Biostatistics (2009), **10**, 4, *pp.* 744–755
doi:10.1093/biostatistics/kxp028
Advance Access publication on July 23, 2009

A mixed model framework for teratology studies

JOHAN BRAEKEN*

Department of Methodology and Statistics, Tilburg University, Tilburg, the Netherlands and CITO, Arnhem, the Netherlands
j.braeken@uvt.nl

FRANCIS TUERLINCKX

Research Group Quantitative Psychology and Individual Differences, University of Leuven, Leuven, Belgium

SUMMARY

A mixed model framework is presented to model the characteristic multivariate binary anomaly data as provided in some teratology studies. The key features of the model are the incorporation of covariate effects, a flexible random effects distribution by means of a finite mixture, and the application of copula functions to better account for the relation structure of the anomalies. The framework is motivated by data of the Boston Anticonvulsant Teratogenesis study and offers an integrated approach to investigate substantive questions, concerning general and anomaly-specific exposure effects of covariates, interrelations between anomalies, and objective diagnostic measurement.

Keywords: Conditional independence; Copula; Finite mixtures; Item response theory; Mixed model; Teratology.

1. INTRODUCTION

Several external agents, such as chemicals and viruses, can cause abnormalities during the development of a fetus. These agents play a key role in the domain of teratology, where researchers are diagnosing birth defects and investigating the causal processes or etiology behind them. About 10% of all children will require extensive medical care to diagnose or treat a birth defect. Significant progress has been made in identifying etiological causes of some birth defects, yet approximately 65% still have no known cause (O'Rahilly and Muller, 1996). This last fact is due to the inherent complexity of the domain: few human teratogens have 1 single well-defined effect, but rather generate a set of (possibly partly overlapping) birth defects out of a variety of deficiencies and anomalies. Therefore, multiple outcomes have to be assessed in teratology studies.

The data set motivating this manuscript is the known Boston Anticonvulsant Teratogenesis study (BAT; Holmes *and others*, 2001) in which the effect of anticonvulsant drug use during pregnancy is evaluated. Each infant in the study was assessed on the presence or absence of several anomalies, resulting in a multivariate binary data string for each infant. Of the 687 infants, there were 168 (24%) whose

*To whom correspondence should be addressed.

© The Author 2009. Published by Oxford University Press. All rights reserved. For permissions, please e-mail: journals.permissions@oxfordjournals.org.

Downloaded from http://biostatistics.oxfordjournals.org/ at KU Leuven University Library on January 22, 2015

Table 1. *Summary statistics for 10 anomalies from the BAT*

Anomaly	Percentage ($n = 687$)
Hypoplastic toenails (TNHYP)	0.179
Hypoplastic fingernails (FNHYP)	0.108
Depressed nasal bridge (DBN)	0.138
Tapered fingernails (TAPF)	0.108
Anteverted nostrils (NATV)	0.102
Small head (SMHEAD)	0.112
Short birth length (SHORT)	0.058
Low birth weight (LOWBT)	0.131
Clinodactyly of the hand (CLIN)	0.128
Broad nasal bridge (BRBRDG)	0.133

mothers took anticonvulsants during pregnancy, 73 (11%) whose mothers suffered from epileptic seizures but did not continue to take the anticonvulsant medication during pregnancy, and 446 (65%) infants from a control group. The data set under consideration consists of 10 physical anomalies, ranging from facial malformations to growth indicators (Table 1). Although most of these anomalies are not of clinical importance themselves, they are of interest due to their potential predictive power as markers of more serious, but not yet emerged developmental problems (Holmes *and others*, 1987). Note that the data set is based on the same clinical study as analyzed by Legler and Ryan (1997).

To solve the measurement problem in this study, a quantification has to be made of how severely an infant is affected. This severity measure will serve as an objective assessment that can be of use for public health and welfare support, and policy makers. Prior research (Legler and Ryan, 1997) introduced a latent variable model based upon item response theory, in which the latent variable reflects the unobserved severity at which an infant is affected. We will propose a more general nonlinear mixed model (NLMM) alternative. Borrowing from recent advances in linear mixed models in biostatistics (Verbeke and Lesaffre, 1996; Muthén *and others*, 2002), we will adopt a finite mixture distribution for the unobserved severity to avoid possibly stringent parametric restrictions for the distribution of the random effect, that is, the latent severity. Furthermore, such a mixture distribution may reveal hidden groups (latent classes) of infants in the data, while maintaining quantitative interinfant differences.

To answer the etiological question whether the fact that newborns of epileptic mothers have more chance of showing adverse birth outcomes is due to the mere presence of maternal epilepsy or simply an artifact of the related anticonvulsant drug therapy, covariate information on the infants can be incorporated in the mixed model. This avoids the need for separate global tests and methods for multiple comparison between covariate groups (see e.g. Lefkopoulu and Ryan, 1993), and, borrowing from the differential item functioning tradition in educational measurement (Holland and Wainer, 1993), one can even proceed in differentiating between a main effect and an anomaly-specific effect of the covariate.

Furthermore, because anomalies may cluster together for various reasons, known and unknown, it can be expected that the commonly made assumption of conditional independence will be violated in this case study. Thus, even after accounting for the dependency among the anomalies due to the underlying latent severity dimension by a random infant-specific effect, residual dependencies among specific anomalies are expected. This type of model misspecification has a long tradition in psychometrics, in which it is referred to as local item dependency (Lord and Novick, 1968). To account for this problem, the copula-based method by Braeken *and others* (2007), developed in an educational measurement context for the simple Rasch model, will be generalized to the broader setting of NLMMs. Correctly accounting for the interrelations between anomalies will lead to more robust model inferences and can also be helpful in identifying a syndrome or providing new insights in potential underlying biological processes.

