

# A D-vine copula mixed model for joint meta-analysis and comparison of diagnostic tests

Aristidis K. Nikoloulopoulos\*

## Abstract

For a particular disease there may be two diagnostic tests developed, where each of the tests is subject to several studies. A quadrivariate generalized linear mixed model (GLMM) has been recently proposed to joint meta-analyse and compare two diagnostic tests. We propose a D-vine copula mixed model for joint meta-analysis and comparison of two diagnostic tests. Our general model includes the quadrivariate GLMM as a special case and can also operate on the original scale of sensitivities and specificities. The method allows the direct calculation of sensitivity and specificity for each test, as well as, the parameters of the summary receiver operator characteristic (SROC) curve, along with a comparison between the SROCs of each test. Our methodology is demonstrated with an extensive simulation study and illustrated by meta-analysing two examples where 2 tests for the diagnosis of a particular disease are compared. Our study suggests that there can be an improvement on GLMM in fit to data since our model can also provide tail dependencies and asymmetries.

*Keywords:* Copula mixed model; generalized linear mixed model; sensitivity/specificity; SROC, vines.

## 1 Introduction

So far, work on multivariate methods and models for meta-analysis of diagnostic studies has mainly focused on a single test.<sup>1-3</sup> However, for a particular disease there may be two (or more) diagnostic tests developed, where each of the tests is subject to several studies; e.g., Takwoingi et al.<sup>4</sup> found a considerable large number of systematic reviews which compared the accuracy of two (or more) tests. Hence, one may want to combine all such studies to see how the competing tests are performing with respect to each other, and choose the best for clinical practice.<sup>5-7</sup>

To compare the accuracy of two (or more) tests to a common gold standard, reviewers typically take one of two approaches: perform a separate bivariate meta-analysis, e.g., fit a generalized linear mixed model<sup>8</sup>

---

\*A.Nikoloulopoulos@uea.ac.uk, School of Computing Sciences, University of East Anglia, Norwich NR4 7TJ, UK

(GLMM) to synthesize information for each test and then compare the meta-analytic summaries, or perform a bivariate GLMM meta-regression using the type of test as a categorical predictor. Both methods assume that all individuals studies are independently sampled (i.e., that the tests study two distinct sets of individuals, each containing individuals with disease and individuals without disease). But when the same individuals receive both tests, their results are correlated and one has to take into account any possible correlations between the 2 tests. Because the two common approaches do not take these correlations into account, neither is a valid method for comparing tests performed on the same individuals.<sup>9</sup>

A valid statistical model that accounts for this dependence, namely a quadrivariate GLMM, has been recently proposed for the comparison of two diagnostic tests.<sup>10</sup> In this paper, we propose a vine copula mixed model as an extension of the quadrivariate GLMM by rather using a vine copula representation of the random effects distribution with normal and beta margins. Our general model (a) includes the quadrivariate GLMM as a special case, (b) can also operate on the original scale of sensitivities and specificities, and (c) can also provide tail dependencies and asymmetries since the random effects distribution is expressed via vine copulas that allow for flexible dependence modelling, different from assuming simple linear correlation structures, normality and tail independence as in GLMM. In fact, we extend the vine copula mixed model proposed by Nikoloulopoulos<sup>11</sup> to the quadrivariate case. To do so, we employ a drawable vine (D-vine) which can nicely capture the dependence within and between the diagnostic tests.

A vine copula approach for meta-analysis for the comparison of two diagnostic tests was recently proposed by Hoyer and Kuss<sup>12</sup> who explored the use of a quadrivariate vine copula model for observed discrete variables (number of true positives and true negatives for both tests) which have beta-binomial margins. This approach is actually an approximated likelihood method for estimating a D-vine copula mixed model with beta margins for the latent vector of sensitivities and specificities.

The remainder of the paper proceeds as follows. Section 2 has a brief overview of relevant copula theory and then introduces the D-vine copula mixed model for the comparison of two diagnostic tests and discusses its relationship with existing models. Section 3 contains small-sample efficiency calculations to investigate the effect of misspecifying the random effects distribution on parameter estimators and standard errors and compare the proposed methodology with existing methods. Section 4 illustrates our methodology with two examples where 2 tests for the diagnosis of a particular disease are compared. We conclude with some discussion in Section 5.

## 2 The vine copula mixed model for diagnostic test accuracy studies

In this section, we introduce the D-vine copula mixed model for the comparison of two diagnostic tests and discuss its relationship with existing models. Before that, the first subsection has some background on copula models. We complete this section with details on maximum likelihood estimation.

### 2.1 Overview and relevant background for vine copulas

A copula is a multivariate cdf with uniform  $U(0, 1)$  margins.<sup>13–15</sup> If  $F$  is a  $d$ -variate cdf with univariate margins  $F_1, \dots, F_d$ , then Sklar's<sup>16</sup> theorem implies that there is a copula  $C$  such that

$$F(x_1, \dots, x_d) = C\left(F_1(x_1), \dots, F_d(x_d)\right).$$

The copula is unique if  $F_1, \dots, F_d$  are continuous. If  $F$  is continuous and  $(Y_1, \dots, Y_d) \sim F$ , then the unique copula is the distribution of  $(U_1, \dots, U_d) = (F_1(Y_1), \dots, F_d(Y_d))$  leading to

$$C(u_1, \dots, u_d) = F\left(F_1^{-1}(u_1), \dots, F_d^{-1}(u_d)\right), \quad 0 \leq u_j \leq 1, j = 1, \dots, d,$$

where  $F_j^{-1}$  are inverse cdfs<sup>17</sup>. In particular, if  $\Phi_d(\cdot; \mathbf{R})$  is the multivariate normal (MVN) cdf with correlation matrix

$$\mathbf{R} = (\rho_{jk} : 1 \leq j < k \leq d)$$

and  $N(0,1)$  margins, and  $\Phi$  is the univariate standard normal cdf, then the MVN copula is

$$C(u_1, \dots, u_d) = \Phi_d\left(\Phi^{-1}(u_1), \dots, \Phi^{-1}(u_d); \mathbf{R}\right). \quad (1)$$

A copula  $C$  has reflection symmetry if  $(U_1, \dots, U_d) \sim C$  implies that  $(1 - U_1, \dots, 1 - U_d)$  has the same distribution  $C$ . This is apparently the case for the MVN copula. When it is necessary to have copula models with reflection asymmetry and flexible lower/upper tail dependence, then vine copulas are the best choice.<sup>18</sup> The  $d$ -dimensional vine copulas are built via successive mixing from  $d(d - 1)/2$  bivariate linking copulas on trees and their cdfs involve lower-dimensional integrals. Since the densities of multivariate vine copulas can be factorized in terms of bivariate linking copulas and lower-dimensional margins, they are computationally tractable.

For the  $d$ -dimensional D-vine, the pairs at level 1 are  $j, j + 1$ , for  $j = 1, \dots, d - 1$ , and for level  $\ell$  ( $2 \leq \ell < d$ ), the (conditional) pairs are  $j, j + \ell | j + 1, \dots, j + \ell - 1$  for  $j = 1, \dots, d - \ell$ . That is, for the D-vine, conditional copulas are specified for variables  $j$  and  $j + \ell$  given the variables indexed in between<sup>19</sup>. In Figure 1 a D-vine with 4 variables and 3 trees/levels is depicted.

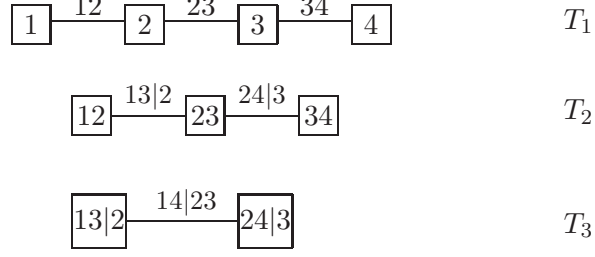


Figure 1: 4-dimensional D-vine with 3 trees/levels ( $T_j$ ,  $j = 1, \dots, 4$ ).

The quadrivariate D-vine density is decomposed in a simple manner by multiplying the nodes of the nested set of trees

$$\begin{aligned}
f_{1234}(x_1, x_2, x_3, x_4) &= f_1(x_1) f_2(x_2) f_3(x_3) f_4(x_4) c_{12}(F_1(x_1), F_2(x_2)) c_{23}(F_2(x_2), F_3(x_3)) \times \\
&\quad c_{34}(F_3(x_3), F_4(x_4)) c_{13|2}(F_{1|2}(x_1|x_2), F_{3|2}(x_3|x_2)) c_{24|3}(F_{2|3}(x_2|x_3), \\
&\quad F_{4|3}(x_4|x_3)) c_{14|23}(F_{1|23}(x_1|x_2, x_3), F_{4|23}(x_4|x_2, x_3)) \\
&= f_1(x_1) f_2(x_2) f_3(x_3) f_4(x_4) c_{1234}(F_1(x_1), F_2(x_2), F_3(x_3), F_4(x_4)), \quad (2)
\end{aligned}$$

where  $F_{j|k}(x_j|x_k) = \partial C_{jk}(F_j(x_j), F_k(x_k)) / \partial F_k(x_k)$ <sup>20</sup>.

## 2.2 The D-vine copula mixed model for the latent pair of transformed sensitivity and specificity

We first introduce the notation used in this paper. The focus is on two-level (within-study and between-studies) cluster data. The data are  $(y_{ij}, n_{ij})$ ,  $i = 1, \dots, N$ ,  $j = 1, 2, 3, 4$ , where  $j$  is an index for the within study measurements and  $i$  is an index for the individual studies. The data, for study  $i$ , can be summarized in two  $2 \times 2$  tables with the number of true positives ( $y_{i1}$ ), true negatives ( $y_{i2}$ ), false negatives ( $n_{i1} - y_{i1}$ ), and false positives ( $n_{i2} - y_{i2}$ ) for the first diagnostic test, and the number of true positives ( $y_{i3}$ ), true negatives ( $y_{i4}$ ), false negatives ( $n_{i3} - y_{i3}$ ), and false positives ( $n_{i4} - y_{i4}$ ) for the second diagnostic test. We assume that the gold standard is the same for both tests, i.e.,  $n_{i1} = n_{i3}$  and  $n_{i2} = n_{i4}$ .

