



Universiteit
Leiden
The Netherlands

To fail or not to fail : clinical trials in depression

Sante, G.W.E.

Citation

Sante, G. W. E. (2008, September 10). *To fail or not to fail : clinical trials in depression*. Retrieved from <https://hdl.handle.net/1887/13091>

Version: Corrected Publisher's Version

License: [Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden](#)

Downloaded from: <https://hdl.handle.net/1887/13091>

Note: To cite this publication please use the final published version (if applicable).

Chapter

7

Heterogeneity in patient response in depression:

The relevance of functional data analysis

Gijs Santen, Erik van Zwet, Meindert Danhof, Oscar Della Pasqua

Submitted for publication

ABSTRACT

Many misconceptions exist about the behaviour of 'typical' patients in clinical trials in depression. Physicians, for example, are taught that it takes 2-4 weeks for the effect of antidepressants to manifest, which has been shown to be untrue. On the other hand, average curves show a time course which is also not representative of many patients, since considerable heterogeneity exists. In this work the focus is on the analysis of the heterogeneity between patients, rather than on the mean behaviour, using methodology from functional data analysis.

Data from five double-blind, randomised, placebo-controlled, clinical studies were used in which the HAM-D was measured as efficacy endpoint. All analyses were performed in the language and environment for statistical computing, R. The package *pcaMethods*, which includes various methods to deal with censored data, was used to carry out the principal component analysis.

The results of the functional data analysis showed that most variation (~65%) was present in a vertical shift of the curve, as was also evident from previous analyses using a linear mixed model. The main principal components of the HAM-D₁₇ were constant over studies and also the same for responders and non-responders. The main principal components were also identified in the HAM-D₇ subscale.

Our analysis enables identification of individual response patterns over time. It also shows that the principal components explaining heterogeneity are constant across clinical studies and in responders *versus* non-responders, although the mean curves do differ between these subpopulations. This finding indicates that responders and non-responders do not constitute two different populations. It is also shown that individual differences in response can be characterised by the use of a subscale, even though it contains only 7 items, as compared to the full HAM-D₁₇. This strengthens the relevance of subscales in clinical research.

INTRODUCTION

Psychiatric and neurological diseases have been fraught with difficulties in the identification of drug treatment which is efficacious for the majority of the patient population. Among other factors, genetic differences, ontogeny and external triggers have been used as explanations for treatment failure. However, such a conclusion about the potential causes of failure has often been derived from measures of efficacy based on the assessment of the so-called change relative to baseline, which ignores the relevance of the time course of response.

From a clinical perspective, the ability to capture the time course of response in individual patients enables the identification of disease patterns and covariates, which can

have prognostic value. It also provides valuable information on the underlying rate of disease progression. The taxonomy associated with such patterns has, however, created a major challenge in clinical research: what to do about typical, atypical and non-typical patients? Hence, variability has been traditionally treated as a nuisance factor in the evaluation of drug efficacy, rather than as a means to understand drug response as well as individual differences in terms of pathophysiology.

Usually, statistical analysis of longitudinal clinical data takes no interest in the particular shapes of the curves that are analysed. For example, a linear mixed model (LMM) such as the mixed model for repeated measures (MMRM) (Mallinckrodt *et al.*, 2001), which is often used to analyse longitudinal data from clinical trials, regards time as a factor. Repeated measures are taken into account in order to provide information about patients that have not completed the clinical trial, but otherwise inference usually only takes into account the last observation. Mean profiles of hundreds of patients do not give a good impression of the variability in curve shapes between patients. Although hierarchical linear mixed models do take inter-individual variability into account, they usually only allow for one additive random component (i.e., patients are on average above or below the mean population values).

In the current investigation we show how individual patterns can provide insight into the nature of response and how this knowledge may improve the evaluation of treatment effect.

The field of functional data analysis provides analysis methods such as functional principal component analysis to investigate the variability between patients with respect to individual curve shapes. Rather than treating data as individual points, it considers functions or curves as 'individuals'. This field is not as focused on inference as traditional longitudinal data analysis, but has a more exploratory nature. In general, principal component analysis (PCA) aims at reducing the number of dimensions in the data, whilst retaining relevant information. For example, the Hamilton depression rating scale has been subjected to several principal component analyses (Trivedi *et al.*, 2005; Fleck *et al.*, 1995, 2004), which attempted to reduce the number of items. Similar investigations have been performed on other psychiatric endpoints (Lam *et al.*, 2004; Goekoop *et al.*, 2007).

In contrast, functional principal component analysis (FPCA) is a continuous version of a principal component analysis where the different dimensions are not separate observations (such as the different items of the HAMD), but rather measurements separated by time. Since data which is continuous in time is rare, various methods have been developed to smooth raw data before applying FPCA methodology (Ramsay and Dalzell, 1991). Another approach is to perform a classical PCA on the raw data and smooth the resulting components (Rice and Silverman, 1991). It may also be possible to draw conclusions from the unsmoothed principal components if interpolation is not a main objective of the investigation. FPCA has been applied to many fields (see for example Newell *et al.* (2006); Ramsay *et al.* (1995, 1994)) and seems a good candidate approach to expand our insight into individual differences during the course of a clinical trial.

Here, we present the results of a functional principal component analysis to explain the heterogeneity in the response patterns of patients with major depressive disorder. It may be possible to correlate the outcomes of the analysis to different patient sub-populations, and link specific patterns to typical and atypical patients. A second objective is to determine the heterogeneity in response patterns for responders and non-responders separately, which may provide insight as to whether these patient groups are in fact truly different. In this context, another aim is to assess whether the degree or nature of the heterogeneity between patients is altered by drug treatment, as compared to placebo. The third objective is to establish whether the inter-curve variability is modified or modulated in the HAM-D₇ subscale (chapter 3). This can be achieved by comparing results from the FPCA on the HAM-D₁₇ and the reduced subscale. It is anticipated that preservation of the principal components in the HAM-D₇ is sufficient to warrant an accurate description of the heterogeneity in individual response patterns.

