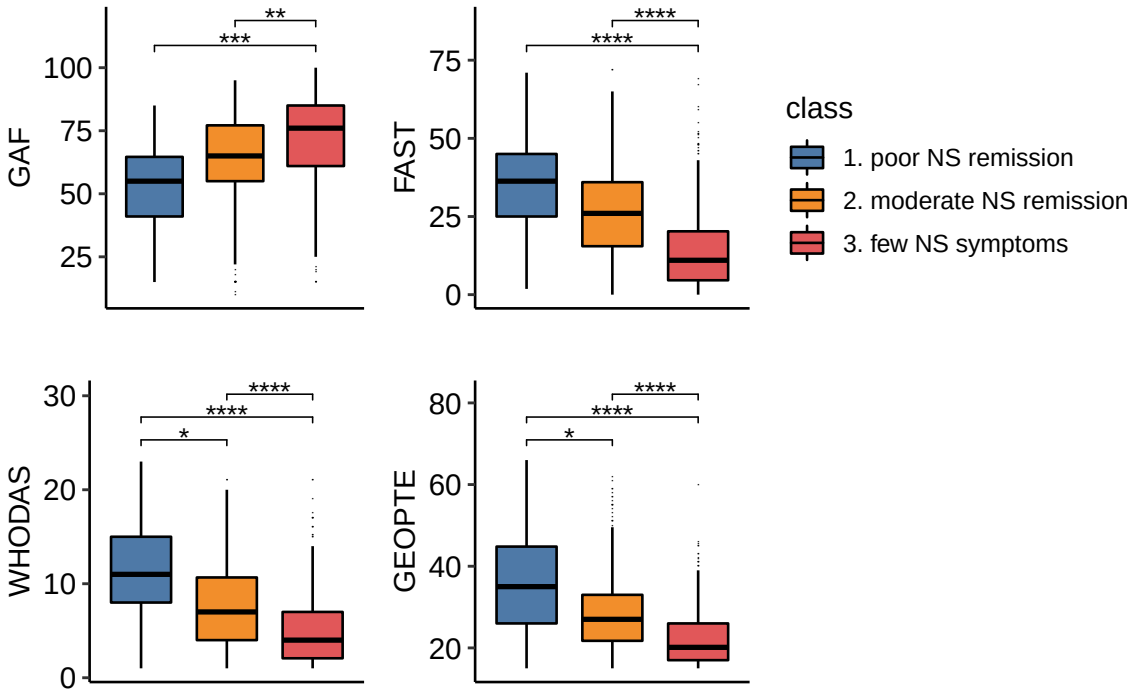


Figure 4.24: Pairwise differences between the 3 class LCGA PANSS negative (NS) symptom model for the baseline Global Assessment of Function (GAF), the Functioning Assessment Short Test (FAST), the WHO Disability Assessment Schedule (WHODAS) and the GEOPTE Scale of social cognition for psychosis.

Table 4.30: Univariate differences of socio-demographic and baseline functional and symptomatic assessment variabels within PANSS general symptoms LCGA classes

Variable	1. slow GS remission, N = 84 (26%)	2. moderate GS remission, N = 182 (57%)	3. few GS symptoms, N = 51 (16%)	pooled p-value	pooled sd	corrected p-value
Age	23.0 (20.0, 27.0)	22.0 (19.0, 28.0)	22.0 (19.0, 28.0)	>0.9	0.00	>0.9
Woman	21 (25%)	62 (34%)	20 (39%)	0.2	0.01	0.2
Education				0.027	0.00	**0.039**
Primary	23 (27%)	42.213 (23%)	10 (20%)			
Secondary	53 (63%)	112.61 (62%)	25 (49%)			
University studies	8 (9.5%)	27.176 (15%)	15 (29%)			
Other	0 (0%)	0 (0%)	1 (2.0%)			
Diagnosis				0.3	0.01	0.3
Toxic psychosis	4 (4.8%)	8 (4.4%)	2 (3.9%)			
Schizophrenia disorders	71 (85%)	152 (84%)	37 (73%)			
Affective disorders	9 (11%)	22 (12%)	12 (24%)			
IQ	90 (75, 100)	95 (80, 105)	100 (85, 110)	<0.001	0.00	**0.001**
DUP	128 (36, 247)	61 (26, 136)	27 (5, 45)	<0.001	0.00	**<0.001**
Subst. abuse	62 (74%)	147.801 (81%)	38 (75%)	0.3	0.02	0.3
PAS	58 (40, 80)	42 (30, 60)	30 (17, 43)	<0.001	0.00	**<0.001**
FES	43 (37, 49)	43 (38, 47)	46 (41, 50)	0.032	0.02	**0.044**
CGI	5.00 (4.00, 5.00)	5.00 (4.00, 5.00)	4.00 (3.00, 5.00)	0.008	0.00	**0.014**
PANSS (PS)	22 (17, 26)	18 (12, 24)	11 (7, 22)	<0.001	0.00	**<0.001**
PANSS (NS)	23 (19, 27)	18 (12, 24)	9 (7, 17)	<0.001	0.00	**<0.001**
PANSS (GS)	47 (41, 53)	35 (28, 43)	24 (20, 31)	<0.001	0.00	**<0.001**
MADRS	18 (12, 25)	10 (5, 17)	5 (1, 8)	<0.001	0.00	**<0.001**
YMRS	8 (1, 17)	6 (0, 14)	2 (0, 14)	0.080	0.00	0.10
GAF	50 (40, 60)	50 (31, 61)	60 (40, 80)	0.013	0.00	**0.020**
FAST	34 (18, 43)	27 (17, 38)	16 (7, 23)	<0.001	0.00	**<0.001**
WHODAS	11.0 (7.8, 15.0)	8.0 (6.0, 12.0)	5.0 (2.0, 9.0)	<0.001	0.00	**<0.001**
GEOPTE	35 (27, 45)	30 (23, 37)	20 (17, 28)	<0.001	0.00	**<0.001**

[1] Median (IQR); n (%)
[2] Kruskal-Wallis test; Fisher's exact test; Pooled values from imputed datasets
[3] False discovery rate correction for multiple testing

The PANSS general symptom scores classes presented a similar pattern of differences as observed for the other two PANSS sub-scores. Firstly differences in Educational level were also seen with a lower level for the more impaired classes, although the differences were less evident that the ones seen in th FAST and both PANSS(PS) and PANSS(NS) scores. Nonetheless there were no significant differences between the diagnosis categories within classes.

Moreover, significant differences within classes were seen in the IQ, the DUP and in the premorbid adjusted scale. In what regards baseline symptom scales, there were significant differences in all of PANSS sub-scores, the CGI and the MADRS. Additionally, the baseline functional profile was different between the PANSS(GS) classes 4.30.

Pairwise comparison only were significant between the "3. few GS symptoms" and the "1. slow GS remission" class (Figure 4.25).

