INSTITUT DE STATISTIQUE

UNIVERSITÉ CATHOLIQUE DE LOUVAIN

DISCUSSION PAPER

0229

STATISTICAL MODELS FOR THE ANALYSIS OF CONTROLLED TRIALS ON ACUTE MIGRAINE

F. VANDENHENDE, P. LAMBERT and N. RAMADAN

http://www.stat.ucl.ac.be

Statistical Models for the Analysis of Controlled Trials on Acute Migraine

François Vandenhende[1,2], Philippe Lambert[2] and Nabih Ramadan[3,4]

[1] Statistics and Information Sciences, Eli Lilly & Company.
[2] Institute of Statistics, Université Catholique de Louvain.
[3] Medical-Neuroscience Research, Eli Lilly and Company.
[4] Indiana University School of Medicine.

November 12, 2002

SUMMARY. Specific efficacy criteria were defined by the International Headache Society for controlled clinical trials on acute migraine. They are derived from the pain profile and the timing of rescue medication intake. We present a methodology to improve the analysis of those trials. Instead of analysing each endpoint separately, we model the joint distribution and derive success rates in any criteria as predictions. We use cumulative regression models for each response at a time and a multivariate normal copula to model the dependence between responses. Parameters are estimated using maximum likelihood. Benefits of the method include a reduction in the number of tests performed and an increase in their power. The method is well suited to dose-response trials from which predictions can be used to select doses and optimise the design of subsequent trials. More generally, our method permits a very flexible modelling of longitudinal series of ordinal data.

KEY WORDS: Migraine; Ordinal Data; Repeated Measures; GEE; Multivariate Probit; Copula.

1. Introduction

Standards for the design, conduct and analysis of controlled trials on acute migraine were developed in 1991 by the International Headache Society (IHS) and revised 10 years later (Tfelt et al., 2000). Efficacy and safety of novel

email: francois@lilly.com. Correspondence to: Francois Vandenhende, Eli Lilly and Company, 11 rue Granbompre, B-1348 Mont-Saint-Guibert, Belgium. Phone: +32-10476469; Fax: +32-10476475.

anti-migraine therapies are generally assessed in double-blind, randomised, placebo-controlled trials treating migraine crises of moderate to severe intensity. After treatment, severity of headache and migraine-related symptoms (nausea, photophobia, phonophobia) are measured at several occasions relative to dosing (e.g. 0, 1, 2, 4, 24 hour) in patient diaries. Liker-type scales counting 4 grades of intensity (0-no, 1-mild, 2-moderate, 3-severe) are most often used. If needed, selected rescue therapies are available upon request from 2-hour post-dose.

The primary efficacy endpoint is *pain-free* (score ≤ 0) at 2-hour. Pain *relief*, being defined as a score ≤ 1 (no or mild pain), is a secondary response. Intake of *rescue* medications (RM) up to an observational time limit t (rescue $\leq t$) is another secondary measure. Pain-free and pain-relief are also assessed at other times than 2-hour. Only outcomes occurring prior to the first intake of RM are used in the evaluation. Hence, pain-free and pain-relief are observed at any time t if $(\text{score}_t \leq 0, \text{rescue} > t)$ and $(\text{score}_t \leq 1, \text{rescue} > t)$, respectively. The second part of the condition (rescue $> t$) only matters when $t >$2-hour, as RM is usually not allowed prior to that time.

Sustained pain-free (relief) is another secondary endpoint defined as pain-free (relief) without relapse through a score > 0 (> 1) for another period of observation (48 hours is an IHS standard). Sustained pain-free from any time t up to time t_{last} can thus be expressed as

$$\text{score}_t \leq 0, ..., \text{score}_{t_{\text{last}}} \leq 0, \text{rescue} > t_{\text{last}},$$

where pain scores recorded at all time points separating t and t_{last} are accounted for in the expression. Sustained pain-relief is defined in the same way using the corresponding score.

Thus, several efficacy outcomes are built as a joint condition on the series of pain scores up to some time and on the timing of RM. The distinction between pain-free and pain-relief is when no (0) or mild (1) scores are considered in the condition.

In practice, those efficacy outcomes are most often compared between treatments at each observation time using non-parametric statistical tests for binary responses. Chi-Square or Cochran-Mantel-Haenszel tests are commonly used in parallel trials (see e.g. Goldstein et al., 2001). McNemar's test is employed in Latin Square cross-overs (see e.g. Goldstein et al., 1997). The number of statistical tests performed can be very substantial, raising the problematic issue of the adjustment for multiple comparisons. The recommended IHS approach is to assess efficacy based on a single primary response, being 2-hour pain-free. In early phase dose-response trials, the evaluation of a unique primary outcome might turn too conservative because of the limited sample size involved.

In a recent submission package for acute migraine, the FDA has even requested for multiple primary endpoints (being all IHS criteria) to be considered in pivotal trials. An alternative inferential procedure that jointly assesses all IHS responses without a need to increase the sample size is presented here. Cumulative regression models for longitudinal ordinal measures are developed. We show how they yield predictions for the 5 main efficacy outcomes being (sustained) pain-free, (sustained) pain-relief and intake of rescues, across all times of measurements.