METHODS

Data from 5 placebo-controlled clinical trials in patients diagnosed with major depressive disorder was extracted from GlaxoSmithKline's clinical trial database. Study 1, a trial comparing a 12.5 mg and 25 mg dose of a controlled formulation of paroxetine, is reported throughout this paper. Data from another four clinical trials were used to evaluate the reproducibility and consistency of the method in different study designs. The characteristics of the 5 trials are shown in table 1. The overall dropout rate in these trials was similar for placebo and active treatment (33% and 31%, respectively).

As endpoints, the full HAM-D₁₇ and the HAM-D₇ (chapter 3) were selected. The HAM-D₇ subscale consists of the following items: *depressed mood, feelings of guilt, suicide, work and interests, retardation, psychic anxiety and somatic symptoms general*.

All data manipulation, analysis and graphical summaries were performed in R, the language and environment for statistical computing (R Development Core Team, 2007). The principal component analysis was carried out using the package *pcaMethods* (Stacklies and Redestig, 2007).

Our analysis included not only an evaluation of the complete datasets (all patients), but also separate analyses of the data by treatment (placebo *versus* active treatment) and by response (responders *versus* non-responders). In the context of this paper, response was defined as a decrease in the HAMD at the last observation of $\geq 50\%$ relative to baseline. First, the classical PCA implementation (singular value decomposition, SVD) was used, which ignores any curves with missing data. The *pcaMethods* package offers a selection of methods to impute missing data. The probabilistic PCA method (Roweis, 1997) and the SVD imputation (Troyanskaya *et al.*, 2001) approach were chosen because the amount of missing data (about 30% in study 1). The other methods available within *pcaMethods* are not suitable for such a large percentage of missing data. Briefly, the SVD imputation imputes missing data until changes in the expected solution exceed a predefined

threshold. Default values were used for all parameters. In contrast, the probabilistic PCA method uses an expectation-maximisation (EM) algorithm in combination with a probabilistic model. All analyses were carried out on centered data (using the option center=T), allowing the principal components to be interpreted as deviations from the mean.

Whereas a classical PCA is aimed at reducing multidimensional data to fewer dimensions, e.g., multivariate or datasets with more than one statistical variable, the objective of a functional PCA is to reduce time dimensions. Each principal component is therefore a time-dependent function. Such a time dependency captures indirectly the progression or rate of change in response, which is not accounted for in the mixed model for repeated measures (MMRM) or in other statistical methods in which the last observation is subtracted from baseline values. For the sake of clarity, only the principal components which were deemed sufficiently explanatory of the heterogeneity will be presented. The explanatory degree of each component was assessed by calculating its contribution to the overall variance.

Table 1. Characteristics of the 5 clinical trials used in this paper. Further details can be found in the references. For unpublished studies, see GlaxoSmithKline’s clinical trial register (<http://ctr.gsk.co.uk>). Study 5 only included elderly patients

no. pat.	Active treatments (dose)	Visits (T=titration design)	Dropout rate	Reference (NP=not published)
1 447	paroxetine CR (12.5 mg) paroxetine CR (25 mg)	1,2,3,4 6,8	22%	Trivedi <i>et al.</i>
2 691	paroxetine (max 50 mg) fluoxetine (max 80 mg)	1,2,3,4, 6,9,12(T)	38%	protocol 115 (NP)
3 848	paroxetine (max 50 mg) fluoxetine (max 80 mg)	1,2,3,4 6,9,12(T)	37%	protocol 128 (NP)
4 315	paroxetine IR (max 50 mg) paroxetine CR (max 62.5 mg)	1,2,3,4 6,8,12(T)	30%	Golden <i>et al.</i>
5 319	paroxetine IR (max 40 mg) paroxetine CR (max 50 mg)	1,2,3,4,6 8,10,12(T)	24%	Rapaport <i>et al.</i>

RESULTS

The loadings, i.e., the deviations from the mean for each observation, of the first four principal components which emerged from the classical principal component analysis (SVD) of only complete curves from study 1, using the HAM-D₁₇ as endpoint, are shown in figure 1A. Figure 1b shows the mean profile.

As the analysis was being performed, it became apparent that the main principal components were in fact fairly smooth. This enhances the interpretability of the components and after smoothing the components may allow interpolation of data. However, since interpolation was not an aim in this analysis, smoothing techniques were not deemed necessary. The shape of the main principal components was consistent enough to make appropriate inferences.

