To fail or not to fail : clinical trials in depression

Sante, G.W.E.

SECTION III

DATA ANALYSIS

Chapter

6

Evaluation of treatment response in depression studies using a Bayesian parametric cure rate model

Gijs Santen, Meindert Danhof, Oscar Della Pasqua

ABSTRACT

Efficacy trials with anti-depressant drugs often fail to show significant treatment effects even though efficacious treatments are investigated. This failure can, amongst other factors, be attributed to the lack of sensitivity of the statistical methods and the endpoints to pharmacological activity. For regulatory purposes the most widely used efficacy endpoint is still the mean change in HAMD score at the end of the study, despite evidence from literature showing that the HAMD scale might not be a sensitive tool to assess drug effect and that changes from baseline at the end of treatment may not reflect the extent of response. In the current study, we evaluate the prospect of applying a Bayesian parametric cure rate model (CRM) to analyse anti-depressant effect in efficacy trials with paroxetine. The model is based on a survival approach, which allows for a fraction of surviving patients indefinitely after completion of treatment.

Data was extracted from GlaxoSmithKline's clinical database. Response was defined as a 50% change from baseline HAMD at any assessment time after start of therapy. Survival times were described by a lognormal distribution and drug effect was parameterised as a covariate on the fraction of non-responders.

The model was able to fit the data from different studies accurately and results show that response to treatment does not lag for two weeks, as is mythically believed. In conclusion, we demonstrate how parameterisation of a survival model can be used to characterise treatment response in depression trials. The method contrasts with the long-established snapshot on changes from baseline, as it incorporates the time course of response throughout treatment.

INTRODUCTION

The evaluation of the efficacy of new antidepressant compounds is an increasing challenge. For every antidepressant drug on the market several large clinical trials had to be conducted to obtain the minimum number of positive trials which are required for a successful registration. Indeed, it is currently assumed that up to 50% of the trials evaluating an antidepressant drug with proven efficacy could fail to show statistically significant separation of the active treatment from placebo (Khan *et al.*, 2000). Numerous reasons have been provided to explain this finding, ranging from large variability in placebo effect to poor sensitivity of the clinical endpoint up to the difficulties in identifying and enrolling patients who are truly sensitive to treatment (Montgomery, 1999; Thase, 2002). Regardless of the reasons, the consequence of such a high failure rate is that several large clinical studies are required in the early stage of clinical development to allow conclusive evaluation of the potential antidepressant activity of a new compound.

The Hamilton depression rating scale (Hamilton, 1960) (HAMD) is the gold standard in depression research. Currently, the analysis of depression studies is based on the difference between placebo and active treatment at the end of the study (usually 6-12

weeks) corrected for baseline HAMD. This approach neglects information yielded during the course of treatment. Furthermore, strong evidence exists that the variability in the HAMD is large and that some of its dimensions are non-specific to treatment response. Whilst it is a well-established means to describe disease condition, the HAMD scale cannot accurately capture the dynamics of state changes that occur during the course of treatment. In a previous investigation, we have shown that the use of subscales from the HAM-D$_{17}$, consisting only of items that were identified as sensitive to changes in response over time, improves the assessment of drug effect, as compared to the full scale (chapter 3).

In the current study, we propose the use of a *time to event* approach to characterise in each single patient the time required to reach a clinical response level which reflects the transition between clinically identifiable states, rather than relying on changes in the HAMD scores following a fixed treatment period. Despite the apparent dichotomisation of the clinical scale, assessing whether a patient has had at least a 50% decrease in the 17-item HAMD score from baseline reflects how response to treatment is evaluated clinically. In fact, the survival-analytic approach has previously been advocated for the analysis of treatment response in depression (Montgomery *et al.*, 2002). In the past years, numerous modifications to survival analysis have been proposed to address methodological requirements. The addition of a cure rate in a non-parametric sense has been suggested for the determination of onset of response (Laska and Siegel, 1995) and the use of sustained response instead of immediate response has also been recommended (Stassen *et al.*, 1993).

In order to account for the presence of non-responders in a study, i.e., patients who are refractory to treatment within the overall patient population, we advocate the use of a so-called cure rate model (CRM). The CRM provides a parametrical description of the survival process with an asymptote that encompasses the fraction of non-responders. Advantages of the use of a parametric model as opposed to a non-parametric approach in general include the possibility to extend beyond the range of observations and narrower confidence intervals. In the specific case of the presence of a cured fraction, a parametric approach has been shown to be more appropriate (Gamel and Vogel, 1997).

In our approach, model parameterisation is performed within a Bayesian context. The advantages of the Bayesian statistics over classical methods include the incorporation of prior knowledge and the possibility to make direct probability statements, both of which are required for the purposes of estimating and predicting response to treatment during an efficacy trial.