The layout of the paper is as follows. In Section 2, we analyse the pain severity (PS) response at a single time using cumulative regression models. Section 3 presents a GEE analysis of the whole PS profile prior to rescue availability. In Section 4, we model the joint distribution of the PS profile and the timing of RM using a multivariate probit model. An extension of that model is given in Section 5, using a copula representation of the joint distribution. We conclude with a discussion of key results.

2. One Pain Score at a Single Time

In a first instance, we analyse PS recorded at a single time (let say 2-hour). We consider the 4 possible events (score ≤ 0), (score ≤ 1), (score ≤ 2), and (score ≤ 3). The first two refer to pain-free and pain-relief, respectively. The third condition (score ≤ 2) expresses *pain-not-severe*. The last event (score ≤ 3) is always true in the considered scale, so that its probability is equal to 1. Events are nested into one another, as the score increases; hence, one has the following ordering of probabilities: $0 \leq \Pr(\text{score} \leq 0) \leq \Pr(\text{score} \leq 1) \leq \Pr(\text{score} \leq 2) \leq \Pr(\text{score} \leq 3) = 1$.

The expression $\Pr(\text{score} \leq k)$, for $k = 0, 1, 2, 3$ is the cumulative distribution of PS. We base our efficacy analysis at a single time on a generalised linear model (McCullagh and Nelder, 1989) for that cumulative distribution. As the distribution is bounded into the $[0, 1]$ interval, we first transform it to an unbounded function covering the whole real line. One such transformation is obtained with the well-known *logit* link: $\text{logit}(p) = \log[p/(1-p)]$, for $p \in [0, 1]$. For any k, $\text{logit}[\Pr(\text{score} \leq k)]$ can take any value (let say a_k) on the real line. As the cumulative distribution increases when k increases and since the logit function is also increasing, one has the ordering $-\infty \leq a_0 \leq a_1 \leq a_2 \leq a_3 = \infty$. The larger (smaller) a_k, the larger (smaller) the probability Pr(score $\leq k$) will be.

We consider the following logistic regression model for the analysis of pain scores at a single time:

$$\text{logit}[\Pr(\text{score} \leq k)] = a_k + \delta_d, \tag{1}$$

having a distinct parameter δ_d for each treatment or dose d. The a_k thresholds estimate the pain-free ($k = 0$), pain-relief ($k = 1$) and pain-not-severe($k = 2$) rates under placebo ($d = 0$) provided that we constrain $\delta_0 = 0$. The difference in pain distribution between patients treated with placebo and with drug d is captured in the estimate of the δ_d parameter. A positive value indicates an increase in probability and thus a better efficacy of drug d.

The above model assumes that the drug induces the same shift (on the logit scale) at each level of the pain distribution. This means that treatment-induced changes on the pain-free and pain relief responses are identical. More explicitly, $\exp(\delta_d)$ is equal to the ratio of the odds to be pain-free under treatment d versus placebo. As a single δ_d effect is included in the model, the treatment odds-ratio is assumed to be the same for the pain-free, pain-relief and pain-not-severe responses. As all odds are proportional between treatments, the model is referred as the *proportional-odds* model (McCullagh and Nelder, 1989, p. 153).

More general models that assume a treatment-specific odds-ratio for the pain-free, pain-relief and pain-not-severe responses are possible. To do so, we should add a subscript k to δ_{dk}. The validity of the reduced model having a single δ_d effect can then be assessed by testing that δ_{dk} parameters are equal for all k in the full model. This corresponds to a lack of fit test for the proportional odds assumption.

Let us illustrate the method in a real example. The considered trial is a double-blind, randomised, parallel, phase II trial assessing the efficacy of 3 oral doses of LY334370 (20, 60 and 200mg) against placebo in the treatment of an acute migraine crisis (Goldstein et al., 2001). LY334370 is a selective agonist of serotonin 1F receptors, which lacks vasoconstrictive effects in vitro and in vivo, unlike other migraine-specific acute therapies. Table 1 presents incidences of 2-hour pain responses in the full analysis set. For each response level, incidences are compared between groups using an ordinary chi-square test, producing three highly significant p-values. Alternatively, the whole

Table 1
Pain responses 2-hour after a single oral dose of LY334370 or a matching placebo in patients with acute migraine.

Group	Pain-Free N (%)	Pain-Relief N (%)	Pain-Not-Severe N (%)	All patients N
Placebo	1 (3)	7 (23)	18 (60)	30
LY 20mg	0 (0)	6 (22)	11 (41)	27
LY 60mg	8 (24)	16 (48)	28 (85)	33
LY 200mg	9 (38)	16 (67)	21 (88)	24
p-value	.001*	.002*	.001*	$< .001^{\$}$

*: ordinary chi-square test; $: Proportional odds model chi-square test.

pain distribution can be compared between groups using the proportional odds model ($p = .0003$). The hypothesis of proportional odds is not rejected ($p = .370$). The estimated odds-ratios (95% confidence intervals) for the 20, 60 and 200mg dose groups versus placebo are $0.55(0.21, 1.49)$, $3.43(1.38, 8.52)$, and $6.73(2.39, 18.93)$ respectively. LY334370 60mg and 200mg doses are significantly superior to placebo in relieving pain at 2-hour, unlike the 20mg dose for which the odds-ratio is not significantly different from 1.

