



CopulaDTA: An R Package for Copula-Based Bivariate Beta-Binomial Models for Diagnostic Test Accuracy Studies in a Bayesian Framework

Victoria N. Nyaga
Scientific Institute
of Public Health
Hasselt University

Marc Arbyn
Scientific Institute
of Public Health

Marc Aerts
Hasselt University

Abstract

The current statistical procedures implemented in statistical software packages for pooling of diagnostic test accuracy data include HSROC regression (Rutter and Gatsonis 2001) and the bivariate random-effects meta-analysis model (BRMA; Reitsma *et al.* 2005; Arends *et al.* 2008; Chu and Cole 2006; Riley *et al.* 2007b). However, these models do not report the overall mean but rather the mean for a central study with random-effect equal to zero and have difficulties estimating the correlation between sensitivity and specificity when the number of studies in the meta-analysis is small and/or when the between-study variance is relatively large (Riley *et al.* 2007a).

This tutorial on advanced statistical methods for meta-analysis of diagnostic accuracy studies discusses and demonstrates Bayesian modeling using the R package **CopulaDTA** (Nyaga 2017) to fit different models to obtain the meta-analytic parameter estimates. The focus is on the joint modeling of sensitivity and specificity using a copula based bivariate beta distribution. Essentially, we extend the work of Nikoloulopoulos (2015) by: (i) presenting the Bayesian approach which offers the flexibility and ability to perform complex statistical modeling even with small data sets and (ii) including covariate information, and (iii) providing an easy to use code. The statistical methods are illustrated by re-analyzing data of two published meta-analyses.

Modeling sensitivity and specificity using the bivariate beta distribution provides marginal as well as study-specific parameter estimates as opposed to using the bivariate normal distribution (e.g., in BRMA) which only yields study-specific parameter estimates. Moreover, copula based models offer greater flexibility in modeling different correlation structures in contrast to the normal distribution which allows for only one correlation structure.

Keywords: diagnostic test accuracy, meta-analysis, Bayesian, random-effects, copula, R.

1. Introduction

In a systematic review of diagnostic test accuracy, the statistical analysis section aims at estimating the average (across studies) sensitivity and specificity of a test and the variability thereof, among other measures. There tends to be a negative correlation between sensitivity and specificity, which postulates the need for correlated data models. The analysis is statistically challenging because the user (i) deals with two summary statistics, (ii) has to account for correlation between sensitivity and specificity, (iii) has to account for heterogeneity in sensitivity and specificity across the studies and (iv) should be allowed to incorporate covariates.

Currently, the HSROC regression (Rutter and Gatsonis 2001) or the bivariate random-effects meta-analysis model (BRMA; Reitsma *et al.* 2005; Arends *et al.* 2008; Chu and Cole 2006) is recommended for pooling of diagnostic test accuracy data. These models fit a bivariate normal distribution which allows for only one correlation structure to the logit transformed sensitivity and specificity. The resulting distribution has no closed form and therefore the mean sensitivity and specificity is only estimated after numerical integration or other approximation methods.

When the number of studies in the meta-analysis is small and/or when the data are sparse (very low counts or even zero cells), maximum likelihood estimation of hierarchical models such as the BRMA and HSROC model encounter computational difficulties (non-convergence) or may give no or unreliable estimates for the between-study correlation (Takwoingi *et al.* 2017). When the correlation is close to the boundary of its parameter space, the between-study variance estimates from the BRMA are upwardly biased as they compensate for the range restriction on the correlation parameter (Riley *et al.* 2007a). According to Riley *et al.* (2007b) this occurs because the maximum likelihood estimator truncates the between-study covariance matrix on the boundary of its parameter space, and this often occurs when the within-study variation is relatively large or the number of studies is small.

For sensitivity and specificity and sample proportions in general, the mean and variance both depend on the underlying probability. Therefore, any factor affecting the probability will change the mean and the variance. This implies that models where the predictors affect the mean but assume a constant variance will generally not be adequate. Both the BRMA and HSROC assume that the transformed sensitivity and specificity is approximately normal with constant variance.