Downloaded from http://biostatistics.oxfordjournals.org/ at KU Leuven University Library on January 22, 2015

2. METHOD

2.1 *Mixed model framework*

Let $\boldsymbol{Y}_p = (Y_{p1}, \ldots, Y_{pi}, \ldots, Y_{pI})^T$ and $\boldsymbol{Z}_p = (Z_{p1}, \ldots, Z_{pj}, \ldots, Z_{pJ})^T$ be the binary anomaly outcome vector for infant p and the covariate vector containing additional information on infant p, respectively. Consider an NLMM(see e.g. Molenberghs and Verbeke, 2005) in which, conditional upon θ_p, the unobserved latent severity of affect of infant p, anomalies occur independently with probabilities $\Pr(Y_{pi} = 1|\theta_p) = \psi(\eta_{pi})$, where $\psi(\cdot)$ is the inverse logit link function and η_{pi} the predictor component defined as

$$\eta_{pi} = \alpha_i(\theta_p - \beta_i + \sum_{j=1}^{J} Z_{pj}\lambda_{ji}). \tag{2.1}$$

The infant-specific (random) effect θ_p is assumed to follow a normal distribution $\phi(\mu, \sigma^2)$, where the mean and variance are set at 0 and 1, respectively ($\mu = 0, \sigma^2 = 1$), to fix the location and scale of the latent severity θ_p, and thereby identify the model. The parameter β_i is the anomaly main effect or severity degree and corresponds to the location on the latent scale θ_p where the probability of having the anomaly i is equal to 0.5. The multiplicative parameter α_i controls the steepness of the logistic curve, indicating the diagnostic value of anomaly i in discriminating between severely affected infants and less severely affected infants (see e.g. the 2-parameter logistic model; Birnbaum, 1968).

The inclusion of covariate information $\boldsymbol{Z}_p$ in the model offers a way of exploring possible risk and protective factors for teratogenesis (i.e. the development of anomalies). For instance, it may be plausible that infants exposed in utero to anticonvulsant drugs will have on average a higher severity of effect θ_p than infants not belonging to this group. This would correspond to a general main effect of the covariate Z_{pj}, such that $\lambda_{ji} = \lambda_j$ for all anomalies. Hence, when the covariate Z_{pj} indicates membership to an exposure group, λ_j will indicate whether or not the exposure group differs in mean severity of effect from the control group. Notice that to quantify the effect on a specific anomaly i, the main effect is in this case scaled following the diagnostic value of the anomaly i for the severity θ_p, leading to a multiplicative change in the odds of having birth defect i by $\exp(\alpha_i\lambda_j)$. If one is not interested in a simple main effect of the covariate, the double index can remain, resulting in unique anomaly-specific covariate effects λ_{ji}.

To identify anomaly-specific effects that deviate from the common main effect for a covariate Z_{pj}, a sequential likelihood ratio test procedure can be followed where a covariate main effect model (λ_j) is compared to a covariate main effect with 1 additional anomaly-specific effect ($\lambda_j + \lambda_{ji}$ for the item under investigation, λ_j otherwise). Note that the inclusion of λ_j and all λ_{ji} leads to an unidentfied model. To control the type I error rate in this multiple testing procedure, a Bonferroni correction can be used. This sequential approach has been successfully used in the study of differential item functioning in item response theory (see e.g. Holland and Wainer, 1993).

2.2 *Specification of the random effect distribution*

The standard practice of choosing $\theta_p \sim \phi(0, 1)$ puts an important constraint on the shape of the distribution of the random effect and a misspecification of the distribution can lead to biased parameter estimates in the model (Agresti *and others*, 2004; Hartford and Davidian, 2000; Neuhaus *and others*, 1992). In a teratogenesis study, the majority of the infants are at most mildly affected and thus scoring positively on few outcomes. Given that the overall incidence of adverse birth defects is likely to be low, one can expect that the common normality assumption of the latent trait will be questionable.

To accommodate such specification issues, we choose to rely on finite mixture distributions, which allow for a more flexible and data-driven distribution (Verbeke and Lesaffre, 1996; Aitkin, 1999), such

Downloaded from http://biostatistics.oxfordjournals.org/ at KU Leuven University Library on January 22, 2015

that $\theta_p \sim \sum_{g=1}^{G} \pi_g \phi_g(\mu_g, 1)$, where for the component weights π_g it holds that $\sum_{g=1}^{G} \pi_g = 1$ and $\pi_g \in [0, 1]$. For identification reasons, the first component mean μ_1 is restricted to 0 so that the origin of the latent severity scale remains fixed.

Besides an increased flexibility, a finite mixture distribution also allows to investigate and reveal possible hidden grouping in the data. In such a case, infants can be ascribed to the component for which they have the highest posterior membership probability, resulting in a classification of infants over components. Classifying infants in this way could yield new insights in the data and the teratogenic processes involved.

2.3 *Specification of the joint distribution*

Traditionally, the joint distribution of the outcome vector $\boldsymbol{Y}_p$ in mixed or latent variable models is formulated under conditional independence:

$$\Pr(\boldsymbol{Y}_p = \boldsymbol{y}_p|\theta_p) = \prod_{i=1}^{I} \Pr(Y_{pi} = y_{pi}|\theta_p), \tag{2.2}$$

The anomaly outcomes are assumed to be independent realizations conditional upon the latent severity θ_p and covariate effects. Although a common assumption, this might not always be plausible. In a teratology context for instance, it is likely that even after accounting for the common dependence due to an infant's severity of effect, specific anomalies will show some extra association. This can be the case when they have a similar origin in common, like for instance the same body part, a genetic link, or some kind of environmental factor. In the BAT, the growth indicators can be expected to show this type of extra association.