The first four components were found to describe approximately 91% of the variability in the data. The remaining components that were identified in the analysis were therefore considered unimportant (figure 2). Interestingly, each of the principal components has a clear shape or profile. The first component is essentially a horizontal line, indicating that the deviation from the mean is consistent over time, both in magnitude and direction. In other words, it represents a vertical shift, i.e., patients that show higher or smaller response than the population average. The deviation resulting from this first component at week 0 and 1 (where week 0 is the start of treatment) implies that the response curve is likely to deviate further from the mean at later time points. The second principal component represents patients which start the trial more depressed as compared to the average, but improve later in the study and *vice versa*. The third component

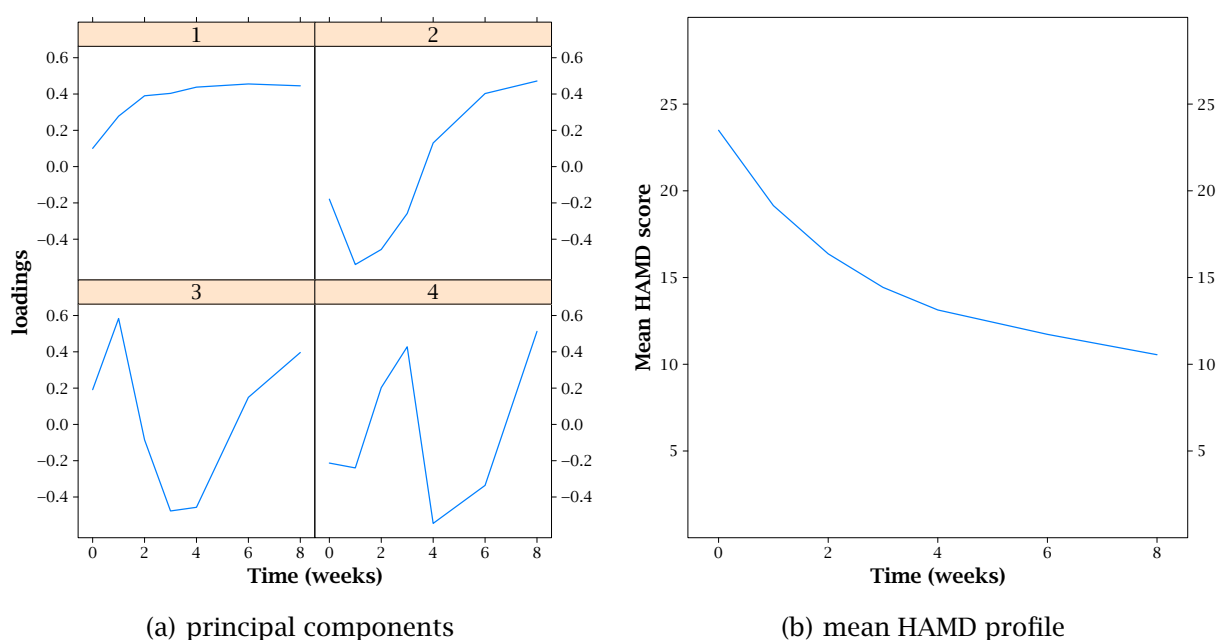


Figure 1. (a) The first four principal components as analysed using singular value decomposition on the complete dataset of study 1. (b) Mean HAM-D profile of all patients

represents patient that have a decline or an improvement halfway during the trial (weeks 3 and 4), whereas the 4th component represents patients which exhibit an oscillatory behaviour around the mean profile. The higher components did not have the same degree of smoothness as those shown in figure 1a and were therefore considered as noise. This can also be deduced from figure 2, where the percentage of variance described by each component is shown (<4% for components PC5, PC6 and PC7).

It is important to note that most of the variance is described by the first two components, with the last components explaining only a small fraction. Since similar results were obtained for the principal components which were identified from the analysis of the data by treatment group, all data reported here refers to pooled data (active treatment and placebo).

To illustrate the characteristics of the principal components in the population, the mean profile and the patients with the highest and lowest score for each of the four main principal components are presented. As can be seen in figure 3, the individual response profiles in each panel illustrate how each component constrains the response pattern. For the first component, it implies that the larger this parameter estimate, the larger the deviation from the average, with patients showing higher or lower than the average HAM-D₁₇ scores over the course of the study. In contrast, the second component accounts for the shift in response which switches patients from improvement to decline and *vice versa*. Here again, the larger the parameter estimate for this component, the wider the amplitude of the shift with respect to the average response. Since the variance in response is mostly explained by the first two components, one can anticipate that the influence of the third component is somewhat obscured. The main feature in this case is a temporary sway from decline to improvement or *vice versa*. Clearly, the contribution of the first

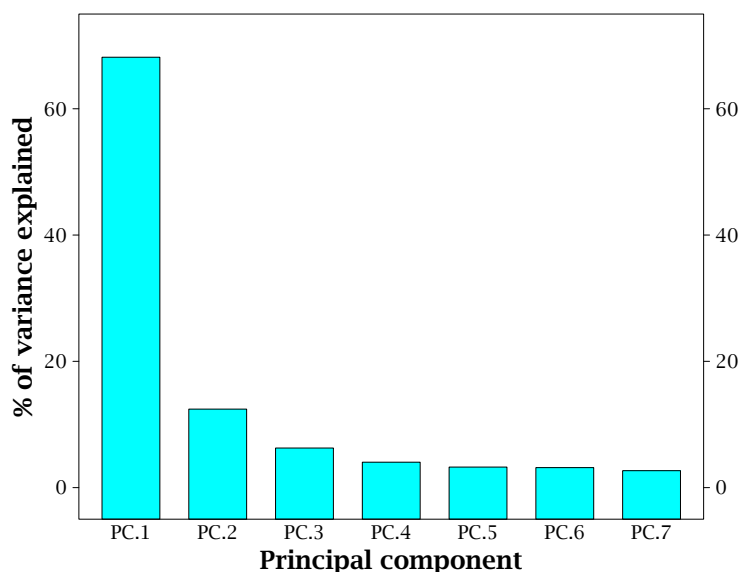


Figure 2. Percentage of the variance associated with each of the principal components as analysed using singular value decomposition on the completer dataset of study 1

three components causes further smoothing of the oscillatory behaviour associated with the fourth component.