Further difference for IQ, the DUP and the PAS score were significant for all class comparison within the PANSS(GS) classes (Figure 4.26)

In what regards to the baseline symptom scales, a significant difference between all clases was seen for both the PANSS sub scores, and the MADRS, while for the CGI, a signfficative difference was seen only between the "3. few GS symptoms" and both "2. moderate GS remission" and "1. slow GS remission" classes

Finally, significant distinctions within the groups were also kept for the FAST, and WHODAS

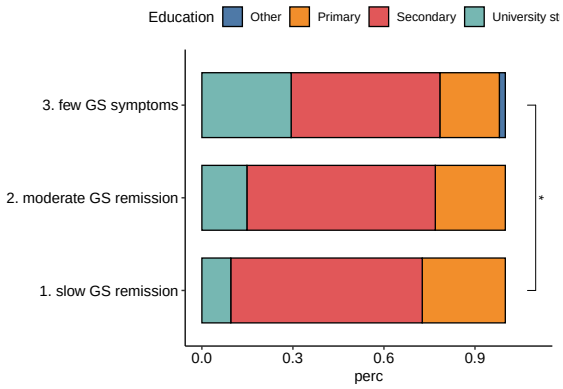


Figure 4.25: Pairwise differences in Educational level within the 3 class LCGA PANSS general symptoms model.

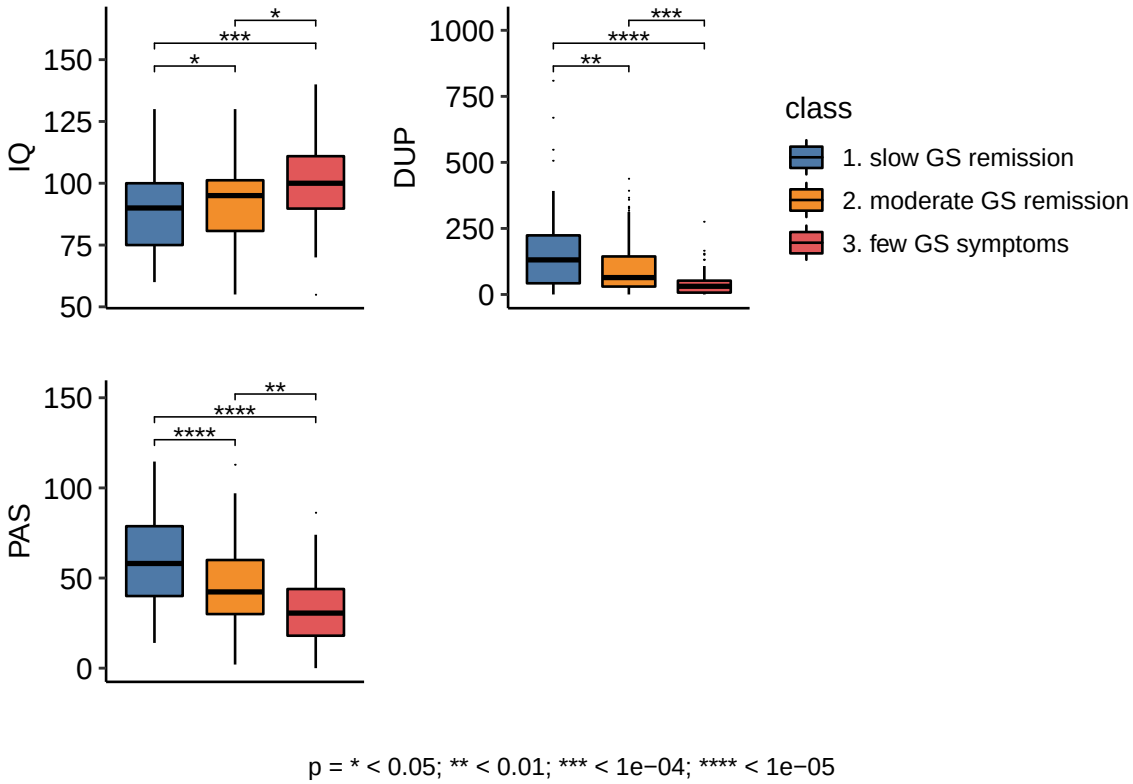


Figure 4.26: Pairwise differences between the 3 class LCGA PANSS general (GS) model, for IQ Premorbid adjustment scale (PAS) and Days of untreated psychosis (DUP).

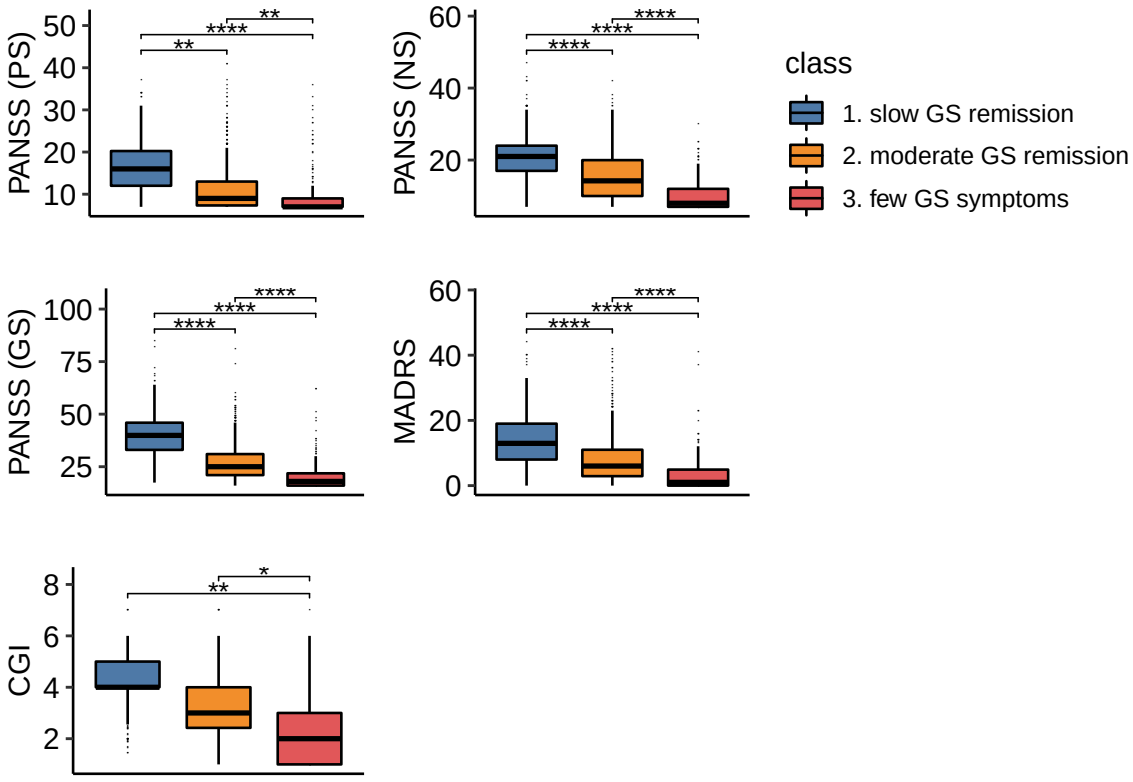


Figure 4.27: Pairwise differences between the 3 class LCGA PANSS general symptoms model for the baseline PANSS positive (PS), negative (NS) and general (GS) symptom scale, and the Montgomery-Asberg Depression Rating scale (MADRS)