METHODS

Study data

Response data from 1286 patients from two double blind, placebo-controlled studies in Major Depressive Disorder (MDD) was available. In study 1, paroxetine and fluoxetine were

compared over a 12-week treatment period. The 17-item HAMD was measured at weeks 0, 1, 2, 3, 4, 6, 9 and 12. The randomisation ratio between placebo, paroxetine and fluoxetine was 1:2.5:2.5. Further details about this study can be found at GlaxoSmithKline's clinical trial register (CTR), http://ctr.gsk.co.uk (protocol PAR128).

Study 2 involved the comparison of two doses of paroxetine in a modified release formulation during the course of an 8-week treatment period. The HAM-D_{17} was measured at weeks 0, 1, 2, 3, 4, 6 and 8. Details about this study have been described by Trivedi *et al.* (2004). In addition, a historical data set was constructed consisting of 850 placebo-treated patients from published and unpublished trials in unipolar depression (DeVeaugh-Geiss *et al.*, 2000; Dunner and Dunbar, 1992; Golden *et al.*, 2002; Rapaport *et al.*, 2003). Information on the placebo response was incorporated into the analysis as *informative priors* on the model parameters of the distribution of the response times.

Clinical response

Treatment effect was evaluated at each scheduled visit. HAMD scores from individual patients were converted into survival data. Time of response was defined as the first occasion at which a decrease of at least 50% from baseline HAMD score was observed. Since the clinical visits occurred on a weekly basis, it is conceivable that treatment response might effectively have taken place at an earlier time point, prior to the visit. Such a potential censoring effect has not been factored in the current analysis, as it is assumed not to be clinically relevant.

Model

The CRM is a modified survival model with a residual surviving fraction. A Bayesian adaptation to the standard mixture model is used, as proposed by Chen *et al.* (1999). The fraction of subjects not responding at time *t* is given by equation 1:

$$S_{pop}(t) = \pi + (1 - \pi)S^*(t) \tag{1}$$

Where S_{pop} represents the population survival function, π is the fraction of non-responders, and $S^*(t)$ is the proper survival function for the responders, defined as:

$$S^*(t) = \frac{\pi^{F(t)} - \pi}{1 - \pi} \tag{2}$$

Here, *F(t)* is a cumulative distribution function of choice. It is important to note that the cure rate parameter π does not only determine the fraction of non-responders, but also exerts influence on the shape of the curve due to the cumulative distribution function, as can be seen from equation 2.

Bayesian framework

In Bayesian hierarchical modelling, all parameters are defined as random variables and their conditional independence can be represented by directed acyclic graphs. Figure 1 summarises the parameterisation for the cure rate model.

Priors

To integrate existing prior knowledge of survival rates, historical placebo data was used to specify a strong informative prior on the treatment-independent parameters. This informative prior was determined by fitting the model to the historical data only and summarising the posterior distributions for treatment-independent parameters using normal and lognormal distributions. Subsequently, we have established that the parametric form of the resulting posterior distribution did not influence model results. A normal distribution was therefore used as informative prior in the final model. Although different prior distributions for variance were specified in these fits (inverse gamma, uniform, half-normal), they had little effect on the resulting posterior estimates. Non-informative flat normal priors were used for the treatment-specific parameter (fraction of non-responders, see below).

Goodness of fit & modelling diagnostics

Data fitting was performed in WinBUGS (version 1.4.1) (Lunn *et al.*, 2000). Appropriateness of parametric distributions and drug model was assessed using the deviance information criterion (DIC) as a measure of the goodness-of-fit (Spiegelhalter *et al.*, 2002). For all runs
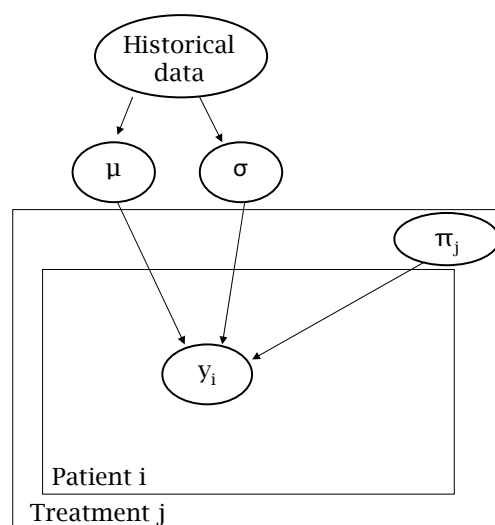

Figure 1. A directed acyclic graph (DAG) representing the conditional independence between random variables in the cure rate model. The historical data is used as an informative prior on the distribution of the log survival times, namely the mean (μ) and standard deviation (σ). Treatment effect is described by the proportion of non-responders π. Y_i indicates the observed individual response variable.

presented in this paper, 50,000 burn-in iterations were performed followed by 100,000 iterations. The number of iterations was deliberately chosen this high to exclude any residual correlation. Convergence was determined using the Gelman-Rubin test statistic (Brooks and Gelman, 1998; Gelman and Rubin, 1992) and visually using history and autocorrelation plots.

