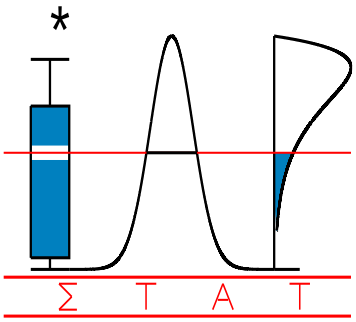
T E C H N I C A L

R E P O R T

0686

A LATENT VARIABLE FRAMEWORK FOR TERATOLOGY STUDIES

BRAEKEN, J., TUERLINCKX, F. and P. DE BOECK



I A P S T A T I S T I C S

N E T W O R K

INTERUNIVERSITY ATTRACTION POLE

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

A latent variable framework for teratology studies.

Johan Braeken
Francis Tuerlinckx
Paul De Boeck

Departement of Psychology
University of Leuven

Abstract

A latent variable framework is presented to model the characteristic rare multivariate binary anomaly data as provided by some teratology studies. The framework allows for a relaxation of two key assumptions underlying traditional latent variable models. To acquire a more flexible and datadriven way of specifying the distribution of the latent variable, finite mixture distributions and the inclusion of covariates are suggested to replace the more standard, but restrictive, assumption of normality. A copula approach is formulated to take into account possible violations of the conditional independence assumption (i.e., residual dependencies), and hereby specifying a more appropriate joint distribution of outcomes. It will be shown that these more technical elaborations of the traditional latent variable models also provide extra information and insight in the date, allowing the proposed latent variable framework to answer substantive questions about for instance general and anomaly-specific exposure effects of covariates, interrelations between anomalies, and individual objective diagnostic measurement.

KEYWORDS: latent variable model; finite mixture; conditional independence; copula; teratology.

1 Introduction

Several external agents such as chemicals, hyperthermia, radiation, or viruses, can cause abnormalities during the development of a fetus. These external agents (a.k.a., teratogens) play an important role in the domain of teratology, where doctors and other researchers are diagnosing birth defects and investigating the causal processes or etiology behind it (tera being the Greek word for monster). About 7 to 10 % of all children will require extensive medical care to diagnose or treat a birth defect. Although significant progress has been made in identifying etiologic causes of some birth defects, approximately 65% still have no known or identifiable cause [1]. This last fact can be ascribed to the inherent complexity of the domain: Few human teratogens have one single well-defined effect, but rather generate a set of (possibly partly overlapping) birth defects out of a variety of deficiencies, malformations, and anomalies. Therefore, multiple outcomes have to be assessed in teratology studies, typically resulting in characteristic rare multivariate binary data (e.g., defects present/absent).

One such example, which motivates this manuscript, is the Boston Anticonvulsant Teratogenesis study (BAT;[2]). From the 687 infants in our BAT dataset, there were 168 (24%) whose mothers took anticonvulsants during pregnancy, 73 (11%) whose mothers did not continue the anticonvulsant drug therapy (but were known to have a history of epileptic seizures), and 446 (65%) infants functioning as a control group (i.e., whose mothers didnt suffer from epileptic seizures, nor took anticonvulsant drugs). Each infant in the study was assessed on the presence or absence of several anomalies. The data set under consideration consists of 10 anomalies, going from facial anomalies (e.g., a depressed nasal bridge) to growth indicators (e.g., a small head) and other fysical features such as hypoplastic finger- and toenails. Summary statistics are presented in Table 1. Although most of these anomalies are not of clinical importance themselves, they are of clinical interest due to their possible predictive power as indicators or markers of more serious, but not yet emerged, anomalies and further developmental problems [3]. Note that this dataset is very similar to the one analyzed by Legler and Ryan[4], and based on the same clinical study.

This prototypical teratology study was set up with three goals in mind: (1) The first goal is situated on an individual-specific diagnostic level and based upon clinical experience indicating that the impact of teratogenic exposure varies over a dimension from extremely severe (i.e., where diagnosis is obvious) to mild (i.e., without major malformations, but still with an impact on functioning and development). From this point of view each infant can be given an underpinned position on this severity dimension. The patterns of anomalies typical for severely affected infants can be helpful in identifying or defining a syndrome, as well as providing new insights in possible underlying biological processes. The severity measure can also serve

as an objective quantified assessment that can be of later use for public health and welfare support, and policy makers. (2) The second goal is situated on a general etiologic level and motivated by the fact that newborns of epileptic mothers have more chance of showing adverse birth outcomes. Scientific interest goes out to differentiating two possible causes: Is this effect due to the mere presence of maternal epilepsy or an artefact of the related anti-convulsant drug therapy, and thus ascribed to in utero exposure to the medication? (3) The third goal is situated on a general level as there is also some interest in how these adverse birth outcomes relate to each other and whether some particular pattern of anomalies can be grouped together (and possibly be identified as a syndrome).

Legler et.al. [5] already partly discussed how the appropriate statistical analysis depends on the specific context and goals of the clinical study. The quantification of the unobserved severity at which an infant is affected, based upon the assessment of multiple birth outcomes (cfr., the third goal of the BAT), can be regarded as a measurement scale problem and is suited for a latent variable modeling approach. The latent variable will be exactly this unobserved severity dimension and allows for inference on the infant level (see e.g, [4]). If the goal is comparing a control group and several exposed groups (cfr., the second goal of the BAT) then a multiple comparison testing approach with a global test for multiple outcomes is adequate (see e.g., [6, 7, 8]). To explore relations between the different outcomes (cfr., the third goal of the BAT) cluster analysis techniques are the most common approach.

In the next section we will propose a unified latent variable model framework to tackle these three goals simultaneously. This by extending and finetuning a basic latent variable model: (1) A finite mixture distribution will be adopted both to identify groups of infants with similar patterns of anomalies, and to allow for quantitative inter-infant differences in order to obtain a suitable latent variable measurement scale. It should be stressed that the finite mixture approach can be seen as a technique to reveal hidden groups of infants in the data, but that it can also be considered as a way of avoiding the standard normality assumption for the distribution of the latent variable. A normal latent distribution is often too restrictive and chosen rather out of convenience, then based on substantive reasons/theory. (2) Person covariates will be added to the model in a regression-like fashion to comply to the interest in exposure group effects; (3) Because anomalies may cluster together for various reasons, known and unknown, it can be expected that the commonly made assumption of conditional independence in latent variable models will be violated in this case study. Specific clusters of anomalies will cause dependence over and above the one explained by the latent variable. As will be explained below, when not handled carefully, this conditional dependence can introduce bias into the model estimates and may yield quite misleading information. Copula functions will be incorporated

into the model to account for these specific associations between anomalies. In the Application section this latent variable modeling framework will be applied to the above-presented BAT study. Results and conclusions will be presented, before turning to a more general discussion.

2 Method

2.1 Latent variable framework

Let $\boldsymbol{Y}_p = (Y_{p1}, Y_{p2}, \ldots, Y_{pI})^T$ represent the $(I \times 1)$ binary anomaly outcome vector for the pth infant, and $\boldsymbol{Z}_p = (Z_{p1}, Z_{p2}, \ldots, Z_{pJ})^T$ the $(J \times 1)$ covariate vector for that same infant p. Assume Y_{pi} is an indicator of the event that some unobserved latent continuous variable X_{pi} (following a standard logistic distribution, i.e., location 0 and scale 1) exceeds a threshold, which can be taken to be zero without loss of generality. If $Y_{pi} = 1$, it indicates the presence of the ith anomaly in infant p and $Y_{pi} = 0$ indicates the abscence. Specifically let

$$Y_{pi} = I(X_{pi} > 0), \;\; X_{pi} = \alpha_i(\theta_p - \beta_i) + \varepsilon_{pi}, \tag{1}$$

such that

$$\Pr(Y_{pi} = 1|\theta_p) = \Pr(X_{pi} > 0|\theta_p) = \Pr(\varepsilon_{pi} > -\alpha_i(\theta_p + \beta_i));$$

with θ_p being an infant specific intercept representing the unobserved severity at which an infant p has been affected (higher being more severely affected), the parameters α_i and β_i representing the discrimination ability and threshold rate of the ith anomaly respectively. Figure 1 illustrates the function of these parameters with respect to the logistic response curve $\Pr(Y_{pi} = 1|\theta_p)$ over the range of the latent variable θ_p. The smaller the threshold rate β_i is, the higher the chance of having the ith anomaly for an infant with severity θ_p. Notice that the value of β_i is exactly the location on the latent scale θ_p where the probability of having the anomaly i is equal to a half (i.e., β_i is the location of the logistic curve). The larger the discrimination α_i is, the higher the effect of the severity of affect θ_p on the occurence of anomaly i. In a way, α_i indicates how effectively the anomaly i can discriminate between highly affected infants and less severely affected infants.