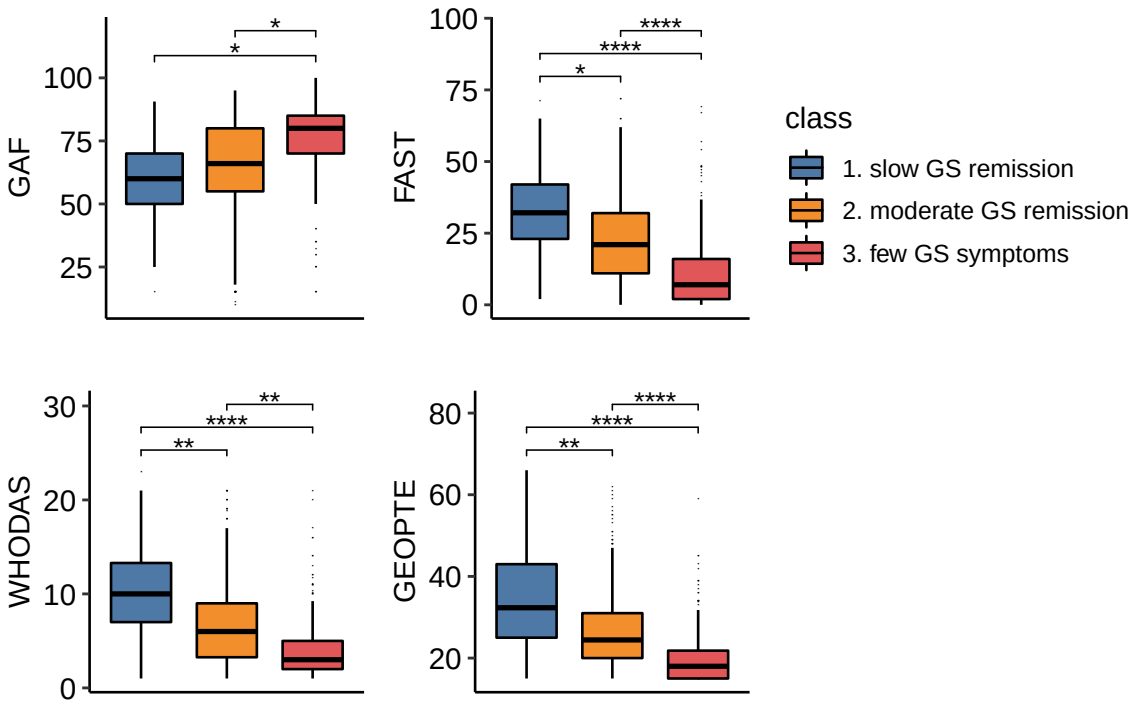


Figure 4.28: Pairwise differences between the 3 class LCGA PANSS general symptoms model for the baseline Global Assessment of Function (GAF), the Functioning Assessment Short Test (FAST), the WHO Disability Assessment Schedule (WHODAS) and the GEOPTE Scale of social cognition for psychosis.

functional scale. Likewise the GEOPTE social cognition score was also different for all thee classes. The GAF baseline score showed differences between the "3. few GS symptoms" and both "1. slow GS remission" and "2. moderate GS remission"groups.

Chapter 5

Discussion

A first point addressed by this work concerns the benefit that medical sciences can find by using methodologies that portray an accurate characterization of the longitudinal patterns of disease progression. This kind of profiling can provide means to early detect and personalize the medical approach and manage resources with an evidence based attitude. The longitudinal modelling accounted on this work, whereas by latent mixed models (LCMM), or growth analysis (LCGA), supports a valid method to model the trajectories of FEP, with a focus on both functional and symptomatic outcomes. Furthermore, longitudinal modelling is capable of appropriate discrimination of heterogeneous trajectories in an unsupervised fashion. Besides, a further characterization of the predicted classes in what regard to socio demographic and other baseline variables provides a means to further differentiate and describe classes, though it needs to be kept in mind that this type of modelling can account only for associative relations and not a causal link.

Secondly, with regards on the missing data problem, there was not an straightforward solutions found. On the one hand, missing data in health science is pervasive, and moreover, in longitudinal studies frequently suffer from more loss of data along time. Currently, imputation of missing data is an accepted strategy specially when missing data follows a MCAR or MAR pattern, as it can contribute to raise the power of the study and eventually can help surpass possible selection bias. In addition, multiple imputation methods are regarded as an adequate paradigm for missing data as it provide a means to measure the uncertainty added by the imputed data.

Nevertheless, an early problem encountered in the development of this work, was that there was no straightforward answer found with regard of the use of imputed data with longitudinal modelling. While there is current work in some specific imputation mechanism, multiple imputation was found to be discourage, as one of its main premises is that the population from where the imputed data needs to be calculated is homogeneous. On the contrary, multi class longitudinal modelling assumes just the contrary; that the population is heterogeneous with regard to their trajectories.

The strategy followed was in line with some other similar works (Colder et al., 2001). Firstly, the longitudinal model can properly deal with incomplete cases, so all available information was used to build the longitudinal models. Secondly, missing data was imputed for the univariate analysis, using the classifications found in the longitudinal modelling to account for the different populations. In any case, the univariate analysis was focused on the socio-demographic and the sudy entry data, which as seen in the EDA (chapter 4.5), had few missing data.

This study was able to describe and characteristic different distinct functional and symptomatic trajectories for FEP in a 2 year follow-up period. Moreover, both primary and secondary outcome trajectories provided clinically feasible and comprehensive progressions. Furthermore, cross tabulation of functional and symptomatic classes showed a strong relation within the symptomatic and functional domains, specially with regard on the relation between sustained negative psychotic symptoms and a poor functional outcome. Moreover, baseline analysis of the predicted functional and symptomatic outcomes indicated a strong relations between socio demographic and premorbid variables within classes, as well as with distinct severity of both depressive and psychotic symptoms at baseline assessment. Although this study can not prove a causal relation, the strong association rises the need for further analysis to properly establish causality.