All previous analyses were performed on the completer dataset, i.e., including only the patients without any missing data. Yet, it is possible to analyse all patients by imputation of missing data, which can be done using SVD imputation and probabilistic PCA. The first four components identified by these methods are shown in figure 4.

A comparison of figures 1 and 4 reveals a great degree of correspondence for the four main principal components estimated by the SVD, PPCA and SVDImpute methods.

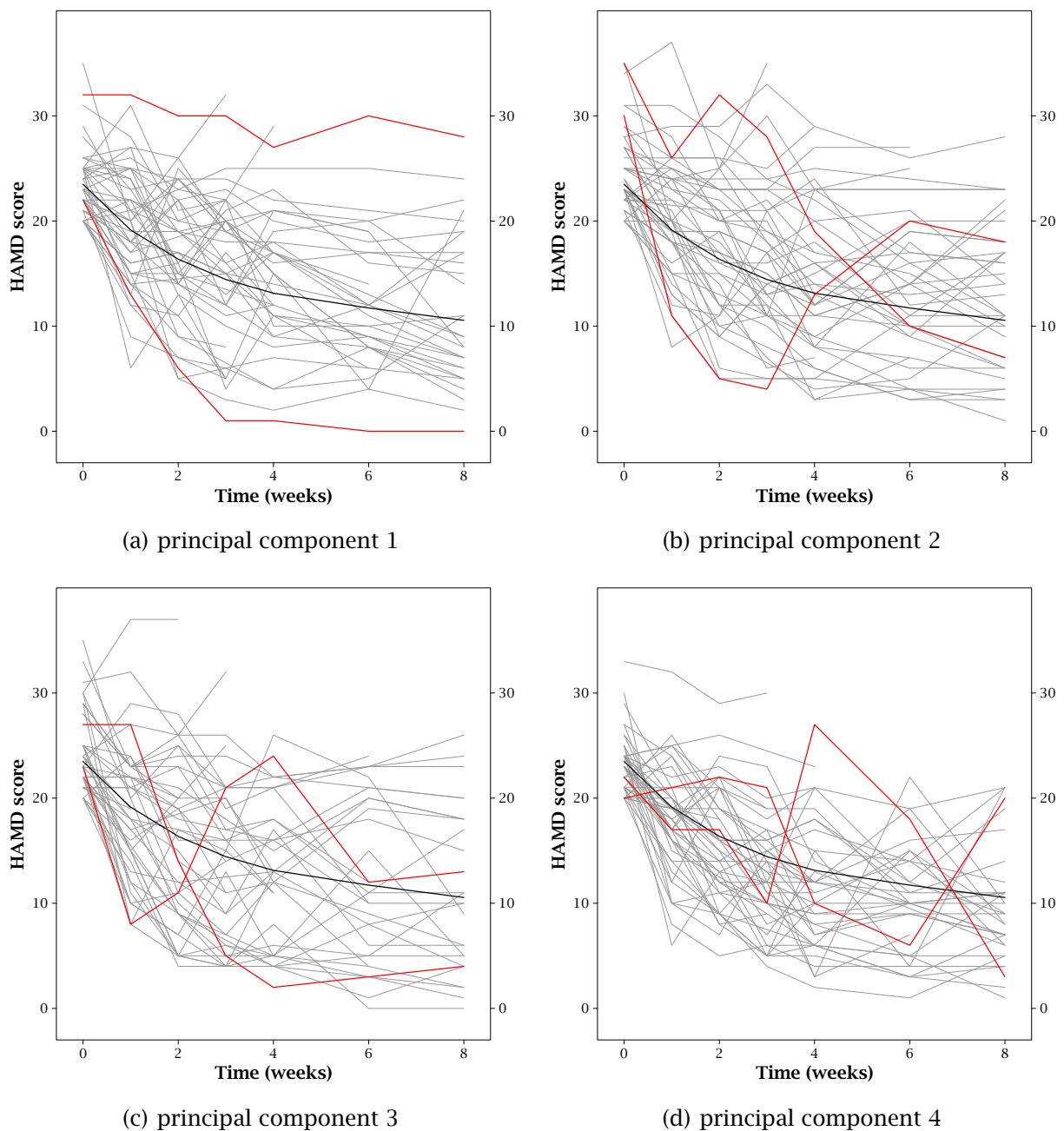


Figure 3. Mean profile and patients with the highest and lowest score for each of the four main principal components as analysed using singular value decomposition on the completer dataset of study 1. Ten percent of the patient population was randomly selected and plotted in grey lines for reference purposes