The resulting model for the joint outcome vector $\boldsymbol{Y}_p = (Y_{p1}, Y_{p2}, \ldots, Y_{pI})^T$ is in fact a Non-Linear Mixed Model (NLMM) that takes the dependency within the multivariate binary outcome vector $\boldsymbol{Y}_p$ of an infant p into account by introducing θ_p as a random effect and using a logit link function,

leading to the following joint marginal probablity:

$$\begin{aligned}
\Pr(\boldsymbol{Y}_p = \boldsymbol{y}_p) &= \int_{\theta_p} f(\boldsymbol{y}_p|\theta_p) h(\theta_p;\boldsymbol{\zeta}) d\theta_p \\
&= \int_{\theta_p} \prod_{i=1}^{I} f(y_{pi}|\theta_p) h(\theta_p;\boldsymbol{\zeta}) d\theta_p \\
&= \int_{\theta_p} \prod_{i=1}^{I} \frac{\exp(y_{pi}\alpha_i(\theta_p - \beta_i))}{1+\exp(\alpha_i(\theta_p-\beta_i))} h(\theta_p;\boldsymbol{\zeta}) d\theta_p,
\end{aligned}$$

where $h(\theta_p;\boldsymbol{\zeta})$ is the density of the latent distribution of the severity of affect θ_p. This distribution is parameterized by the vector $\boldsymbol{\zeta}$. Note that because of identfication reasons, the model should be restricted to fix the scale of the latent variable. One option is to fix one α_i, let's say α_1 at value 1. Hence α_2 until α_I have to be interpreted in reference to the first anomaly. However, we choose to put a restriction on the parametrization of the latent distribution $h(\theta_p;\boldsymbol{\zeta})$ leaving α_1 until α_I free and avoiding the hassle of a relative interpretation.

In the next few subsections we will go into more detail about the assumptions behind this basic model, highlight possible problems and outline model extensions to capture specific features of the data from the BAT study.

2.2 Specification of the latent distribution

The random intercept θ_p follows a density function h, which has a certain form, generally assumed to be a Normal with mean zero and unknown variance σ^2. The standard practice of choosing the specific parametric form of a Normal distribution is mainly motivated by mathematical convenience, tradition and the fact that it is provided by most commercial statistical software packages. However, this assumption 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 (see e.g., [9, 10, 11, 12]). Furthermore this assumption is difficult to check, because the random effect is a latent variable, and consequently unobservable.

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 random effect will be questionable. A positively skewed distribution for the random intercept (i.e., the severity of affect dimension) will be more suitable for these data. Note that this is the exact reason why Legler and Ryan [4] turned to a Poisson distribution for θ_p instead of the Normal distribution.

To accommodate for possible specification problems, one could choose to rely on finite mixture distributions, which are a more data-driven semi-

parametric way (thus, avoiding the specification of a parametric form) allowing for a more flexible and appropriate latent distribution (see e.g., [13]). Let the mixture distribution $G(\theta_p)$ consist of G normal distributions $H(\theta_p, \sigma^2)$, with mean vector $\boldsymbol{\mu} = [\mu_1, \ldots, \mu_G]^T$, common variance σ^2, and component probabilities $\pi_1, \ldots, \pi_g, \ldots, \pi_G$ such that $\sum_{g=1}^{G} \pi_g = 1$. Note that the common variance constraint is needed to avoid infinite likelihoods, as we are working with a unidimensional latent variable ([14], p.199). Each component distribution of the mixture will cover a local area of the true distribution, hereby enabling the modeling of a quite complex distribution. The choice of normal component distributions is arbitrary and chosen here for ease of implementation, but in principle any other parametric distribution can be an option. Extending our model with a finite mixture distribution of normals for the random effect θ_p results in the following marginal probability of the outcome vector:

$$\Pr(\boldsymbol{Y}_p = \boldsymbol{y}_p) = \sum_{g=1}^{G} \pi_g \int_{\theta_p} f(\boldsymbol{y}_p|\theta_p) N_g(\mu_g, \sigma^2) d\theta_p,$$

with μ_1 restricted to be zero (to set the scale of the latent variable).

Additionally the use of mixture distributions also allows the clinical team to investigate and reveal possible hidden grouping in the data. In this mixture setting infants can be ascribed to the component for which they have the highest posterior probability to belong to, resulting in a classification of infants over components. By characterizing the typical infants belonging to a specific component the researchers might attempt to interprete the underlying reason of the component separation and hereby possibly gaining new unexpected insights in the data and the teratogenic processes involved. Thus, besides the technical advantages, a finite mixture distribution for the latent variable can also provide new and more information at a general etiologic and specific diagnostic level when trying to answer the goals of a teratology study.

2.3 Specification of covariates

With the availability of additional covariate information as represented by $\boldsymbol{Z}_p = (Z_{p1}, Z_{p2}, \ldots, Z_{pJ})^T$, it is of interest to estimate the extent to which the severity of affect θ_p of an infant is determined by these infant characteristics. For instance, we know that 24% of the infants have a mother that took anticonvulsants during pregnancy, let us call these the drug exposure group. One can hypothesise that this exposure group will have on average a higher severity of affect θ_p than infants not belonging to this group. The connection with the previously discussed finite mixtures is easily made. Instead of hidden groups and unknown group membership, we now have known groups and known group membership (Z_{pj}): if $Z_{pj} = 1$ then $H(\theta_p; \boldsymbol{\zeta}) \sim N(\lambda_j, \sigma + \sigma_j^2)$,

else $H(\theta_p;\boldsymbol{\zeta}) \sim N(0,\sigma^2)$. This results in the following marginal probablity for the joint outcome vector:

$$\Pr(\boldsymbol{Y}_p = \boldsymbol{y}_p) = \int_{\theta_p} f(\boldsymbol{y}_p|\theta_p) N_g \left(\sum_{j=1}^{J} Z_{pj}\lambda_j, \sigma^2 + \sum_{j=1}^{J} Z_{pj}\sigma_j^2 \right) d\theta_p. \qquad (2)$$

This can be seen as a form of multi-level modeling, where one decomposes the severity of affect θ_p in a fixed part λ_j, representing the effect of the known covariate, and a random part θ_p^*, representing the remaining latent infant specific severity:

$$Y_{pi} = I(X_{pi} > 0),\ X_{pi} = \alpha_i \left([\theta_p^* + \sum_{j=1}^{J} Z_{pj}\lambda_j] - \beta_i \right) + \varepsilon_{pi}. \qquad (3)$$

The covariate effect λ_j can be interpreted as the change on the logit scale one would expect when a newborn has the jth covariate or infant characteristic, relative to when that characteristic is absent; this is equivalent to a multiplication of the odds ratio by $\exp(\lambda_j)$. In the above model, it is assumed that there is a general effect of the covariate; however, it can be that even after controlling for this general effect on an infants severity of defect the covariate still has an additional differential effect for some of the anomalies.

To accommodate for these additional anomaly-specific covariate effects, the model can be relaxed by constructing the covariate effect as the sum of a general effect λ_j and a specific effect λ_{ji} :

$$Y_{pi} = I(X_{pi} > 0),\ X_{pi} = \alpha_i \left([\theta_p^* + \sum_{j=1}^{J} Z_{pj}\lambda_j] - [\beta_i + \sum_{j=1}^{J} Z_{pj}\lambda_{ji}] \right) + \varepsilon_{pi}. \qquad (4)$$

In this model λ_{ji} can be interpreted as the change in log odds ratio for the ith anomaly when a newborn has the jth covariate or infant characteristic, relative to when that characteristic is absent. Note that to identify this model and keep the scale of the latent variable comparable between the covariate groups, one λ_{ji} has to be fixed for each covariate j. Here λ_{ji} can be interpreted as the additional change (above the general effect λ_j) in log odds ratio for the ith anomaly when a newborn has the jth covariate or infant characteristic, relative to when that characteristic is absent (cfr., differential item functioning in item response theory; see e.g., [15]).

If one is not interested in the overall general effect of the covariates, but only in anomaly-specific effects one can opt for the following formulation:

$$Y_{pi} = I(X_{pi} > 0),\ X_{pi} = \alpha_i \left(\theta_p - [\beta_i + \sum_{j=1}^{J} Z_{pj}\lambda_{ji}] \right) + \varepsilon_{pi}. \qquad (5)$$

In this last case, λ_{ji} can be interpreted as the log odds ratio for the ith anomaly when a newborn has the jth covariate or infant characteristic, relative to when that characteristic is absent (see also, [16]). This last formulation has the slight interpretational disadvantage that one can not formulate a general θ_p, because an infant's total severity of affect is now also anomaly specific: $\theta_{pi} = \theta_p^* + \lambda_{ji}$. One can still consider θ_p^* as being the infant's residual severity of affect not accounted for by the known covariates.

The inclusion of such covariate information on infants in the model has the statistical advantage that more precise estimates for both the fixed and random model parameters can be made [17], avoids multiple comparisons (hence, being statistically more efficient), and allows for general and anomaly-specific exposure effects. From a clinical perspective this offers a way of exploring possible risk and protective factors for teratogenesis (i.e., the development of anomalies).

2.4 Specification of the joint distribution

Traditionally latent variables model the joint distribution of the outcome vector $\boldsymbol{Y}_p$ as:

$$\Pr(\boldsymbol{Y}_p = \boldsymbol{y}_p|\theta_p) = \prod_{i=1}^{I} f(y_{pi}|\theta_p), \tag{6}$$

indicating that there is conditional independence between the different outcomes. This means that the anomaly outcomes are assumed to be independent realizations conditional upon the latent severity variable θ_p (in addition to the covariates and other fixed effects). and thus, the dependency in the data is ascribed to the fact that repeated measurements were taken from the same infant p. Equivalently, this assumption can be written as $F_{\boldsymbol{\varepsilon}_p}(\varepsilon_{p1}, \ldots, \varepsilon_{pI}) = \prod_{i=1}^{I} F_{\varepsilon_{pi}}(\varepsilon_{pi})$. Thus, another way to look at the conditional independence assumption is having uncorrelated error terms ε_{pi} (over items).