5.1 Longitudinal models

In what regard to the longitudinal modelling methodology, the link function used by the latent modelling was able to properly fit both longitudinal outcomes. For both primary and secondary measures, a custom 5 I-spline link function was used. This approach allowed the extension of longitudinal modelling to account for left skewed non-Gaussian outcomes, as was the case for both outcomes (Chapter (4.3.4)

Secondly, the selection of the best number of classes for the many-class models was achieved by following several strategies. On the one hand, in what regards the FAST scale, the selection of the best model for the LCMM approach was straightforward as the 2 class model had the lowest BIC, was clinically feasible, and provided a good post probability differentiation between classes. On the other hand, for the LCGA approach models with classes that accounted for less than 5% were discarded, resulting in a 4 class model, which was both clinically relevant and provided good post probabilities of class membership. Moreover, this 4 functional classes are coherent with similar studies (Abdin et al., 2017; Levine et al., 2011).

The FAST modelling proves the complimentary nature of both LCMM and LCGA modelling. While the LCMM resulted in less classes, with a more parsimonious and simple model, the LCGA provides a more detailed classification which in this case produced clinical relevant classes. While the 2 class LCMM accounts for a bigger “A. poor function” class, the LCGA further differentiates it into two groups, which in addition, have distinct baseline characteristics.

Nevertheless, a 2 class model provides advantages, specially from a statistical point of view, as its simplicity support straightforward extension capability.

For the secondary outcomes notwithstanding, the complementation of both modelling strategies was less apparent. For the PANSS (PS) a 3 class LCGA model was selected using a similar approach as de one described earlier. Although the model had a low BIC, the posterior probability classification was less accurate. When depicted in the "spaghetti" plot, the third class identified was poorly differentiated from the other two, and was deemed irrelevant from a clinical stand (Figure 4.13).

On the other hand, for PANSS(NS) and PANSS(GS) both LCMM and LCAG provided 2 and 3 class models with good posterior classification profiles, and clinical feasibility. While the 2 class model provides a more simple and parsimonious solution, the 3 class model had a lower BIC and better differentiated the trajectories along time.

This two scenarios endorse the importance of a thorough domain knowledge. Thou the statistical criteria for model selection has straightforward value, the underlying domain knowledge should drive the analysis and model selection.

5.2 Clinical relevance

From a clinical standpoint the results of the study show that overall, most FEP have a slow functional recovery along time. Furthermore, the result suggest heterogeneous outcomes that can be associated with baseline characteristics. In line with other works, a poor functional outcome was strongly related with the premorbid adjustment, in this cases, measured with the PAS score (Jordan et al., 2017; A. Malla, 2005). Furthermore, even thou the study was not able to find causation, the relation depicted was a direct one, meaning that the more impaired the premorbid functioning, the association with poor functional outcomes was stronger.

Functional outcome was also associated with a prolonged DUP (Craig et al., 2000; Valmaggia et al., 2015), as well as for the IQ, thou less strongly. In what regards to the clinical onset and he functional outcome, this study points to a relation between the severity of symptoms at onset and a poor functional outcome, as poorer functional classes where more associated to a more severe clinical picture specially in what regards the PANSS(NS), PANSS(GS) and the MADRS. This findings are coherent with the prevalent evidence in regard to the relation between persistence of negative symptoms and a poor functional outcome. (Jordan et al., 2017; Tamminga, Buchanan, & Gold, 1998). Moreover the baseline functional assessment, in this case determined with the FAST scale, the WHODAS, and the GAF, exhibited a clear relation with the functional outcomes.

A first interpretation of this finding could relate to a rather deterministic assumption, were functionality is "fixed" at illness onset, which is counter-intuitive and not what is usually seen on a clinical setting. On the other hand, it is probably related to one of the main limitation of the study which has to do with the established fact that functional recovery varies in a time frame longer the the one measured in this study. Nevertheless, the findings regarding functional outcome support the neurodevelopmental hypothesis of schizophrenia (McGrath, Féron, Burne, Mackay-Sim, & Eyles, 2003) which states that clinical symptoms onset of the PEP is preceded

by a process of brain development disruption. This disturbances could be expressed earlier as a poor adjustment.

Furthermore there is cumulative evidence that tend to support the relationship between symptom remission and functional recovery (Dazzan et al., 2019). Within this matter, on the first hand the symptom trajectory found for all of positive, negative and general symptoms are coherent with concurrent evidence.

For the positive symptoms slightly more than half of the cases belonged to the "B. fast PS remission" class, which in turn, were classified as "3. good function recovery" or "4. very good function" in 62% of the cases. While being just associative in nature, this fact support the ideas that an early remission is related to better functional outcome.

This is highly more noticeable towards the relation between negative symptoms and a poor functional outcome, where the "1. poor NS remission" group was associated to both the "2. poor function recovery" and "1. sustained poor function" functional outcome classes on 88% percent of the named negative symptoms group. Likewise, in what regards to the general symptoms, which gathers symptoms from different domains such as anxiety or depressive syndromes, the association between a slow symptoms remission with a poor functional outcome was also unmistakeable.

In what regards to PANSS symptoms trajectories, the findings are also in line with other works. Positive symptoms, in contrast with both negative an general symptoms, tend to remit fast in time in almost half of the cases. Negative and general symptoms had more "fixed" trajectory , probably with a slower rate of change. Nevertheless, in what regards the negative symptoms trajectories there was an apparent relation, firstly in what regards to a poor pre-onset adjusted by means of both the PAS and the IQ, which conformed a relation with the rate of remission. Secondly a more sever onset, specially in what regards negative, general an depressive symptoms was associated to a poor remission rate.

The causality behind the relation between symptom remission and functional outcome is ought to be complex in nature and beyond the aims of this study. Nevertheless, there seems to be a clear association between a more a sever developmental disruption with further more difficulties in adjustment, more educational challenges which can result in a lower education level. Furthermore, there is evidence that suggest that symptoms that are onsets that are gradual, with more negative symptoms are associated with longer DUP, a fact which has also been associated with worst functional outcomes. Moreover, a lower educational levels is often associate to a lower social status which itself can be related to a worst access to mental health services further prolonging the DUP. this last point, is fundamentally important as it is one of the few variables which can be addressed and modified. In this regard, early identification of "poor recovery profiles" could make it possible to improve their functional outcomes.

Finally, one of the pervasive associations in both functional and symptoms trajectory classification was related to a baseline functional assessment. The functional assessment that in this study comprised the FAST scale, the WHODAS scale, the GAF and the GEOPTE social cognition scale, maintained a clear association and difference within both functional and symp-

tomatic outcomes. This instrument s are pragmatic in nature; they utilization requires less training than clinical symptom scales. they could provide a good measure of assessment are associated to the functional outcome, so they may prove some means stratify and operationalize early detection and personalized interventions.

5.3 Study limitations

As introduced earlier, the main limitation of the study relates to the short follow up period of the study. A difficult obstacle in what regards functional outcome studies of FEP relates to the extensive time frame that the different functional trajectories seem to occupy. One the hand hand, there are studies that depict that an early remission is related to a long-term better functional recovery, a model detailed trajectory modelling can only be achieved with long term follow.up studies.