Here we generalize the quadrivariate GLMM<sup>10</sup> by proposing a model that links the four random effects using a D-vine copula rather than the quadrivariate normal (QVN) distribution. The within-study model assumes that the number of true positives  $Y_{i1}$  and true negatives  $Y_{i2}$  for the first test and the number of true positives  $Y_{i3}$  and true negatives  $Y_{i4}$  for the second test are conditionally independent and binomially distributed given  $(\mathbf{X}_1, \mathbf{X}_2) = (\mathbf{x}_1, \mathbf{x}_2)$ , where  $\mathbf{X}_1 = (X_1, X_2)$  denotes the bivariate latent pair of (transformed) sensitivity and specificity for the first test and  $\mathbf{X}_2 = (X_3, X_4)$  denotes the bivariate latent pair of (transformed) sensitivity and

specificity for the second test. That is

$$\begin{aligned}
Y_{i1}|X_1 = x_1 &\sim \text{Binomial}(n_{i1}, l^{-1}(x_1)); \\
Y_{i2}|X_2 = x_2 &\sim \text{Binomial}(n_{i2}, l^{-1}(x_2)); \\
Y_{i3}|X_3 = x_3 &\sim \text{Binomial}(n_{i1}, l^{-1}(x_3)); \\
Y_{i4}|X_4 = x_4 &\sim \text{Binomial}(n_{i2}, l^{-1}(x_4)),
\end{aligned} \tag{3}$$

where  $l(\cdot)$  is a link function.

The stochastic representation of the between studies model takes the form

$$\left( F(X_1; l(\pi_1), \delta_1), F(X_2; l(\pi_2), \delta_2), F(X_3; l(\pi_3), \delta_3), F(X_4; l(\pi_4), \delta_4) \right) \sim C(\cdot; \boldsymbol{\theta}), \tag{4}$$

where  $C(\cdot; \boldsymbol{\theta})$  is a D-vine copula with dependence parameter vector  $\boldsymbol{\theta} = (\theta_{12}, \theta_{23}, \theta_{34}, \theta_{13|2}, \theta_{24|3}, \theta_{14|23})$  and  $F(\cdot; l(\pi), \delta)$  is the cdf of the univariate distribution of the random effect.

The models in (3) and (4) together specify a D-vine copula mixed model with joint likelihood

$$\begin{aligned}
L(\pi_1, \pi_2, \pi_3, \pi_4, \delta_1, \delta_2, \delta_3, \delta_4, \boldsymbol{\theta}) = \\
\prod_{i=1}^N \int_0^1 \int_0^1 \int_0^1 \int_0^1 \prod_{j=1}^4 g(y_{ij}; n_{ij}, l^{-1}(F^{-1}(u_j; l(\pi_j), \delta_j))) c_{1234}(u_1, u_2, u_3, u_4; \boldsymbol{\theta}) du_1 du_2 du_3 du_4, \tag{5}
\end{aligned}$$

where  $g(y; n, \pi) = \binom{n}{y} \pi^y (1 - \pi)^{n-y}$ ,  $y = 0, 1, \dots, n$ ,  $0 < \pi < 1$ , is the binomial probability mass function (pmf).

The copula parameter vector  $\boldsymbol{\theta}$  has parameters of the random effects model and they are separated from the univariate parameters  $(\pi_j, \delta_j)$ ,  $j = 1, \dots, 4$ . The parameters  $\pi_1$  and  $\pi_2$  are the meta-analytic parameters for the sensitivity and specificity,  $\delta_1$  and  $\delta_2$  express the between-study variabilities for the first test, and  $\pi_3$  and  $\pi_4$  are the meta-analytic parameters for the sensitivity and specificity,  $\delta_3$  and  $\delta_4$  express the between-study variabilities for the second test. The choices of the  $F(\cdot; l(\pi), \delta)$  and  $l$  are given in Table 1. If the Beta( $\pi, \gamma$ ) distribution is used for the marginal modelling of the latent proportions, then one does not have to transform the latent sensitivities and specificities and can work on the original scale. Our general statistical model allows for selection of copulas and margins independently, i.e., there are no constraints in the choices of parametric copulas and margins.

Table 1: The choices of the  $F(\cdot; l(\pi), \delta)$  and  $l$  in the copula mixed model.

$F(\cdot; l(\pi), \delta)$	$l$	$\pi$	$\delta$
$N(\mu, \sigma)$	logit, probit, cloglog	$l^{-1}(\mu)$	$\sigma$
$\text{Beta}(\pi, \gamma)$	identity	$\pi$	$\gamma$

## 2.3 Relationship with existing models

### 2.3.1 Relationship with the quadrivariate GLMM

In this subsection, we show what happens when all the bivariate copulas are bivariate normal (BVN) and the univariate distribution of the random effects is the  $N(\mu, \sigma)$  distribution. One can easily deduce that the within-study model in (3) is the same as in the quadrivariate GLMM.

Furthermore, when all the bivariate copulas are BVN copulas with correlation parameters  $\rho_{12}, \rho_{23}, \rho_{34}$  (1st tree/level) and partial correlation parameters  $\rho_{13|2}, \rho_{24|3}, \rho_{14|23}$  (2nd and 3rd tree/level), the resulting distribution is the QVN with mean vector  $\boldsymbol{\mu} = (l(\pi_1), l(\pi_2), l(\pi_3), l(\pi_4))^\top$  and variance covariance matrix  $\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_1^2 & \rho_{12}\sigma_1\sigma_2 & \rho_{13}\sigma_1\sigma_3 & \rho_{14}\sigma_1\sigma_4 \\ \rho_{12}\sigma_1\sigma_2 & \sigma_2^2 & \rho_{23}\sigma_2\sigma_3 & \rho_{24}\sigma_2\sigma_4 \\ \rho_{13}\sigma_1\sigma_3 & \rho_{23}\sigma_2\sigma_3 & \sigma_3^2 & \rho_{34}\sigma_3\sigma_4 \\ \rho_{14}\sigma_1\sigma_4 & \rho_{24}\sigma_2\sigma_4 & \rho_{34}\sigma_3\sigma_4 & \sigma_4^2 \end{pmatrix}$ , where  $\rho_{13} = \rho_{13|2}\sqrt{1-\rho_{12}^2}\sqrt{1-\rho_{23}^2} + \rho_{12}\rho_{23}$ ,  $\rho_{24} = \rho_{24|3}\sqrt{1-\rho_{23}^2}\sqrt{1-\rho_{34}^2} + \rho_{23}\rho_{34}$ ,  $\rho_{14} = \rho_{14|23}\sqrt{1-\rho_{12}^2}\sqrt{1-\rho_{24}^2} + \rho_{12}\rho_{24}$ ,  $\rho_{14|2} = \rho_{14|23}\sqrt{1-\rho_{13|2}^2}\sqrt{1-\rho_{34|2}^2} + \rho_{13|2}\rho_{34|2}$ ,  $\rho_{13|2} = (\rho_{13} - \rho_{12}\rho_{23})/\sqrt{1-\rho_{12}^2}/\sqrt{1-\rho_{23}^2}$  and  $\rho_{34|2} = (\rho_{34} - \rho_{23}\rho_{24})/\sqrt{1-\rho_{23}^2}/\sqrt{1-\rho_{24}^2}$ . Therefore, the between-studies model in (4) assumes that  $(\mathbf{X}_1, \mathbf{X}_2)$  is QVN distributed, i.e.,  $\mathbf{X} \sim \text{QVN}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ .

With some algebra it can be shown that the joint likelihood in (5) becomes

$$L(\pi_1, \pi_2, \pi_3, \pi_4, \sigma_1, \sigma_2, \sigma_3, \sigma_4, \rho_{12}, \rho_{13}, \rho_{14}, \rho_{23}, \rho_{24}, \rho_{34}) = \prod_{i=1}^N \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \prod_{j=1}^4 g(y_{ij}; n_{ij}, l^{-1}(x_j)) \phi_{1234}(x_1, x_2, x_3, x_4; \boldsymbol{\mu}, \boldsymbol{\Sigma}) dx_1 dx_2 dx_3 dx_4.$$

where  $\phi_{1234}(\cdot; \boldsymbol{\mu}, \boldsymbol{\Sigma})$  is the QVN density with mean vector  $\boldsymbol{\mu}$  and variance covariance matrix  $\boldsymbol{\Sigma}$ . Hence, this model is the same as the quadrivariate GLMM<sup>10</sup>.

### 2.3.2 Relationship with a vine copula model with beta-binomial margins

Hoyer and Kuss<sup>12</sup> proposed a vine copula model with beta-binomial margins in this context. This model is actually an approximated likelihood method for the estimation of the D-vine copula mixed model with beta margins for the latent pair of sensitivity and specificity in (3) and (4). They attempt to approximate the likelihood in (5) with the likelihood of a copula model for observed discrete variables which have beta-binomial margins.

The approximation that they suggest is

$$L(\pi_1, \pi_2, \pi_3, \pi_4, \gamma_1, \gamma_2, \gamma_3, \gamma_4, \boldsymbol{\theta}) \approx \prod_{i=1}^N c_{1234} \left( H(y_{i1}; n_{i1}, \pi_1, \gamma_1), H(y_{i2}; n_{i2}, \pi_2, \gamma_2), H(y_{i3}; n_{i3}, \pi_3, \gamma_3), H(y_{i4}; n_{i4}, \pi_4, \gamma_4); \boldsymbol{\theta} \right) \prod_{j=1}^4 h(y_{ij}; n_{ij}, \pi_j, \gamma_j),$$
 where  $H(\cdot; n, \pi, \gamma)$  and  $h(\cdot; n, \pi, \gamma)$  is the cdf and the density, respectively, of the the Beta-Binomial( $n, \pi, \gamma$ ) distribution. In their approximation the authors also treat the observed variables which have beta-binomial distributions as being continuous, and model them under the theory for copula models with continuous margins.

