Comparing treatment effect in depression trials:
Mixed Model for Repeated Measures vs Linear Mixed Model

Gijs Santen(1), Meindert Danhof(1) and Oscar Della Pasqua(1,2)

Leiden/Amsterdam Center for Drug Research
P.O. Box 9502, 2300 RA Leiden, The Netherlands
Universiteit Leiden, Vrije Universiteit Amsterdam

(1) Division of Pharmacology, LACDR, Leiden University, The Netherlands
(2) Clinical Pharmacology and Discovery Medicine, GSK, UK

GlaxoSmithKline

BACKGROUND

Since 50% of depression trials fail, even for marketed anti-depressants, the analysis of depression trials is of great interest. One of the possible causes of failure is the high drop-out rate in these trials, which can be as high as 50%. To account for this factor, the Last Observation Carried Forward (LOCF) approach has been used to impute missing values in longitudinal studies. Despite the evidence of bias in data analysis, LOCF has been used because of its supposedly conservative nature.

In recent years a new methodology has become a common standard for the analysis of longitudinal data, the Mixed Model for Repeated Measures (MMRM)[1] (see box 1). A Google search on 'MMRM analysis' returned 9640 hits. The uptake of this method is based on its ability to deal with missingness at random, as opposed to LOCF. However, little evidence has been gathered about its performance in describing longitudinal data relatively to a linear mixed model (LMM) to fit the data.

Box 1

MMRM is rapidly being adopted as the method of choice for clinical trial data sets;

most statisticians view MMRM as superior to LOCF for handling the data of subjects who do not complete the treatment protocol.[2]

OBJECTIVES

The objective of this investigation was to:
- Explore the suitability of the MMRM and the LMM to fit longitudinal depression data using several goodness-of-fit (GOF) criteria
- Compare the results of data analysis based on the MMRM, LMM and LOCF approaches.

METHODS

Data & Data Management

Data from 8-week to 12-week double blind, placebo-controlled clinical studies in major depression were used. All data management and plotting was performed in R, the language and environment for statistical computing[3].

MMRM

The Mixed Model for Repeated Measures was implemented using *proc mixed* in SAS, version 9.1.
The HAMD scale was modelled using the interactions between baseline-time and treatment-time as fixed effects. A multivariate normal distribution with a common unstructured covariance matrix between individuals was used to fit the model to the observations (see formula below).

$$HAMD = time_i * treatment_j + time_i * baseline_k$$

LMM

The Linear Mixed Model was implemented in WinBUGS[4] version 1.41. The HAMD was modelled using the interactions between baseline-time and treatment-time as fixed effects. An additive random effect was also included in this model (see formula below). Missing values were imputed using the posterior predictive distribution. The mean of the posterior distributions was used for the GOF plots.

$$HAMD = time_i * treatment_j + time_i * baseline_k + \eta_k$$

STRUCTURE OF THE DATA

Longitudinal data sampled from a subject are highly correlated. This correlation is shown in figure 1.

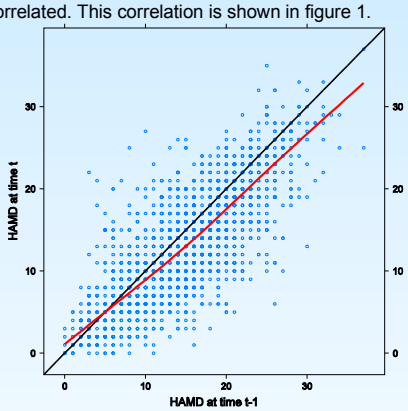

Figure 1. A scatter plot of each observation against the preceding observation in the same subject.

RESULTS

Figure 2 illustrates that the LMM fits the observed data better than the MMRM.

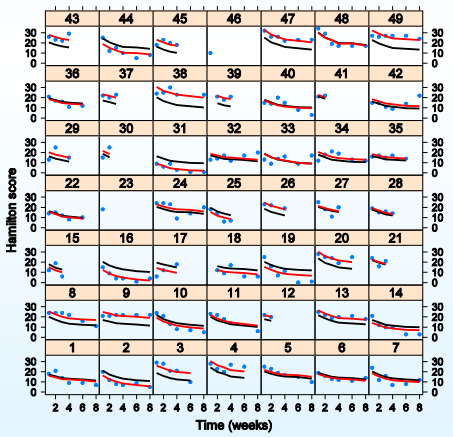

Figure 2. Observed (blue dots) and predicted HAMD vs. time (weeks) using the Linear Mixed Model (black) and the Mixed Model for Repeated Measures (red) (n=49).