Failing to correctly account for the association in the data implies a misspecified model. The most prominent effects of not acknowledging the redundant information in the data are that the standard error of the estimated infants severity $\hat{\theta}_p$ is underestimated (Junker, 1991) and that discrimination parameters are overestimated or even diverge to an unrealistic large value, thereby falsely considering anomalies to be better indicators of θ_p (Masters, 1988; Tuerlinckx and De Boeck, 2001). In this manuscript, we propose the use of copula functions to account for residual dependencies in the data by formulating a more proper joint distribution than in 2.2, thereby correcting for the abovementioned problems.

Formally, an I-dimensional copula is a function $C : [0, 1]^I \to [0, 1]$ with the following 3 properties: First, for every vector $\boldsymbol{u} \in [0, 1]^I$, $C(\boldsymbol{u})$ is increasing in each component u_i with $i \in \{1, \ldots, I\}$; second, $C(\boldsymbol{u}) = 0$ if at least 1 coordinate of the vector is 0 and $C(\boldsymbol{u}) = u_i$ if all the coordinates of the vector are equal to 1 except the i-th one; third, for every $\boldsymbol{a}, \boldsymbol{b} \in [0, 1]^I$ with $\forall i, a_i \leqslant b_i$, given a hypercube $\boldsymbol{B} = [\boldsymbol{a}, \boldsymbol{b}] = [a_1, b_1] \times \ldots \times [a_I, b_I]$ whose vertices lie in the domain of C, and with $V_C(\boldsymbol{B}) \geqslant 0$. The volume $V_C(\boldsymbol{B})$ is defined as:

$$V_C(\boldsymbol{B}) = \sum_{k_1=1}^{2} \ldots \sum_{k_I=1}^{2} (-1)^{k_1+\ldots+k_I} C\left(d_1^{(k_1)}, \ldots, d_I^{(k_I)}\right)$$

where $d_i^{(1)} = a_i$ and $d_i^{(2)} = b_i$ $(i = 1, \ldots, I)$.

For a thorough overview of copula theory, see Nelsen (1998), and Joe (1997). The application possibilities of copulas in multivariate modeling are mainly due to Sklar's (1959) theorem. The theorem states that for any I-dimensional distribution function $F_{\boldsymbol{X}}$ with univariate margins $F_{X_1}, \ldots, F_{X_I}$, there exists a copula function C such that this multivariate distribution $F_{\boldsymbol{X}}$ can be represented as a function of its margins through this copula: $F_{\boldsymbol{X}} = C(F_{X_1}, \ldots, F_{X_I})$. In the continuous case, this copula reformulation of an existing multivariate distribution is unique, for discrete random variables the copula C

Downloaded from http://biostatistics.oxfordjournals.org/ at KU Leuven University Library on January 22, 2015

is uniquely determined on $\text{Ran}(F_{X_1}) \times \ldots \times \text{Ran}(F_{X_I})$, the Cartesian product of ranges of the margins. Furthermore, the converse of the theorem also holds. Thus, given a set of univariate margins $F_{X_1}, \ldots, F_{X_I}$, a multivariate (i.e. joint) distribution $F_{\boldsymbol{X}}$ can be constructed by adopting a specific copula C, such that $F_{\boldsymbol{X}} = C(F_{X_1}, \ldots, F_{X_I})$. Given the second property in the copula definition, it can easily be deduced that the univariate marginal distribution for X_i equals $C\left(1, \ldots, 1, F_{X_i}, 1, \ldots, 1\right) = F_{X_i}$. Hence, in this way an association between the I random variables is allowed through the copula C, while still preserving the original univariate margins we started from. The copula approach introduced by Braeken *and others* (2007), makes use of exactly this theorem to formulate a more proper joint distribution that accounts for the association structure in the data.

We extend and generalize this approach to the setting of nonlinear mixed models. Consider S disjoint subsets of the set of anomalies $\{1, \ldots, I\}$ denoted as $J_1, \ldots, J_S$, where J_s has cardinality I_s. While all anomalies are indicators of the unobserved severity θ_p, anomalies within a subset J_s show some residual dependency. The different subsets are assumed independent given θ_p, and the anomaly outcomes in a subset $\boldsymbol{Y}_p^{(s)}$ are assumed exchangeable. The model we propose is then

$$\Pr(\boldsymbol{Y}_p = \boldsymbol{y}_p|\theta_p) = \prod_{s=1}^{S} \Pr_s(\boldsymbol{Y}_p^{(s)} = \boldsymbol{y}_p^{(s)}|\theta_p), \tag{2.3}$$

where $\Pr_s(\boldsymbol{Y}_p^{(s)} = \boldsymbol{y}_p^{(s)}|\theta_p)$ is the joint probability of the anomaly outcome vector $\boldsymbol{y}_p^{(s)}$ in subset J_s.