Nikoloulopoulos<sup>21</sup> has extensively studied the small-sample and theoretical efficiency of this approximation in the bivariate case<sup>22</sup>. It was clearly shown that this approximation leads to substantial downward bias for the estimates of the dependence and bias for the meta-analytic parameters for fully specified copula mixed models. This evolves because there are serious problems on modelling assumptions under the case of heterogeneous study sizes. If the number of true positives and true negatives do not have a common support over different studies, then one cannot conclude that there is a copula. The copula is not common if the mixing distribution for binomials is common. Nevertheless, it is a stronger assumption to assume a common copula for beta-binomial random variables with different parameters (the study size  $n_{ij}$ ), compared with a common copula for the random effects. It is more natural to join the random effects with a copula.<sup>23</sup>

## 2.4 Maximum likelihood estimation and computational details

Estimation of the model parameters  $(\pi_1, \pi_2, \pi_3, \pi_4, \delta_1, \delta_2, \delta_3, \delta_4, \boldsymbol{\theta})$  can be approached by the standard maximum likelihood (ML) method, by maximizing the logarithm of the joint likelihood in (5). The estimated parameters can be obtained by using a quasi-Newton<sup>24</sup> method applied to the logarithm of the joint likelihood. This numerical method requires only the objective function, i.e., the logarithm of the joint likelihood, while the gradients are computed numerically and the Hessian matrix of the second order derivatives is updated in each iteration. The standard errors (SE) of the ML estimates can be also obtained via the gradients and the Hessian computed numerically during the maximization process.

For D-vine copula mixed models of the form with joint likelihood as in (5), numerical evaluation of the joint pmf can be achieved with the following steps:

1. Calculate Gauss-Legendre<sup>25</sup> quadrature points  $\{u_q : q = 1, \dots, n_q\}$  and weights  $\{w_q : q = 1, \dots, n_q\}$  in terms of standard uniform.
2. Convert from independent uniform random variables  $\{u_{q1} : q_1 = 1, \dots, n_q\}$ ,  $\{u_{q2} : q_2 = 1, \dots, n_q\}$ ,

$\{u_{q_3} : q_3 = 1, \dots, n_q\}$ , and  $\{u_{q_4} : q_4 = 1, \dots, n_q\}$  to dependent uniform random variables that have a D-vine distribution  $C(\cdot; \theta)$ :

- 1: Set  $v_{q_1} = u_{q_1}$
- 2:  $v_{q_2|q_1} = C_{2|1}^{-1}(u_{q_2}|u_{q_1}; \theta_{12})$
- 3:  $t_1 = C_{1|2}(v_{q_1}|v_{q_2|q_1}; \theta_{12})$
- 4:  $t_2 = C_{3|1;2}^{-1}(u_{q_3}|t_1; \theta_{12}), \theta_{13|2})$
- 5:  $v_{q_3|q_1; q_2} = C_{3|2}^{-1}(t_2|v_{q_2|q_1}; \theta_{23})$
- 6:  $t_3 = C_{2|3}(v_{q_2|q_1}|v_{q_3|q_1; q_2}; \theta_{23})$
- 7:  $t_4 = C_{1|3;2}(t_1|t_2; \theta_{13|2})$
- 8:  $t_5 = C_{4|1;2,3}(u_{q_4}|t_4; \theta_{14|23})$
- 9:  $t_6 = C_{4|2;3}^{-1}(t_5|t_3; \theta_{24|3})$
- 10:  $v_{q_4|q_1; q_2, q_3} = C_{4|3}^{-1}(t_6|v_{q_3|q_1; q_2}; \theta_{34})$

The simulation algorithm of a D-vine copula in Joe<sup>14</sup> is used to achieve this.

### 3. Numerically evaluate the joint pmf

$$\int_0^1 \int_0^1 \int_0^1 \int_0^1 \prod_{j=1}^4 g(y_{ij}; n_{ij}, l^{-1}(F^{-1}(u_j; l(\pi_j), \gamma_j))) c_{1234}(u_1, u_2, u_3, u_4; \theta) du_1 du_2 du_3 du_4$$

in a quadruple sum:

$$\sum_{q_1=1}^{n_q} \sum_{q_2=1}^{n_q} \sum_{q_3=1}^{n_q} \sum_{q_4=1}^{n_q} w_{q_1} w_{q_2} w_{q_3} w_{q_4} g\left(y_1; n, l^{-1}(F^{-1}(v_{q_1}; l(\pi_1), \gamma_1))\right) g\left(y_2; n, l^{-1}(F^{-1}(v_{q_2|q_1}; l(\pi_2), \gamma_2))\right) g\left(y_3; n, l^{-1}(F^{-1}(v_{q_2q_3|q_1}; l(\pi_3), \gamma_3))\right) g\left(y_4; n, l^{-1}(F^{-1}(v_{q_4|q_1; q_2, q_3}; l(\pi_4), \gamma_4))\right).$$

With Gauss-Legendre quadrature, the same nodes and weights are used for different functions; this helps in yielding smooth numerical derivatives for numerical optimization via quasi-Newton.<sup>24</sup> The conditional copula cdfs  $C(v|u; \theta)$  and their inverses  $C^{-1}(v|u; \theta)$  are given in Table 2 for the sufficient list of parametric families of copulas for meta-analysis of diagnostic test accuracy studies.<sup>11;21;26;27</sup> Since the copula parameter  $\theta$  of each family has different range, in the sequel we reparametrize them via their Kendalls  $\tau$ ; that is comparable across families. The functional relationships between  $\theta$  and  $\tau$  for each family are also given in Table 2.



Table 2: Parametric families of bivariate copulas and their Kendall's  $\tau$  as a strictly increasing function of the copula parameter  $\theta$ .

Copula	$C(v u; \theta)$	$C^{-1}(v u; \theta)$	$\tau$
BVN	$\Phi\left(\frac{\Phi^{-1}(v) - \theta\Phi^{-1}(u)}{\sqrt{1-\theta^2}}\right)$	$\Phi\left(\sqrt{1-\theta^2}\Phi^{-1}(v) + \theta\Phi^{-1}(u)\right)$	$\frac{2}{\pi} \arcsin(\theta)$ , $-1 \leq \theta \leq 1$
Frank	$e^{-\theta u} \left[ \frac{1-e^{-\theta}}{1-e^{-\theta v}} - (1 - e^{-\theta u}) \right]^{-1}$	$-\frac{1}{\theta} \log \left[ 1 - \frac{1-e^{-\theta}}{(v^{-1}-1)e^{-\theta u}+1} \right]$	$1 - 4\theta^{-1} - 4\theta^{-2} \int_{\theta}^0 \frac{t}{e^t-1} dt$ , $\theta < 0$ $1 - 4\theta^{-1} + 4\theta^{-2} \int_0^{\theta} \frac{t}{e^t-1} dt$ , $\theta > 0$
Clayton	$\left[ 1 + u^{\theta}(v^{-\theta} - 1) \right]^{-1-1/\theta}$	$\left[ (v^{-\theta/(1+\theta)} - 1)u^{-\theta} + 1 \right]^{-1/\theta}$	$\theta/(\theta + 2)$ , $\theta > 0$
Cln90°	$\left[ 1 + (1 - u)^{\theta}(v^{-\theta} - 1) \right]^{-1-1/\theta}$	$\left[ (v^{-\theta/(1+\theta)} - 1)(1 - u)^{-\theta} + 1 \right]^{-1/\theta}$	$-\theta/(\theta + 2)$ , $\theta > 0$
Cln180°	$1 - \left[ 1 + (1 - u)^{\theta}((1 - v)^{-\theta} - 1) \right]^{-1-1/\theta}$	$1 - \left[ \{(1 - v)^{-\theta/(1+\theta)} - 1\}(1 - u)^{-\theta} + 1 \right]^{-1/\theta}$	$\theta/(\theta + 2)$ , $\theta > 0$
Cln270°	$1 - \left[ 1 + u^{\theta}\{(1 - v)^{-\theta} - 1\} \right]^{-1-1/\theta}$	$1 - \left[ \{(1 - v)^{-\theta/(1+\theta)} - 1\}u^{-\theta} + 1 \right]^{-1/\theta}$	$-\theta/(\theta + 2)$ , $\theta > 0$

Cln $\omega$ °: Clayton rotated by  $\omega$  degrees.

### 3 Small-sample efficiency – Misspecification of the random effects distribution

An extensive simulation study is conducted (a) to gauge the small-sample efficiency of the ML and approximated likelihood in Hoyer and Kuss<sup>12</sup> (hereafter HK) methods, and (b) to investigate in detail the misspecification of the parametric margin or family of copulas of the random effects distribution.

We set the sample size  $N$ , the study size  $n$ , the true univariate and Kendall's  $\tau$  parameters, and the disease prevalence  $\pi$  to mimic both the arthritis<sup>28</sup> and the diabetes<sup>29;30</sup> data sets. These example meta-analyses have different disease prevalences as depicted in Figure 2. To quantify this we fit the model ignoring the dependence among the random effects, that is we assume independence. Figure 3 depicts the size of the estimated beta-binomial discrete probabilities under the independence assumption for both datasets. As revealed the diabetes data poses very small individual-study discrete probabilities.