Optimisation of model structure

Study 2 was used for model optimisation. The lognormal, Weibull, exponential, normal and loglogistic distributions were fitted to the data using the cure rate model. Although the DIC was similar for most distributions, the lognormal distribution was selected since it showed the lowest value. Parameterisation of drug effect was explored by evaluating its influence on the mean and/or variance of the distribution of survival times and on the cure rate (i.e., fraction of non-responders). According to the DIC, the latter approach was the most appropriate one to incorporate drug effect into the model. Therefore, for the remainder of this paper, drug action will be assumed to be mediated only through a change in the fraction of non-responders.

Differences in treatment effect

Treatment effect was estimated by computing the posterior distribution for the difference in the fraction of non-responders between active treatment(s) and placebo or between active treatment and comparator. If the 95% credible interval of this distribution included zero, the treatment difference was not considered statistically significant. This was further summarised in the posterior probability of superiority (PPS), which is a measure of the probability that the drug is superior to placebo or comparator treatment, i.e., the surface of the posterior distribution of the difference between treatments which favours the drug. It offers the possibility to capture drug effect in a single number, enhancing the clarity in the interpretation of the findings. For comparison purposes, *p*-values for differences in treatment effect were also calculated using the Cox proportional hazards model (COXPH), the mixed model for repeated measures (MMRM) (Mallinckrodt *et al.*, 2004) with baseline-time and treatment-time interactions as fixed effects and last observation carried forward (LOCF) imputation followed by a standard *t*-test.

RESULTS

Model fits

The model fits for data from both clinical studies are shown in figure 2. As indicated by the prediction lines, the model can accurately describe the Kaplan-Meier curve. It is important to note that discrepancies between data and prediction are partially due to the sampling frequency (weekly visit).

An overview of the goodness-of-fit is presented in figure 3. The diagnostic plots display the observed *versus* model-predicted surviving fractions and the corresponding residual fractions over time. For model predictions, a response assessment was simulated every 5 days.

The final parameter estimates and 95% credible intervals are summarised in table 1.
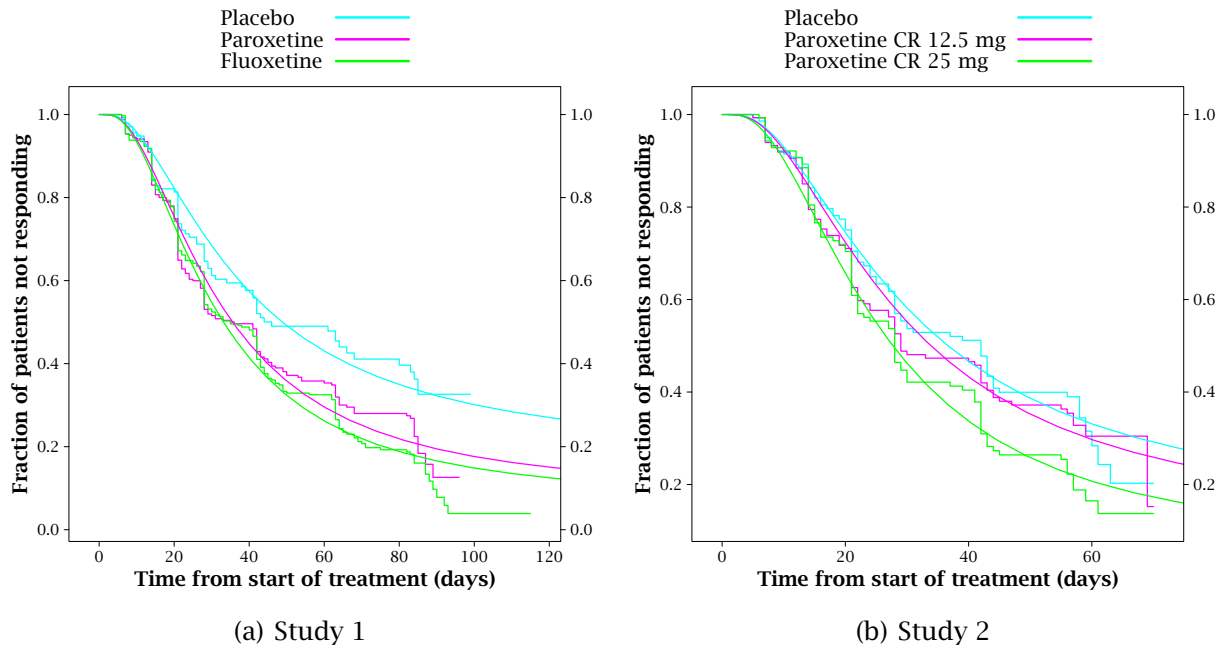