If the cardinality of subset J_s is larger than 1 ($I_s > 1$), $\Pr_s(\boldsymbol{Y}_p^{(s)} = \boldsymbol{y}_p^{(s)}|\theta_p)$ is the joint probability of the outcome vector of subset s evaluated from the copula function C_s for all $Y_{pi} \in J_s$ under the following formula:

$$\begin{aligned}
&\Pr_s(Y_{p1} = y_{p1}, \ldots, Y_{pI_s} = y_{pI_s}|\theta_p) \\
&\quad = \Pr_s(d_{p1}^{(1)} < Y_{p1} \leqslant d_{p1}^{(2)}, \ldots, d_{pI_s}^{(1)} < Y_{pI_s} \leqslant d_{pI_s}^{(2)}|\theta_p) \\
&\quad = \sum_{k_1=1}^{2} \cdots \sum_{k_{I_s}=1}^{2} (-1)^{k_1+\ldots+k_{I_s}} C_s\left(F_{Y_{p1}|\theta_p}(d_{p1}^{(k_1)}), \ldots, F_{Y_{pI_s}|\theta_p}(d_{pI_s}^{(k_{I_s})})\right),
\end{aligned} \tag{2.4}$$

where $d_{pi}^{(1)}$ and $d_{pi}^{(2)}$ stem from the definition of the distribution functions $F_{Y_{pi}|\theta_p}(y_{pi})$

$$F_{Y_{pi}|\theta_p}(y_{pi}) = \begin{cases} 0 & \text{for } y_{pi} < 0 \\ \Pr(Y_{pi} = 0|\theta_p) & \text{for } y_{pi} = 0 \\ 1 & \text{for } y_{pi} = 1 \end{cases}$$

such that for $y_{pi} = 1$, $d_{pi}^{(1)} = 0$ and $d_{pi}^{(2)} = \infty$, and for $y_{pi} = 0$, $d_{pi}^{(1)} = -\infty$ and $d_{pi}^{(2)} = 0$.

The regular conditional independence model arises as a special case when $S = 1$ and C_s is the independence copula $\Pi(F_{Y_{p1}|\theta_p}, \ldots, F_{Y_{pI}|\theta_p}) = \prod_{i=1}^{I} F_{Y_{pi}|\theta_p}$. In general, a copula function C is typically a member of a more general family of copulas that are capable of inducing a particular type of association structure. In this study, 3 copula functions, presented below, will be used to model residual dependencies within subsets of anomalies. These 3 copulas were chosen because they have a simple functional form, with only a single parameter (δ_s) capturing the whole range of positive association (from independence, denoted by Π, to absolute positive dependence, denoted by M), known multivariate extensions, and together comprise a broad range of possible association structures for the residual dependencies.

Downloaded from http://biostatistics.oxfordjournals.org/ at KU Leuven University Library on January 22, 2015

1. Frank copula (Frank, 1979)

$$C_s(F_{Y_{p1}|\theta_p}, \ldots, F_{Y_{pI_s}|\theta_p}) = \frac{-1}{\delta} \log\left(1 - \frac{\prod_{i=1}^{I_s}(1 - \exp(-\delta F_{Y_{pi}|\theta_p}))}{\prod_{i=1}^{I_s-1}(1 - \exp(-\delta))}\right)$$

with $\delta > 0$; if $\delta \to 0$ then $C \to \Pi$, and if $\delta \to \infty$, $C \to M$.

2. Cook–Johnson copula (Clayton, 1978; Cook and Johnson, 1981)

$$C_s(F_{Y_{p1}|\theta_p}, \ldots, F_{Y_{pI_s}|\theta_p}) = \left(\sum_{i=1}^{I_s} F_{Y_{pi}|\theta_p}^{-\delta} - I_s + 1\right)^{\frac{-1}{\delta}}$$

with $\delta > 0$; if $\delta \to 0$ then $C \to \Pi$, and if $\delta \to \infty$ then $C \to M$.

3. Gumbel–Hougaard copula (Gumbel, 1960; Hougaard, 1986)

$$C(F_{Y_{p1}|\theta_p}, \ldots, F_{Y_{pI_s}|\theta_p}) = \exp\left(-\left[\sum_{i=1}^{I_s}(-\log(F_{Y_{pi}|\theta_p}))^{\delta}\right]^{\frac{1}{\delta}}\right)$$

with $\delta > 1$; if $\delta \to 1$ then $C \to \Pi$, and if $\delta \to \infty$ then $C \to M$.

The advantage of the copula formulation is that the parameterization of each univariate margin (i.e. anomaly) preserves its natural interpretation. This is called the reproducibility property (Ip, 2002; Liang *and others*, 1992), which allows for changes to the joint model without having to leave the attractive modeling framework as described by the latent variable model for the univariate margins proposed in 2.1. A wide variety of association structures can be accommodated by applying different copula functions to the anomaly subsets. Besides their use as a vehicle to take into account violations of the conditional independency assumption, copulas can also be used as an indicator of association between specific anomalies and thereby provide additional information about potential common underlying pathways. Note that most alternative modeling approaches to capture these residual dependencies (see e.g. Hoskens and De Boeck, 1997; Wainer *and others*, 2007; Tuerlinckx and De Boeck, 2004) do not retain reproducibility. For instance, the testlet models of Wainer *and others* (2007) add an extra subset-specific and infant-specific (random) ζ_{ps} to the formulation of the margins to capture the residual dependency, such that $\eta_{pi} = \alpha_i(\theta_p + \zeta_{ps} - \beta_i + \sum_{j=1}^{J} Z_{pj}\lambda_{ji})$ and the severity degree of an anomaly β_i is not anymore the location on the latent severity dimension θ_p where the probability of having the anomaly i is equal to 0.5. In other words, this approach changes the individual model characteristics of an anomaly to model the joint behavior of an anomaly set, whereas the copulas take a more natural approach by directly adapting the joint model and leaving the marginal model part intact.