This assumption allows for a mathematical convenient way of modeling the joint probability of the anomaly outcome vector, but might not always be that plausible. In a teratology context it is likely that even after accounting for the common dependence due to a person's severity of affect, specific anomalies will show some extra association or residual dependency. 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, for instance, the growth indicators can be expected to show this type of extra association.

When ignored, violations of the conditional independence assumption may lead to biased estimates for both fixed and random effects [18, 19, 20]. To illustrate this intuitively, consider the extreme case wherein several slightly differently phrased screening questions are used in one test; this

is almost equivalent to assessing a single anomaly. This redundancy situation will lead to an inflation or double-counting of information. One of the consequences are for instance the underestimation of the standard error of the infants severity estimate θ_p, resulting in a less reliable measuring instrument. The estimation of the discrimination and baseline parameter may be affected as well by unmodeled residual dependencies. The main empirical finding (see e.g., [21, 20]) is that the discrimination parameters are overestimated or even explode to an extreme value when positive residual dependencies between anomalies are not taken into account.

Various types of diagnostic tools have already been developed to detect residual dependencies (see e.g., [22, 23]; for a comparison see [24]). In this manuscript the Mantel-Haenszel procedure ([25]; see also [26]) will be used as a data-analytical tool to identify pairs of anomaly outcomes that exhibit residual dependency beyond random chance. The main idea is to test for equal odds ratio between groups of infants; the group division is based upon a proxy for the severity of affect θ_p. Usually the somscore over anomalies is taken as a rough approximation of θ_p, or one could also opt to use a model-based estimate. The Mantel-Haenszel test will give an indication of the presence of residual dependency between the anomalies while controlling for the severity dimension θ_p. It is recommended to regard these Mantel-Haenszel tests mainly as an exploratory screening tool and not as a formal statistical test. In the current context, these Mantel-Haenszel tests and other types of diagnostic tests are cumbersome due to the multiple testing on the same interrelated data and the fact that they don't take into account the uncertainty as implied by the proxy $\hat{\theta_p}$.

Figure 2 shows a matrix made up by a pairwise crossing of all anomalies and contains a color representation of the value of the Mantel-Haenszel test statistic for each of these pairs. The cell at row 4 and collumn 9 contains the Mantel-Haenszel test statistic for the anomalies 4 and 9 (i.e., tapered fingers and clinodactyly of the hand, respectively; cfr. Table 1). Dark colors represent high values and indicate much residual dependency, light colors represent low values and indicate little residual dependency; Because the order of the anomaly pair is not relevant, as the residual dependency is considered symmetrical, the lower triangle of the matrix is left open as it only mirrors the values of the upper triangle. Upon visual inspection we pick out the pair $\{1, 2\}$, i.e., the anomalies involving hypoplasia. Their Mantel-Haenszel statistic is extremely high (M-H $Z = 4.57$), indicating that extra association above the dependency explained by the common severity of defect dimension is present between these two anomalies. Note that the average expected value of the Mantel-Haenszel test statistic is zero, as the standardized test statistic is reported. As expected the anomalies 6, 7, and 8, a.k.a. the growth indicators, form a residual dependent subset of anomalies (M-H: $Z > 6.88$ for the 3 pairwise combinations). Most other cells formed by a combination of an anomaly with one of the growth indicators also seem

to be influenced by this cluster. However, the values of the Mantel-Haenszel statistic for these cells are rather low compared to the two prominent clusters (exception should be made for cell $\{3,6\}$, M-H $Z = -5.08$) and the pattern of residual dependencies is not consistent (in direction or over the different combinations).

These preliminary findings clearly imply that further analysis of the BAT study has to take into account the presence of these residual dependencies. Consequently a more appropriate dependence structure will be required than the one proposed by the conditional independency assumption. In this manuscript we propose the use of copula functions to overcome the abovementioned bias problems by modeling detected (or theoretical) residual dependencies in the data, and thus providing a more proper formulation of the joint distribution. In order that this manuscript should be reasonably self-contained, an Appendix has been added for people less familiar with copula functions. A more thorough overview of copula theory can be found in the reference works by Nelsen[27], and Joe[28].

The main idea in our approach is to use copula functions to construct a more appropriate joint distribution function $F_{\boldsymbol{\varepsilon}_p}(\varepsilon_{p1}, \ldots, \varepsilon_{pI})$ that takes into account the association between the errorterms ε_{pi}, hereby modeling the residual dependencies. The marginal distributions $F_{\varepsilon_{pi}}$ are distributed following the general model proposed in Equation 1. The unknown I-variate joint distribution $F_{\boldsymbol{\varepsilon}_p}(\varepsilon_{p1}, \ldots, \varepsilon_{pI})$ will be constructed from these margins by means of copula functions.

Consider S disjoint subsets of $\{1, \ldots, I\}$ denoted as $J_1, \ldots, J_S$, where J_s has cardinality I_s. The vector of errorterms $\boldsymbol{\varepsilon}_p$ is similarly divided into subsets $\boldsymbol{\varepsilon}_p^{(1)}, \ldots, \boldsymbol{\varepsilon}_p^{(S)}$ where $\boldsymbol{\varepsilon}_p^{(s)} = (\varepsilon_{pi}, i \in J_s)$. The different subsets are independent, and the variables in a subset $\boldsymbol{\varepsilon}_p^{(s)}$ are assumed exchangeable. Subsets of anomalies can be chosen based upon diagnostic tests for residual dependency, or substantive theory.

The joint probability of the outcome vector of an infant p is:

$$\Pr(\boldsymbol{Y}_p|\theta_p) = \prod_{s=1}^{S} \Pr(Y_{pi} = y_{pi}, i \in J_s|\theta_p),$$

with the joint probability of responses on anomaly subset J_s equal to:

$$\Pr(Y_{pi} = y_{pi}, i \in J_s|\theta_p) = \Pr(d_{pi}^{(1)} < \varepsilon_{pi} \leq d_{pi}^{(2)}, i \in J_s|\theta_p),$$

where for $y_{pi} = 1$, $d_{pi}^{(1)} = -\theta_p + \beta_i$ and $d_{pi}^{(2)} = \infty$, and for $y_{pi} = 0$, $d_{pi}^{(1)} = -\infty$ and $d_{pi}^{(2)} = -\theta_p + \beta_i$. If the cardinality of subset J_s is larger than one ($I_s > 1$), $\Pr(d_{pi}^{(1)} < \varepsilon_{pi} \leq d_{pi}^{(2)}, i \in J_s|\theta_p)$ is evaluated from the copula $C_S(.;\delta_s)$ for

$(\varepsilon_{pi}, i \in J_s)$ as:

$$\Pr(d_{pi}^{(1)} < \varepsilon_{pi} \leq d_{pi}^{(2)}, i \in J_s|\theta_p) = \sum_{k_1=1}^{2} \cdots \sum_{k_{I_s}=1}^{2} (-1)^{k_1+\ldots+k_{I_s}} C_s \left(F_{\varepsilon_{p1}}(d_{p1}^{(k_1)}), \ldots, F_{\varepsilon_{pi}}(d_{pi}^{(k_i)}), \ldots, F_{\varepsilon_{pI_s}}(d_{pI_s}^{(k_{I_s})}) \right).$$

For clarity, assume $I = 2$ and $J_s = \{1, 2\}$, the equations above simplify as follows for a $(0, 0)$-outcome:

$$\begin{aligned}
&\Pr(Y_{p1} = 0, Y_{p2} = 0|\theta_p) \\
&\quad = \Pr(d_{pi}^{(1)} < \varepsilon_{pi} \leq d_{pi}^{(2)}, i \in J_s|\theta_p) \\
&\quad = \Pr(d_{p1}^{(1)} < \varepsilon_{p1} \leq d_{p1}^{(2)}, d_{p2}^{(1)} < \varepsilon_{p2} \leq d_{p2}^{(2)}|\theta_p) \\
&\quad = \Pr(-\infty < \varepsilon_{p1} \leq -\theta_p + \beta_1, -\infty < \varepsilon_{p2} \leq -\theta_p + \beta_2) \\
&\quad = C_s \left(F_{\varepsilon_{p1}}(-\infty), F_{\varepsilon_{p2}}(-\infty) \right) - C_s \left(F_{\varepsilon_{p1}}(-\theta_p + \beta_1), F_{\varepsilon_{p2}}(-\infty) \right) \\
&\quad - C_s \left(F_{\varepsilon_{p1}}(-\infty), F_{\varepsilon_{p2}}(-\theta_p + \beta_2) \right) + C_s \left(F_{\varepsilon_{p1}}(-\theta_p + \beta_1), F_{\varepsilon_{p2}}(-\theta_p + \beta_2) \right) \\
&\quad = 0 - 0 - 0 + C_s \left(F_{\varepsilon_{p1}}(-\theta_p + \beta_1), F_{\varepsilon_{p2}}(-\theta_p + \beta_2) \right) \\
&\quad = C_s \left(F_{X_{p1}|\theta_p}(0|\theta_p), F_{X_{p2}|\theta_p}(0|\theta_p) \right).
\end{aligned}$$