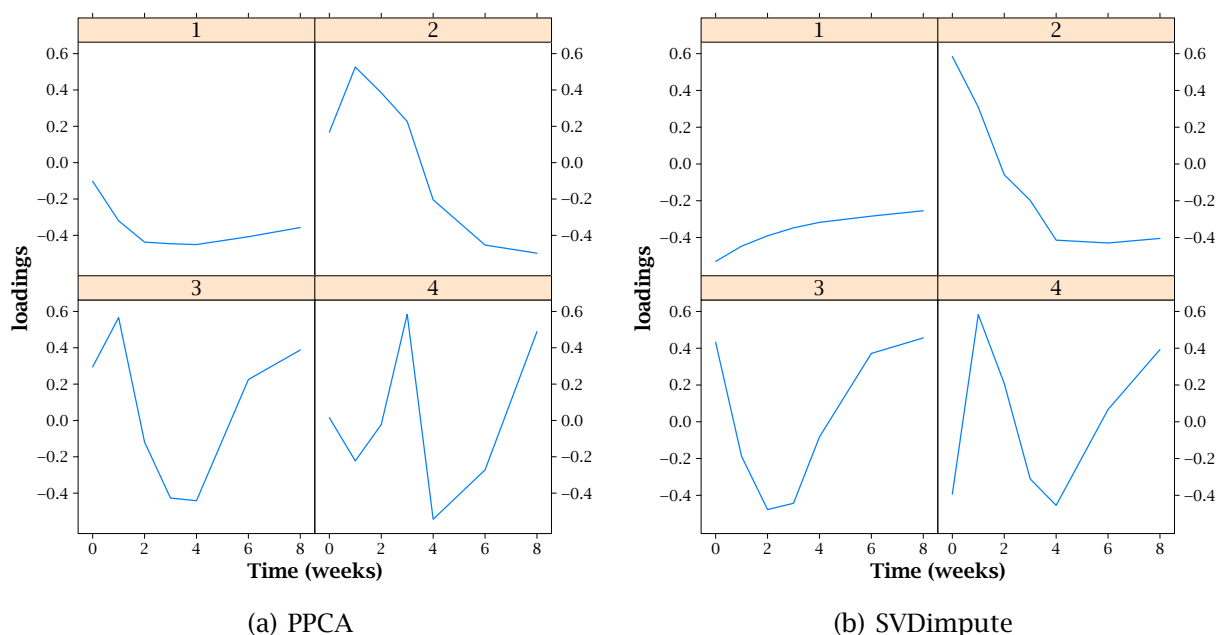


Figure 4. Main principal components as computed using PPCA and SVDImpute methods on the full dataset of study 1

Moreover, the principal components identified in the remaining studies that were available for the analysis closely match those obtained in study 1 (*data not shown*). The distribution of the scores for each component was also examined for differences between treatment groups. In all studies, a significant difference between placebo and active treatments was found for the first and, in most studies, also for the second component, if one of the missing data imputation methods was used. Upon omission of missing data, especially in those studies with higher dropout rates, the classical SVD method was more conservative.

Another important aspect of the disease was also explored, namely whether specific differences in the response patterns exist between responders and non-responders. These findings are summarised in figure 5a and 5b, which depict the first four components in responders and non-responders, respectively.

The mean profiles of responders and non-responders are shown in figure 6. It is important to note that the patterns of response in responders and non-responders is independent of treatment type (i.e., active *versus* placebo).

On the other hand, despite these differences in the mean profiles, the main principal components are very similar, which indicates a comparable degree of heterogeneity in either group. The profiles are less smooth than those with all data pooled together, probably because less data was used per group.

Finally, we have assessed whether the differences in the sensitivity of the full HAM-D₁₇ scale and the HAM-D₇ also affected the heterogeneity in response patterns. Interestingly, it was found that the principal components in the HAM-D₇ subscale are undistinguishable from those of the full HAM-D₁₇ (figure 7).

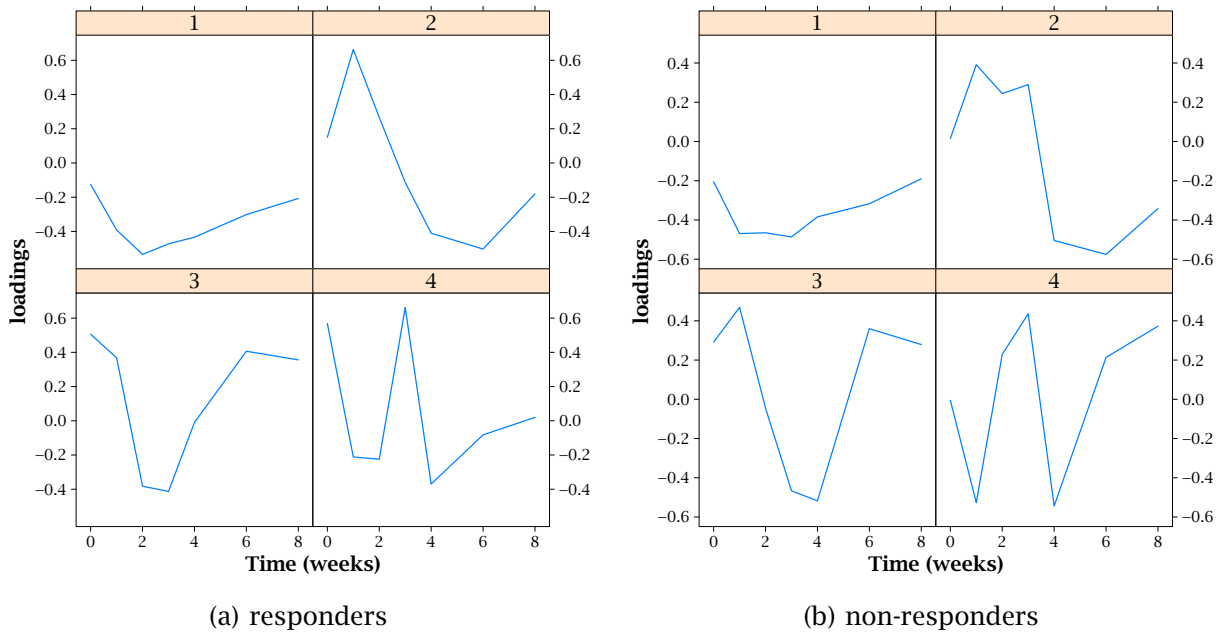


Figure 5. Main principal components in responders and non-responders in study 1 as analysed using PPCA