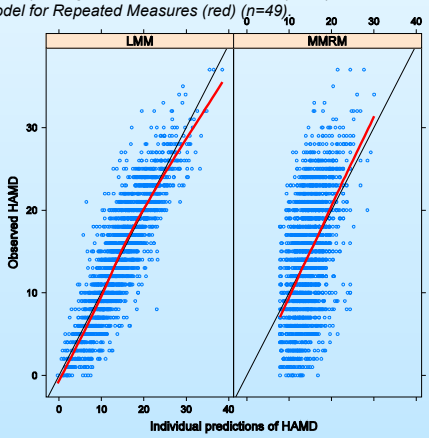

Figure 3. Observed vs. individual predictions for each model. The red line is a smoothing function. The HAMD values estimated by LMM are nearly unbiased. In contrast, MMRM predictions accurately capture only a limited range of values.

RESULTS CTND

As clearly shown in figure 4, the LMM accounts for the within-subject correlation whereas the MMRM does not.

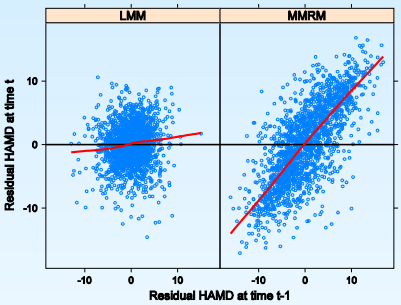

Figure 4. Residuals for each observation plotted against the residuals for the preceding observation in the same subject.

CONSEQUENCES

What are the consequences of the use of different models for the assessment of efficacy?
Table 1 shows the results of data analysis using each of the methods for 6 depression trials .

St	Arm	n	Δ effect	LMM	MMRM	LOCF
6	2	100	-3.34	0.0001	0.0006	0.0018
3	1	300	-2.58	0.0008	0.0009	0.0048
1	2	100	-2.65	0.003	0.0039	0.0138
2	2	100	-2.54	0.0054	0.0163	0.0076
3	2	300	-2.06	0.0094	0.0022	0.016
1	1	100	-2.88	0.0102	0.0452	0.2887
4	1	200	-1.86	0.0270	0.0366	0.1140
6	1	100	-1.62	0.034	0.0703	0.0713
5	2	100	-1.87	0.0498	0.0483	0.1618
2	1	100	-1.80	0.0768	0.0236	0.0197
4	2	200	-1.21	0.1442	0.1372	0.417
5	1	100	-0.13	0.1496	0.1020	0.0267

Table 1. Δ effect indicates mean difference between placebo and treatment at the end of treatment. Data is sorted by statistical significance as estimated by the LMM. The p-values for the MMRM, LMM and LOCF methods are against placebo. Red indicates p>0.05, green p<0.05. Discrepancies between methods reveal the differences in the sensitivity of methods.

The consequences of the differences between the LMM and the MMRM seem to be limited. However, there are instances in which a a treatment becomes statistically significant when a cut-off value of 0.05 is used. LOCF generally performs worse.

CONCLUSIONS

The MMRM does not describe longitudinal depression data accurately

- Our findings illustrate how the differences in the sensitivity of the methodology for data analysis may alter the conclusions drawn about the statistical significance of treatment effect.
- Graphical and statistical assessment of the goodness of fit is required for appropriate model evaluation.
- In the absence of drug exposure data, the LMM should be considered as first choice method for the analysis of depression trials.

1. Mallinckrodt CH, Clark WS, David SR: Accounting for dropout bias using mixed-effects models. J Biopharm.Stat. 2001; 11:9-21
2. Thase M., http://www.furiousseasons.com/archives/2007/01/more_on_serocquels_bolders.html. 2007
3. R Development Core Team. R: A language and environment for statistical computing. 2005. Vienna, Austria, R Foundation for Statistical Computing.
4. Gilks W.R. TAS, D.J.: A language and program for complex Bayesian modelling. The Statistician 2007; 43:169-178