3. Pain Score Profile over Time

Model (1) can easily be extended to analyse PS over a time period and not just after 2 hours. To do so, one can add a term θ_t for the time t effect and the treatment by time interaction (γ_{dt}) in the model. The assessment of the effect of treatment d on the probabilities $\Pr(\text{score}_t \leq k)$, $k = 0, 1, 2$ at any time t and the associated odds ratios are easily derived from the estimates of the overall (δ_d) and time-specific (γ_{dt}) treatment effects.

The estimation of model parameters and of their standard errors is more challenging when dealing with longitudinal series of ordinal responses than if they were all independent. Since PS collected at successive times on the same patient are correlated, this dependence needs to be accounted for in the estimation of parameters. We shall elaborate more on the topic of dependence in the following section. For now and for the purpose of inference, we use a statistical algorithm known as generalised estimating equations (GEE) that produces consistent (unbiased) parameter estimates in a longitudinal framework. The GEE method that we use to analyse ordinal pain scores and related properties are described in Lipsitz, Kim and Zhao (1994) and Miller, Davis and Landis (1993). We used the `genmod` procedure in version 8 of the `SAS` package for the estimation.

Figure 1 presents the distribution of PS at various times relative to dosing. Of the 116 treated patients, some PS were missing at selected times. Under placebo, there were one patient with no score at 0.5 hour and three at 4 hours. Two patients had no score at 4 hour in the 20 mg group. One patient in the 60 mg group had missing scores at 0.5 and 1 hours. In the 200 mg group, one patient did not respond before 4 hour, another after 0.5 hour and a last one had a missing score at 1.5 hour. Proportions presented in Figure 1 use all available cases. Missing data are mostly caused by patient unavailability at the time of recording. They are not related to the actual PS profile of the patient.

The GEE analysis of pain scores up to 2 hour (before any rescue) is presented in Table 2. Since missing data appear completely at random in the sense of Rubin (1976), a GEE analysis is justified (Park and Lee, 1997). As shown in Table 2, pain severity decreases with time in all groups ($p < .001$). The change in pain over time is different between treatments ($p = .013$). Indeed, pain relief increases with dose, confirming the overall efficacy of the drug ($p < .001$). Looking at treatment odds ratios over time, the 60 and 200 mg doses are superior to placebo from 1-hour after dosing. Efficacy of the 20mg dose is not demonstrated.

4. Sustained Responses

Sustained responses are responses that do not worsen over time. They are more robust markers of efficacy than direct responses because potential point-wise artifacts in individual assessments of pain are removed. They also help to identify and penalise possible relapses of the migraine. We have shown in Section 1 that the probability of sustained responses could be derived from the joint distribution of PS and RM. GEE methods are not of a direct use here because a full specification of the joint distribution is required to enable predictions (Lesaffre and Kaufmann, 1992). In this section, we construct a model for that joint distribution based on cumulative regressions for the marginal distributions and a function of those margins that models the dependence between them.

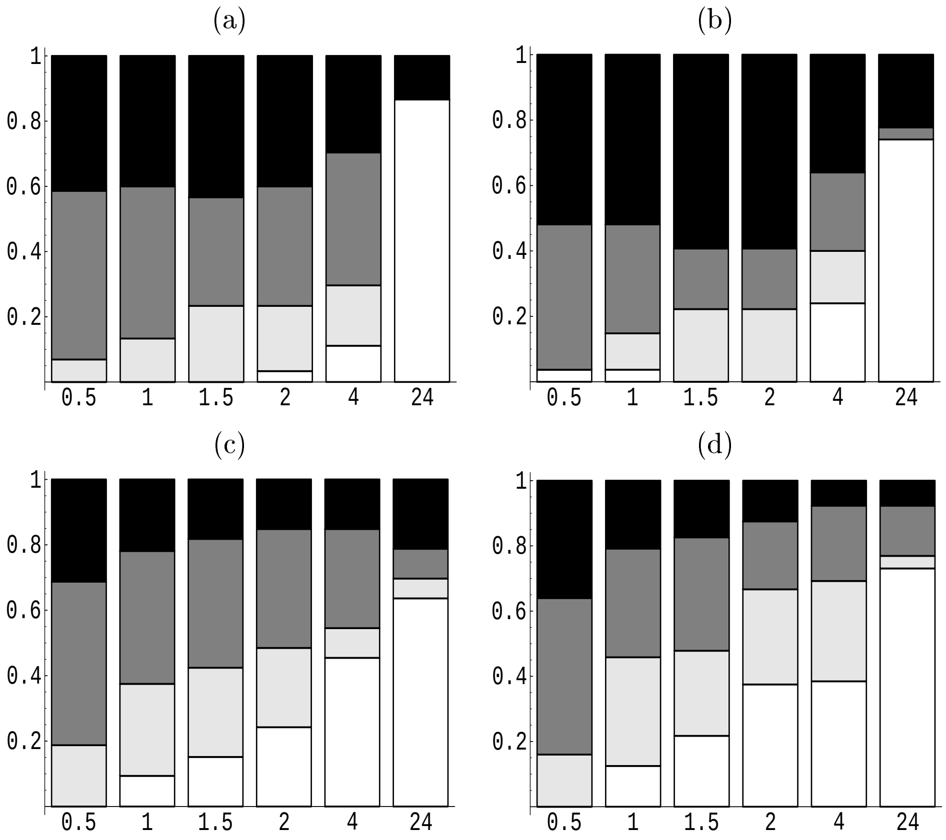

Figure 1. Incidence of pain severity (no: white, mild: light, moderate: dark, severe: black) over time after single oral doses of a) placebo, b) 20, c) 60 and d) 200 mg LY334370 in patients with acute migraine.