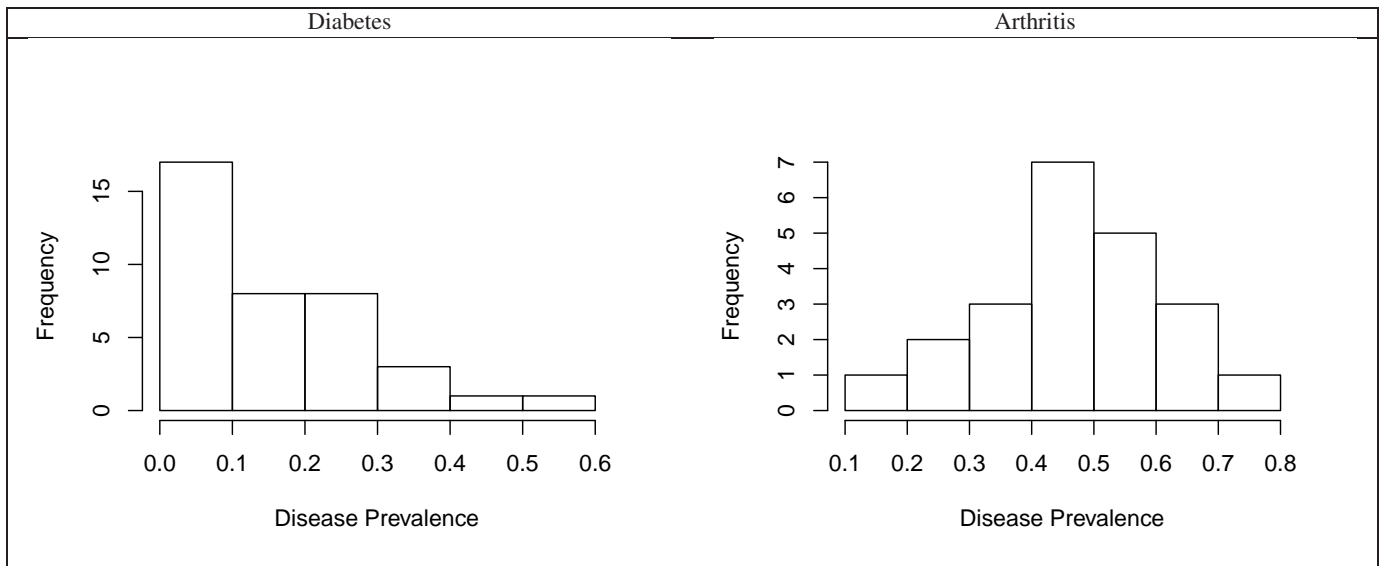


Figure 2: Disease prevalences for the arthritis and diabetes data.

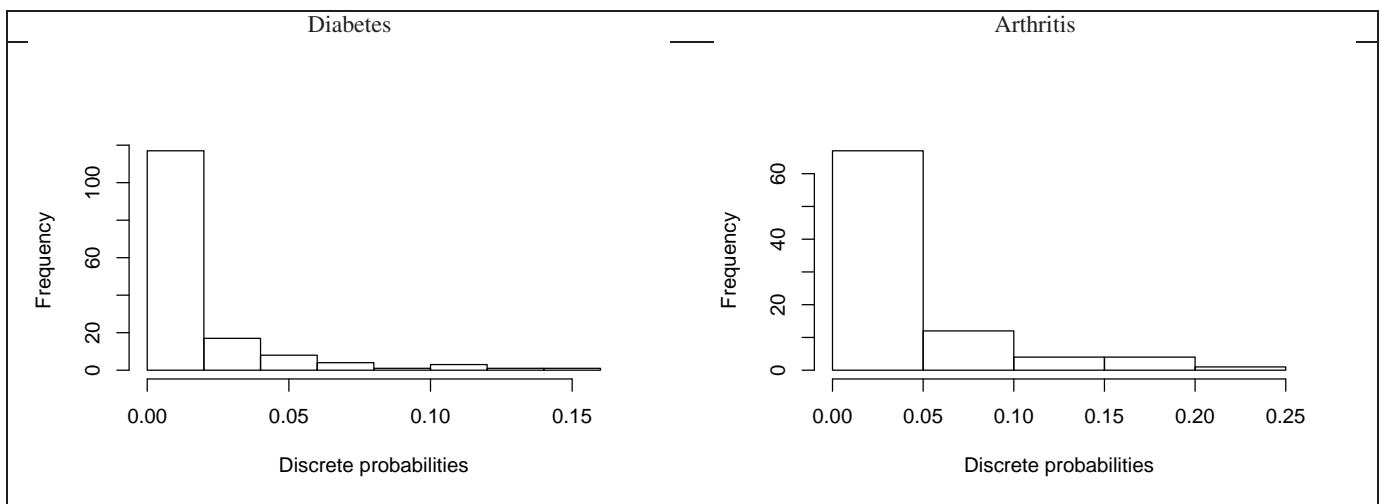


Figure 3: The size of the discrete probabilities assuming independence for the arthritis and diabetes data.

More specifically, we randomly generate samples of size  $N = 22, 38$  from the D-vine copula mixed model with BVN linking copulas and both normal and beta margins. The simulation process is as below:

1. Simulate  $(u_1, u_2, u_3, u_4)$  from a D-vine distribution  $C(; \tau)$ ;  $\tau$  is converted to the dependence parameters  $\theta$  via the relations in Table 2.
2. Convert to proportions via  $x_j = l^{-1}\left(F_j^{-1}(u_j, l(\pi_j), \delta_j)\right)$ .
3. Draw the number of diseased  $n_1$  from a  $B(n, \pi)$  distribution.
4. Set  $n_2 = n - n_1$  and generate  $y_j$  from a  $B(n_j, x_j)$  for  $j = 1, \dots, 4$ .

Tables 3 and 4 contain the resultant biases, root mean square errors (RMSEs) and standard deviations (SDs), along with average theoretical variances, for the MLEs under different copula choices and margins. The theoretical variances of the maximum likelihood estimates (MLEs) are obtained via the gradients and the Hessian that were computed numerically during the maximization process. We also provide biases, RMSEs and SDs for the HK estimates under the true model, that is, the BVN copula mixed model with normal (Table 3) and beta (Table 4) margins.

Conclusions from the values in the 2 tables and other computations that we have done are the following:

- ML with the true copula mixed model is highly efficient according to the simulated biases and standard deviations.
- The MLEs of the meta-analytic parameters are slightly underestimated under copula misspecification.
- The SDs are rather robust to the copula misspecification.
- The meta-analytic MLEs and SDs are not robust to the margin misspecification, while the MLE of  $\tau$  and its SD is.
- The HK method is not efficient for the univariate marginal parameters when there is more discretization (larger individual-study discrete probabilities).
- The efficiency of the HK method is low for the Kendall's  $\tau$  association parameters. The parameters are substantially underestimated when there is more discretization (larger individual probabilities).
- The HK method yields estimates that are almost as good as the MLEs for the univariate parameters when there is less discretization (small individual probabilities) and the true distribution used for the marginal modelling of the latent proportions is the Beta( $\pi, \gamma$ ).
- When there is less discretization the number of quadrature points should be at least  $n_q = 30$  rather than  $n_q = 15$  which is a sufficient number of quadrature points when there are large individual probabilities.

Table 3: Small sample of sizes  $N = 22$  simulations ( $10^3$  replications;  $n_q = 15$ ) from the D-vine copula mixed model with BVN copulas and normal margins and resultant biases, root mean square errors (RMSE) and standard deviations (SD), along with the square root of the average theoretical variances ( $\sqrt{V}$ ), for the MLEs under different copula choices and margins. We also provide biases, RMSEs and SDs for the HK estimates under the true model. We set the disease prevalence  $\pi$  to mimic the arthritis data.