Joint modeling of study specific sensitivity and specificity using a known or copula based bivariate beta distribution overcomes the above mentioned difficulties. Since both sensitivity and specificity take values in the interval space $(0, 1)$, it is a more natural choice to use a beta distribution to describe their distribution across studies, without the need for any transformation. The beta distribution is conjugate to the binomial distribution and therefore it is easy to integrate out the random-effects analytically giving rise to the beta-binomial marginal distributions. Moreover no further integration is needed to obtain the meta-analytically pooled sensitivity and specificity. Previously, Cong *et al.* (2007) fitted separate beta-binomial models to the number of true positives and the number of false positives. While the model ignores correlation between sensitivity and specificity, Cong *et al.* (2007) reported that the model estimates are comparable to those from the SROC model (Moses *et al.* 1993), the predecessor of the HSROC model.

According to Riley (2009), ignoring the correlation would have negligible influence on the

meta-analysis results when the within-study variability is large relative to the between-study variability. It is generally known that full parametric specification of a (hierarchical) model, including the specification of an (existing) correlation structure, increases the efficiency of the estimation of the parameters, resulting in smaller standard errors. The use of copula based mixed models within the frequentist framework for meta-analysis of diagnostic test accuracy was recently introduced by [Nikoloulopoulos \(2015\)](#) who evaluated the joint density numerically.

This tutorial presents and demonstrates hierarchical mixed models for meta-analysis of diagnostic accuracy studies. In the first level of the hierarchy, given sensitivity and specificity for each study, two binomial distributions are used to describe the variation in the number of true positives and true negatives among the diseased and healthy individuals, respectively. In the second level, we model the unobserved sensitivities and specificities using a bivariate distribution. While hierarchical models are used, the focus of meta-analysis is on the pooled average across studies and rarely on a given study estimate.

The methods are demonstrated using datasets from two previously published meta-analyses: a) on diagnostic accuracy of telomerase in urine as a tumor marker for the diagnosis of primary bladder cancer from [Glas *et al.* \(2003\)](#) previously used by [Riley *et al.* \(2007b\)](#) and [Nikoloulopoulos \(2015\)](#) since it is a problematic dataset that has convergence issues caused by the correlation parameter being estimated to be -1 and has no covariates and b) on the comparison of the sensitivity and specificity of human papillomavirus testing (using the HC2 assay) versus repeat cytology to triage women with minor cytological cervical lesions to detect underlying cervical precancer from [Arbyn *et al.* \(2013\)](#). The second dataset is used to demonstrate meta-regression with one covariate which can be naturally extended to include several covariates.

The layout of this tutorial is as follows: Section 2 introduces the concept of copula theory and different bivariate distributions for sensitivity and specificity. The software implementation and model selection in a Bayesian framework are discussed in Section 3. The two aforementioned datasets are introduced in Section 4. Application of software, code examples and the results of the models fitted to the data are presented in Sections 5 and 6. The complete code is available with this article at the Journal of Statistical Software web site alongside this article. A brief discussion is found in Section 7 and a conclusion in Section 8.

2. Statistical methods for meta-analysis

2.1. Definition of the copula function

A bivariate copula function describes the dependence structure between two random variables. Two random variables X_1 and X_2 are joined by a copula function C if their joint cumulative distribution function can be written as

$$F(x_1, x_2) = C(F_1(x_1), F_2(x_2)), \quad -\infty \leq x_1, x_2 \leq +\infty, \quad (1)$$

where F_1 and F_2 denote the univariate cumulative distribution functions of X_1 and X_2 respectively.

According to the theorem of [Sklar \(1959\)](#), there exists for every bivariate (multivariate in extension) distribution a copula representation C which is unique for continuous random

variables. If the joint cumulative distribution function and the two marginals are known, then the copula function can be written as

$$C(u, v) = F(F_1^{-1}(u), F_2^{-1}(v)), 0 \leq u, v \leq 1. \quad (2)$$

A 2-dimensional copula is in fact simply a 2-dimensional cumulative distribution function restricted to the unit square with standard uniform marginals. A comprehensive overview of copulas and their mathematical properties can be found in [Nelsen \(2006\)](#). To obtain the joint probability density, the joint cumulative distribution in Equation 1 is differentiated to yield

$$f(x_1, x_2) = f_1(x_1) f_2(x_2) c(F_1(x_1), F_2(x_2)), \quad (3)$$

where f_1 and f_2 denote the marginal density functions and c the copula density function corresponding to the copula cumulative distribution function C . Therefore from Equation 3, a bivariate probability density can be expressed using the marginal densities and the copula density, given that the copula function is absolutely continuous and twice differentiable.