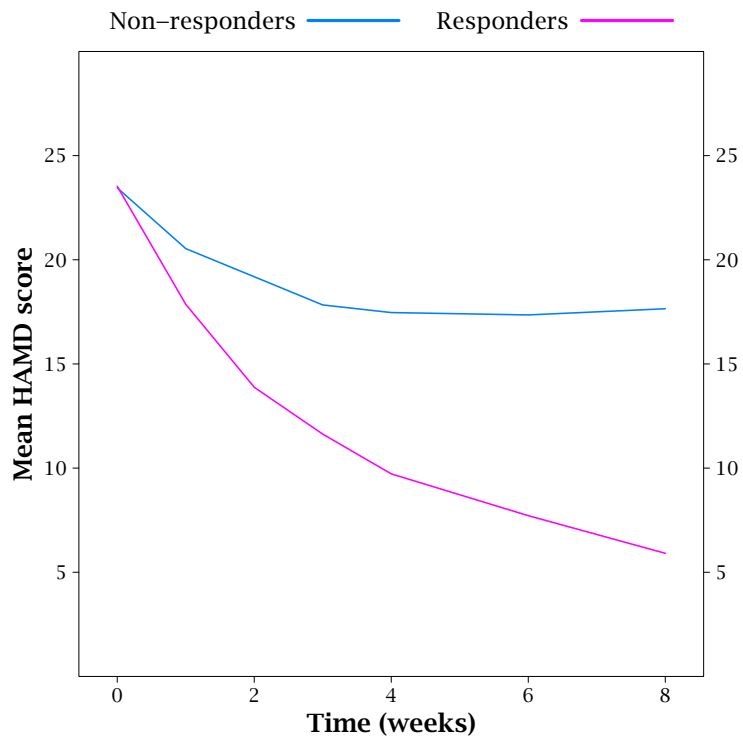


Figure 6. Mean HAMD profiles for responders and non-responders in study 1

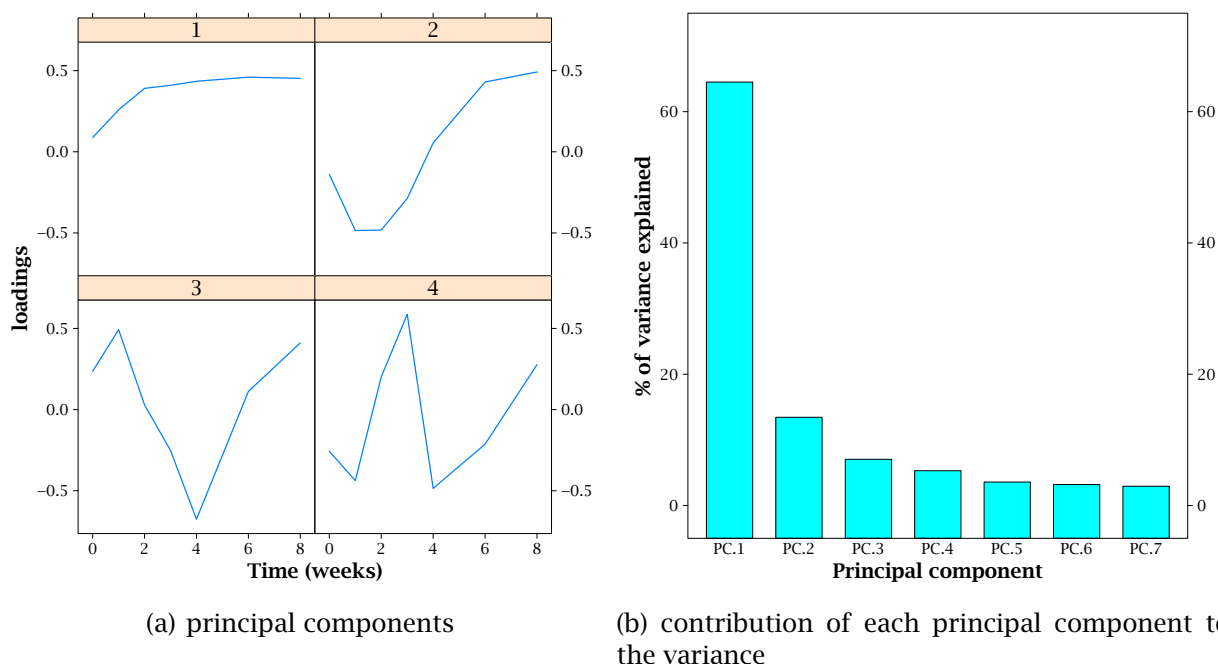


Figure 7. Loadings of the 4 principal components and contribution of each component to the variance for study 1, based on the HAM-D₇ subscale

DISCUSSION

Relevance of FPCA and implications of the current findings

There are important differences in the aims of the application of a principal component analysis in the context of functional data analysis, as we have performed here, or in the context of dimension reduction. In the context of this paper, the aim was to understand the heterogeneity between patients with respect to the time course of response in depression. A 'classical' principal component analysis aims at dissecting a multi-dimensional endpoint in single dimensions, as has been performed repeatedly for the HAMD (Trivedi *et al.*, 2005; Fleck *et al.*, 1995, 2004). These two techniques can be considered complementary. Principal component analysis of the HAMD has revealed its multidimensionality, which has important consequences in the interpretation of the HAMD itself. In addition, it has led to the development of unidimensional subscales which are more sensitive to detect drug effect. In contrast, the functional data analysis reported here implicates the *time course* of the HAMD and indirectly the differences in response rate.