In what regards the missing information, although we were able to analysis the patterns and mechanism of missing data, the missing information compromised the findings by reducing the sample and potentially introducing selection bias, as the missingness could be associated in an indirect and complex way to our studied variables. On the other hand, algorithms for imputing data for methodologies which imply an heterogeneous populations are being developed, but we were not able to use them in this study, as they are still in an early period of development.

In what regards to the findings, these methodologies can only account to association but not causation. Even thou it is suggested that is an timely relation between described socio demographic and clinical variables with a certain type of progression, a causal link can not be established. Nevertheless, the description of these relationships is a necessary step for the proper evidencing the causal link.
Finally, comparable to other modelling studies, longitudinal modelling is sensible to the sample size. All thought the sample size in this study is not small, dropouts, a missing data along time can be problematic. A larger sample size and a more rigurous treatmetn of the follow up data would be advantageous for both statistical power and more solid clinical description.

Bibliography

Abdin, E., Chong, S. A., Vaingankar, J. A., Peh, C. X., Poon, L. Y., Rao, S., … Subramaniam, M. (2017). Trajectories of positive, negative and general psychopathology symptoms in first episode psychosis and their relationship with functioning over a 2-year follow-up period. *PLOS ONE*, *12*, e0187141.

Álvarez-Jiménez, M., Gleeson, J. F., Henry, L. P., Harrigan, S. M., Harris, M. G., Killackey, E., … McGorry, P. D. (2012). Road to full recovery: Longitudinal relationship between symptomatic remission and psychosocial recovery in first-episode psychosis over 7.5 years. *Psychological Medicine*, *42*, 595–606.

Amoretti, S., Verdolini, N., Mezquida, G., Rabelo-da-Ponte, F. D., Cuesta, M. J., Pina-Camacho, L., … Bernardo, M. (2021). Identifying clinical clusters with distinct trajectories in first-episode psychosis through an unsupervised machine learning technique. *European Neuropsychopharmacology*, *47*, 112–129.

Andreasen, N. C., Carpenter, W. T., Kane, J. M., Lasser, R. A., Marder, S. R., & Weinberger, D. R. (2005). Remission in Schizophrenia: Proposed Criteria and Rationale for Consensus. *American Journal of Psychiatry*, *162*, 441–449.

Austin, S. F., Mors, O., Budtz-Jørgensen, E., Secher, R. G., Hjorthøj, C. R., Bertelsen, M., … Nordentoft, M. (2015). Long-term trajectories of positive and negative symptoms in first episode psychosis: A 10year follow-up study in the OPUS cohort. *Schizophrenia Research*, *168*, 84–91.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society: Series B (Methodological)*, *57*, 289–300.

Bernardo, Miguel, & Bioque, M. (2014). ¿Qué hemos aprendido de la investigación en primeros episodios psicóticos? *Revista de Psiquiatría y Salud Mental*, *7*, 61–63.

Bernardo, Miguel, Bioque, M., Cabrera, B., Lobo, A., González-Pinto, A., Pina, L., … Mas, S. (2017). Modelling gene-environment interaction in first episodes of psychosis. *Schizophrenia Research*, *189*, 181–189.

Bernardo, Miquel, Bioque, M., Parellada, M., Ruiz, J. S., Cuesta, M. J., Llerena, A., … Cabrera, B. (2013). Assessing clinical and functional outcomes in a gene–environment interaction study in first episode of psychosis (PEPs). *Revista de Psiquiatría y Salud Mental (English Edition)*, *6*, 4–16.

Bernardo, Miquel, Bioque, M., Parellada, M., Saiz Ruiz, J., Cuesta, M. J., Llerena, A., … PEPs Group. (2013). Assessing clinical and functional outcomes in a gene-environment

interaction study in first episode of psychosis (PEPs). *Revista De Psiquiatria Y Salud Mental*, *6*, 4–16.

Buuren, S. van. (2018). *Flexible imputation of missing data. Second edition.* Boca Raton, FL.: CRC Press.

Chang, Wing Chung, Kwong, V. W. Y., Chan, G. H. K., Jim, O. T. T., Lau, E. S. K., Hui, C. L. M., … Chen, E. Y. H. (2016). Prediction of functional remission in first-episode psychosis: 12-month follow-up of the randomized-controlled trial on extended early intervention in Hong Kong. *Schizophrenia Research*, *173*, 79–83.

Chang, W. C., Tang, J. Y., Hui, C. L., Lam, M. M., Chan, S. K., Wong, G. H., … Chen, E. Y. (2012). Prediction of remission and recovery in young people presenting with first-episode psychosis in Hong Kong: A 3-year follow-up study. *Australian and New Zealand Journal of Psychiatry*, *46*, 100–108.

Colder, C. R., Mehta, P., Balanda, K., Campbell, R. T., Mayhew, K., Stanton, W. R., … Flay, B. R. (2001). Identifying trajectories of adolescent smoking: An application of latent growth mixture modeling. *Health Psychology*, *20*, 127–135.

Craig, T. J., Bromet, E. J., Fennig, S., Tanenberg-Karant, M., Lavelle, J., & Galambos, N. (2000). Is There an Association Between Duration of Untreated Psychosis and 24-Month Clinical Outcome in a First-Admission Series? *American Journal of Psychiatry*, *157*, 60–66.

Cuesta, M. J., Sánchez-Torres, A. M., Cabrera, B., Bioque, M., Merchán-Naranjo, J., Corripio, I., … PEPs Group. (2015). Premorbid adjustment and clinical correlates of cognitive impairment in first-episode psychosis. The PEPsCog Study. *Schizophrenia Research*, *164*, 65–73.

Dazzan, P., Lappin, J. M., Heslin, M., Donoghue, K., Lomas, B., Reininghaus, U., … Morgan, C. (2019). Symptom remission at 12-weeks strongly predicts long-term recovery from the first episode of psychosis. *Psychological Medicine*, *50*, 1452–1462.

Emsley, R., Chiliza, B., Asmal, L., & Lehloenya, K. (2011). The concepts of remission and recovery in schizophrenia. *Current Opinion in Psychiatry*, *24*, 1.

Genolini, C., Alacoque, X., Sentenac, M., & Arnaud, C. (2015). kmlandkml3d:RPackages to Cluster Longitudinal Data. *Journal of Statistical Software*, *65*. https://doi.org/10.18637/jss.v065.i04

Genolini, C., & Falissard, B. (2009). KmL: k-means for longitudinal data. *Computational Statistics*, *25*, 317–328.

González-Ortega, I., Rosa, A., Alberich, S., Barbeito, S., Vega, P., Echeburúa, E., … González-Pinto, A. (2010). Validation and use of the functioning assessment short test in first psychotic episodes. *The Journal of Nervous and Mental Disease*, *198*, 836–840.

Hall, M.-H., Holton, K. M., Öngür, D., Montrose, D., & Keshavan, M. S. (2019). Longitudinal trajectory of early functional recovery in patients with first episode psychosis. *Schizophrenia Research*, *209*, 234–244.