The last equality follows from the definition of X_{pi}. The other probabilities are then:

$$\begin{aligned}
\Pr(Y_{p1} = 1, Y_{p2} = 1|\theta_p) &= 1 - F_{X_{p1}|\theta_p}(0|\theta_p) - F_{X_{p2}|\theta_p}(0|\theta_p) + C_s \left(F_{X_{p1}|\theta_p}(0|\theta_p), F_{X_{p2}|\theta_p}(0|\theta_p) \right) \\
\Pr(Y_{p1} = 1, Y_{p2} = 0|\theta_p) &= F_{X_{p2}|\theta_p}(0|\theta_p) - C_s \left(F_{X_{p1}|\theta_p}(0|\theta_p), F_{X_{p2}|\theta_p}(0|\theta_p) \right) \\
\Pr(Y_{p1} = 0, Y_{p2} = 1|\theta_p) &= F_{X_{p1}|\theta_p}(0|\theta_p) - C_s \left(F_{X_{p1}|\theta_p}(0|\theta_p), F_{X_{p2}|\theta_p}(0|\theta_p) \right)
\end{aligned}$$

Figure 3 offers an intuitive insight in these calculations by presenting the contour lines of bivariate logistic densities constructed by means of Frank copula, Cook-Johnson copula, and Gumbel-Hougaard copula. Each contour plot is divided into quadrants made up by the solid lines drawn at the latent thresholds (the dashed lines indicating the marginal means $\theta_p - \beta_i$). In order to calculate the joint probabilities from the joint distribution functions of the latent random variables $X_{p1}|\theta_p$ and $X_{p2}|\theta_p$, the volume under the density for the corresponding quadrant needs to be calculated (see e.g., [29], and see also Appendix: Equation 7).

The regular conditional independence model arises as a special case when $S = I$ and each subset J_s has size 1; or when the different C are assumed to be the product copula Π (equivalent to independence). As an example, consider a teratology study with $I = 7$ anomalies where anomalies 1 and 2 exhibit some symmetric residual dependence, the set of anomalies 3 to 5 also form a dependent subset (independent of the first) and anomalies 6 and 7 do not show any violation of the general conditional independence assumption and are independent of the first two subsets. Thus,

$\{1, \ldots, 7\}$ is partitioned as: $J_1 = \{1, 2\}, J_2 = \{3, 4, 5\}, J_3 = \{6\}$ and $J_4 = \{7\}$. The proposed I-variate distribution for the error component vector $\boldsymbol{\varepsilon}_p = (\varepsilon_{p1}, \ldots, \varepsilon_{pI})^T$ is then: $F_{\boldsymbol{\varepsilon}_p}(\varepsilon_{p1}, \ldots, \varepsilon_{pI}) = C_1\left(F_{\varepsilon_{p1}}(\varepsilon_{p1}), F_{\varepsilon_{p2}}(\varepsilon_{p2})\right) \times C_2\left(F_{\varepsilon_{p3}}(\varepsilon_{p3}), F_{\varepsilon_{p4}}(\varepsilon_{p4}), F_{\varepsilon_{p5}}(\varepsilon_{p5})\right) \times F_{\varepsilon_{p6}}(\varepsilon_{p6}) \times F_{\varepsilon_{p7}}(\varepsilon_{p7})$.

Because a copula model is a form of marginal modeling, a broad range of association structures (by means of different copula functions) for the subsets showing residual dependency can be compared without fundamentally changing the base model of the marginal probabilities. In the previous example, C_1 could be either Frank copula or Cook-Johnson copula, and the same holds for C_2. The parametrisation of each margin (i.e., anomaly) preserves its natural interpretation independent of other observed anomaly outcomes. This reproducibility property (see [30, 31, 32]), or what McCullagh[33] calls upward compatibility' allows for complex changes to the joint (i.e., multivariate) model without having to leave the attractive modeling framework as described by the latent variable model proposed in Equation 1. Figure 3 illustrates the degree and kind of dependence for several latent bivariate distributions $F_{\boldsymbol{X}_p|\theta_p}$ with standard logistic margins $F_{X_{p2}|\theta_p}$ and $F_{X_{p1}|\theta_p}$ constructed by means of Frank copula (upper 4 panels), Cook-Johnson copula (middle 4 panels), and Gumbel-Hougaard copula (lower 4 panels) for varying values of the association parameter δ. Notice that the copulas can capture a broad range of dependency and differ in the type of dependence they induce; for instance, Cook-Johnson copula has a prominent lower tail (i.e., more formally, $C(u, \ldots, u)/u$ converges to a constant c in $[0, 1]$ as $u \to 0$; [34, 27]), while the Gumbel-Hougaard copula has a prominent upper tail. On the other hand Frank copula leads to a similar kind of dependence in both tails.

In order to illustrate that the introduction of the copula can take residual dependencies into account, the odds ratio (conditional on θ_p) for anomalies 1 and 2 involved in the copula C can be computed as follows:

$$\begin{aligned} \mathrm{OR}(\theta_p) &= \frac{\Pr(Y_{p1} = 1, Y_{p2} = 1|\theta_p)\Pr(Y_{p1} = 0, Y_{p2} = 0|\theta_p)}{\Pr(Y_{p1} = 1, Y_{p2} = 0|\theta_p)\Pr(Y_{p1} = 0, Y_{p2} = 1|\theta_p)} \\ &= \frac{\left(1 - F_{X_{p1}|\theta_p}(0|\theta_p) - F_{X_{p2}|\theta_p}(0|\theta_p) + C\right)C}{\left(F_{X_{p2}|\theta_p}(0|\theta_p) - C\right)\left(F_{X_{p1}|\theta_p}(0|\theta_p) - C\right)}, \end{aligned}$$

with $C = C\left(F_{X_{p1}|\theta_p}(0|\theta_p), F_{X_{p2}|\theta_p}(0|\theta_p)\right)$. Using Frank copula, the value of the log odds ratio is then computed for several values of δ and for θ_p ranging from -4 to 4 and the result is shown in the upper panel of Figure 4. For Cook-Johnson copula, and the Gumbel-Hougaard copula, the same procedure is followed and this result is shown in the middle and lower panel of the same figure, respectively. For ease of demonstration the two margins were set equal to one another, with anomaly parameters $\alpha_i = 1$ and $\beta_i = 0$ $(i = 1, 2)$ and no covariate information, so that the log odds ratio (conditional on θ_p) was only a function of the copula's association parameter and

the marginal probabilities as determined by θ_p. From all three panels it can be seen that when the value of the copula parameter δ rises the log odds ratio also increases, indicating the copula parameter's function as an association measure. Furthermore it appears that for a fixed value of the copula parameter δ, there is a dependency between the log odds ratio and the value of the latent trait. Frank copula shows a more static residual dependence between the two anomalies (the log odds ratio is more or less constant, unless for large values of δ). On the contrary, Cook-Johnson copula has a dimensional residual dependence structure because the log odds ratio increases as θ_p increases, whereas for the Gumbel-Hougaard copula the log odds ratio increases as θ_p decreases. Our attention is restricted in this manuscript to these three Archimedean copulas because they are comprehensive (i.e., they can capture the whole range from independency towards absolute positive dependency), have existing multivariate extensions, simple form and parsimonious parametrization (i.e., only one parameter), and because they show such distinct, interesting, and contrasting patterns of association.

On a more general level, one might consider the copula function capturing anomalies containing similar additional information above the part they provide to the common severity of affect of an infant (similar, i.e., when restricting our attention to the case of positive residual dependency). Hence, besides its use as a technical vehicle to take into account severe violations of the conditional independency assumption, the copula can also be used as a measure or indicator of association between specific anomalies. The specific type of association induced by the copula can provide extra information about the underlying processes behind the occurence of these residual dependencies.

3 Model inference

Once the partitioning of anomalies into subsets is given, and the copula families are given, the following set of parameters has to be estimated: the anomaly parameters α_i and β_i $(i = 1, \ldots, I)$, the distributional parameters of the latent trait, and the association parameters $\delta_1, \ldots, \delta_S$. Over all anomalies and infants the marginal maximum likelihood under the copula latent variable model is:

$$\prod_{p=1}^{P} \int_{\theta_p} \prod_{s=1}^{S} \left[\Pr(d_{pi}^{(1)} < \varepsilon_{pi} \leq d_{pi}^{(2)}, i \in J_s | \theta_p) \right] \phi(\theta_p | \sigma^2) d\theta_p.$$

Usually the log is taken for numerical reasons. The negative of this log-likelihood is minimized using a quasi-newton optimization algorithm. The approximation of the intractable integral with respect to the distribution of θ_p will be carried out with a Gauss-Hermite quadrature. Estimates of the latent variable θ_p are acquired in a second step using empirical Bayes. In

case of a finite mixture specification of the latent variable θ_p optimization is performed using a generalized Expectation-Maximization algorithm with a quasi-newton iteration to solve the M-step (see e.g., [35, 36]). Note that one can best apply a multiple-rerun strategy for the EM algorithm as it is highly sensitive to initial starting values in this type of context. A program in Matlab has been written to estimate these copula models.