Our findings provide evidence that the heterogeneity in the response patterns of individual patients is real heterogeneity and not random noise, as indicated by the consistent findings across different subgroups and endpoints. The principal components that were found throughout this investigation clarify one important thing: apparently, the heterogeneity in the response patterns in depression is very similar in all subgroups tested. That is, the heterogeneity found in the overall study population is comparable to that in the

non-responder and responder subgroups. Moreover, evidence was found that the heterogeneity in the response patterns is primarily determined by the seven items present in the HAM-D₇ subscale, indicating that the use of a subscale will not limit the characterisation of the time course of response in patients. The importance of this finding is that it may lead to a consensus that subscales do tell us everything we need to know about the time course of depression without the noise added by the remaining items of the HAM-D₁₇. We do not advocate abolishing the remaining items altogether, but rather recommend that data analysis in clinical trials, both graphical and statistical, should be primarily based upon the HAM-D₇ or any of the other available subscales. Eventually, this knowledge may also contribute to the development of better longitudinal models for the assessment of drug effect.

Methodological aspects of the FPCA and longitudinal modelling of response

The main objective of a statistical model for longitudinal data in clinical development is hypothesis testing. Consequently, the main focus is on the estimation of mean profiles, discarding information about the individual profiles. Most clinicians think in terms of patients rather than in terms of treatment arms, clinical studies and statistical significance. Furthermore, individual patient data may be directly linked to relevant covariates; it may support the detection of specific subgroups in the population or allow extrapolation in a more meaningful way than aggregate data. Statistical models which require only baseline and last observation cannot capture individual differences. It is part of the biostatistician's job to bring these different levels together, and it is our firm belief that model parameterisation plays an important role in this context.

The current approach, based on functional data analysis, is inspired by this same idea. Why not investigate the main differences between individual curves rather than "binning" data? Results may lead to better statistical models, possibly solving some of the problems that drug development in depression is facing.

The first principal component indicated that the main difference between patients is a vertical shift, providing strong evidence for the parameterisation of the linear mixed model with an additive random effect. The percentage of variability that was explained by this component was between 54-69% (SVD) or 63-73% (PPCA) for the different studies. T-tests found that statistically significant differences exist in the estimate values of this component between different treatment groups. One may expect that an efficacious treatment will cause a more pronounced reduction in the HAMD than placebo. Hence, it is not surprising that a large portion of this response is mediated through the first principal component.

The second principal component indicated that some patients were better off than the population mean at the start of the trial, but ended worse, and *vice versa*. This includes the patients who have a later but more accelerated onset of response and those who show an initial response to treatment but then experience worsening of the symptoms (12-16% SVD, 10-13% PPCA). Significant differences were found between the treatment groups in

some studies for the scores of this component. These differences indicate that patients on placebo treatment tended to be less depressed than the average at the earlier phases of the trial, but more depressed in later stages. An investigation of the consequences of the incorporation of this second principal component as an additional random effect in a linear mixed model is part of an ongoing investigation by our group.

The third component consisted of a sudden amelioration or deterioration in weeks 3 and 4 of the study, a feature which was also observed in some typical patients (5-8% SVD, 5-7% PPCA).

The fourth principal component, only explaining 4% of the variance, has an oscillatory component with patients first performing better and subsequently worse than the average patient (and *vice versa*). This oscillatory behaviour was the basis for a set-point model that was used to describe the interaction of paroxetine and pindolol (Gruwez *et al.*, 2005), and the interaction between clomipramine and lithium (Gruwez *et al.*, 2007). A closer inspection of this model using the available data revealed that some model parameters were not identifiable. Lack of identifiability is also seen in the aforementioned paper describing application of the set-point model to the clomipramine-lithium interaction, since the rate of disappearance of clomipramine was estimated at the lower boundary of this parameter. The fact that only 4% of the variance was explained by an oscillatory principal component may explain the identifiability issues with this otherwise interesting model.

Heterogeneity between patients and groups

No differences were found for the heterogeneity in the response of patients receiving active treatment or placebo. This suggests that drug treatment modulates the rate of change in response patterns rather than modifies the actual course of disease, as indicated by the range of parameter estimates for each component.

Another interesting finding is the similarity of the main principal components across responders and non-responders. This may indicate that there is no dichotomous difference between patients who respond and those who do not, but rather a continuum of responsiveness to drugs. Indeed, the difference between responders and non-responders may be explained by other important factors such as drug exposure and disease severity.

The consistency of the same principal components across the use of different endpoints, in this case the full HAM-D₁₇ and the HAM-D₇ subscale, indicates that the main characteristics of the profiles appear to be maintained in spite of the removal of 10 items. This is a strong argument for further use of the HAM-D₇ in clinical trials. This would allow a considerable reduction in the number of patients required to demonstrate efficacy (chapters 3 and 4). Unfortunately, decision-makers and regulatory agencies continue to maintain a conservative attitude with regard to endpoints. This conservatism is illustrated by the Bech subscale (Bech and Rafaelsen, 1980), which was published in 1980 but is still not used as primary endpoint in many trials. The current work shows that the information on the variability of the profiles is preserved. The items which were removed

do in fact contribute mostly to noise.

Another important finding with regard to the heterogeneity of response is that the *typical patient* does not exist. Each patient has a specific time course of depression severity, and the 'typical' average profile is not representative of any single individual patient. Therefore it is more informative to display individual patient curves alongside average curves. In depression, where many misconceptions exist about the behaviour of individual patients (Stassen and Angst, 1998), a closer look on individual data may be very useful. Many clinicians are under the impression that an antidepressant effect is only discernable after 2 to 4 weeks of treatment. Our data show that a considerable number of patients respond much faster. Mean profiles from clinical trials in depression confirm this: although *separation* between drug and placebo usually occurs after 2-4 weeks, the improvement of each treatment group is observed as rapidly as after 1 week of treatment, albeit only a few points on the HAMD scale.