	margin	copula	$\pi_1 =$ 0.7	$\pi_2 =$ 0.8	$\pi_3 =$ 0.7	$\pi_4 =$ 0.95	$\sigma_1 =$ 0.7	$\sigma_2 =$ 1	$\sigma_3 =$ 0.7	$\sigma_4 =$ 0.8	$\tau_{12} =$ -0.3	$\tau_{23} =$ 0.1	$\tau_{34} =$ -0.4	$\tau_{13 2} =$ 0.5	$\tau_{24 3} =$ 0.4	$\tau_{14 23} =$ 0.2
Bias	† normal	BVN	-0.003	-0.004	-0.003	-0.001	-0.017	-0.013	-0.023	-0.033	-0.001	0.009	-0.008	0.038	0.058	0.116
	beta		-0.018	-0.042	-0.018	-0.015					-0.001	0.003	-0.003	0.033	0.040	0.095
	normal	Frank	-0.001	-0.005	-0.002	-0.001	-0.016	0.001	-0.014	-0.031	0.002	0.023	-0.013	0.042	0.064	0.123
	beta		-0.016	-0.041	-0.017	-0.015					0.001	0.017	-0.015	0.039	0.046	0.092
	normal	Cln90°	-0.004	-0.001	-0.002	0.000	-0.012	0.011	-0.020	-0.022	0.050	0.066	0.071	0.006	0.066	0.083
	beta		-0.019	-0.041	-0.017	-0.015					0.053	0.076	0.076	0.006	0.067	0.082
	normal	Cln270°	-0.001	-0.007	-0.004	-0.001	-0.011	0.044	-0.005	-0.026	0.057	0.065	0.019	0.016	0.074	0.076
	beta		-0.018	-0.047	-0.020	-0.015					0.062	0.050	0.027	0.005	0.038	0.032
‡ beta	BVN	-0.010	-0.048	-0.012	-0.014					0.081	0.020	0.249	-0.124	-0.244	-0.178	
SD	† normal	BVN	0.036	0.037	0.035	0.010	0.145	0.197	0.134	0.173	0.166	0.168	0.173	0.162	0.206	0.447
	beta		0.034	0.038	0.033	0.012	0.030	0.042	0.028	0.017	0.163	0.166	0.168	0.161	0.204	0.424
	normal	Frank	0.036	0.038	0.036	0.010	0.146	0.202	0.139	0.177	0.173	0.170	0.189	0.162	0.213	0.465
	beta		0.035	0.038	0.034	0.013	0.030	0.043	0.029	0.017	0.168	0.167	0.185	0.164	0.208	0.446
	normal	Cln90°	0.036	0.039	0.036	0.010	0.148	0.219	0.147	0.182	0.167	0.174	0.204	0.180	0.219	0.468
	beta		0.034	0.038	0.034	0.013	0.029	0.047	0.030	0.019	0.162	0.164	0.204	0.178	0.217	0.442
	normal	Cln270°	0.036	0.040	0.035	0.010	0.146	0.228	0.154	0.182	0.166	0.184	0.234	0.177	0.222	0.479
	beta		0.035	0.039	0.034	0.013	0.032	0.048	0.034	0.017	0.175	0.192	0.228	0.179	0.216	0.452
‡ beta	BVN	0.033	0.038	0.032	0.012	0.024	0.037	0.024	0.015	0.132	0.128	0.114	0.128	0.121	0.109	
$\sqrt{V}$	† normal	BVN	0.027	0.030	0.031	0.009	0.101	0.148	0.117	0.166	0.114	0.138	0.155	0.138	0.237	0.571
	beta		0.025	0.028	0.028	0.011	0.021	0.030	0.025	0.014	0.120	0.138	0.153	0.131	0.159	0.678
	normal	Frank	0.027	0.029	0.030	0.009	0.070	0.149	0.113	0.165	0.112	0.136	0.162	0.135	0.165	0.729
	beta		0.025	0.027	0.028	0.011	0.021	0.030	0.026	0.015	0.132	0.140	0.161	0.127	0.194	0.452
	normal	Cln90°	0.027	0.030	0.032	0.009	0.099	0.155	0.120	0.167	0.094	0.144	0.159	0.137	0.174	0.625
	beta		0.025	0.027	0.029	0.011	0.020	0.031	0.025	0.015	0.097	0.136	0.157	0.136	0.174	0.584
	normal	Cln270°	0.027	0.028	0.031	0.009	0.101	0.142	0.123	0.167	0.099	0.141	0.183	0.138	0.205	0.479
	beta		0.026	0.024	0.029	0.010	0.021	0.027	0.027	0.014	0.104	0.143	0.188	0.137	0.067	0.319
‡ beta	BVN	0.030	0.034	0.030	0.011	0.024	0.033	0.023	0.013	0.113	0.117	0.080	0.104	0.082	0.091	
RMSE	† normal	BVN	0.036	0.037	0.035	0.010	0.146	0.197	0.136	0.176	0.166	0.168	0.173	0.166	0.214	0.461
	beta		0.038	0.056	0.038	0.019					0.163	0.166	0.168	0.164	0.208	0.434
	normal	Frank	0.036	0.039	0.036	0.010	0.147	0.202	0.140	0.180	0.173	0.172	0.189	0.168	0.223	0.481
	beta		0.038	0.056	0.038	0.020					0.168	0.168	0.185	0.168	0.213	0.455
	normal	Cln90°	0.036	0.039	0.036	0.010	0.149	0.220	0.149	0.183	0.174	0.187	0.216	0.180	0.228	0.475
	beta		0.039	0.056	0.038	0.020					0.171	0.181	0.217	0.178	0.227	0.450
	normal	Cln270°	0.036	0.041	0.036	0.010	0.147	0.232	0.154	0.184	0.176	0.196	0.235	0.178	0.234	0.485
	beta		0.039	0.060	0.039	0.020					0.186	0.199	0.230	0.179	0.220	0.453
‡ beta	BVN	0.034	0.061	0.034	0.018					0.154	0.129	0.274	0.178	0.273	0.208	

Cln $\omega^\circ$ : Clayton rotated by  $\omega$  degrees; †: The resulting model is the same as the GLMM; ‡: HK approximated likelihood method.

Table 4: Small sample of sizes  $N = 22$  simulations ( $10^3$  replications;  $n_q = 15$ ) from the D-vine copula mixed model with BVN copulas and beta margins and resultant biases, root mean square errors (RMSE) and standard deviations (SD), along with the square root of the average theoretical variances ( $\sqrt{V}$ ), for the MLEs under different copula choices and margins. We also provide biases, RMSEs and SDs for the HK estimates under the true model. We set the disease prevalence  $\pi$  to mimic the arthritis data.

	margin	copula	$\pi_1 =$ 0.7	$\pi_2 =$ 0.8	$\pi_3 =$ 0.7	$\pi_4 =$ 0.95	$\gamma_1 =$ 0.1	$\gamma_2 =$ 0.15	$\gamma_3 =$ 0.1	$\gamma_4 =$ 0.02	$\tau_{12} =$ -0.3	$\tau_{23} =$ 0.1	$\tau_{34} =$ $\tau_{34} = -0.4$	$\tau_{13 2} =$ 0.5	$\tau_{24 3} =$ 0.4	$\tau_{14 23} =$ 0.2
Bias	† normal	BVN	0.020	0.044	0.019	0.008					0.000	0.008	-0.006	0.042	0.070	0.135
	beta		0.000	-0.001	0.000	0.000	-0.004	-0.006	-0.004	-0.001	-0.003	0.002	-0.015	0.038	0.052	0.110
	normal	Frank	0.019	0.041	0.018	0.007					0.002	0.020	-0.016	0.043	0.082	0.137
	beta		0.000	-0.002	0.000	-0.001	-0.004	-0.003	-0.002	-0.001	-0.002	0.017	-0.025	0.041	0.063	0.109
	normal	Cln90°	0.019	0.047	0.019	0.008					0.044	0.043	0.057	0.000	0.069	0.061
	beta		-0.002	0.000	0.000	0.000	-0.004	0.000	-0.002	0.000	0.054	0.062	0.043	0.007	0.071	0.060
	normal	Cln270°	0.021	0.044	0.018	0.007					0.042	0.070	0.015	0.025	0.099	0.101
	‡ beta	BVN	0.000	-0.003	0.000	0.000	-0.001	0.006	0.000	0.000	0.049	0.050	0.019	0.016	0.071	0.064
SD	† normal	BVN	0.038	0.037	0.037	0.008	0.177	0.265	0.154	0.161	0.165	0.176	0.186	0.154	0.247	0.486
	beta		0.035	0.035	0.034	0.008	0.033	0.045	0.030	0.009	0.162	0.167	0.183	0.153	0.240	0.477
	normal	Frank	0.038	0.038	0.038	0.008	0.177	0.269	0.157	0.164	0.173	0.176	0.203	0.154	0.260	0.501
	beta		0.035	0.036	0.036	0.009	0.033	0.048	0.032	0.010	0.168	0.166	0.199	0.153	0.252	0.485
	normal	Cln90°	0.038	0.038	0.037	0.008	0.185	0.277	0.172	0.169	0.169	0.190	0.229	0.178	0.261	0.514
	beta		0.035	0.036	0.035	0.008	0.033	0.049	0.035	0.010	0.164	0.173	0.229	0.173	0.258	0.509
	normal	Cln270°	0.038	0.039	0.038	0.008	0.176	0.298	0.168	0.169	0.160	0.180	0.249	0.171	0.267	0.512
	‡ beta	BVN	0.035	0.037	0.035	0.008	0.035	0.052	0.036	0.010	0.166	0.180	0.242	0.172	0.259	0.505
$\sqrt{V}$	† normal	BVN	0.026	0.026	0.038	0.007	0.109	0.181	0.177	0.170	0.101	0.141	0.177	0.129	0.252	0.781
	beta		0.024	0.025	0.029	0.008	0.022	0.033	0.028	0.009	0.110	0.134	0.168	0.128	0.245	1.056
	normal	Frank	0.026	0.025	0.031	0.007	0.109	0.181	0.128	0.161	0.097	0.133	0.177	0.131	0.223	0.845
	beta		0.024	0.025	0.029	0.008	0.022	0.034	0.028	0.009	0.111	0.133	0.175	0.128	0.245	0.597
	normal	Cln90°	0.026	0.027	0.032	0.007	0.110	0.191	0.133	0.161	0.083	0.159	0.196	0.133	0.297	0.772
	beta		0.024	0.025	0.030	0.008	0.022	0.035	0.028	0.010	0.092	0.137	0.156	0.129	0.296	0.724
	normal	Cln270°	0.026	0.024	0.031	0.007	0.108	0.173	0.134	0.163	0.085	0.133	0.196	0.128	0.310	0.825
	‡ beta	BVN	0.024	0.023	0.030	0.008	0.023	0.032	0.030	0.010	0.097	0.142	0.185	0.133	0.227	0.826
RMSE	† normal	BVN	0.043	0.058	0.041	0.011					0.165	0.176	0.187	0.160	0.257	0.505
	beta		0.035	0.035	0.034	0.008	0.034	0.046	0.031	0.009	0.162	0.167	0.184	0.157	0.246	0.490
	normal	Frank	0.042	0.056	0.042	0.011					0.173	0.178	0.203	0.160	0.273	0.519
	beta		0.035	0.036	0.036	0.009	0.033	0.048	0.032	0.010	0.168	0.167	0.201	0.159	0.259	0.497
	normal	Cln90°	0.042	0.060	0.042	0.011					0.175	0.195	0.236	0.178	0.270	0.517
	beta		0.035	0.036	0.035	0.008	0.033	0.049	0.035	0.010	0.173	0.183	0.233	0.173	0.268	0.513
	normal	Cln270°	0.043	0.059	0.042	0.011					0.165	0.193	0.250	0.173	0.285	0.522
	‡ beta	BVN	0.035	0.037	0.035	0.008	0.035	0.053	0.036	0.010	0.173	0.187	0.243	0.173	0.268	0.509
			0.036	0.036	0.035	0.008	0.032	0.044	0.033	0.009	0.171	0.113	0.289	0.188	0.289	0.205

Cln $\omega^\circ$ : Clayton rotated by  $\omega$  degrees; †: The resulting model is the same as the GLMM; ‡: HK approximated likelihood method.