Figure 2. Fits of the cure rate model to data from study 1 and 2. Crooked lines show Kaplan-Meier plots whereas the corresponding mode fits are represented by smooth lines

Table 1. Median and 95% credible intervals for estimated parameters in study 1 and 2. μ and σ are the mean and standard deviation of the log survival times. π represents the percentage of non-responders for each treatment

Study	Parameter	median	95% Credible interval
1	μ	4.115	3.95-4.30
	σ	0.931	0.860-1.01
	$\pi_{placebo}$	17.8	10.4-27.2
	$\pi_{paroxetine}$	8.3	4.6-13.0
	$\pi_{fluoxetine}$	6.5	3.5-10.5
2	μ	3.96	3.76-4.18
	σ	0.923	0.838-1.01
	$\pi_{placebo}$	13.7	7.0-22.5
	$\pi_{paroxetine\ 12.5\ mg}$	11.3	5.5-19.0
	$\pi_{paroxetine\ 25\ mg}$	5.9	2.3-11.8
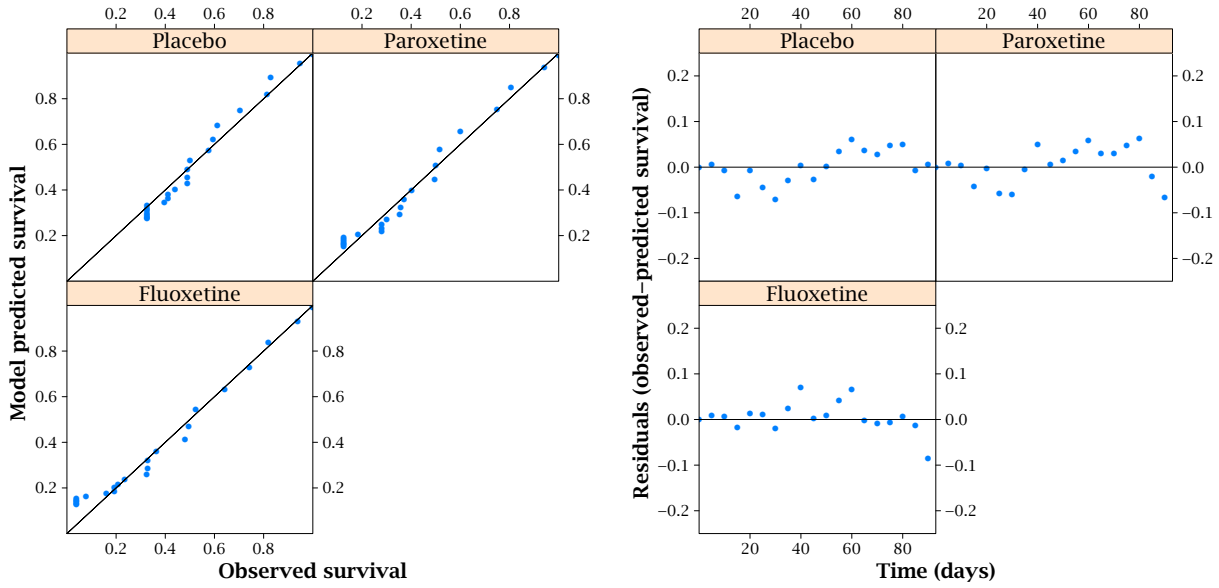
(a) Observed *versus* Predicted for study 1