Harvey, P. D., & Bellack, A. S. (2009). Toward a Terminology for Functional Recovery in Schizophrenia: Is Functional Remission a Viable Concept? *Schizophrenia Bulletin*, *35*, 300–306.

Hodgekins, J., Birchwood, M., Christopher, R., Marshall, M., Coker, S., Everard, L., … Fowler, D. (2015). Investigating trajectories of social recovery in individuals with first-episode

psychosis: A latent class growth analysis. *The British Journal of Psychiatry: The Journal of Mental Science*, *207*, 536–543.

Jordan, G., Veru, F., Lepage, M., Joober, R., Malla, A., & Iyer, S. N. (2017). Pathways to functional outcomes following a first episode of psychosis: The roles of premorbid adjustment, verbal memory and symptom remission. *Australian & New Zealand Journal of Psychiatry*, *52*, 793–803.

Jung, T., & Wickrama, K. A. S. (2007). An Introduction to Latent Class Growth Analysis and Growth Mixture Modeling. *Social and Personality Psychology Compass*, *2*, 302–317.

Keshavan, M. S., Haas, G., Miewald, J., Montrose, D. M., Reddy, R., Schooler, N. R., & Sweeney, J. A. (2003). Prolonged untreated illness duration from prodromal onset predicts outcome in first episode psychoses. *Schizophrenia Bulletin*, *29*, 757–769.

Lambert, M., Karow, A., Leucht, S., Schimmelmann, B. G., & Naber, D. (2010). Remission in schizophrenia: Validity, frequency, predictors, and patients' perspective 5 years later. *Dialogues in Clinical Neuroscience*, *12*, 393–407.

Law, H., & Morrison, A. P. (2014). Recovery in psychosis: A Delphi study with experts by experience. *Schizophrenia Bulletin*, *40*, 1347–1355.

Levine, S. Z., Lurie, I., Kohn, R., & Levav, I. (2011). Trajectories of the course of schizophrenia: From progressive deterioration to amelioration over three decades. *Schizophrenia Research*, *126*, 184–191.

Liberman, R. P., & Kopelowicz, A. (2005). Recovery from schizophrenia: A concept in search of research. *Psychiatric Services (Washington, D.C.)*, *56*, 735–742.

Lieberman, J., Jody, D., Geisler, S., Alvir, J., Loebel, A., Szymanski, S., … Borenstein, M. (1993). Time course and biologic correlates of treatment response in first-episode schizophrenia. *Archives of General Psychiatry*, *50*, 369–376.

Lucas, S., Redoblado-Hodge, M. A., Shores, A., Brennan, J., & Harris, A. (2008). Predictors of outcome three years after diagnosis of first episode psychosis. *Psychiatry Research*, *161*, 11–18.

Maechler, M., Rousseeuw, P., Struyf, A., & Hubert, M. (2021). *Cluster: "Finding groups in data": Cluster analysis extended rousseeuw et al.* Retrieved from https://svn.r-project.org/R-packages/trunk/cluster/

Malla, A. (2005). First-Episode Psychosis: Psychopathology, Quality of Life, and Functional Outcome. *Schizophrenia Bulletin*, *31*, 650–671.

Malla, A. K., Norman, R. M. G., Manchanda, R., & Townsend, L. (2002). Symptoms, cognition, treatment adherence and functional outcome in first-episode psychosis. *Psychological Medicine*, *32*, 1109–1119.

Mausbach, B. T., Moore, R., Bowie, C., Cardenas, V., & Patterson, T. L. (2009). A review of instruments for measuring functional recovery in those diagnosed with psychosis. *Schizophrenia Bulletin*, *35*, 307–318.

McGrath, J. J., Féron, F. P., Burne, T. H., Mackay-Sim, A., & Eyles, D. W. (2003). The neurodevelopmental hypothesis of schizophrenia: a review of recent developments. *Annals of Medicine*, *35*, 86–93.

Menezes, N. M., Arenovich, T., & Zipursky, R. B. (2006). A systematic review of longitudinal outcome studies of first-episode psychosis. *Psychological Medicine*, *36*, 1349–1362.

Nagin, D. S., Jones, B. L., Passos, V. L., & Tremblay, R. E. (2016). Group-based multi-trajectory modeling. *Statistical Methods in Medical Research*, *27*, 2015–2023.

Nest, G. van der, Lima Passos, V., Candel, M. J. J. M., & Breukelen, G. J. P. van. (2020). An overview of mixture modelling for latent evolutions in longitudinal data: Modelling approaches, fit statistics and software. *Advances in Life Course Research*, *43*, 100323.

Norman, R. M. G. G., Carr, J., & Manchanda, R. (2015). Cognition and the prediction of functioning in patients with a first treated episode of psychosis: A prospective study. *Schizophrenia Research*, *162*, 138–142.

Pelayo-Terán, J. M., Diaz, F. J., Pérez-Iglesias, R., Suárez-Pinilla, P., Tabarés-Seisdedos, R., de León, J., & Crespo-Facorro, B. (2014). Trajectories of symptom dimensions in short-term response to antipsychotic treatment in patients with a first episode of non-affective psychosis. *Psychological Medicine*, *44*, 37–50.

Proust-Lima, Cecile, Philipps, V., Diakite, A., & Liquet, B. (2022). *Lcmm: Extended mixed models using latent classes and latent processes*.

Proust-Lima, Cécile, Philipps, V., & Liquet, B. (2017). Estimation of Extended Mixed Models Using Latent Classes and Latent Processes: The *R* Package **Lcmm**. *Journal of Statistical Software*, *78*. https://doi.org/10.18637/jss.v078.i02

R Core Team. (2021). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.

Rubin, D. B. (1976). Inference and missing data. *Biometrika*, *63*, 581–592.

Schafer, J. L., & Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological Methods*, *7*, 147–177.

Schennach, R., Meyer, S., Seemüller, F., Jäger, M., Schmauss, M., Laux, G., … Riedel, M. (2012). Response trajectories in "real-world" naturalistically treated schizophrenia patients. *Schizophrenia Research*, *139*, 218–224.

Tamminga, C. A., Buchanan, R. W., & Gold, J. M. (1998). The role of negative symptoms and cognitive dysfunction in schizophrenia outcome. *International Clinical Psychopharmacology*, *13*, S21–S26.

Valmaggia, L. R., Byrne, M., Day, F., Broome, M. R., Johns, L., Howes, O., … McGuire, P. K. (2015). Duration of untreated psychosis and need for admission in patients who engage with mental health services in the prodromal phase. *British Journal of Psychiatry*, *207*, 130–134.

van Buuren, S., & Groothuis-Oudshoorn, K. (2011). mice: Multivariate imputation by chained equations in r. *Journal of Statistical Software*, *45*, 1–67.

van Buuren, S., & Groothuis-Oudshoorn, K. (2021). *Mice: Multivariate imputation by chained equations*.