These results are in line with our previous studies<sup>11;21;26;27</sup> and also reveal that the size of the discrete probabilities is crucial for the efficiency of the HK method. For small prevalence, the probability of any given event will decrease, and as a result, one would be approximating ‘smaller steps’ in the cdf, hence the HK method at this case appears to be reliable as long as the true margin is beta.

## 4 Illustrations

We illustrate the use of the vine copula mixed model for the meta-analysis and for comparison of diagnostic accuracy studies by using the data of three published meta-analyses<sup>28–30</sup>. These data have been previously meta-analysed in the quadrivariate case.<sup>10;12;31</sup>

We fit the D-vine copula mixed model for all choices of parametric families of copulas for the bivariate margins of interest and margins. For ease of interpretation, we do not mix margins for a single test; hence, we allow two different margins, one for test 1 and one for test 2. For the bivariate copula  $C_{23}(\cdot)$  that links the data between the two tests and the copulas at the second and third tree we use the BVN copulas as on the one hand they provide a wide range of dependence to account for the conditional dependence and on the other hand allow for intermediate tail dependence<sup>32</sup>. We summarize the model in terms of largest likelihood along with the quadrivariate GLMM. We also perform a separate bivariate meta-analysis for each test fitting the bivariate copula mixed model<sup>21</sup> to synthesize information assuming independence between the tests. Once again we summarize the best fit in terms of the likelihood principle.

Finally, we demonstrate summary receiver operating characteristic (SROC) curves and summary operating points (a pair of average sensitivity and specificity) with a confidence region and a predictive region for each test. Hence, we provide a direct and visual comparison between the two competing diagnostic tests.

In the first subsection we apply the ML and HK methods to the rheumatoid arthritis data<sup>28</sup>, also analysed in Dimou et al.<sup>31</sup> using all the available studies and imputation to fill in the missing data. In this dataset, there are large individual probabilities, i.e., more discretization. In the second subsection the methods are applied to the diabetes data<sup>29;30</sup> for which there apparently exist small individual probabilities. As emphasized in the preceding section, the size of the discrete probabilities can substantially influence the efficiency of the HK approximate likelihood method.

### 4.1 Rheumatoid arthritis

The methods are applied to the data obtained from a meta-analysis that aimed to determine whether anti-cyclic citrullinated peptide (anti-CCP) antibody identifies more accurately patients with rheumatoid arthritis than

rheumatoid factor (RF) does.<sup>31</sup> We include  $N = 22$  studies that assessed both RF and anti-CCP2 antibody for diagnosing rheumatoid arthritis.

The log-likelihood showed that a D-vine copula mixed model with Clayton rotated by 270 degrees and Frank copulas both with normal margins to join the (latent) sensitivity and specificity of the first and second test, respectively, provides the best fit (Table 5).

Table 5: Maximized ML and HK log-likelihoods, estimates and standard errors (SE) of the D-vine copula mixed models for the arthritis data.

test	¶ copula	margin	§ copula	margin	† copula	margin	test	‡ copula	margin
RF	Cln270°	normal	Cln270°	normal	BVN	normal	RF	Cln270°	beta
Anti-CCP2	Frank	normal	Frank	normal	BVN	normal	Anti-CCP2	Frank	beta
Param.	Estimate	SE	Estimate	SE	Estimate	SE	Param.	Estimate	SE
$\pi_1$	0.679	0.064	0.683	0.039	0.674	0.030	$\pi_1$	0.660	0.029
$\pi_2$	0.826	0.027	0.824	0.026	0.821	0.025	$\pi_2$	0.772	0.034
$\pi_3$	0.680	0.056	0.681	0.026	0.677	0.032	$\pi_3$	0.669	0.030
$\pi_4$	0.959	0.008	0.959	0.008	0.959	0.008	$\pi_4$	0.949	0.009
$\sigma_1$	0.711	0.127	0.715	0.115	0.705	0.093	$\gamma_1$	0.064	0.021
$\sigma_2$	1.033	0.197	1.037	0.193	1.054	0.176	$\gamma_2$	0.120	0.034
$\sigma_3$	0.698	0.129	0.653	0.073	0.708	0.113	$\gamma_3$	0.075	0.024
$\sigma_4$	0.784	0.174	0.815	0.183	0.791	0.179	$\gamma_4$	0.025	0.012
$\tau_{12}$	-0.156	0.113	-0.162	0.135	-0.126	0.134	$\tau_{12}$	-0.156	0.112
$\tau_{23}$	-0.115	0.145	-	-	-0.077	0.148	$\tau_{23}$	0.051	0.083
$\tau_{34}$	-0.243	0.185	-0.255	0.185	-0.186	0.176	$\tau_{34}$	-0.143	0.132
$\tau_{13 2}$	0.597	0.095	-	-	0.597	0.096	$\tau_{13 2}$	0.308	0.077
$\tau_{24 3}$	0.331	0.178	-	-	0.357	0.174	$\tau_{24 3}$	0.117	0.065
$\tau_{14 23}$	0.127	0.223	-	-	0.111	0.224	$\tau_{14 23}$	0.088	0.070
$-\log L$	319.4		329.5		320.3		$-\ell$	322.6	

Cln270°: Clayton rotated by 270 degrees; ¶: Best fit; §: Separate bivariate meta-analysis for each test; †: The resulting model is the same as the quadrivariate GLMM; ‡: HK approximated likelihood method.

Furthermore, a quadrivariate copula mixed model leads to better inferences than two bivariate copula mixed models with independence between the two diagnostic tests, since the likelihood has been improved by  $10.1 = -319.4 - (-329.5)$ . This indicates that there is strong evidence of dependence between the two diagnostic tests. The fact that the best-fitting bivariate copula for the first diagnostic test is Clayton rotated by 270 degrees (instead of say, BVN) indicates that there is also negative tail dependence (see Figure 4). In Figure 5 is revealed the anti-CCP2 antibody is better compared with RF.

It is also revealed that the HK method leads to biased estimates for the meta-analytic parameters, their variabilities and Kendall's tau associations as the individual discrete probabilities are large and the margins of the random effects are misspecified (the HK method restricts itself to beta margins).

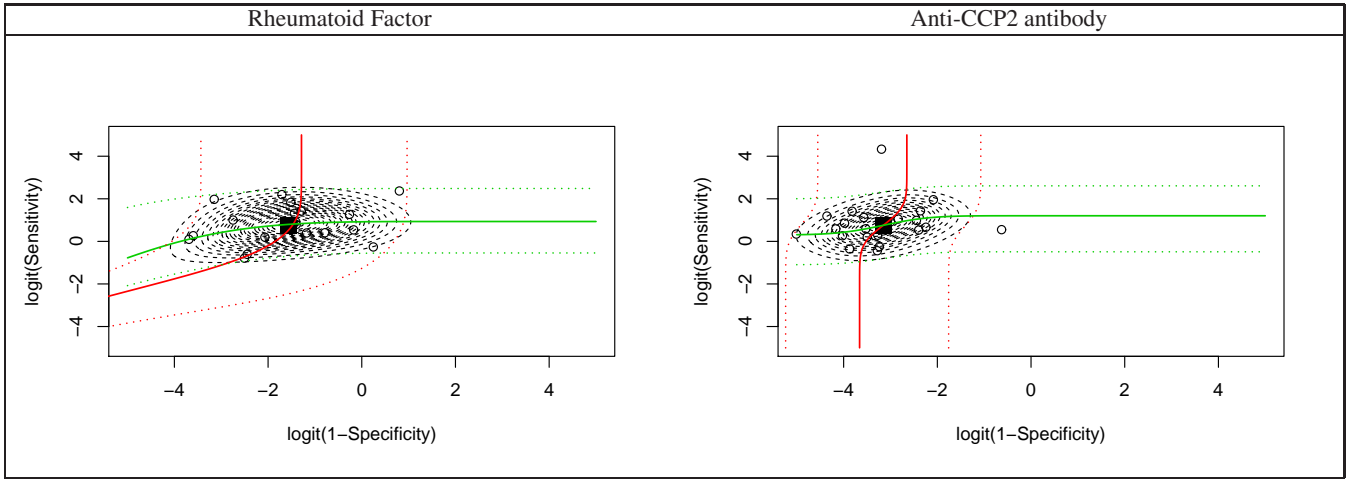


Figure 4: Contour plots (predictive region) and quantile regression curves from the best fitted D-vine copula mixed model for the arthritis data. Red and green lines represent the quantile regression curves  $x_1 := \tilde{x}_1(x_2, q)$  and  $x_2 := \tilde{x}_2(x_1, q)$ , respectively; for  $q = 0.5$  solid lines and for  $q \in \{0.01, 0.99\}$  dotted lines (confidence region). median regression curve for each model. The axes are in logit scale since we also plot the estimated contour plot of the random effects distribution as predictive region; this has been estimated for the logit pair of (Sensitivity, Specificity).