Comparing a conditional independence model against it's copula counterpart has a straightforward solution, as the independence copula is a special case of all copulas considered in this manuscript. Hence, likelihood tools usually applied in mixed models are available (i.e., Wald, score and likelihood ratio tests). Compared to the independence case, the only additional parameter in the copula functions applied here is the association parameter that defines the degree of dependency within a set of anomalies (conditional on θ_p). Note that for Frank copula Meester [37] showed that once R exceeds 2, the lower bound of the copula parameter δ needs to be adapted in function of R. For any R this adapted lower-bound is always strictly less than zero, thus not restricting the positive association range and technically leaving the position of the independence point in the interior of the parameter space. The range of negative dependency in the case that $R > 2$ is rather limited and attaching a meaningfull interpretation to negative dependency between three or more anomalies is not very likely. Hence, we do not consider the use of Frank copula for negative residual dependency between more than 2 anomalies for reasons of clarity. However for Cook-Johnson copula the independence case lies on the boundary of the parameter space since δ cannot be lower than zero. Therefore the appropriate reference distribution for the likelihood ratio test statistic comparing the independence and the Cook-Johnson model is not a chi-square with 1 degree of freedom but a mixture of two chi-square distributions, with 0 and 1 degree of freedom respectively. When more copulas and consequently more association parameters are involved, deducing the appropriate mixture of chi-square distributions to function as reference distribution may get very complicated (see e.g., [38]). Therefore, we will rely in all cases on the traditional reference distribution, which yields a more conservative test of the null hypothesis. Attention has to be given to the selection of a particular copula function. The 3 copulas considered in this manuscript do not have a nested relation and consequently, the selection is best based on methods such as the AIC [39] or BIC [40]. Given the comprehensive nature of these copulas and their shared boundary cases (i.e., the Fréchet bounds and the independence copula, see Appendix), one can expect that differentiation between the types of copula functions will get more difficult near the extremes of independence or deterministic positive dependence.

4 Application: Boston Anticonvulsant Teratogenesis study

As a start the base model in Equation 1 was fitted on the BAT data. The results, as shown in Table 2, clearly indicate that there are some estimation problems: The standard error of the threshold rate (i.e., β_i) for anomalies 3 and 5 is large, and the discrimination parameter (i.e., α_i) of anomalies 6, 7 and 8 takes quite extreme values. This last fact, together with the results of the Mantel-Haenszel tests, leads to the conclusion that the problem of residual dependencies in the data can not be ignored and have to be taken into account in our model. Of course these model estimates can still be influenced by another factor, being a possible misspecification of the distribution of the latent variable θ_p. To verify this and clear up any possible confounding between the two misspecification factors (i.e., joint distribution and latent distribution) a finite mixture version of the base model was fitted to acquire a more flexible and appropriate specification of the latent variable's distribution. The mixture model results in a better fit, but the previously mentioned problems remain. However, this part of our modeling effort can already give some more information and insight in the data. A finite mixture with two components, with means 0 and -5.14, and component probabilities 0.61 and 0.39 respectively, gave the best fit (see Table 2). The infants were classified into these two components based upon their maximum posterior component probability and a characterization of the resulting classification in terms of the available data is given in Table 3. Notice that the second component mainly contains unaffected infants and also has a lower percentage of infants belonging to the anticonvulsants-exposed group relative to the infants classified in the first component.

Furthermore, when adding the available covariates (i.e., exposed in utero to anticonvulsants, gender, and seizure-history of the mother) to the mixture model, it reduced to the regular one component case. The results of the one-component model with general covariate effects included are presented in Table 2 (Note that adding covariate effects to the variance of the latent distribution did not improve the model fit, and thus were excluded from the presented model). Only one of the available covariates resulted in an overall effect on an infant's severity of affect. Being exposed in utero to anticonvulsant drugs, relative to not being exposed to such drugs, raises the odds of having an anomaly by 3.67 (i.e., $\exp(\lambda_1)$). Gender or having a mother with a seizure history had no significant influence on an infant's severity of affect. We checked for the possibility of noticeable anomaly-specific effects of the covariates, but these models did not provide a better fit to the data compared to the more parsimonious model with general anomaly effects. Thus, these results suggest that the higher prevalence of anomalies in children of epileptic mothers is mainly due to their anticonvulsant medication therapy.

In comparison with the variance of the latent variable θ_p that was fixed at 1, the variance explained by anticonvulsant exposure is equal to: the variance of anticonvulsant exposure (0.19) multiplied by it's squared effect λ_1^2 is 0.27, which is 21% when added to latent variable's variance. These results indicate that for this data set incorporating the known covariates into the model, results in an appropriate specification of the latent variable's distribution, and suggest that a relatively large part of the between-infant variability is accounted for by these known covariates.

To take into account the detected residual dependencies in the data, a copula model was fitted. A Gumbel-Hougaard copula for for anomaly subset $\{1, 2\}$, and a Frank copula for anomaly subset $\{6, 7, 8\}$ resulted in the best fitting models. Both copula association parameters were significant ($p < 0.0001$). This suggest that the infants who are less severely affected (low θ_p), have a higher tendency to have both finger- and toenail hypoplasia anomalies present. For the growth indicators the residual dependency is more equally spread out over the severity of affect scale, but the high value of the association parameter δ_2 implies that the general tendency is that these growth indicators co-occur. If one type of growth deficiency is present, it is very likely that the other two are present as well.

To illustrate that ignoring residual dependencies leads to a less reliable measurement scale, a plot was made from the empirical Bayes standard errors of θ_p under the conditional independence model and the corresponding copula model (see Figure 6). To ease interpretation the difference between the standard errors under both model was set out on the Y-axis. If residual dependencies would not have an effect, one would expect no significant difference and all points in Figure 6 should be roughly lying around the zero-point reference line of the Y-axis. Clearly this is not the case, and one can see that in the presence of positive residual dependencies the double counting of information under a conditional independence model leads to an artificial, but systematic, underestimation of the standard error of θ_p. When comparing the discrimination parameter estimates (α_i) under the covariate model with conditional independence and the model with copula functions, one can observe that the α_i of the residual dependency subsets $\{1, 2\}$ and $\{6, 7, 8\}$ are corrected downwards under the copula model. This corresponds with [20], who showed that ignoring the presence of positive residual dependency leads to an overestimation of the discrimination parameters α_i.

When looking at the results of our final copula model, it shows that the anomalies with the highest threshold rate (i.e., they are 'hardest' to get) are the growth indicators and a broad nasal bridge (i.e., anomalies 6, 7, 8 and 10, respectively). Hypoplastic and tapered fingernails and anteverted nostrils (i.e., anomalies 2, 4 and 5, respectively) are the anomalies that best differentiate between less and more severly affected infants. Typical anomaly patterns for the less severly affected infants in this dataset as indicated by our measurement scale ($\theta_p <= 1; n = 561$) are having at most one anomaly

present or a combination of growth indicators. The most severly affected infants in this dataset as indicated by our measurement scale ($\theta_p > 1; n = 126$) show anomaly patterns with on average 4 anomalies present. Figure 5 shows the observed distribution of empirical bayes estimates of the infants severity of affect.

5 Discussion

The more general goal of this manuscript was to highlight the strong relation between statistically 'biased' models and substantively 'unrealistic' models. The specification of the latent distribution and the joint distribution given a latent variable model received focus. Instead of rashly accepting the standard approach of a normal distributed latent variable, a more flexible data-driven way of specifying the latent distribution can be obtained by adopting a finite mixture approach and by incorporating covariates into the model. Besides its statistical efficiency advantages, this approach can also provide new insights and more information to the clinical researchers. The same can be said about the traditional conditional independence assumption in the joint specification of latent variable models. A copula approach to take into account specific violations of this assumption, leads to more solid models and provides at the same time a measure of association between specific subsets of anomalies. Both these extensions of the standard approach to latent variable modeling, allow to work further within the interpretational attractive latent variable framework for teratology studies, and, should greatly improve the quality of the constructed measurement scale and model-based inferences. In other words, it is the believe of this authors that the here presented framework offers a viable alternative available to the standard approach taken. In principle, the approach suggested in this manuscript can also be applied to categorical and continuous variables. A limitation that has to be taken into account as the copula approach, as presented here, considers only a symmetrical type of dependencies.

Appendix: copula theory

In this section we will introduce the mathematical concept of copula functions and indicate their use for the modeling of dependency, and the construction of multivariate distributions. In mathematics, a copula (Latin for link or tie) defines a function that relates, or couples' a multivariate distribution function to its univariate margins.

An R-dimensional copula is a function $C : [0,1]^R \rightarrow [0,1]$ with the following properties:

1. For every vector $\boldsymbol{u} \in [0,1]^R, C(\boldsymbol{u})$ is increasing in each component u_r with $r \in 1, 2, \ldots, R$.

2. For every vector $\boldsymbol{u} \in [0,1]^R, C(\boldsymbol{u}) = 0$ if at least one coordinate of the vector is 0 and $C(\boldsymbol{u}) = u_r$ if all the coordinates of the vector are equal to one except the r-th one.