Velthorst, E., Fett, A.-K. J., Reichenberg, A., Perlman, G., van Os, J., Bromet, E. J., & Kotov, R. (2017). The 20-Year Longitudinal Trajectories of Social Functioning in Individuals With Psychotic Disorders. *American Journal of Psychiatry*, appi.ajp.2016.1.

Verma, S., Subramaniam, M., Abdin, E., Poon, L. Y., & Chong, S. A. (2012). Symptomatic and functional remission in patients with first-episode psychosis. *Acta Psychiatrica Scandinavica*, *126*, 282–289.

von Hippel, P. T. (2018). How Many Imputations Do You Need? A Two-stage Calculation

Using a Quadratic Rule. *Sociological Methods & Research*, *49*, 699–718.

White, C., Stirling, J., Hopkins, R., Morris, J., Montague, L., Tantam, D., & Lewis, S. (2009). Predictors of 10-year outcome of first-episode psychosis. *Psychological Medicine*, *39*, 1447–1456.

Wickham, H. (2021). *Tidyverse: Easily install and load the tidyverse*.

Wickham, H., & Bryan, J. (2022). *Readxl: Read excel files*.

Wunderink, L., Sytema, S., Nienhuis, F. J., & Wiersma, D. (2009). Clinical Recovery in First-Episode Psychosis. *Schizophrenia Bulletin*, *35*, 362–369.

Appendix A

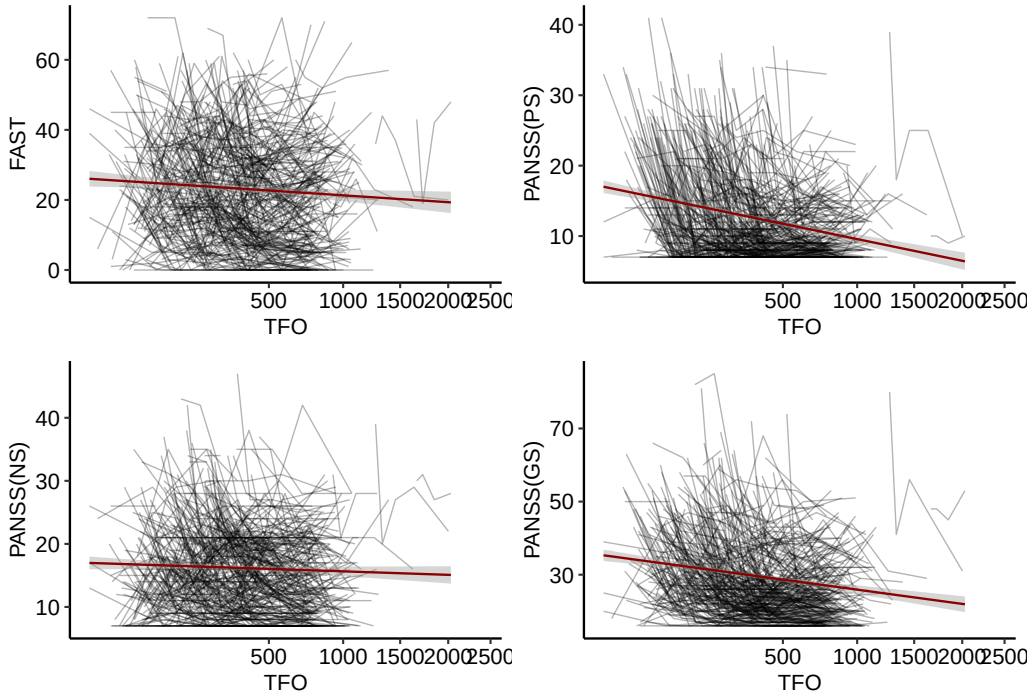
Additional figures and tables

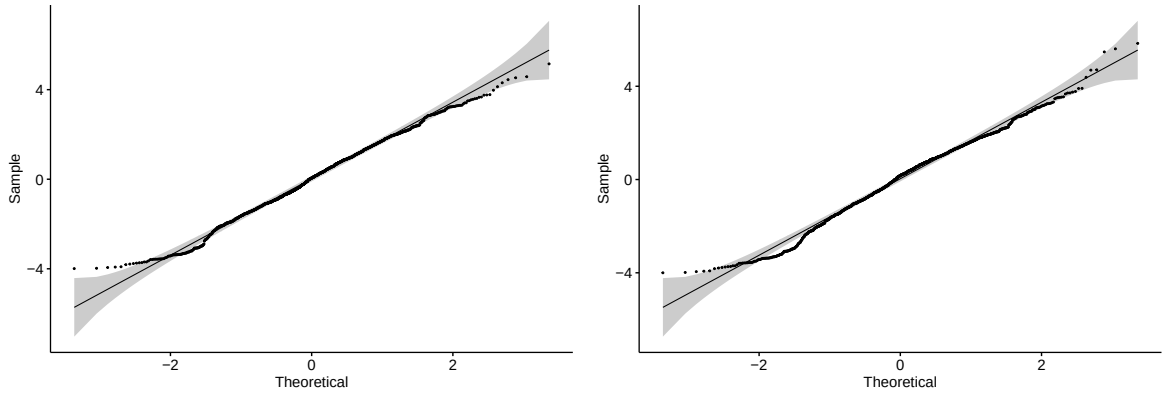
Table A.1: Shapiro-Wilk normality test for baseline continuous variables

name	statistic	p.value	method
ID	0.92	**<0.001*** **	Shapiro-Wilk normality test
Age	0.97	**<0.001*** **	Shapiro-Wilk normality test
IQ	0.98	**0.0028** **	Shapiro-Wilk normality test
DUP	0.78	**<0.001*** **	Shapiro-Wilk normality test
LEWIS	0.59	**<0.001*** **	Shapiro-Wilk normality test
BW	0.96	**<0.001*** **	Shapiro-Wilk normality test
PAS	0.97	**<0.001*** **	Shapiro-Wilk normality test