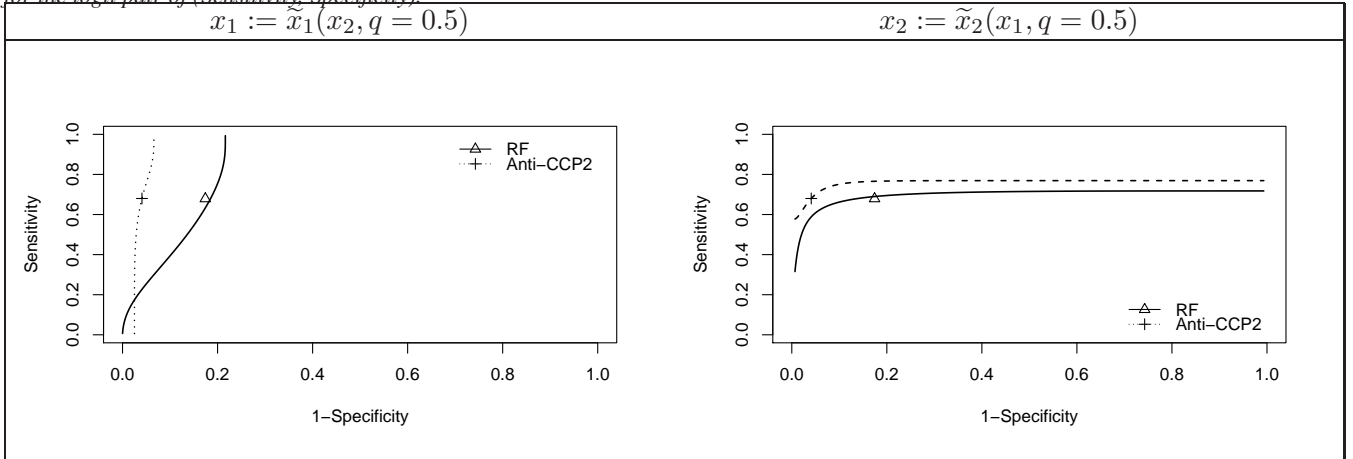


Figure 5: Median regression curves for each test backtransformed to the original scale of sensitivity and specificity for the arthritis data.

## 4.2 Diabetes

The diagnostic tests under investigation are glycated HbA<sub>1c</sub> and FPG<sup>29;30</sup>. Both diagnostic tests are potential alternatives to the oral glucose tolerance test (OGTT), which is commonly used as the gold standard. While the OGTT and FPG require that probands refrain from eating and drinking any liquids, this is not mandatory for the measurement of HbA<sub>1c</sub>.<sup>12</sup> Furthermore, HbA<sub>1c</sub> and FPG are less expensive than the OGTT.<sup>12</sup> The two systematic reviews include in total 38 different studies.

The log-likelihood showed that a D-vine copula mixed model with Clayton rotated by 90 degrees copula with normal margins for test 1 and beta margins for test 2, to join the (latent) sensitivities and specificities, provides the best fit (Table 6).

Furthermore, a quadrivariate copula mixed model leads to better inferences than two bivariate copula mixed



Table 6: Maximized ML and HK log-likelihoods, estimates and standard errors (SE) of the D-vine copula mixed models for the diabetes data.

test	¶ copula	margin	§ copula	margin	test	† copula	margin	test	‡ copula	margin
HbA <sub>1c</sub>	Cln90°	normal	Cln90°	normal	HbA <sub>1c</sub>	BVN	normal	HbA <sub>1c</sub>	Cln90°	beta
FPG	Cln90°	beta	BVN	beta	FPG	BVN	normal	FPG	Cln90°	beta
Param.	Estimate	SE	Estimate	SE	Param.	Estimate	SE	Param.	Estimate	SE
$\pi_1$	0.726	0.024	0.731	0.024	$\pi_1$	0.724	0.024	$\pi_1$	0.706	0.023
$\pi_2$	0.810	0.019	0.808	0.013	$\pi_2$	0.807	0.014	$\pi_2$	0.779	0.021
$\pi_3$	0.703	0.029	0.689	0.025	$\pi_3$	0.737	0.029	$\pi_3$	0.691	0.029
$\pi_4$	0.804	0.022	0.803	0.013	$\pi_4$	0.839	0.012	$\pi_4$	0.804	0.022
$\sigma_1$	0.748	0.087	0.748	0.092	$\sigma_1$	0.762	0.099	$\gamma_1$	0.081	0.018
$\sigma_2$	0.825	0.088	0.836	0.068	$\sigma_2$	0.846	0.076	$\gamma_2$	0.095	0.017
$\gamma_3$	0.137	0.027	0.133	0.022	$\sigma_3$	1.049	0.112	$\gamma_3$	0.143	0.033
$\gamma_4$	0.124	0.023	0.125	0.021	$\sigma_4$	1.055	0.118	$\gamma_4$	0.112	0.023
$\tau_{12}$	-0.307	0.033	-0.287	0.021	$\tau_{12}$	-0.378	0.060	$\tau_{12}$	-0.210	0.078
$\tau_{23}$	0.090	0.094	-	-	$\tau_{23}$	0.016	0.094	$\tau_{23}$	0.053	0.100
$\tau_{34}$	-0.282	0.094	-0.401	0.053	$\tau_{34}$	-0.401	0.072	$\tau_{34}$	-0.245	0.115
$\tau_{13 2}$	0.375	0.080	-	-	$\tau_{13 2}$	0.319	0.083	$\tau_{13 2}$	0.247	0.094
$\tau_{24 3}$	0.400	0.085	-	-	$\tau_{24 3}$	0.289	0.091	$\tau_{24 3}$	0.236	0.071
$\tau_{14 23}$	0.223	0.112	-	-	$\tau_{14 23}$	0.318	0.072	$\tau_{14 23}$	0.156	0.080
$-\log L$	772.6		783.4		$-\ell$	779.9		$-\ell$	778.5	

Cln90°: Clayton rotated by 90 degrees; ¶: Best fit; §: Separate bivariate meta-analysis for each test; †: The resulting model is the same as the quadrivariate GLMM; ‡: HK approximated likelihood method.

models with independence between the two diagnostic tests since the likelihood has been improved by  $10.8 = -772.6 - (-783.4)$ . This indicates that there is strong evidence of dependence between the two diagnostic tests. The fact that the best-fitting bivariate copula for both diagnostic tests is Clayton rotated by 90 (instead of say, BVN) indicates that there is also negative tail dependence (see Figure 6). In Figure 7 is revealed the FPG antibody is slightly better compared with HbA<sub>1c</sub>.

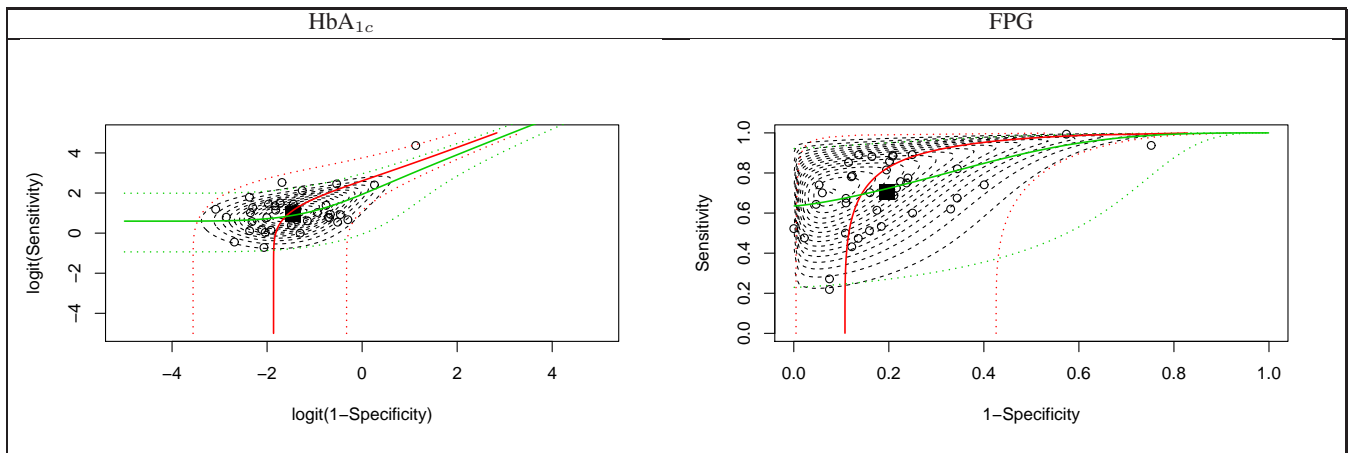


Figure 6: Contour plots (predictive region) and quantile regression curves from the best fitted D-vine copula mixed model for the diabetes data. Red and green lines represent the quantile regression curves  $x_1 := \hat{x}_1(x_2, q)$  and  $x_2 := \hat{x}_2(x_1, q)$ , respectively; for  $q = 0.5$  solid lines and for  $q \in \{0.01, 0.99\}$  dotted lines (confidence region). median regression curve for each model. In case of HbA<sub>1c</sub> the axes are in logit scale since we also plot the estimated contour plot of the random effects distribution as predictive region; this has been estimated for the logit pair of (Sensitivity, Specificity).

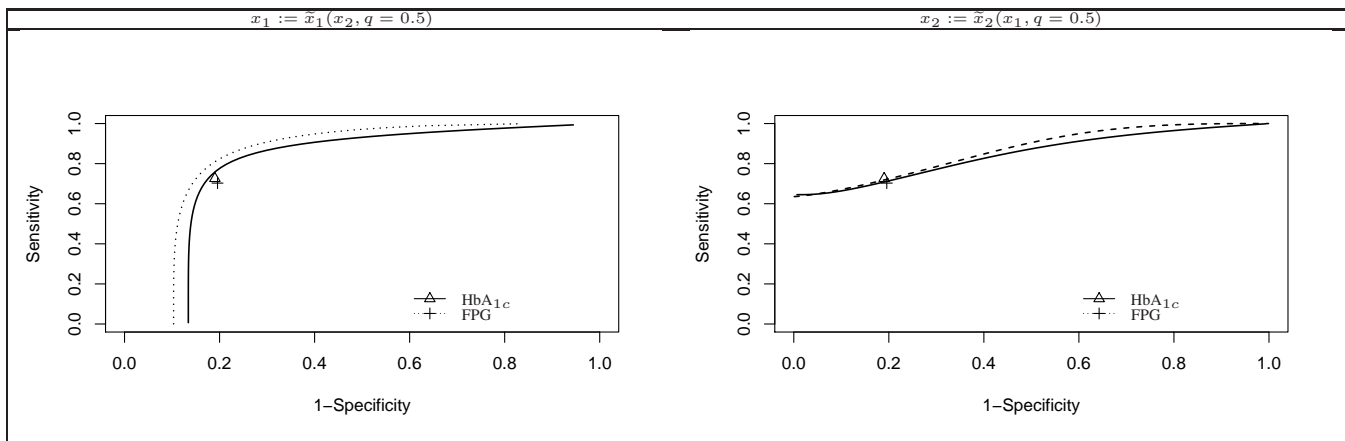


Figure 7: Median regression curves for each test back-transformed to the original scale of sensitivity and specificity for the diabetes data.