3. Model inference

The parameters to be estimated can be divided into 3 groups: The anomaly-specific parameters and covariate regression weights (α_i, β_i, and λ_{ji} where $i = 1, \ldots, I$ and $j = 1, \ldots, J$), the copula dependency parameters ($\delta_1, \ldots, \delta_S$, where S is the number of disjoint anomaly subsets) and the parameters of the latent distribution (σ^2, $\mu_2, \ldots, \mu_G$ and $\pi_1, \ldots, \pi_{G-1}$). Note that some of the δ parameters drop out because not all subsets will be modeled by means of a copula (i.e. the independence subsets). The marginal likelihood under the full model is then:

$$L(\boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{\lambda}, \boldsymbol{\delta}, \boldsymbol{\mu}, \boldsymbol{\pi}, \sigma^2) = \prod_{p=1}^{P}\sum_{g=1}^{G}\pi_g \int_{\theta_{pg}} \prod_{s=1}^{S} \Pr_s(\boldsymbol{Y}_p^{(s)} = \boldsymbol{y}_p^{(s)}|\theta_{pg})\phi(\theta_{pg}; \mu_g, \sigma^2)\mathrm{d}\theta_{pg}.$$

Downloaded from http://biostatistics.oxfordjournals.org/ at KU Leuven University Library on January 22, 2015

A full information marginal maximum likelihood estimation approach is followed, in which the intractable integral with respect to the distribution of θ_p is approximated using nonadaptive Gauss–Hermite quadrature. If $G = 1$, the applied optimization algorithm is a quasi-Newton method. In case $G > 1$, the optimization is performed using a generalized expectation–maximization (EM) algorithm (McLachlan and Krishnan, 1997). Multiple starting points were used for the EM algorithm to safeguard against the danger of local minima. Making use of Bayes theorem, empirical Bayes estimates of the latent severity θ_p can be computed (Bock and Lieberman, 1970) as

$$\hat{\theta}_p = \sum_{g=1}^{G} \pi_g \frac{\int_{\theta_{pg}} \theta_{pg} \prod_{s=1}^{S} \Pr_s(\boldsymbol{Y}_p^{(s)} = \boldsymbol{y}_p^{(s)} | \theta_{pg}) \phi(\theta_{pg}; \mu_g, \sigma^2) \mathrm{d}\theta_{pg}}{\int_{\theta_{pg}} \prod_{s=1}^{S} \Pr_s(\boldsymbol{Y}_p^{(s)} = \boldsymbol{y}_p^{(s)} | \theta_{pg}) \phi(\theta_{pg}; \mu_g, \sigma^2) \mathrm{d}\theta_{pg}}.$$

With respect to model checking and selection, the tools usually applied in the context of mixed (e.g. Wald, score, or likelihood ratio tests) and mixture models (e.g. information criteria) are also applicable here.

4. APPLICATION: BAT STUDY

As a start of the analysis, a reference model without the covariate information $\boldsymbol{Z}_p$, with a regular standard normal distribution for θ_p, and without copulas, was fitted to the BAT data. Although the model reached convergence, the estimates and standard errors of the discrimination parameter α_i of anomalies 6, 7, and 8 take relatively large values, and severity degree β_i of anomalies 3 and 10 are also affected (see Model 1 in Table 2). These observations indicate, as expected, that the model is hampered by model specification issues.

A first attempt to deal with the specification issues was to extend the standard model with a finite mixture distribution for the severity of effect θ_p. This resulted in an improved fit and the previously mentioned estimation problems occur to a lesser degree (see Model 2 in Table 2). The finite mixture consists of 2 components with means 0 and -10.96, common variance equal to 1, and component weights 0.54 and 0.46, respectively. This configuration leads to a clear bimodality in the distribution of an infant's severity of effect θ_p. After classifying the infants into these 2 components based upon their maximum posterior component probability, the second component can be characterized as gathering unaffected infants (i.e. all infants without any anomaly and a few infants with a single anomaly present, being either a depressed or a broad nasal bridge), while the first component gathers the affected infants, which contains the majority (68%) of infants belonging to the anticonvulsant exposure group. There is no difference between the 2 components with respect to seizure history and gender.

Adding the 3 covariates (i.e. exposed in utero to anticonvulsants, gender, and maternal seizure history) to both the standard and the finite mixture model further improves the fit. Of all 4 fitted models, the model without a mixture but with covariates, is preferred based upon the BIC. Looking in detail to the mixture model with covariates (Model 4), it can be seen why this is the case: The second component's mean increased to -3.76 and the difference in presence of treatment groups leveled out. This is an example of a finite mixture model that picks up an unmodeled covariate effect and where the need for more than 1 component disappears when accounting for the relevant covariate.

The extreme values of some discrimination parameters indicated the potential presence of residual dependencies. To explore such dependencies, various diagnostic tools have been developed (Holland and Rosenbaum, 1986; Yen, 1984; Tate, 2003). In this manuscript, the Mantel–Haenszel (MH) procedure (Mantel and Haenszel, 1959) will be used to identify pairs of anomaly outcomes that exhibit residual dependency that cannot be explained by a single latent trait θ_p. The main idea is to test for an equal odds ratio deviating from one across different groups of infants (i.e. conditional independence). The group division in the MH procedure is based upon a proxy for the severity of effect (e.g. the sum score over anomalies or a preliminary model-based estimate $\hat{\theta}_p$). We regard the MH test procedure in this context

Downloaded from http://biostatistics.oxfordjournals.org/ at KU Leuven University Library on January 22, 2015