3. For every $\boldsymbol{a}, \boldsymbol{b} \in [0,1]^R$ with $\forall r \in \{1, 2, \ldots, R\}, a_r \leq b_r$, given a hypercube $\boldsymbol{B} = [\boldsymbol{a}, \boldsymbol{b}] = [a_1, b_1] \times [a_2, b_2] \times \ldots \times [a_R, b_R]$ whose vertices lie in the domain of C, and with $V_C(\boldsymbol{B}) \geq 0$. The volume $V_C(\boldsymbol{B})$ is defined as:

$$V_C(\boldsymbol{B}) = \sum_{k_1=1}^{2} \sum_{k_2=1}^{2} \cdots \sum_{k_R=1}^{2} (-1)^{k_1+k_2+\ldots+k_R} C\left(d_1^{(k_1)}, d_2^{(k_2)}, \ldots, d_R^{(k_R)}\right), \tag{7}$$

where $d_r^{(1)} = a_r$ and $d_r^{(2)} = b_r$ $(r = 1, \ldots, R)$.

An important result from the theory of copulas is Sklar's Theorem [41]. The theorem states that for any R-dimensional distribution function $F_{\boldsymbol{X}}$ with univariate margins $F_{X_1}, F_{X_2}, \ldots, F_{X_R}$ 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: $F_{\boldsymbol{X}} = C\left(F_{X_1}, F_{X_2}, \ldots, F_{X_R}\right)$. While the key idea behind Sklar's theorem is that any existing multivariate distribution can be reformulated according this copula presentation, we used in this manuscript the fact that the opposite also holds. Starting from given univariate distributions, we will construct a multivariate distribution using copulas. As a result the univariate marginals of this newly constructed multivariate distribution are still the original univariate distributions we started from. In this way the copula truly couples several univariate distributions. Thus, given the univariate margins $F_{X_1}, F_{X_2}, \ldots, F_{X_R}$ and a choice for the copula function C, we construct a multivariate distribution $F_{\boldsymbol{X}} = C\left(F_{X_1}, F_{X_2}, \ldots, F_{X_R}\right)$. Making use of the second property of the copula definition, it can easily be deduced that the univariate marginal distribution for X_r equals $C\left(1, \ldots, 1, F_{X_r}, 1, \ldots, 1\right) = F_{X_r}$. Hence, in this way an association between the R random variables is allowed while preserving the univariate margins.

For each joint distribution with margins $F_{X_1}, F_{X_2}, \ldots, F_{X_R}$, and constructed by means of a copula the following applies:

$$W\left(F_{X_1}, \ldots, F_{X_R}\right) \leq C\left(F_{X_1}, \ldots, F_{X_R}\right) \leq M\left(F_{X_1}, \ldots, F_{X_R}\right),$$

where $W\left(F_{X_1}, \ldots, F_{X_R}\right) = \max\left(F_{X_1} + \ldots + F_{X_R} - R + 1, 0\right)$, and $M\left(F_{X_1}, \ldots, F_{X_R}\right) = \min\left(F_{X_1}, \ldots, F_{X_R}\right)$. The functions $W\left(F_{X_1}, \ldots, F_{X_R}\right)$ and $M\left(F_{X_1}, \ldots, F_{X_R}\right)$ correspond to the Fréchet-Hoeffding lower and upper bound [42, 43] and define the maximum negative and positive dependency of a joint distribution that can be obtained given fixed margins. These bounds can be used to indicate the range of dependency a copula function can capture.

A variety of possible copula functions exist which allow for fitting a wide range of dependency types (see e.g., [34, 27]). Our modeling approach will focus mainly (but not exclusively) on the class of Archimedean copulas [44, 34, 27]. Archimedean copulas have a simple structure and can be written as:

$$C(u_1, \ldots, u_R) = \psi^{-1}\left(\sum_{r=1}^{R} \psi(u_r)\right), \tag{8}$$

where $\psi : [0,1] \rightarrow [0,\infty]$ is a continuous strictly decreasing function, called the generator function, such that $\psi(0) = \infty$ and $\psi(1) = 0$, and ψ^{-1} is completely monotonic on $[0,\infty)$, such that $(-1)^k \frac{d^k}{dt^k}\psi^{-1}(t) \geq 0 \ \forall t \in [0,\infty)$ and $k \in \mathbb{N}$. Archimedean copulas have nice symmetry properties, as there are permutation symmetry, $C(u_1, u_2) = C(u_2, u_1)$, and associativity, $C(u_1, u_2, u_3) = C(u_1, C(u_2, u_3)) = C(C(u_1, u_2), u_3)$; which makes them especially attractive for modeling symmetrically dependent data. Furthermore, notice that the independence case can also be rewritten in a convenient way under an Archimedean copula representation:

$$F_{\boldsymbol{X}} = \prod_{r=1}^{R} F_{X_r} = C(F_{X_1}, \ldots, F_{X_R}) = \exp\left(-\sum_{r=1}^{R}[-\log(F_{X_r})]\right),$$

with $\psi(u) = -\log(u)$ and $\psi^{-1}(t) = \exp(-t)$. This copula is also known as the product copula, denoted by Π [45, 27]. Table 4 presents three important instantiations of the class of Archimedean copulas: Frank copula [46], Cook-Johnson copula [47, 48], and Gumbel-Hougaard copula [49, 50]. For each of the presented copula functions in Table 4, the parameter δ captures the degree of association between the random variables. For instance, when the value of δ increases, the dependence captured by all copulas in Table 4 gets closer to the theoretical maximum positive dependence for $F_{\boldsymbol{X}}$ (i.e., M, the Fréchet-Hoeffding upper-bound). The table also mentions possible constraints on this dependency parameter δ, the range of the dependency that can be captured by the copula, and their generator functions are given as well.

Acknowledgements

The authors wish to thank Lewis Holmes, Louise Ryan and Mary Sammel for kindly providing data from the Boston Anticonvulsant Teratogenesis study.

References

[1] O'Rahilly R, Muller F, *Human embryology and teratologgy.*, John Wiley & Sons., New York, 1996.

[2] Holmes LB, Harvey EA, Brown KS, Hayes AM, Khoshbin S, Anticonvulsant teratogenesis: 1. a study design for newborn infants. *Teratology* 1994; 49: 202–207.

[3] Holmes LB, Harvey EA, Kleiner BC, Leppig KA, Cann CI, Munoz A, Polk BF, Predictive value of minor anomalies: Ii. use in cohort studies to identify teratogens. *Teratology* 1987; 36: 291–297.

[4] Legler JM, Ryan LM, Latent variable models for teratogenesis using multiple binary outcomes. *Journal of the American Statistical Association* 1997; 92: 13–20.

[5] Legler JM, Ryan LM, Harvey EA, Holmes LB, Anticonvulsant teratogenesis:2. statistical methods for multiple birth outcomes. *Teratology* 1994; 50: 74–79.

[6] Legler JM, Lefkopoulu M, Ryan LM, Efficiency and power of tests for multiple binary outcomes. *Journal of the American Statistical Association* 1995; 90: 680–693.

[7] Lefkopoulu M, Ryan LM, Global tests for multiple binary outcomes. *Biometrics* 1993; 49: 975–988.

[8] Lefkopoulu M, Moore D, Ryan LM, The analysis of multiple correlated binary outcomes. *Journal of the American Statistical Association* 1989; 84: 810–815.

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

[10] Hartford A, Davidian M, Consequences of misspecifying assumptions in nonlinear mixed effects. *Computational Statistics and Data Analysis* 2000; 34: 139–164.

[11] Heckman JJ, Singer B, A method for minimizing the impact of distributional assumptions in econometric models of duration. *Econometrica* 1984; 52: 271–320.

[12] Neuhaus JM, Hauck WW, Kalbfleisch JD, The effects of mixture distribution misspecification when fitting mixed-effect logistic models. *Biometrika* 1992; 79: 755–762.

[13] Aitkin M, A general maximum likelihood analysis of variance components in generalized linear models. *Biometrics* 1999; 55: 117–128.

[14] Böhning D, *Computer-assisted analysis of mixtures and applications: meta-analysis, disease mapping and others.*, Chapman & Hall, London, 1999.

[15] Holland PW, Wainer H, *Differential item functioning*, Lawrence Erlbaum Associates., Hillsdale, 1993.

[16] Das A, Poole WK, Bada HS, A repeated measurement approach for simultaneous modeling of multiple neurobehavioral outcomes in newborns exposed to cocaine in utero. *American Journal of Epidemiology* 2004; 159: 891–899.

[17] Mislevy RJ, Exploiting auxiliary information about examinees in the estimation of item parameters. *Aplied Psychological Measurement* 1987; 11: 81–91.

[18] Chen W, Thissen D, Local dependence indexes for item pairs using item response theory. *Journal of Educational and Behavioral Statistics* 1997; 22: 265–289.

[19] Junker BW, Essential independence and likelihood-based ability estimation for polytomous items. *Psychometrika.* 1991; 56: 255–278.

[20] Tuerlinckx F, De Boeck P, The effect of ignoring item interactions on the estimated discrimination parameters in item response theory. *Psychological Methods* 2001a; 6: 181–195.

[21] Masters GN, Item discrimination: when more is worse. *Journal of Educational Measurement* 1988; 25: 15–29.

[22] Holland PW, Rosenbaum PR, Conditional association and unidimensionality in monotone latent variable models. *Annals of Statistics.* 1986; 14: 1523–1543.

[23] Yen WM, Effects of local item dependence on the fit and equating performance of the three-parameter logistic model. *Applied Psychological Measurement* 1984; 8: 125–145.

[24] Tate R, A comparison of selected empirical methods for assessing the structure of responses to test items. *Applied Psychological Measurement.* 2003; 27: 159–203.