Table 2

GEE analysis of pain severity up to 2-hours post-dose.

Group	Pain	0.5h	1h	1.5h	2h
Placebo	No	0	0	0	1
	Mild	2	4	7	6
	Moderate	15	14	10	11
	Severe	12	12	13	12
LY20mg	No	1	1	0	0
	Mild	0	3	6	6
	Moderate	12	9	5	5
	Severe	14	14	16	16
	OR*	.71	.74	.59	.52
	(95% CI)	(.29, 1.77)	(.28, 1.93)	(.20, 1.69)	(.18, 1.50)
LY60mg	No	0	3	5	8
	Mild	6	9	9	8
	Moderate	16	13	13	12
	Severe	10	7	6	5
	OR*	1.59	2.75	3.28	4.17
	(95% CI)	(.70, 3.63)	(1.13, 6.68)	(1.28, 8.44)	(1.58, 11.02)
LY200mg	No	0	3	5	9
	Mild	4	8	6	7
	Moderate	12	8	8	5
	Severe	9	5	4	3
	OR*	1.34	3.56	4.24	9.24
	(95% CI)	(.55, 3.27)	(1.34, 9.41)	(1.46, 12.32)	(3.08, 27.73)
p-value$	Treatment	$< .001$			
	Time	$< .001$			
	Treatment*Time	.013			

*: Odds ratio and 95% confidence interval for the comparison to placebo.

$: Type III GEE chi-square test.

4.1 *Dependence and Copula*

We first present our developments in a simple example. Considering only 2 time points: $t = 1$ and $t_{\text{last}} = 2$ hours and assuming no intake of rescues, sustained relief at 1 hour is when $(\text{score}_1 \leq 1, \text{score}_2 \leq 1)$. The joint probability $\Pr(\text{score}_1 \leq k_1, \text{score}_2 \leq k_2)$ turns to be an increasing function of the two individual (called marginal) probabilities $\Pr(\text{score}_1 \leq k_1)$ and $\Pr(\text{score}_2 \leq k_2)$. It is also a function of the agreement in individual scores between 1 and 2 hours. To fix ideas, let us assume that pain relief rates are 50% both at 1 and 2-hours.

1. If the same patients relief at 1 and 2 hours, then the proportion of patients who relief both at 1 and 2-hours is also 50%.

2. If those who relief at 1 hour do not relief at 2 hours, the proportion that relieves both at 1 and at 2-hours is 0.

3. If 50% of those who relief at 1 do not relief at 2-hour and vice versa, then the joint proportion is 25%.

This last scenario is well known as it illustrates independence between relieves at 1 and 2 hours: the joint probability is equal to the product of the marginal probabilities. Scenario 1 illustrates positive, and even complete positive dependence between the two events. Positive dependence induces a joint probability that is superior to the corresponding probability under independence. Conversely, as illustrated in Scenario 2, when relieves at 1 and 2 hours are negatively correlated, the joint probability is lower than the product of independence. It was even equal to 0 in the second example of complete negative dependence.

Sklar (1959) showed that any joint distribution can be expressed as a function of the univariate marginal distributions. Hence, one can write

$$\Pr(\text{score}_1 \leq k_1, \text{score}_2 \leq k_2) = C[\Pr(\text{score}_1 \leq k_1), \Pr(\text{score}_2 \leq k_2)], \qquad (2)$$

where C is uniquely defined on the range of the margins. As C couples the margins and quantifies their dependence, it has been called *copula* (Sklar, 1959). One has $C[\Pr(score_1 \leq k_1), \Pr(\text{score}_2 \leq k_2)] > (<) \Pr(\text{score}_1 \leq k_1) \Pr(\text{score}_2 \leq k_2)$ when scores at 1 and 2-hours are positively (negatively) dependent, and it is equal to the product of the margins under independence.

The strength of dependence in bivariate random vectors with given marginal distributions is characterized by the copula. It can be assessed using rank-based association measures such as Kendall's tau-b or Spearman's rho (Kruskal, 1958).

In the example, out of all 112 patients having non-missing scores at 1 and 2-hours, 13 (11.6%) have pain relief at 1 hour and 43 (38.39%) relief at 2-hour. The proportion of pain relieves both at 1 and 2 hours is 9.82% (N=11). That proportion is slightly superior to the product of the individual proportions (8.24%), suggesting positive association between successive measures. This is confirmed by the positive Kendall's tau-b estimate of 0.34 (s.e.=0.08).

4.2 *Multivariate Probit Model*

The above copula representation extends to distributions of higher-order dimensions. We model the joint distribution $\Pr(\text{score}_{t_1} \leq k_{t_1}, ..., \text{score}_{t_{\text{last}}} \leq k_{t_{\text{last}}}, \text{rescue} \leq t_{\text{r}})$ as a function $C(\cdot)$ of the univariate margins. The true function C is generally unknown but it can be modelled using parametric functions which parameters measure the dependence between pairs of margins. When marginal distributions are all univariate standard normal, i.e. when a *probit* link is used to regress marginal distributions on a set of covariates, the joint distribution can be modelled as a multivariate standard normal distribution with correlation matrix Σ. This creates a multivariate probit model (Ashford and Sowden, 1970) for ordinal data.