(b) Residuals *versus* time for study 1
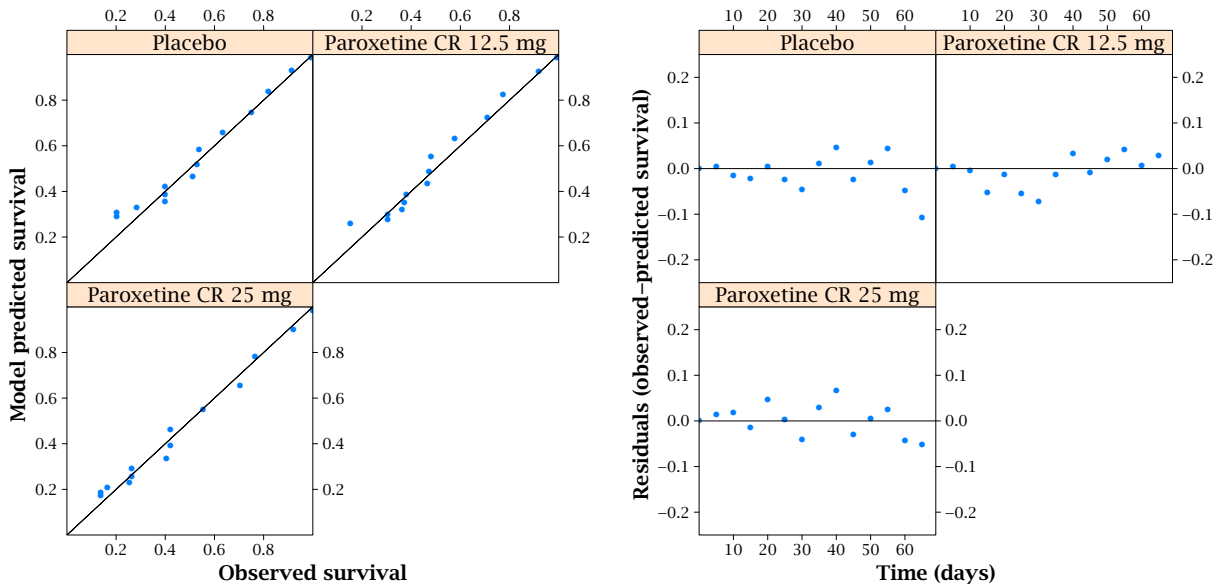
(c) Observed *versus* Predicted for study 2

(d) Residuals *versus* time for study 2

Figure 3. Goodness-of-fit plots of the fit for both studies.

Treatment effect

The estimates for the posterior probability of significance (PPS) for the differences between treatments are shown in table 2. The statistical significance levels for the same treatment comparisons are also presented for the COXPH, MMRM and LOCF methods. Further evidence of the significance level of treatment differences is provided in figure 4 by graphical representation of posterior probability distribution.

The density plots for study 1 illustrate that both active treatments are superior to placebo. In study 2, the density plots reveal that paroxetine CR formulation (25 mg) is clearly a better treatment than placebo, and also that it seems to differ from the 12.5 mg

dose. Such a separation was not detectable by the COXPH, MMRM or LOCF methods (table 2). In this study, the 12.5 mg dose is found not to be statistically different from placebo by any statistical endpoint.

Table 2. Significance level of the treatment effects for different study arms using the cure rate model (CRM), the Cox proportional hazard (COXPH) model, the mixed model for repeated measures (MMRM) and a *t*-test based on last observation carried forward imputation (LOCF). *p*-values are given for the COXPH, MMRM and LOCF models, whilst posterior probabilities of superiority (PPS) and the complementary Bayesian one-sided "*p*-values" (1-PPS) are presented for the CRM. Usually the threshold for statistical significance in clinical trials is set at *p*<0.05. Treatment comparisons for which statistical significance was not reached are shown in bold.

Study	Comparison	PPS	Survival data		HAMD data	
			CRM (1-PPS)	Cox	MMRM	LOCF
1	placebo-paroxetine	0.9975	0.0025	0.0067	0.0022	0.0162
	placebo-fluoxetine	0.9999	0.0001	0.0006	0.0009	0.0046
	paroxetine-fluoxetine	**0.8420**	**0.1580**	**0.34**	**0.7592**	**0.5232**
2	placebo-12.5 mg paroxetine	**0.7303**	**0.2697**	**0.540**	**0.0703**	**0.0713**
	placebo-25 mg paroxetine	0.9914	0.0086	0.016	0.0006	0.0018
	12.5 mg-25 mg paroxetine	0.9640	**0.0360**	**0.07**	**0.0831**	**0.1647**