It is also demonstrated that the HK method leads to biased estimates for the meta-analytic parameters, as although the individual discrete probabilities are small, the margins of the random effects are misspecified (the HK method restricts itself to beta margins).

## 5 Discussion

We have proposed a D-vine copula mixed model for joint meta-analysis and comparison of diagnostic tests. This is the most general meta-analytic model, with univariate parameters separated from dependence parameters. Our general model includes the quadrivariate GLMM as a special case and can provide an improvement over the latter based on log-likelihood and thus, can provide a better statistical inference for the SROC and the meta-analytic parameters of interest.

For the quadrivariate D-vine copula mixed model the model parameters (including dependence parameters), the choice of the bivariate copulas to link the sensitivity and specificity, and the choice of the margin for each test affect the shape of the SROC curves. The HK<sup>12</sup> approximation method cannot be used to produce the SROC curves, since the dependence parameters affect the shape of the SROC curve and these are generally underestimated. Note in passing that the quadrivariate GLMM<sup>10</sup> can produce SROC curves but these are restricted to the elliptical shape.

We have proposed a numerically stable ML estimation technique based on Gauss-Legendre quadrature; the crucial step is to convert from independent to dependent quadrature points. Although there is an issue of computational burden as the number of quadrature points increase, this will subside, as computing technology is advancing rapidly. Any comparison with the HK method in terms of computing time is a digression. It is obvious that the HK method is much faster than the ML method (even when  $n_q = 15$ ), since a numerically more

difficult 4-dimensional integral calculation is replaced with a much simpler computationally copula density value. However, theoretically there are still problems for large individual probabilities, since the HK approximation of ‘large steps’ in the cdf is poor. In fact, we novelty propose simple diagnostics (descriptive statistics such as a histogram) to judge if the HK method is reliable and reduce by its use the computational burden when it is possible (i.e., for small individual probabilities). Nevertheless, even at this case the HK method might lead to biased estimates since it assumes an  $\text{Beta}(\pi, \gamma)$  distribution for the marginal modelling of the latent proportions. This was the case for the diabetes data in Section 4.2.

## Software

R functions to implement the D-vine copula mixed model for meta-analysis and comparison of two diagnostic tests are part of the R package `CopulaREMADA`.<sup>33</sup>

## Acknowledgements

The simulations presented in this paper were carried out on the High Performance Computing Cluster supported by the Research and Specialist Computing Support service at the University of East Anglia.

## References

- [1] D. Jackson, R. Riley, and I. R. White. Multivariate meta-analysis: Potential and promise. *Statistics in Medicine*, 30(20):2481–2498, 2011.
- [2] D. Mavridis and G. Salanti. A practical introduction to multivariate meta-analysis. *Statistical Methods in Medical Research*, 22(2):133–158, 2013.
- [3] X. Ma, L. Nie, S. R Cole, and H. Chu. Statistical methods for multivariate meta-analysis of diagnostic tests: An overview and tutorial. *Statistical Methods in Medical Research*, 25:1596–1619, 2016.
- [4] Y. Takwoingi, M.M.G. Leeflang, and J.J. Deeks. Empirical evidence of the importance of comparative studies of diagnostic test accuracy. *Annals of Internal Medicine*, 158(7):544–554, 2013.
- [5] M. S. Siadaty and J. Shu. Proportional odds ratio model for comparison of diagnostic tests in meta-analysis. *BMC Medical Research Methodology*, 4:27, 2004.
- [6] A. Tatsioni, D.A. Zarin, N. Aronson, D.J. Samson, C.R. Flamm, C. Schmid, and J. Lau. Challenges in systematic reviews of diagnostic technologies. *Annals of Internal Medicine*, 142:1048–1055, 2005.
- [7] M.M.G. Leeflang, J.J. Deeks, C. Gatsonis, and P.M.M. Bossuyt. Systematic reviews of diagnostic test accuracy. *Annals of Internal Medicine*, 149(12):889–897, 2008.

- [8] H. Chu and S. R. Cole. Bivariate meta-analysis of sensitivity and specificity with sparse data: a generalized linear mixed model approach. *Journal of Clinical Epidemiology*, 59(12):1331–1332, 2006.
- [9] T. A. Trikalinos, D. C. Hoaglin, K. M. Small, N. Terrin, and C. H. Schmid. Methods for the joint meta-analysis of multiple tests. *Research Synthesis Methods*, 5(4):294–312, 2014.
- [10] A. Hoyer and O. Kuss. Meta-analysis for the comparison of two diagnostic tests to a common gold standard: A generalized linear mixed model approach. *Statistical Methods in Medical Research*, 27(5):1410–1421, 2018.
- [11] A. K. Nikoloulopoulos. A vine copula mixed effect model for trivariate meta-analysis of diagnostic test accuracy studies accounting for disease prevalence. *Statistical Methods in Medical Research*, 26(5):2270–2286, 2017.
- [12] A. Hoyer and O. Kuss. Meta-analysis for the comparison of two diagnostic tests—a new approach based on copulas. *Statistics in Medicine*, 37(5):739–748, 2018.
- [13] H. Joe. *Multivariate Models and Dependence Concepts*. Chapman & Hall, London, 1997.
- [14] H. Joe. *Dependence Modeling with Copulas*. Chapman & Hall, London, 2014.
- [15] R. B. Nelsen. *An Introduction to Copulas*. Springer-Verlag, New York, 2006.
- [16] M. Sklar. Fonctions de répartition à  $n$  dimensions et leurs marges. *Publications de l’Institut de Statistique de l’Université de Paris*, 8:229–231, 1959.
- [17] A. K. Nikoloulopoulos and H. Joe. Factor copula models for item response data. *Psychometrika*, 80:126–150, 2015.
- [18] H. Joe, H. Li, and A. K. Nikoloulopoulos. Tail dependence functions and vine copulas. *Journal of Multivariate Analysis*, 101:252–270, 2010.
- [19] A. K. Nikoloulopoulos, H. Joe, and H. Li. Vine copulas with asymmetric tail dependence and applications to financial return data. *Computational Statistics & Data Analysis*, 56:659–3673, 2012.
- [20] H. Joe. Families of  $m$ -variate distributions with given margins and  $m(m - 1)/2$  bivariate dependence parameters. In L. Rüschendorf, B. Schweizer, and M.D. Taylor, editors, *Distributions with Fixed Marginals and Related Topics*, volume 28, pages 120–141, Hayward, CA, 1996. Institute of Mathematical Statistics.
- [21] A. K. Nikoloulopoulos. A mixed effect model for bivariate meta-analysis of diagnostic test accuracy studies using a copula representation of the random effects distribution. *Statistics in Medicine*, 34:3842–3865, 2015.
- [22] O. Kuss, A. Hoyer, and A. Solms. Meta-analysis for diagnostic accuracy studies: a new statistical model

- using beta-binomial distributions and bivariate copulas. *Statistics in Medicine*, 33(1):17–30, 2014.
- [23] Aristidis K Nikoloulopoulos. Comment on ‘A vine copula mixed effect model for trivariate meta-analysis of diagnostic test accuracy studies accounting for disease prevalence’. *Statistical Methods in Medical Research*, 25(2):988–991, 2016.
- [24] J.C. Nash. *Compact Numerical Methods for Computers: Linear Algebra and Function Minimisation*. Hilger, New York, 1990. 2nd edition.
- [25] A. H. Stroud and D. Secrest. *Gaussian Quadrature Formulas*. Prentice-Hall, Englewood Cliffs, NJ, 1966.
- [26] A. K. Nikoloulopoulos. Hybrid copula mixed models for combining case-control and cohort studies in meta-analysis of diagnostic tests. *Statistical Methods in Medical Research*, 2016. DOI:10.1177/0962280216682376.
- [27] Aristidis K. Nikoloulopoulos. On composite likelihood in bivariate meta-analysis of diagnostic test accuracy studies. *AStA Advances in Statistical Analysis*, 102:211–227, 2018.
- [28] K. Nishimura, D. Sugiyama, Y. Kogata, G. Tsuji, T. Nakazawa, S. Kawano, K. Saigo, A. Morinobu, M. Koshiha, K.M. Kuntz, I. Kamae, and S. Kumagai. Meta-analysis: Diagnostic accuracy of anti-cyclic citrullinated peptide antibody and rheumatoid factor for rheumatoid arthritis. *Annals of Internal Medicine*, 146(11):797–808, 2007.
- [29] C.M. Bennett, M. Guo, and S.C. Dharmage. Hba1c as a screening tool for detection of type 2 diabetes: A systematic review. *Diabetic Medicine*, 24(4):333–343, 2007.
- [30] S. Kodama, C. Horikawa, K. Fujihara, R. Hirasawa, Y. Yachi, S. Yoshizawa, S. Tanaka, Y. Sone, H. Shimano, K.T. Iida, K. Saito, and H. Sone. Use of high-normal levels of haemoglobin a1c and fasting plasma glucose for diabetes screening and for prediction: A meta-analysis. *Diabetes/Metabolism Research and Reviews*, 29(8):680–692, 2013.
- [31] N. L. Dimou, M. Adam, and P. G. Bagos. A multivariate method for meta-analysis and comparison of diagnostic tests. *Statistics in Medicine*, 35(20):3509–3523, 2016.
- [32] L. Hua and H. Joe. Tail order and intermediate tail dependence of multivariate copulas. *Journal of Multivariate Analysis*, 102(10):1454–1471, 2011.
- [33] A. K. Nikoloulopoulos. *CopulaREMADA: Copula mixed effect models for multivariate meta-analysis of diagnostic test accuracy studies*, 2018. R package version 1.1.