Table 2. *Models for the BAT data*

	Model 1: no mixture, no covariates, no copulas	Model 2: mixture, no covariates, no copulas	Model 3: no mixture, Covariates, no copulas	Model 4: mixture covariates, no copulas	Model 5: no mixture covariates, copulas
β_1	2.46 (0.52)	4.78 (1.71)	1.94 (0.20)	1.95 (0.46)	2.25 (0.28)
β_2	2.43 (0.40)	4.07 (1.16)	1.97 (0.16)	2.02 (0.39)	2.10 (0.19)
β_3	4.97 (2.02)	9.76 (3.29)	2.85 (0.39)	3.75 (0.95)	2.56 (0.31)
β_4	2.54 (0.42)	5.21 (1.87)	2.40 (0.23)	2.85 (0.59)	2.47 (0.25)
β_5	3.37 (0.81)	7.56 (2.68)	2.61 (0.27)	3.30 (0.73)	2.52 (0.26)
β_6	1.32 (0.08)	0.89 (0.11)	3.44 (0.65)	4.63 (1.28)	8.40 (3.97)
β_7	1.65 (0.09)	1.31 (0.13)	3.53 (0.57)	4.80 (1.19)	5.74 (1.69)
β_8	1.30 (0.09)	0.85 (0.12)	3.06 (0.51)	4.00 (1.06)	4.82 (1.22)
β_9	3.14 (0.72)	8.50 (3.09)	3.41 (0.57)	4.84 (1.36)	3.62 (0.66)
β_{10}	18.25 (18.71)	13.54 (4.16)	4.77 (1.26)	7.22 (2.43)	4.26 (0.95)
α_1	0.67 (0.16)	0.19 (0.05)	1.30 (0.20)	0.60 (0.15)	0.99 (0.17)
α_2	1.02 (0.21)	0.35 (0.10)	2.72 (0.58)	1.32 (0.37)	2.22 (0.45)
α_3	0.38 (0.16)	0.14 (0.04)	0.85 (0.16)	0.41 (0.11)	1.02 (0.18)
α_4	0.96 (0.20)	0.28 (0.07)	1.43 (0.22)	0.70 (0.17)	1.39 (0.23)
α_5	0.70 (0.19)	0.21 (0.07)	1.26 (0.21)	0.60 (0.14)	1.40 (0.23)
α_6	3.95 (0.93)	3.49 (0.97)	0.75 (0.19)	0.39 (0.11)	0.26 (0.13)
α_7	4.85 (1.08)	4.14 (1.47)	1.05 (0.26)	0.55 (0.15)	0.54 (0.19)
α_8	3.01 (0.56)	2.31 (0.47)	0.79 (0.18)	0.40 (0.11)	0.44 (0.13)
α_9	0.66 (0.17)	0.16 (0.05)	0.69 (0.15)	0.34 (0.10)	0.65 (0.15)
α_{10}	0.10 (0.11)	0.11 (0.03)	0.44 (0.13)	0.23 (0.07)	0.51 (0.14)
μ_2	. .	−10.96 (2.51)	. .	−3.76 (0.98)	. .
π_2	. .	0.46 (0.05)	. .	0.58 (0.04)	. .
$\lambda_{\text{anticonvulsants}}$	. .	. .	1.19 (0.14)	2.33 (0.48)	1.30 (0.15)
λ_{gender}	. .	. .	0.00 (0.11)	0.00 (0.22)	−0.01 (0.12)
$\lambda_{\text{seizure history}}$	. .	. .	0.46 (0.19)	0.94 (0.40)	0.39 (0.20)
G–H: $J_1\{1, 2\}$; δ_1	. .	. .	. .	. .	1.42 (0.15)
F: $J_2\{6, 7, 8\}$; δ_2	. .	. .	. .	. .	13.51 (1.99)
−LL	2317	2291	2286	2281	2153
Parameters	20	22	23	25	25
BIC	4765	4726	4722	4725	4469

Values are given as estimates (standard errors)

mainly as a crude screening tool, and do not report p values. Two subsets of anomalies were demarcated, the anomaly pair $J_1 = \{1, 2\}$ (MH = 4.6) , and the triplet $J_2 = \{6, 7, 8\}$ (all 3 MH $>$ 6.9). Both subsets appear to make sense from a substantive perspective as well; subset J_1 refers to hypoplastic nails, while subset J_2 gathers the growth indicators.

The latent variable model with covariates was extended with copulas to account for the presence of the anomaly subsets. Selecting the best fitting combination of possible copula functions resulted in the Gumbel–Hougaard copula for subset J_1, and the Frank copula for subset J_2. Results are shown in Model 5 of Table 2.

The chances of both fingernail and toenail hypoplasia to co-occur or be jointly absent are higher than can be expected based upon an infant's severity of effect θ_p ($\hat{\delta}_1 = 1.42$, $p < 0.0001$). The infants who are less severely affected (low θ_p) show a higher tendency to have either both anomalies present/absent, than only 1 of both (Figure 1). The growth indicators also show a very large tendency to co-occur or jointly

Downloaded from http://biostatistics.oxfordjournals.org/ at KU Leuven University Library on January 22, 2015

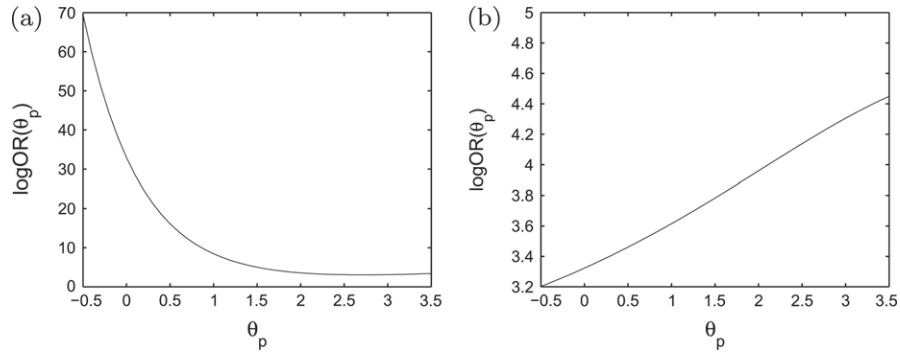

Fig. 1. Conditional log odds ratio plots for the 2 copula subsets in the final model.