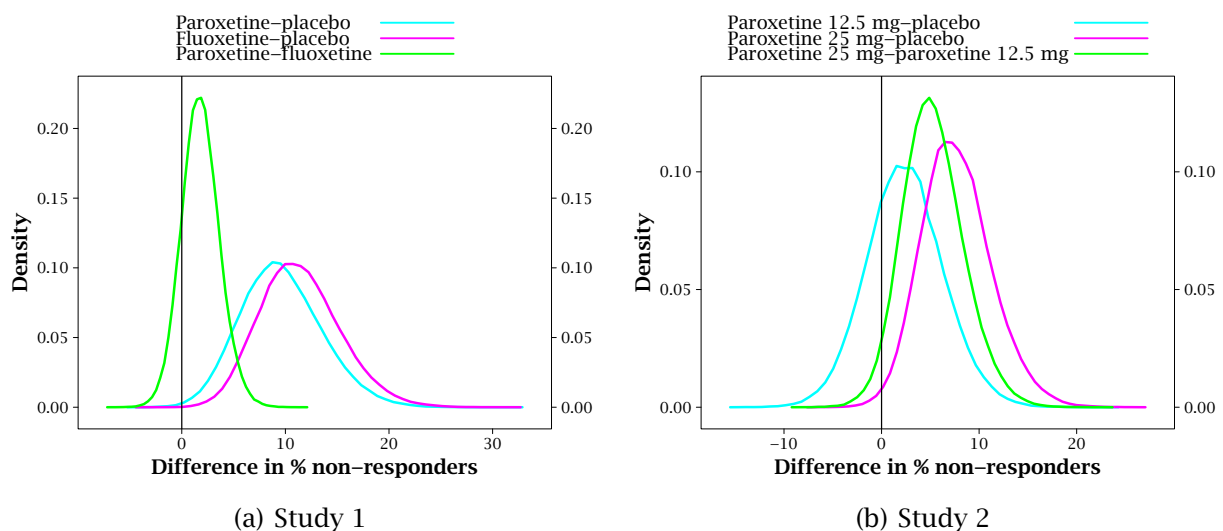(a) Study 1   (b) Study 2

Figure 4. Graphical representation of the posterior probability distribution. The density plots for study 1 illustrate that both active treatments are superior to placebo. In study 2, the density plots reveal that paroxetine CR formulation (25 mg) is clearly a better treatment than placebo, and also that it differs from the 12.5 mg doses. In this study, the 12.5 mg dose is not statistically different from placebo.

DISCUSSION

This paper shows the first application of the Bayesian cure rate model on clinical data from depression studies. In contrast to non-parametric methods, which are usually descriptive in nature, the choice for a parametric approach to characterise time to response was driven by the need to estimate and predict the rate and variability in placebo response during the course of treatment.

Variable placebo response remains one of the major causes of failure in demonstrating statistically significant differences in clinical studies with antidepressant drugs. A model-based approach that takes into account historical evidence as well as allows for updates in prior information may enable accurate prediction of treatment effect size and anticipate deviations of response from estimated distributions. In addition, parameterisation of drug effect on the fraction of non-responders is particularly appealing, as it creates a possibility to evaluate drug effect independently of the mechanism of action. This aspect has been highlighted previously in modelling research in different therapeutic areas, in which parameterisation attempts aim at identifying and separating drug- and system-related parameters (Maas *et al.*, 2006). Such an approach envisages better estimation of treatment effect under non-stationary conditions.

Numerous reviews are available discussing different methodologies to analyse unipolar and bipolar depression trial data (Hennen, 2003; Montgomery *et al.*, 2002). The necessity for improvement in this area is outlined in a summary document of the ECNP consensus meeting (Montgomery, 1999). Although this meeting took place in 1997, the playing field does not appear to have changed much since then. In fact, regulatory authorities still request last observation carried forward (LOCF)-adjusted analyses, where the change from baseline at the end of the trial is compared between treatment and placebo, disregarding all data obtained between the first and last visit for each subject. A second method of analysis, the mixed model for repeated measures (MMRM) (Mallinckrodt *et al.*, 2004) uses all available longitudinal data, but treats the change from baseline as a continuous variable. However, evidence exists that the HAMD scale is multidimensional and it is therefore not plausible to treat it as a continuous response variable (Bech and Rafaelsen, 1980; Moller, 2001). In addition, we have demonstrated in a previous investigation (chapter 3) that the HAMD may not be suitable for this purpose, due to the varying sensitivity of the individual items to treatment response. The use of a subscale is proposed to overcome this problem. Since the HAMD was originally intended as a marker of disease state (Hamilton, 1960), another approach would be to dichotomise the HAMD into response and non-response.

A third possible method for analysis of longitudinal data from depression trials is a survival-analytic approach, which is proposed in this paper. This approach makes use of all data until either response occurs or a subject is censored due to dropout or end of follow-up. As argued above, dichotomising the HAMD is defendable and a loss of information is not necessarily expected. Several reviews indeed advocate this approach as a valid and statistically sound alternative (Hennen, 2003; Montgomery *et al.*, 2002; Thase, 2001).

Furthermore, the use of a parametric model prevents the constraints of proportionality in hazards imposed by non-parametric approaches. Based on a preliminary evaluation, we have also found out that the CRM enables prediction of trial outcome with reasonable accuracy well before trial completion. These findings have prompted us to investigate how to optimally implement the cure rate model within an interim analysis.

The consequences of differences in the sensitivity of a statistical method (i.e., false negative rate and study power) become evident when comparing the results obtained by the COXPH, MMRM and LOCF methods (table 2). Whilst all 4 methods provide evidence of significant differences between active treatment and placebo in study 1, this is not observed for the comparison between the two dose levels of paroxetine in study 2. Even though the 12.5 mg paroxetine dose group does not show separation from the higher dose group according to the standard statistical methods, the survival model seems to indicate that the two treatment arms are different. It is also clear from figure 2 that a survival model will not find any degree of separation between the 12.5 mg treatment arm and placebo. Such discrepancies between methods are not unexpected since the data used in the analyses varies from survival data to continuous longitudinal data. These differences highlight however the need to judge the relevance of a statistical method for the clinical research question under scrutiny.