Table A.2: Shapiro-Wilk normality test for baseline continuous variables

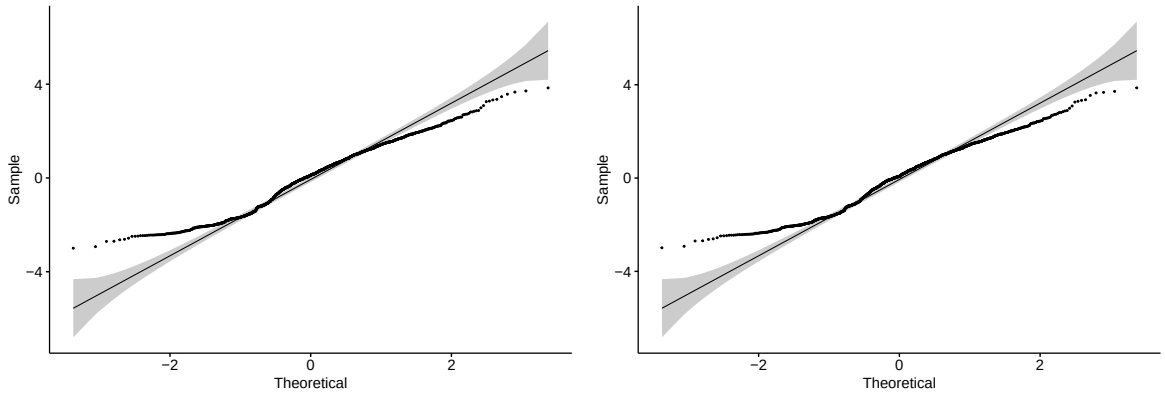
name	statistic	p.value	method
CGI	0.94	**<0.001*** **	Shapiro-Wilk normality test
GAF	0.97	**<0.001*** **	Shapiro-Wilk normality test
MADRS	0.89	**<0.001*** **	Shapiro-Wilk normality test
TFO	0.94	**<0.001*** **	Shapiro-Wilk normality test
WHODAS	0.95	**<0.001*** **	Shapiro-Wilk normality test
YMRS	0.64	**<0.001*** **	Shapiro-Wilk normality test
FAST	0.97	**<0.001*** **	Shapiro-Wilk normality test
PANSS	0.93	**<0.001*** **	Shapiro-Wilk normality test
GEOPTE	0.92	**<0.001*** **	Shapiro-Wilk normality test
FES	0.98	**<0.001*** **	Shapiro-Wilk normality test



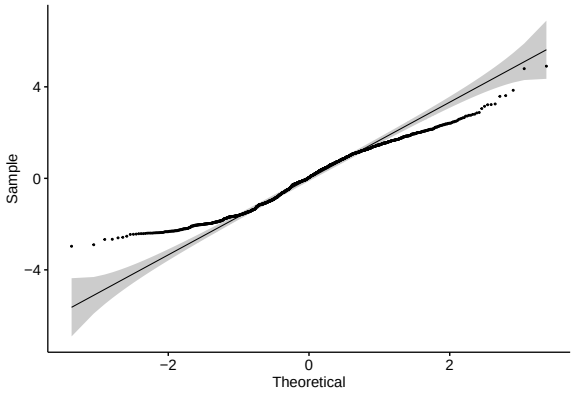


(a) manual 5 I-splines (b) Beta cummulative distribution

Figure A.1: QQ plots of FAST outcome

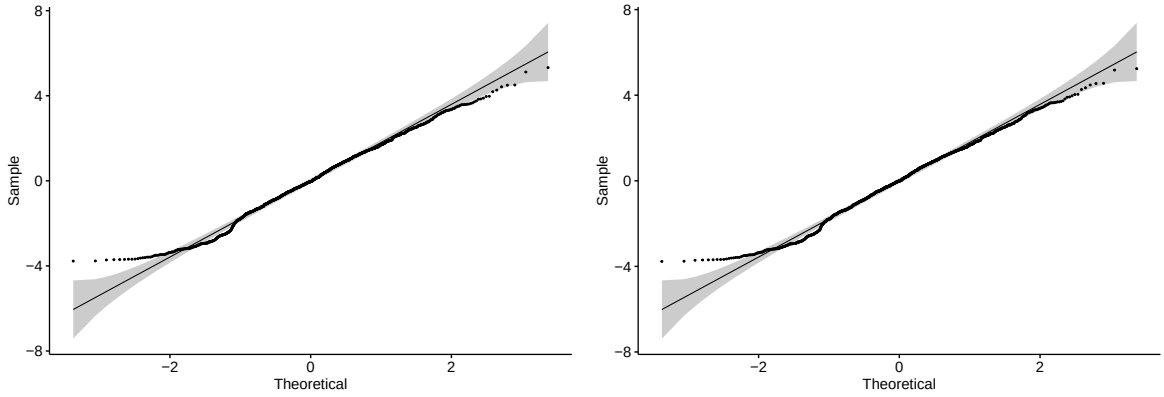


(a) manual 5 I-splines (b) Beta cummulative distribution



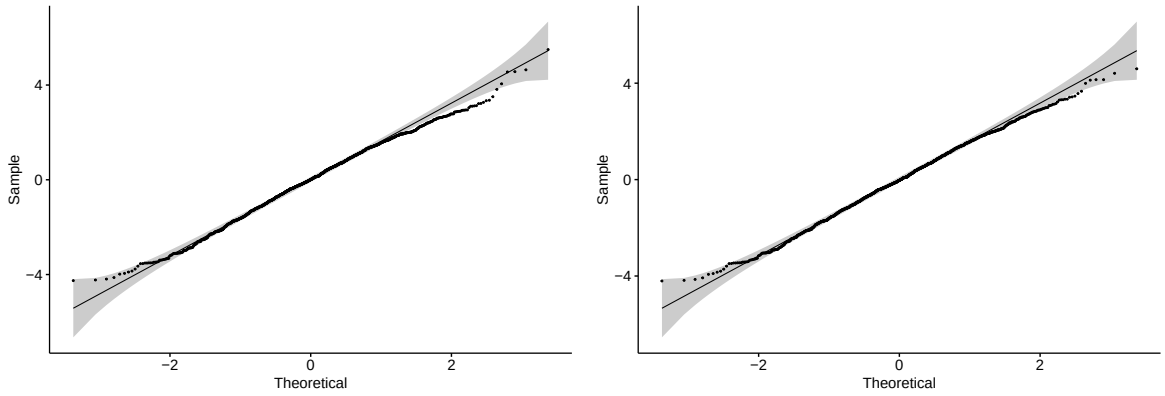
(c) manual 5 I-splines

Figure A.2: QQ plots of PANSS(PS) outcome

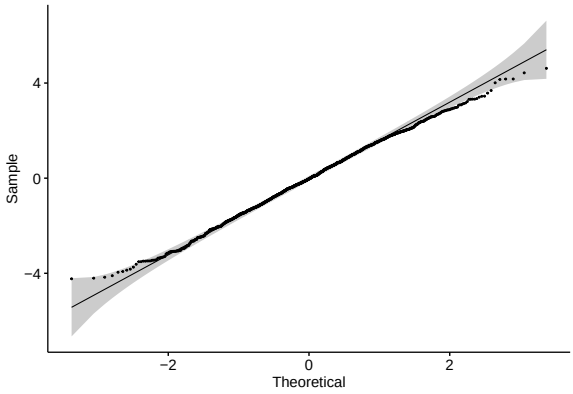


(a) manual 5 I-splines (b) Beta cummulative distribution

Figure A.3: QQ plots of PANSS(NS) outcome



(a) manual 5 I-splines (b) Beta cummulative distribution



(c) manual 5 I-splines

Figure A.4: QQ plots of PANSS(GS) outcome