[25] Mantel N, Haenszel W, Statistical aspects of the analysis of data from retrospective sudies of disease. *Journal of National Cancer Institute* 1959; 22: 719–748.

[26] Rosenbaum PR, Testing the conditional independence and monotonicity assumptions of item response theory. *Psychometrika* 1984; 49: 425–435.

[27] Nelsen RB, *An introduction to copulas*, Springer, New York, 1998.

[28] Joe H, *Multivariate models and dependence concepts*, Chapman & Hall, London, 1997.

[29] Mood AM, Graybill FA, Boes DC, *Introduction to the theory of statistics*, McGraw-Hill, New York, 1974.

[30] Fitzmaurice GM, Laird NM, Rotnitzky AG, Regression models for discrete longitudinal responses. *Statistical Science* 1993; 8: 284–309.

[31] Ip E, Locally dependent latent trait model and the dutch identity revisited. *Psychometrika* 2002; 67: 367–386.

[32] Liang KY, Zeger SL, Qaqish B, Multivariate regression analyses for categorical data. *Journal of the Royal Statistical Society (Series B)* 1992; 54: 3–40.

[33] McCullagh P, Models for discrete multivariate responses. *Bulletin of the International Statistics Institute* 1989; 53: 407–418.

[34] Joe H, Parametric families of multivariate distributions with given margins. *Journal of Multivariate Analysis* 1993; 46: 262–282.

[35] Dempster A, Laird N, Rubin D, Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society, Series B* 1977; 39: 1–38.

[36] McLachlan G, Krishnan T, *The EM algorithm and extensions*, John Wiley & Sons., New York, 1997.

[37] Meester SG, Methods for clustered categorical data. 1991.

[38] Self GH, Liang KY, Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. *Journal of the American Statistical Association* 1987; 82: 605–610.

[39] Akaike H, Information theory and an extension of the maximum likelihood principle, in BN Petrov, C F (Eds.), *2nd International Symposium on Information Theory*, Tsahkadsov, Armenia, USSR, 1973, 267–281.

[40] Schwarz G, Estimating the dimension of a model. *Annals of Statistics* 1978; 6: 461–464.

[41] Sklar A, Fonctions de répartition à n dimension et leurs marges. *Publications Statistiques Université de Paris* 1959; 8: 229–231.

[42] Fréchet M, Sur les tableaux de corrélation dont les marges sont données. *Annales de l'Université Lyon: Série 3* 1951; 14: 53–77.

[43] Hoeffding W, Masstabinvariante Korrelations-Theorie. *Schriften des Matematischen Instituts und des Instituts für angewandte Mathematik der Universität Berlin* 1940; 5:3: 179–223.

[44] Genest C, MacKay J, Copules archimédiennes et familles de lois bidimensionelles dont les marges sont donées. *Canadian Journal of Statistics* 1986; 14: 145–159.

[45] Frees AW, Valdez EA, Understanding relationships using copulas. *Actuarial Research Clearing House* 1998; 1: 5–45.

[46] Frank MJ, On the simultaneaous associativity of $F(x, y)$ and $x + y - F(x, y)$. *Aequationes Mathematica* 1979; 19: 194–226.

[47] Clayton DG, A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence. *Biometrika* 1978; 65: 141–151.

[48] Cook RD, Johnson ME, A family of distributions to modeling nonelliptically symmetric multivariate data. *Journal of the Royal Statistical Society, Series B* 1981; 43: 210–218.

[49] Gumbel EJ, Distributions des valeurs extrêmes en plusieurs dimensions. *Publications de l'Institut de Statistique de l'Université de Paris* 1960; 9: 171–173.

[50] Hougaard P, A class of multivariate failure time distributions. *Biometrika* 1986; 73: 671–678.

Tables

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

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

Table 2: Models for the BAT data.

	base		mixture		covariate		copula	
β_1	2.46	0.52	2.82	0.65	1.94	0.20	2.25	0.28
β_2	2.43	0.40	2.67	0.54	1.97	0.16	2.10	0.19
β_3	4.97	2.02	4.79	1.09	2.85	0.39	2.56	0.31
β_4	2.54	0.42	3.07	0.64	2.40	0.23	2.47	0.25
β_5	3.37	0.81	3.97	0.87	2.61	0.27	2.52	0.26
β_6	1.32	0.08	1.02	0.11	3.44	0.65	8.40	3.97
β_7	1.65	0.09	1.39	0.12	3.53	0.57	5.74	1.69
β_8	1.30	0.09	0.97	0.12	3.06	0.51	4.82	1.22
β_9	3.14	0.72	4.11	0.91	3.41	0.57	3.62	0.66
β_{10}	18.25	18.71	6.51	1.46	4.77	1.26	4.26	0.95
α_1	0.67	0.16	0.37	0.07	1.30	0.20	0.99	0.17
α_2	1.02	0.21	0.63	0.13	2.72	0.58	2.22	0.45
α_3	0.38	0.16	0.30	0.06	0.85	0.16	1.02	0.18
α_4	0.96	0.20	0.55	0.11	1.43	0.22	1.39	0.23
α_5	0.70	0.19	0.44	0.09	1.26	0.21	1.40	0.23
α_6	3.95	0.93	3.06	0.70	0.75	0.19	0.26	0.13
α_7	4.85	1.08	4.12	1.25	1.05	0.26	0.54	0.19
α_8	3.01	0.56	2.35	0.48	0.79	0.18	0.44	0.13
α_9	0.66	0.17	0.36	0.07	0.69	0.15	0.65	0.15
α_{10}	0.10	0.11	0.23	0.05	0.44	0.13	0.51	0.14
μ_2	.	.	-5.14	0.76	.	.	.	.
π_2	.	.	0.39	0.03	.	.	.	.
$\lambda_{anticonvulsants}$	.	.	.	.	1.19	0.14	1.30	0.15
λ_{gender}	.	.	.	.	0.00	0.11	-0.01	0.12
$\lambda_{seizurehistory}$	.	.	.	.	0.46	0.19	0.39	0.20
G-H: $J_1\{1,2\};\delta_1$	.	.	.	.	.	.	1.42	0.15
F: $J_2\{6,7,8\};\delta_2$	.	.	.	.	.	.	13.51	1.99
-LogL	2317		2305		2286		2153	
	20		22		23		25	
AIC	4674		4654		4618		4356	

Table 3: Characterization of the Mixture components.

	percentage	component1 ($n_1 = 355$)	component2 ($n_2 = 332$)
1	TNHYP	0.35	0
2	FNHYP	0.21	0
3	DBN	0.27	0
4	TAPF	0.21	0
5	NATV	0.20	0
6	SMHEAD	0.22	0
7	SHORT	0.11	0
8	LOWBT	0.25	0
9	CLIN	0.25	0
10	BRBRDG	0.15	0.11
	anticonvulsants exposed	0.34	0.14
	male	0.51	0.47
	seizure history mother	0.12	0.10

Table 4: Overview of three Archimedean copulas.

	Frank copula	Cook-Johnson copula	Gumbel-Hougaard copula
$C\left(F_{X_1}, F_{X_2}, \ldots, F_{X_R}\right)$	$\frac{-1}{\delta}\log\left(1-\frac{\prod_{r=1}^{R}(1-\exp(-\delta F_{X_r}))}{\prod_{r=1}^{R-1}(1-\exp(-\delta))}\right)$	$\left(\sum_{r=1}^{R} F_{X_r}^{-\delta} - R + 1\right)^{\frac{-1}{\delta}}$	$\exp\left(-\left[\sum_{r=1}^{R}(-\log(F_{X_r}))^{\delta}\right]^{\frac{1}{\delta}}\right)$
$\psi(t)$	$-\log\left(\frac{1-\exp(-\delta t)}{1-\exp(-\delta)}\right)$	$t^{-\delta}-1$	$(-\log(t))^{\delta}$
Distribution	logarithmic series on positive integers	Gamma	Positive stable
$\psi^{-1}(s)$	$\frac{-1}{\delta}\log\left(1-(1-\exp(-\delta))\exp(-s)\right)$	$(1+s)^{-1/\delta}$	$\exp(-s^{1/\delta})$
parameter constraints	$R=2:\delta\in\mathbb{R}/\{0\}$, $R>2:\delta>0$	$\delta>0$	$\delta>1$
Range of dependency[1]	$\delta\to 0: C\to\Pi$ $\delta\to\infty: C\to M$ if $R=2, \delta\to-\infty: C\to W$	$\delta\to 0: C\to\Pi$ $\delta\to\infty: C\to M$ $\delta\to 0: C\to\Pi$	$\delta\to\infty: C\to M$

Figures

Figure 1: Illustration of the model parameter interpretation.