From a statistical perspective, it is worth providing the reader with further considerations about the use of lognormal distribution for the survival times, the parameterisation of drug effect on the fraction of non-responders and model implementation within a Bayesian framework. The choices were made on practical and clinical grounds. The lognormal distribution has been used in other areas where survival times are modelled parametrically, see for example Tai *et al.* (2005). The percentage of responders is often defined in clinical study protocols as an important measure of drug effect at completion of treatment. This figure is also reported when studies are cited by non-scientific articles and by the general media. In contrast to the observed fraction of (non-)responders, the model-based asymptote proposed in this paper reflects the response rate beyond completion of the clinical trial.

Advantages

Modelling in a Bayesian context has various advantages:

(1) Estimated probabilities can be interpreted more directly. For example, in table 2, the posterior probability of significance is shown to be comparable to the *p*-values determined using a Cox proportional hazard model. However, the interpretation of these quantities is very different. The PPS has a direct interpretation, i.e., the probability that the drug is superior to placebo is 99%. Although a *p*-value is often interpreted in a similar manner, its interpretation is more cumbersome, i.e., 'assuming no difference between placebo and active treatment, the probability of observing the current data is 0.1%'.

(2) Incorporation of historical data is straightforward. In this case, we have implemented an informative prior on two of the three parameters of the model. This means

that less data is required to estimate the third parameter accurately. Another important aspect is the possibility to compare historical placebo response with the placebo effect in the current study. This comparison may flag an unusually large placebo effect. In fact, it has been claimed that a clear trend in placebo effect over time exists (Walsh *et al.*, 2002). This change may be caused by a change in expectancy by the population and the fact that most patients in a trial have already been treated with an anti-depressant. Interestingly, we did not find such an effect in the limited dataset included in this analysis using the cure rate model (8 placebo-arms from studies dating from 1985 to 2002).

(3) The advantages also include the possibility of introducing drug effect on the proportion of non-responders and, as in all survival models, the possibility to incorporate censored data without resorting to last observation carried forward (LOCF) approaches.

Potential limitations

(1) We chose not to use the definition of sustained response (Stassen *et al.*, 1993) as a criterion for response because this work was considered in the context of the search for a new methodology for interim analysis of efficacy trials. The time constraints that are imposed in the definition of sustained response may be difficult to implement with interim data (e.g., patients may be changing from responder to non-responder because of consecutive measurements).

(2) With drug effect only on the asymptotic proportion of non-responders, differences in the onset of effect may be missed. Even though the antidepressant drugs (SSRIs and TCAs) analysed thus far with this model do not appear to differ in terms of the timing of the onset of action, it is conceivable that novel targets may show a faster onset of action. Therefore it would be advisable to test the most appropriate parameterisation of drug effect when antidepressants with new mechanisms of action are included in a trial.

(3) The impact of different randomisation ratios has not been investigated. For instance, the inclusion of fewer placebo patients may decrease the precision of the estimated percentage of non-responders in the placebo arm and therefore decrease the power to detect statistical differences in treatment effect. As a matter of fact, randomisation ratio and stratification rules must be considered carefully in any statistical analysis.

(4) Another issue in the analysis of antidepressant trials is whether fixed or titrated dose regimens are applied or a placebo run-in phase is used. The informative prior used on the distribution of response times leaves enough flexibility to account for small differences in onset of effect due to titrated study designs. A meta-analysis has shown that the effect size of trials with a placebo run-in phase did not differ significantly from trials without such a phase (Lee *et al.*, 2004) and should no longer be applied (Montgomery, 1999). The change in the endpoint is merely postponed and starts when investigator and subject know the placebo run-in has ended. The priors on the distribution of the response times can accommodate this possible delay.

(5) Much computational power is required for the analysis. In order to incorporate historical information on placebo response two methods were evaluated. First the his-

torical data was added to the current study data as an extra study arm. This approach allows direct hypothesis testing on differences between the current placebo arm and the historical placebo arm with regard to the asymptotic fraction of non-responders. The addition of 850 patients to the model did, however, slow the MCMC algorithm considerably. Historical information was therefore summarised as described in the methods section. The results were practically indistinguishable from the pooling method, and run times were reduced by 50% (to approximately 3 h for 150,000 iterations with 3 chains on a fast Pentium IV computer).

(6) Any survival analysis depends upon the criterion used for response. In the current investigation, we have applied a clinically accepted definition of treatment response. It was not our objective to explore model sensitivity to varying degrees of change in HAMD score over time.