Kendall's tau-b coefficients (τ) can yet be used to guide the specification of a dependence structure in Σ. Indeed, each correlation coefficient ρ in Σ is a monotonic function of the related τ, approximately equal to $\sin(\tau\pi/2)$ (see e.g. Joe, 1997, p. 54). The equality is obtained with continuous random variables and the approximation improves as the number of categories for the categorical response increases (Vandenhende and Lambert, 2000).

The method is illustrated with the analysis of the 24-hour PS profile displayed in Figure 1, together with the timing of RM presented in Table 3. Estimates of Kendall's tau-b coefficients (not reported) suggest that the correlation in the PS series decreases with lag-time at a rate close to $\rho_{\text{ps}}^{|t_i - t_j|}$. A negative correlation is observed between PS and discrete time to RM. That correlation does not change much between times of PS evaluation. We use a single parameter $\rho_{\text{rm}-\text{ps}}$ to model the dependence between any PS and time to RM in Σ. By randomisation, baseline PS were comparable between groups. They are also similar at 24-hour (see Figure 1), when the migraine has disappeared in most patients. The rate of PS improvement

Table 3
Incidence of Rescue Medication Intakes.

Group	Rescue in (2h, 4h]	Rescue $\leq$ 24h
Placebo (N=30)	11 (37%)	17 (57%)
LY20mg (N=27)	14 (52%)	17 (63%)
LY60mg (N=33)	5 (15%)	10 (30%)
LY200mg (N=26)	7 (27%)	9 (35%)
p-value*	.020	.028

*: ordinary chi-square test.

over time is not comparable between groups. It tends to accelerate as dose increases. We have therefore considered the following model for the marginal PS distributions:

$$\text{probit}[\Pr(\text{score}_t \leq k)] = a_k^{\text{ps}} + \theta^{\text{ps}}(t/24)^{1-\delta_d^{\text{ps}}}, \tag{3}$$

where t refers to the time in hours. The parameter θ^{ps} is for the linear time effect under placebo. The parameters δ_d^{ps}, d=20, 60, 200 are dose-specific non-linearity indexes for each dose. A positive value indicates an accelerated relapse of the migraine compared to placebo.

The model for discrete timing of RM is, for $t = 4, 24$ hours,

$$\text{probit}[\Pr(\text{rescue} \leq t)] = b_t^{\text{rm}} + \phi_d^{\text{rm}}, \tag{4}$$

where ϕ_d^{rm}, d=20, 60,200 are dose-specific changes versus placebo. A negative ϕ_d^{rm} value indicates a smaller probability of early RM intake at dose d.

4.3 *Maximum likelihood estimation*

We consider a maximum likelihood method to estimate parameters and their variability. To avoid estimation problems at the boundaries of the parametric space, we first transform the threshold and covariance parameters to permit an unconstrained optimization. We set $a_2^{\text{ps}} = a_1^{\text{ps}} + e^{\alpha_2^{\text{ps}}}$, $a_3^{\text{ps}} = a_2^{\text{ps}} + e^{\alpha_3^{\text{ps}}}$ and $b_2^{\text{rm}} = b_1^{\text{rm}} + e^{\beta_2^{\text{rm}}}$ for the cut-off parameters, where $a_1^{\text{rm}}, \alpha_2^{\text{ps}}, \alpha_3^{\text{ps}}, b_1^{\text{rm}}, \beta_2^{\text{rm}} \in \Re$. For the correlation parameters, we follow Lesaffre and Molenberghs (1991) and set $\rho_{\text{ps}} = 1/(1 + e^{-\nu_{\text{ps}}})$ and $\rho_{\text{rm-ps}} = (1 - e^{-\nu_{\text{rm-ps}}})/(1 + e^{-\nu_{\text{rm-ps}}})$, where $\nu_{\text{ps}}, \nu_{\text{rm-ps}} \in \Re$.

Although theoretically appealing, the use of likelihood methods for estimating high-dimensional multivariate probit models is a challenging task. The computation of cumulative multivariate normal distributions involves analytically intractable integrals that need approximations. The computational time and the error in approximations increase with the dimensionality of the problem. The availability of a powerful approximation algorithm implemented in the procedure `pmvnorm`, in the freeware language *R* (Ihaka and Gentleman, 1996) has convinced us of the possibility to use those models in practice. This function comes from the `mvtnorm` add-on package. It approximates multivariate normal probabilities for hyperrectangular integration regions. A randomised quasi-Monte Carlo method (Genz, 1992) is used. The efficiency of this procedure to control the approximation error at little expenses for the computational time was demonstrated by Genz (1993).

In the example, we compute for each patient a 7-variate (6 PS + RM) normal$(0, \Sigma)$ probability for a hyperrectangle delimited by the marginal models (3) and (4). Patient's PS over time and RM timing populate the upper bounds of the hyperrectangle. The preceding response categories populate the lower bounds. When data are missing in any dimension, the bounds of integration range from $-\infty$ to $+\infty$ in that dimension. In such a way, the dimension of the joint distribution is adjusted to the number of non-missing components for each patient. The likelihood function is then written as the product of probabilities across patients.