be absent ($\hat{\delta}_2 = 13.51$, $p < 0.0001$), with a slightly higher tendency for more severely affected infants (high θ_p) (Figure 1). A 2-component mixture model with covariates and copulas resulted in approximately the same minus log likelihood value as for the 1-component alternative. Because it does not provide new information, detailed results on this mixture solution are omitted.

Another model specification issue that has to be assessed is the impact of the covariates on the various anomalies. After running the sequential likelihood ratio procedure for each of the j covariates (Section 2.1), none of the anomaly-specific covariate effects λ_{ji} were retained. Therefore, it can be concluded that a general main effect λ_j for each covariate j suffices for the BAT study.

In the Supplementary Material of this paper (available at Biostatistics online; http://www.biostatistics.oxfordjournals.org), some additional model checks and assessments are reported to further support the model of choice. The possibility that the latent severity θ_p is in fact multidimensional instead of unidimensional is explored, results of a recovery simulation are presented, and some parametric bootstraps are performed to illustrate the importance of accounting for the residual dependency structure. These various model assessment results suggest that the variation and association structure in the data are adequately accounted for by capturing the residual dependencies and the covariate information in the model.

5. DISCUSSION

A mixed model framework for studying teratology problems, that goes beyond standard assumptions and is motivated by scientific questions, has been presented. The core of the model is a nonlinear mixed model with anomaly-specific discrimination, severity degree, and covariate effects. Instead of naively accepting the standard approach of a normally distributed latent variable, a more flexible data-driven way of specifying the random effect distribution is proposed by adopting a finite mixture. Finally, the model can also handle residual dependencies stemming from similarities between anomalies that are not accounted for by traditional conditional independence models. These dependencies are modeled using copulas because this preserves the univariate margins and their interpretation.

The results of our analysis of the teratology data clearly suggest that in utero exposure to anticonvulsants is a high risk factor for having an infant with birth defects. There was no evidence for an effect of having a mother with a seizure history, or an effect of gender on a higher risk for birth defects, unless a misspecified model was used. This last observation highlights the strong relation between statistically biased models and substantively unrealistic models in a latent variable context. The same conclusion on the covariate effects can be found in the review article by Holmes *and others* (2001) on the larger project of which the BAT study was part.

However, the present study does not allow us to make the bold conclusion that epilepsy has no influence on birth defect risks and that the higher risk is solely to be ascribed to the related anticonvulsant

Downloaded from http://biostatistics.oxfordjournals.org/ at KU Leuven University Library on January 22, 2015

medication therapy followed by most epileptic women. To fully exclude the influence of epilepsy, one should control for the strength of the epileptic disease in the respective participants of the study. Clinical studies have suggested that our set of 10 anomalies might even be predictive for a weaker cognitive development (Holmes *and others*, 2005). It would be interesting to further explore the relation between the measurement scale we obtained, the infant severity index θ_p, and other external criteria to further validate the model.

Acknowledgments

The authors thank Geert Molenberghs for helpful comments and Lewis Holmes, Louise Ryan and Mary Sammel for kindly providing data from the BAT study. *Conflict of Interest*: None declared.

Funding

Scientific Research Flanders (FWO) (G.0148.04), IAP research network (P6/03) of the Belgian Government, psychometric knowledge and research center of CITO.

Supplementary material

Supplementary Material is available at http://www.biostatistics.oxfordjournals.org.

References

Agresti, A., Ohman, P. and Caffo, B. (2004). Examples in which misspecification of a random effects distribution reduces efficiency and possible remedies. *Computational Statistics and Data Analysis* **47**, 639–653.

Aitkin, M. (1999). A general maximum likelihood analysis of variance components in generalized linear models. *Biometrics* **55**, 117–128.

Birnbaum, A. (1968). Some latent trait models and their use in inferring an examinee's ability. In: Lord , F. M. and Novick, M. R. (editors), *Statistical Theories of Mental Test Scores*. Reading: Addison-Wesley, pp. 397–497.

Bock, R. D. and Lieberman, M. (1970). Fitting a response model for *n* dichotomously scored items. *Psychometrika* **35**, 179–197.

Braeken, J., Tuerlinckx, F. and De Boeck, P. (2007). Copulas for residual dependency. *Psychometrika* **72**, 393–411.

Clayton, D. G. (1978). A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence. *Biometrika* **65**, 141–151.

Cook, R. D. and Johnson, M. E. (1981). A family of distributions to modeling non-elliptically symmetric multivariate data. *Journal of the Royal Statistical Society, Series B* **43**, 210–218.

Frank, M. J. (1979). On the simultaneous associativity of $F(x, y)$ and $x + y - F(x, y)$. *Aequationes Mathematica* **19**, 194–226.

Gumbel, E. J. (1960). Distributions des valeurs extrêmes en plusieurs dimensions. *Publications de l'Institut de Statistique de l'Université de Paris* **9**, 171–173.

Hartford, A. and Davidian, M. (2000). Consequences of misspecifying assumptions in nonlinear mixed effects. *Computational Statistics and Data Analysis* **34**, 139–164.