Other applications of FPCA

An extension of the analysis to other phenotypes of disease was not possible, as our database contained only patients diagnosed with major depressive disorder. It is conceivable, however, that the different principal components may be linked to specific phenotypes or patient subgroups. For example, it is possible that a given principal component is more pronounced or shows higher parameter values in certain types of patients. In addition, we investigated whether the parameter values for each of the principal components correlated with the contribution that each individual item of the HAMD had in the overall response at completion of the treatment. For instance, patients mostly responding by a change in the anxiety-related items could show higher parameter values for the fourth principal component (oscillatory behaviour). We did not find a compelling relation between any of the components and items: the percentage of variability in each component explained by the contribution of each item to the overall response (r^2) always remained below 5%. However, it is also plausible that a correlation exists between these principal components and genetic phenotypes or other disease covariates. Our dataset did not allow such investigations.

In conclusion, we have shown that characterisation of the heterogeneity in the time course of response is essential to further understanding of disease features and treatment effect. Moreover, we can state that heterogeneity in response is not a random effect or noise and that current therapies do not alter the underlying structure of response. They rather modulate the rate of change over the course of treatment. No specific features differentiate response patterns in responders and non-responders. The differences in treatment outcome are likely to be explained by differences in drug exposure and disease severity. These results should be taken into account in the development of new statistical models for the analysis of treatment efficacy in depression.

REFERENCES

- Bech P and Rafaelsen OJ (1980) The use of rating-scales exemplified by a comparison of the Hamilton and the Bech-Rafaelsen melancholia scale. *Acta Psychiatr Scand* **62**:128-132.
- Fleck MPA, Poirierlittre MF, Guelfi JD, Bourdel MC, and Loo H (1995) Factorial structure of the 17-item Hamilton depression rating-scale. *Acta Psychiatr Scand* **92**:168-172.
- Fleck MPD, Chaves MLF, Poirier-Littre MF, Bourdel MC, Loo H, and Guelfi JD (2004) Depression in France and Brazil - factorial structure of the 17-item Hamilton depression scale in inpatients. *J Nerv Ment Dis* **192**:103-110.
- Goekoop JG, de Beurs E, and Zitman FG (2007) Four-dimensional structure underlying scales for depression anxiety and retardation: emergence of trapped anger and scale improvements. *Compr Psychiatry* **48**:192-198.
- Golden RN, Nemeroff CB, McSorley P, Pitts CD, and Dube EM (2002) Efficacy and tolerability of controlled-release and immediate-release paroxetine in the treatment of depression. *J Clin Psychiatry* **63**:577-584.
- Gruwez B, Dauphin A, and Tod M (2005) A mathematical model for paroxetine antidepressant effect time course and its interaction with pindolol. *J Pharmacokinet Pharmacodyn* **32**:663-683.
- Gruwez B, Poirier MF, Dauphin A, Olie JP, and Tod M (2007) A kinetic-pharmacodynamic model for clinical trial simulation of antidepressant action: Application to clomipramine-lithium interaction. *Contemp Clin Trials* **28**:276-287.
- Lam D, Wright K, and Smith N (2004) Dysfunctional assumptions in bipolar disorder. *J Affect Disord* **79**:193-199.
- Mallinckrodt C, Clark W, and David S (2001) Accounting for dropout bias using mixed-effects models. *J Biopharm Stat* **11**:9-21.
- Newell J, McMillan K, Grant S, and McCabe G (2006) Using functional data analysis to summarise and interpret lactate curves. *Comput Biol Med* **36**:262-275.
- R Development Core Team (2007) *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, ISBN 3-900051-07-0.
- Ramsay JO, Altman N, and Bock RD (1994) Variation in height acceleration in the fels growth data. *Can J Stat-Rev Can Stat* **22**:89-102.
- Ramsay JO and Dalzell CJ (1991) Some tools for functional data-analysis. *J R Stat Soc Ser B-Methodol* **53**:539-572.
- Ramsay JO, Wang X, and Flanagan R (1995) A functional data-analysis of the pinch force of human fingers. *J R Stat Soc Ser C Appl Stat* **44**:17-30.
- Rapaport MH, Schneider LS, Dunner DL, Davies JT, and Pitts CD (2003) Efficacy of controlled-release paroxetine in the treatment of late-life depression. *J Clin Psychiatry* **64**:1065-1074.
- Rice JA and Silverman BW (1991) Estimating the mean and covariance structure nonparametrically when the data are curves. *J R Stat Soc Ser B-Methodol* **53**:233-243.
- Roweis S (1997) EM Algorithms for PCA and SPCA, in *Neural Information Processing Systems 10 (NIPS'97)* (Jordan M, Kearns M, and Solla S, eds.), pp. 626-632, The MIT Press.
- Stacklies W and Redestig H (2007) *pcaMethods: A collection of PCA methods*, R package version 1.2.3.
- Stassen HH and Angst J (1998) Delayed onset of action of antidepressants - fact or fiction? *CNS Drugs* **9**:177-184.
- Trivedi MH, Morris DW, Grannemann BD, and Mahadi S (2005) Symptom clusters as predictors of late response to antidepressant treatment. *J Clin Psychiatry* **66**:1064-1070.

- Trivedi MH, Pigott TA, Perera P, Dillingham KE, Carfagno ML, and Pitts CD (2004) Effectiveness of low doses of paroxetine controlled release in the treatment of major depressive disorder. *J Clin Psychiatry* **65**:1356-1364.
- Troyanskaya O, Cantor M, Sherlock G, Brown P, Hastie T, Tibshirani R, Botstein D, and Altman RB (2001) Missing value estimation methods for DNA microarrays. *Bioinformatics* **17**:520-525.