We have implemented the Davidon, Fletcher and Powell (DFP) quasi-Newton algorithm in *R*, by adapting the `Mathematica` code of Rose and Smith (2001, pp.405–409) to obtain MLEs and an estimate of their covariance matrix. Gradients were calculated numerically based on finite difference approximations. The code is available upon request to the authors. MLEs and standard errors are in Table 4. The first column reports estimates from an independent probit model, in which ρ_{ps} and $\rho_{\text{rm}-\text{ps}}$ were set equal to zero. The joint probit model estimates are in the second column. The modelling of the dependence created a huge drop in the deviance (-444.2; 2df), indicating a better adjustment to the data.

Significance of the dose effects on δ_d^{ps} and ϕ_d^{rm} was assessed using likelihood-ratio tests (results not reported). The joint analysis confirms in a unique inferential framework all previous results: the 60 and 200mg doses are superior to placebo in relieving pain. There is no evidence that the 20mg dose is efficacious. A delay in RM intake is observed at 60 mg only. PS measured 1-hour apart are highly positively correlated ($\widehat{\tau}_{\text{ps}} = 0.74$). It is also more

Table 4
MLEs (s.e.) from the joint analysis of the 24-hour pain severity profile and the timing of rescue medication intakes.

Parameter	Independent/ Probit -2logL=1856.9	Mult. Normal/ Probit -2logL= 1412.7	Mult. Normal/ Loglog -2logL= 1362.8
a_1^{ps}	-1.60 (0.09)	-1.53 (0.16)	-1.53 (0.13)
α_2^{ps}	-0.47 (0.08)	-0.71 (0.11)	-0.54 (0.10)
α_3^{ps}	-0.10 (0.06)	-0.37 (0.09)	-0.06 (0.09)
θ^{ps}	2.01 (0.15)	2.08 (0.17)	2.93 (0.23)
δ_{20}^{ps}	-0.07 (0.20)	0.33 (0.20)	0.35 (0.15)
δ_{60}^{ps}	0.65 (0.06)	0.68 (0.09)	0.59 (0.08)
δ_{200}^{ps}	0.70 (0.05)	0.69 (0.06)	0.67 (0.07)
ν_{ps}	N.A.	2.44 (0.19)	2.03 (0.19)
b_1^{rm}	-0.30 (0.16)	-0.31 (0.14)	0.05 (0.15)
β^{rm}	-0.94 (0.17)	-0.95 (0.19)	-0.90 (0.22)
ϕ_{20}^{rm}	0.28 (0.18)	0.35 (0.18)	0.38 (0.27)
ϕ_{60}^{rm}	-0.62 (0.23)	-0.71 (0.21)	-0.70 (0.21)
ϕ_{200}^{rm}	-0.41 (0.22)	-0.40 (0.22)	-0.45 (0.22)
$\nu_{\text{rm}-\text{ps}}$	N.A.	-0.46 (0.21)	-0.39 (0.21)

likely to take RM early when PS is large ($\hat{\tau}_{\text{rm}-\text{ps}} = -0.15$).

4.4 *Prediction of IHS outcomes*

MLEs and their covariance matrix can also be used to predict the IHS efficacy outcomes. For instance, the probability of sustained pain-free at 1-hour $\Pr(\text{score}_1 \leq 0, \text{score}_{1.5} \leq 0, \text{score}_2 \leq 0, \text{score}_4 \leq 0, \text{score}_{24} \leq 0, \text{rescue} > 24)$ is $\Pr(\text{score}_t \leq 0, t = 1, 1.5, 2, 4, 24) - \Pr(\text{rescue} \leq 24, \text{score}_t \leq 0, t = 1, 1.5, 2, 4, 24)$ with any pain score possible at 0.5 hour ($k_{.5} = 3$).

We use two asymptotic properties of the MLE - namely the convergence in distribution to a Normal distribution and the invariance property - to estimate the asymptotic distribution of any joint probability estimator based on simulations. To do so, we generate a random sample from the asymptotic normal distribution of the parameter MLEs. 1000 random samples were taken using the `rmvnorm` function from the `mvtnorm` package. Then we calculate the proportions of interest for each sample. An histogram of predicted

1-hour sustained pain-free rates after a 60 mg dose is provided in Figure 2 for the selected models.

The observed proportion (9.1%) is under-estimated in the independent model when the joint distribution equals the product of all margins. With the multivariate probit model, predicted rates are closer to their empirical counterpart. This provides an additional evidence of the better adjustment of that model to the data.

5. Extension of the Multivariate Probit Model

In this section, we extend of the multivariate probit model to improve the flexibility of the modelling of longitudinal ordinal responses. Developments are given to consider other link functions than the probit for the marginal distributions. When used in the example, the change in margins permits a better adjustment to the data and therefore a more reliable prediction of the drug effect (see Table 4 and Figure 2).

Using standard notations, let $\Pr(Y_j \leq k) = g_j(\alpha_k + \mathbf{x}\beta)$ be a cumulative distribution model for the latent variable associated to the ordinal response Y_j. The joint distribution of the J-dimensional vector $(Y_1, ..., Y_J)$ can be modelled as

$$\Phi_\Sigma\{\Phi^{-1}[g_1(\alpha_{k_1} + \mathbf{x}\beta)], ..., \Phi^{-1}[g_J(\alpha_{k_J} + \mathbf{x}\beta)]\}, \tag{5}$$

using the J-variate Φ_Σ and univariate Φ standard normal distributions. The function $C(u_1, ..., u_J) = \Phi_\Sigma[\Phi^{-1}(u_1), ..., \Phi^{-1}(u_J)]$ is the multivariate normal copula (Lambert and Vandenhende, 2002; Song, 2000).