Holland, P. W. and Rosenbaum, P. R. (1986). Conditional association and unidimensionality in monotone latent variable models. *Annals of Statistics* **14**, 1523–1543.

Downloaded from http://biostatistics.oxfordjournals.org/ at KU Leuven University Library on January 22, 2015

HOLLAND, P. W. AND WAINER, H. (1993). *Differential Item Functioning.* Hillsdale: Lawrence Erlbaum Associates.

HOLMES, L. B., COULL, B. A., DORFMAN, J. AND ROSENBERGER, P. B. (2005). The correlation of deficits in IQ with midface and digit hypoplasia in children exposed in utero to anticonvulsant drugs. *Journal of Pediatrics* **146**, 118–122.

HOLMES, L. B., HARVEY, E. A., COULL, B. A., HUNTINGTON, K. B., KHOSHBIN, S. K., HAYES, A. M. AND RYAN, L. M. (2001). The teratogenicity of anticonvulsant drugs. *New England Journal of Medicine* **344**, 1132–1138.

HOLMES, L. B., HARVEY, E. A., KLEINER, B. C., LEPPIG, K. A., CANN, C. I., MUNOZ, A. AND POLK, B. F. (1987). Predictive value of minor anomalies: II. Use in cohort studies to identify teratogens. *Teratology* **36**, 291–297.

HOSKENS, M. AND DE BOECK, P. (1997). A parametric model for local item dependencies among test items. *Psychological Methods* **2**, 261–277.

HOUGAARD, P. (1986). A class of multivariate failure time distributions. *Biometrika* **73**, 671–678.

IP, E. (2002). Locally dependent latent trait model and the dutch identity revisited. *Psychometrika* **67**, 367–386.

JOE, H. (1997). *Multivariate Models and Dependence Concepts.* London: Chapman & Hall.

JUNKER, B. W. (1991). Essential independence and likelihood-based ability estimation for polytomous items. *Psychometrika* **56**, 255–278.

LEFKOPOULU, M. AND RYAN, L. M. (1993). Global tests for multiple binary outcomes. *Biometrics* **49**, 975–988.

LEGLER, J. M. AND RYAN, L. M. (1997). Latent variable models for teratogenesis using multiple binary outcomes. *Journal of the American Statistical Association* **92**, 13–20.

LIANG, K.-Y., ZEGER, S. L. AND QAQISH, B. (1992). Multivariate regression analyses for categorical data. *Journal of the Royal Statistical Society, Series B* **54**, 3–40.

LORD, F. M. AND NOVICK, M. R. (1968). *Statistical Theories of Mental Test Scores.* Reading: Addison-Wesley.

MANTEL, N. AND HAENSZEL, W. (1959). Statistical aspects of the analysis of data from retrospective sudies of disease. *Journal of National Cancer Institute* **22**, 719–748.

MASTERS, G. N. (1988). Item discrimination: when more is worse. *Journal of Educational Measurement* **25**, 15–29.

MCLACHLAN, G. AND KRISHNAN, T. (1997). *The EM Algorithm and Extensions.* New York: John Wiley & Sons.

MOLENBERGHS, G. AND VERBEKE, G. (2005). *Models for Discrete Longitudinal Data.* New York: Springer.

MUTHÉN, B. O., BROWN, C. H., MASYN, K., JO, B., KHOO, S. T., YANG, C. C., WANG, C. P., KELLAM, S., CARLIN, J. AND LIAO, J. (2002). General growth mixture modeling for randomized preventive interventions. *Biostatistics* **3**, 459–475.

NELSEN, R. B. (1998). *An Introduction to Copulas.* New York: Springer.

NEUHAUS, J. M., HAUCK, W. W. AND KALBFLEISCH, J. D. (1992). The effects of mixture distribution misspecification when fitting mixed-effect logistic models. *Biometrika* **79**, 755–762.

O'RAHILLY, R. AND MULLER, F. (1996). *Human Embryology and Teratology.* New York: John Wiley & Sons.

SKLAR, A. (1959). Fonctions de répartition à n dimension et leurs marges. *Publications Statistiques Université de Paris* **8**, 229–231.

TATE, R. (2003). A comparison of selected empirical methods for assessing the structure of responses to test items. *Applied Psychological Measurement* **27**, 159–203.

TUERLINCKX, F. AND DE BOECK, P. (2001). The effect of ignoring item interactions on the estimated discrimination parameters in item response theory. *Psychological Methods* **6**, 181–195.

Downloaded from http://biostatistics.oxfordjournals.org/ at KU Leuven University Library on January 22, 2015

TUERLINCKX, F. AND DE BOECK, P. (2004). Models for residual dependencies. In: De Boeck, P. and Wilson, M. (editors), *Explanatory Item Response Models: A Generalized Linear and Nonlinear Approach* New York: Springer, pp. 289–316.

VERBEKE, G. AND LESAFFRE, E. (1996). A linear mixed-effects model with heterogeneity in the random effects population. *Journal of the American Statistical Association* **91**, 217–221.

WAINER, H., BRADLOW, E. AND WANG, X. (2007). *Testlet Response Theory and Its Applications*. Cambridge: Cambridge University Press.

YEN, W. M. (1984). Effects of local item dependence on the fit and equating performance of the 3 parameter logistic model. *Applied Psychological Measurement* **8**, 125–145.

[*Received June 6, 2007; first revision February 4, 2008; second revision March 12, 2009; third revision May 11, 2009 accepted for publication June 27, 2009*]

Downloaded from http://biostatistics.oxfordjournals.org/ at KU Leuven University Library on January 22, 2015