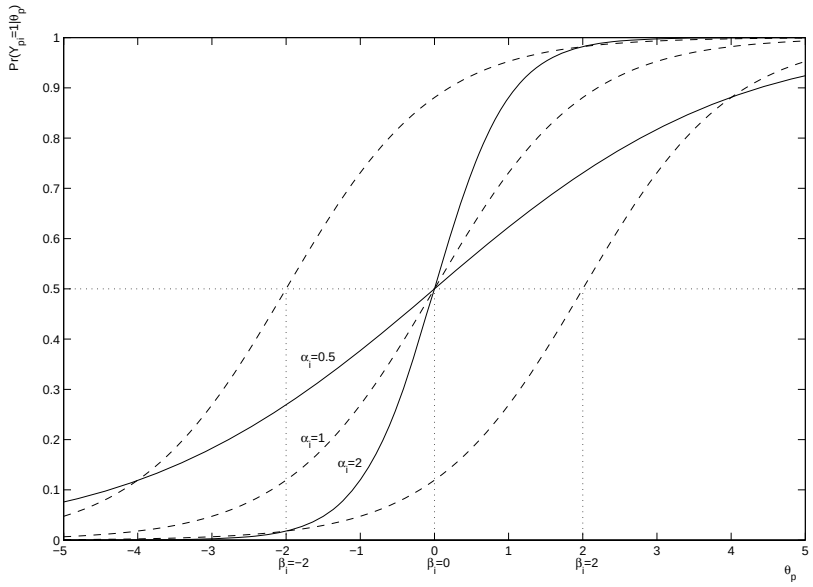


Figure 2: Mantel-Haenszel statistic gray map for the 10-by-10 matrix of anomaly pairs.

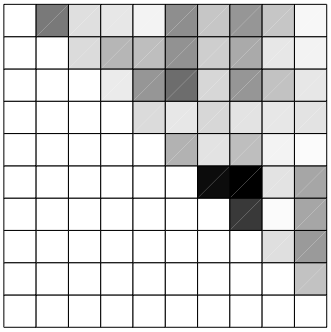
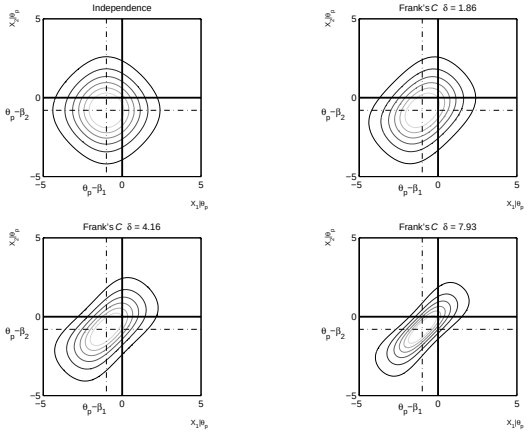
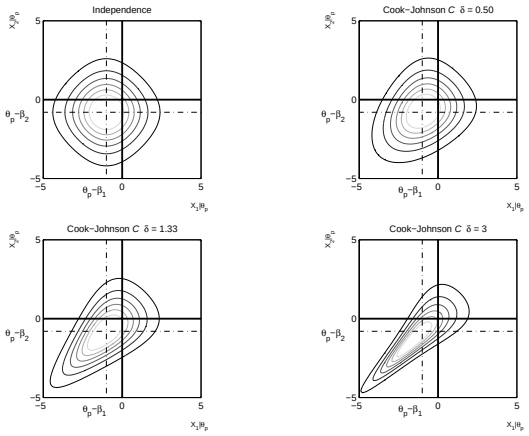


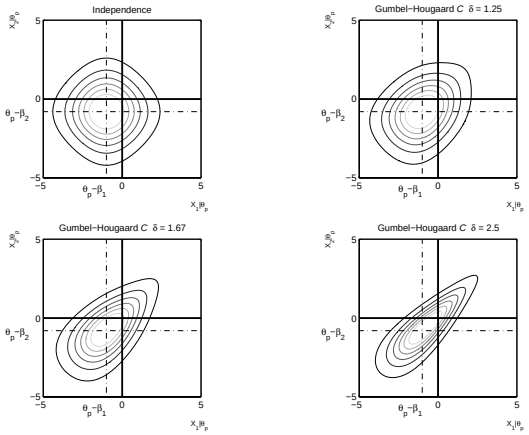
Figure 3: Bivariate logistic density contour plots for 3 Archimedean copulas.



(a) Frank copula

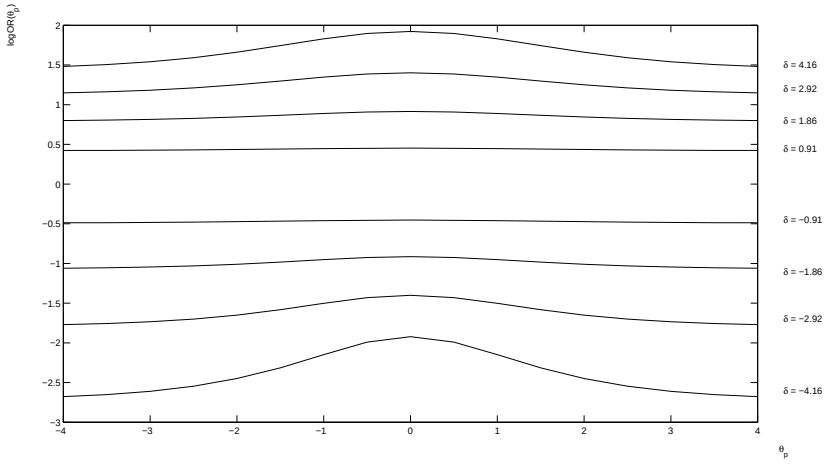


(b) Cook-Johnson copula

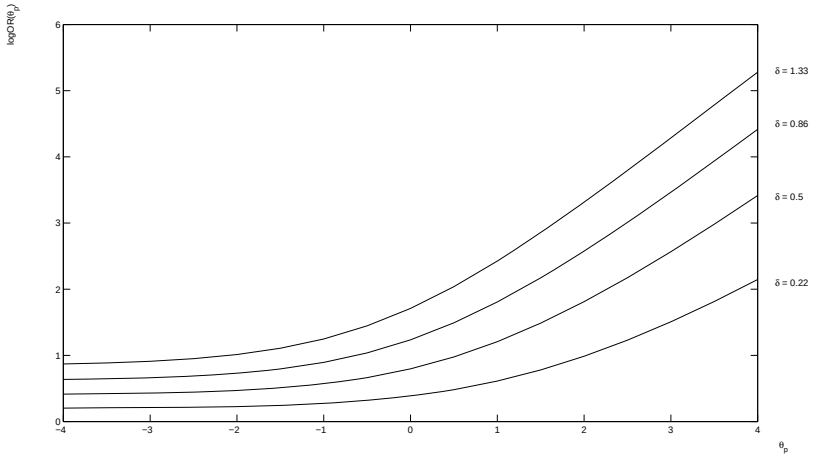


(c) Gumbel-Hougaard copula

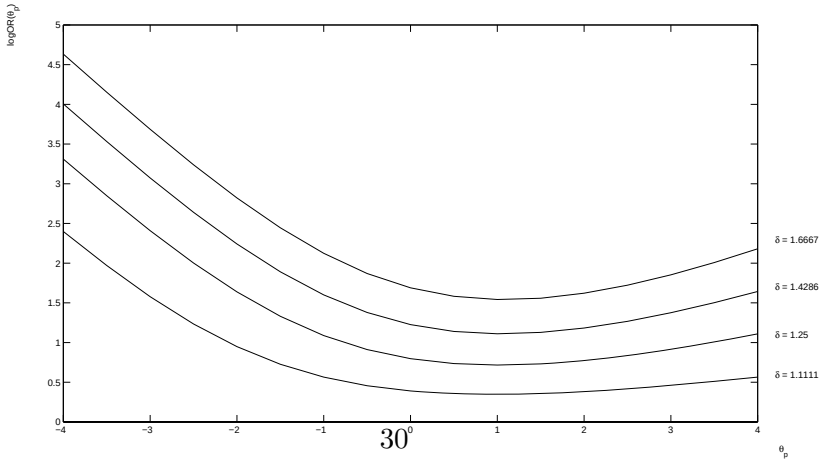
Figure 4: log odds ratio plots for 3 Archimedean copulas in the latent variable model.



(a) Frank copula



(b) Cook-Johnson copula



(c) Gumbel-Hougaard copula

Figure 5: Observed distribution of the severity of affect for infants in the BAT study.

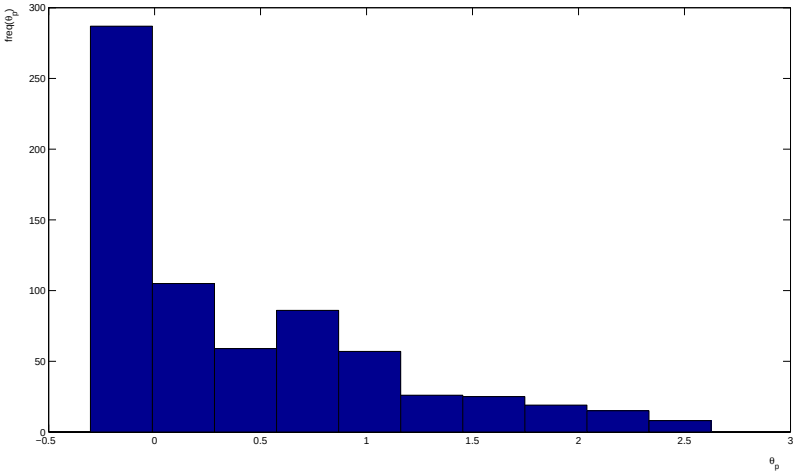


Figure 6: Standard erros of θ_p in the presence of residual dependencies under a conditional independence model and a copula model.