In conclusion, our results show how a parametric Bayesian approach can be used to describe time to response in depression trials and overcome one of the main limitations of current methodology for the analysis of longitudinal data. Moreover, we show how historical data can be integrated into the statistical analysis to improve estimation and prediction of placebo response. Future research into the application of the cure rate model encompasses its use as an interim analysis tool, which may allow for early termination of unsuccessful trials.

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.

Brooks SP and Gelman A (1998) General methods for monitoring convergence of iterative simulations. *J Comput Graph Stat* **7**:434–455.

Chen MH, Ibrahim JG, and Sinha D (1999) A new Bayesian model for survival data with a surviving fraction. *J Am Stat Assoc* **94**:909–919.

DeVeaugh-Geiss J, Ascher J, and Brook S (2000) Safety and tolerability of lamotrigine in controlled monotherapy trials in mood disorders, in *39th ACNP Annual meeting*, San Juan, Puerto Rico.

Dunner DL and Dunbar GC (1992) Optimal dose regimen for paroxetine. *J Clin Psychiatry* **53**:21–26.

Gamel JW and Vogel RL (1997) Comparison of parametric and non-parametric survival methods using simulated clinical data. *Stat Med* **16**:1629–1643.

Gelman A and Rubin D (1992) Inference from iterative simulation using multiple sequences. *Stat Sci* **7**:457–511.

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.

Hamilton M (1960) A rating scale for depression. *J Neurol Neurosurg Psychiatry* **23**:56–62.

Hennen J (2003) Statistical methods for longitudinal research on bipolar disorders. *Bipolar Disord* **5**:156–168.

Khan A, Warner HA, and Brown WA (2000) Symptom reduction and suicide risk in patients treated with placebo in antidepressant clinical trials - an analysis of the food and drug administration database. *Arch Gen Psychiatry* **57**:311–317.

Laska EM and Siegel C (1995) Characterizing onset in psychopharmacological clinical-trials. *Psychopharmacol Bull* **31**:29–35.

Lee S, Walker JR, Jakul L, and Sexton K (2004) Does elimination of placebo responders in a placebo run-in increase the treatment effect in randomized clinical trials? A meta-analytic evaluation. *Depress Anxiety* **19**:10–19.

Lunn DJ, Thomas A, Best N, and Spiegelhalter D (2000) WinBUGS - A Bayesian modelling framework: Concepts, structure, and extensibility. *Stat Comput* **10**:325–337.

Maas HJ, Danhof M, and Pasqua OED (2006) Prediction of headache response in migraine treatment. *Cephalalgia* **26**:416–422.

Mallinckrodt C, Kaiser C, Watkin J, Molenberghs G, and Carroll R (2004) The effect of correlation structure on treatment contrasts estimated from incomplete clinical trial data with likelihood-based repeated measures compared with last observation carried forward ANOVA. *Clin Trials* **1**:477–489.

Moller H (2001) Methodological aspects in the assessment of severity of depression by the Hamilton depression scale. *Eur Arch Psychiatry Clin Neurosci* **251 Suppl 2**:II13–II20.

Montgomery S (1999) The failure of placebo-controlled studies. ECNP consensus meeting, September 13, 1997, Vienna. European College of Neuropsychopharmacology. *Eur Neuropsychopharmacol* **9**:271–276.

Montgomery SA, Bech P, Blier P, Moller HJ, Nierenberg AA, Pinder RM, Quitkin FM, Reimitz PE, Rosenbaum JF, Rush AJ, Stassen HH, and Thase ME (2002) Selecting methodologies for the evaluation of differences in time to response between antidepressants. *J Clin Psychiatry* **63**:694–699.

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.

Spiegelhalter DJ, Best NG, Carlin BR, and van der Linde A (2002) Bayesian measures of model complexity and fit. *J R Stat Soc Ser B-Methodol* **64**:583–616.

Stassen HH, Delini-Stula A, and Angst J (1993) Time course of improvement under antidepressant treatment: a survival-analytical approach. *Eur Neuropsychopharmacol* **3**:127–135.

Tai P, Yu EW, Cserni G, Vlastos G, Royce M, Kunkler I, and Vinh-Hung V (2005) Minimum follow-up time required for the estimation of statistical cure of cancer patients: verification using data from 42 cancer sites in the SEER database. *BMC Cancer* **5**.

Thase ME (2001) Methodology to measure onset of action. *J Clin Psychiatry* **62**:18–21.

Thase ME (2002) Studying new antidepressants: If there were a light at the end of the tunnel, could we see it? *J Clin Psychiatry* **63**:24–28.

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.

Walsh BT, Seidman SN, Sysko R, and Gould M (2002) Placebo response in studies of major depression - Variable, substantial, and growing. *JAMA* **287**:1840–1847.