In a multivariate probit model, $g_j = \Phi$, for all j and (5) simplifies to $\Phi_\Sigma(\alpha_{k_1} + \mathbf{x}\beta, ..., \alpha_{k_J} + \mathbf{x}\beta)$. However, other distributions than the Gaussian can be considered for g_j and, even, different distributions can be specified for every j. For instance, a logit link, as presented in Section 2, could be used for any j.

We follow on Vandenhende and Lambert (2000)'s findings who tested several link functions in the analysis of another trial with the same drug. We use an asymmetric log-log link for all margins. Then, model (5) has an extreme-value (EV) distribution $g(x) = \exp[-\exp(-x)]$ for each univariate margin. Compared to the normal distribution that is symmetrical, the EV is skewed to the right. Its mode, median and mean are equal to 0, 0.337 and 0.577, respectively. The EV thus allows a better coverage of severe pains in the migraine example.

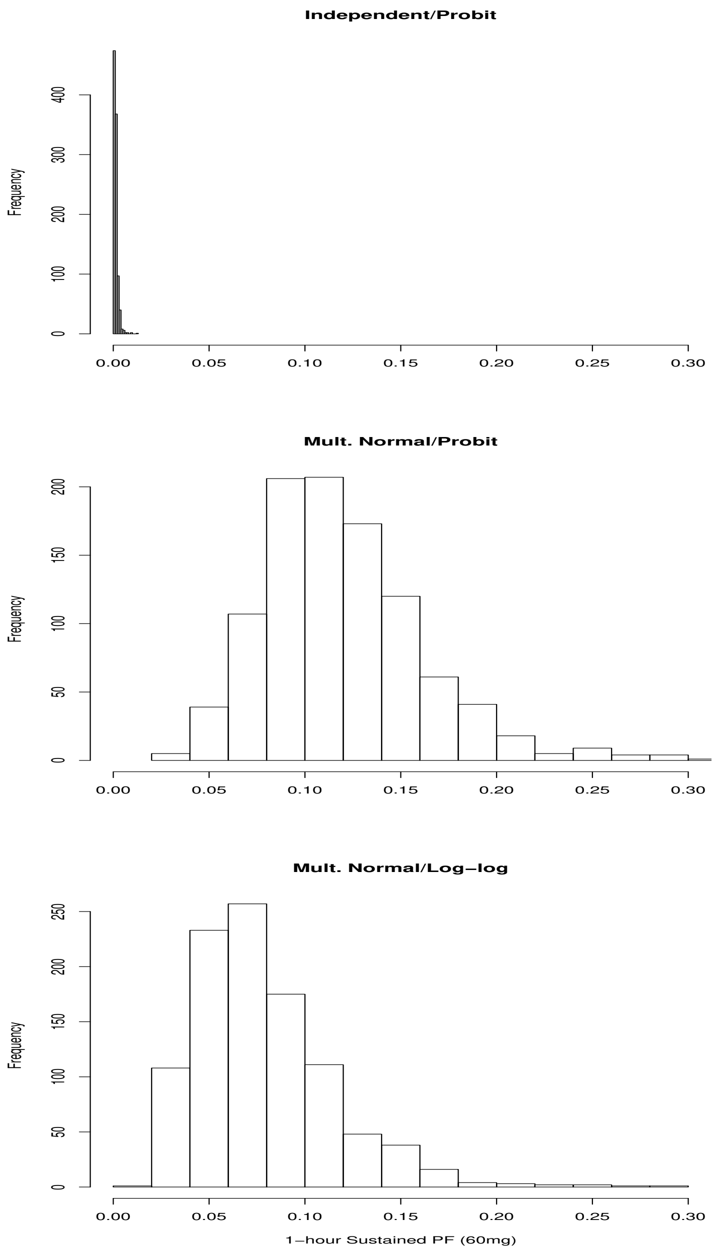

Figure 2. Histogram of predicted 1-hour sustained pain-free rates based on models from Table 4. Three out of 33 patients (9.1%) responded after a 60 mg dose.

Even with non-normal margins, the matrix Σ still quantifies the dependence between marginal distributions, whatever the models g_j. The increasing relationship with rank-based dependence measures is preserved when changing the marginal models, so that MLEs of dependence parameters should not be much affected by a change in link functions.

MLEs are in the third column of Table 4 and predicted 1-hour sustained PF rates are in the third panel of Figure 2. The fit of the EV model (-2logL=1362.8) is superior to that of the probit model (-2logL=1412.7). Likelihood-ratio tests confirmed the significance of the treatment effects at 60 and 200 mg. As shown in Figure 2, the histogram of predictions is not symmetrical anymore with the EV model. A lower sustained PF rate is more likely that with the probit model, whereas rates as large as 25% are yet possible.

6. Discussion

GEE- and a likelihood-based methods were presented for the longitudinal analysis of trials on acute migraine. The GEE method is useful to estimate and test treatment effects on the pain profile prior to the availability of rescue medications (generally 2-hours). We advice its use in pivotal trials for registration. However, we would recommend not using GEE on pain series truncated at time of RM intake because missing data could be dependent on previously observed scores (MCAR assumption not met).

The likelihood method relies on a parametric model for the joint distribution of pain responses over time and of timing of rescue medication intake. Then, success rates in any IHS outcome, including sustained responses and RM intake, are predicted from the fitted model. The method is well suited to dose-response trials from which predictions can be used to select doses and optimise the design of subsequent trials. The assessment of treatment effects and the comparison of different models are also possible using likelihood-ratio tests for nested models or Akaike's information criteria.

The use of the multivariate normal copula allowed the specification of flexible dependence models. The dependence structure could then be parameterized in a Gaussian correlation matrix, whatever the considered model for the marginal distributions. Technical difficulties for a ML estimation of that model were also overcomed using the `mvtnorm` package and a personnal implementation of the DFP optimizer in the R language. Our code was short (2 pages) and speed was acceptable (a few minutes). The good convergence properties of the algorithm have convinced us of the usefulness of the method in high-dimensional real applications.

An immediate extension of the normal copula is possible by copulas generated from the multivariate standard t-distribution. The additional parameter (df) corresponding to the degrees of freedom can then be estimated to better model the dependence in the tails of the joint distribution. The multivariate t-copula model reduces to a normal copula model when df$\rightarrow \infty$. In the example, that extension was not significantly superior (-2logl=1362.4) to that of the reduced normal copula model.

Acknowledgements

We wish to thank Dr. Walt Offen from Eli Lilly and Company for his helpful suggestions on the initial version of the manuscript.

Financial support from the contract 'Projet d'Actions de Recherche Concertées' nr 98/03-217 from the Belgian government, and from the IAP research network nr P5/24 of the Belgian State (Federal Office for Scientific, Technical and Cultural Affairs) is gratefully acknowledged.

References

Ashford, J. R. and Sowden, R. R. (1970). Multivariate probit analysis. *Biometrics* **26**, 535–546.

Genz, A. (1992). Numerical computation of multivariate normal probabilities. *Journal of Computational and Graphical Statistics* **1**, 141– 150.

Genz, A. (1993). Comparison of methods for the computation of multivariate normal probabilities. *Computing Science and Statistics* **25**, 400– 405.

Goldstein, D. J., Roon, K. I., Offen, W. W., Ramadan, N. M., Phebus, L. A., Johnson, K. W., Schaus, J. M. and Ferrari, M. D. (2001). Selective serotonin 1f (5HT1F) receptor agonist LY334370 for acute migraine: a randomized controlled trial. *The Lancet* **358**, 1230–1234.

Goldstein, D. J., Wang, O., Saper, J. R., Stoltz, R., Silberstein, S. D. and Mathew, N. T. (1997). Ineffectiveness of neurokinin-1 antagonist in acute migraine: a crossover study. *Cephalalgia* **17**, 785–790.

Ihaka, R. and Gentleman, R. (1996). R: A language for data analysis and graphics. *Journal of Computational and Graphical Statistics* **5**, 299–314.

Joe, H. (1997). *Multivariate Models and Dependence Concepts*. Chapman & Hall.

Kruskal, W. H. (1958). Ordinal measures of association. *Journal of the American Statistical Association* **53**, 814–861.

Lambert, P. and Vandenhende, F. (2002). A copula based model for multivariate non normal longitudinal data: Analysis of a dose titration safety study on a new antidepressant. *Statistics in Medicine* **21**, in press.

Lesaffre, E. and Kaufmann, H. (1992). Existence and uniqueness of the maximum likelihood estimator for a multivariate probit model. *Journal of the American Statistical Association* **87**, 805–811.

Lesaffre, E. and Molenberghs, G. (1991). Multivariate probit analysis: A neglected procedure in medical statistics. *Statistics in Medicine* **10**, 1391–1403.

Lipsitz, S. R., Kim, K. and Zhao, L. (1994). Analysis of repeated categorical data using generalized estimating equations. *Statistics in Medicine* **13**, 1149–1163.

McCullagh, P. and Nelder, J. A. (1989). *Generalized Linear Models, 2nd Edition.* Chapman & Hall.

Miller, M. E., Davis, C. S. and Landis, J. R. (1993). The analysis of longitudinal polytomous data: Generalized estimating equations and connections with weighted least squares. *Biometrics* **49**, 1033–1044.

Park, T. and Lee, S.-Y. (1997). A test of missing completely at random for longitudinal data with missing observations. *Statistics in Medicine* **16**, 1859–1871.

Rose, C. and Smith, M. D. (2001). *Mathetatical Statistics with Mathematica.* Springer.

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

Sklar, A. (1959). Fonctions de répartition à n dimensions et leurs marges. *Publication Institut Statistiques Université Paris* **8**, 229–231.

Song, P. X.-K. (2000). Multivariate dispersion models generated from gaussian copula. *Scandinavian Journal of Statistics* **27**, 305–320.

Tfelt-Hansen, P., Block, G., Dahlof, C., Diener, H. C., Ferrari, M. D., Goadsby, P. J., Guidetti, V., Jones, B., Lipton, R. B., Massiou, H., Meinert, C., Sandrini, G., Steiner, T. and Winter, P. B. (2000). Guidelines for controlled trials of drugs in migraine: Second edition *Cephalalgia* **20**, 765–786.

Vandenhende, F. and Lambert, P. (2000). Modeling repeated ordered categorical data using copulas. *Discussion Paper* **00-25**, Institut de Statistique, Université Catholique de Louvain, Louvain–la–Neuve, Belgium. (ftp://www.stat.ucl.ac.be/pub/papers/dp/dp00/dp0025.ps).