When the functional form of the marginal and the joint densities are known, the copula density can be derived as follows

$$c(F_1(x_1), F_2(x_2)) = \frac{f(x_1, x_2)}{f_1(x_1) f_2(x_2)}. \quad (4)$$

While our interest does not lie in finding the copula function, Equations 3 and 4 serve to show how one can move from the copula function to the bivariate density or vice-versa, given that the marginal densities are known. The decompositions allow for constructions of other and possibly better models for the variables than would be possible if we limited ourselves to only existing standard bivariate distributions.

We finish this section by mentioning an important implication when Sklar's theorem is extended to a meta-regression setting with covariates. According to [Patton \(2006\)](#), it is important that the conditioning variable remains the same for both marginal distributions and the copula, as otherwise the joint distribution might not be properly defined. This implies that covariate information should be introduced in both the marginals and the association parameters of the model.

2.2. The hierarchical model

Since there are two sources of heterogeneity in the data, the within- and between-study variability, the parameters involved in a meta-analysis of diagnostic accuracy studies vary at two levels. For each study i , $i = 1, \dots, n$, let $Y_i = (Y_{i1}, Y_{i2})$ denote the true positives and true negatives, $N_i = (N_{i1}, N_{i2})$ the diseased and healthy individuals respectively, and $\pi_i = (\pi_{i1}, \pi_{i2})$ represent the "unobserved" sensitivity and specificity respectively.

Given study-specific sensitivity and specificity, two separate binomial distributions describe the distribution of true positives and true negatives among the diseased and the healthy individuals as follows

$$Y_{ij} | \pi_{ij} \sim \text{bin}(\pi_{ij}, N_{ij}), i = 1, \dots, n, j = 1, 2. \quad (5)$$

Equation 5 forms the higher level of the hierarchy and models the within-study variability. The second level of the hierarchy aims to model the between-study variability of sensitivity

and specificity while accounting for the inherent negative correlation thereof, with a bivariate distribution as follows

$$\begin{pmatrix} g(\pi_{i1}) \\ g(\pi_{i2}) \end{pmatrix} \sim f(g(\pi_{i1}), g(\pi_{i2})) = f(g(\pi_{i1})) f(g(\pi_{i2})) c(F_1(g(\pi_{i1})), F_2(g(\pi_{i2}))), \quad (6)$$

where $g(\cdot)$ denotes a transformation that is used to map the $(0, 1)$ range to the whole real line. While it is critical to ensure that the studies included in the meta-analysis satisfy the specified entry criterion, there are study specific characteristics like different test thresholds and other unobserved differences that give rise to the second source of variability, the between-study variability. It is indeed the difference in the test thresholds between the studies that gives rise to the correlation between sensitivity and specificity. Including study level covariates allows us to model part of the between-study variability. The covariate information can and should (Patton 2006) be used to model the mean as well as the correlation between sensitivity and specificity.

In the next section we give more details on different bivariate distributions $f(g(\pi_{i1}), g(\pi_{i2}))$ constructed using the logit or identity link function $g(\cdot)$, different marginal densities and/or different copula densities c . We discuss their implications and demonstrate their application in meta-analysis of diagnostic accuracy studies. An overview of suitable parametric families of copulas for mixed models for diagnostic test accuracy studies was recently given by Nikoloulopoulos (2015). In the following section, a short description of well-known copulas implemented in the package is given. Here, we consider five copula functions which can be plugged in Equation 3 to model negative correlation.

Bivariate Gaussian copula

Given the density and the distribution function of the univariate and bivariate standard normal distribution with correlation parameter $\rho \in (-1, 1)$, the bivariate Gaussian copula function and density are expressed (Meyer 2013) as

$$\begin{aligned} C(u, v, \rho) &= \Phi_2(\Phi^{-1}(u), \Phi^{-1}(v), \rho), \\ c(u, v, \rho) &= \frac{1}{\sqrt{1 - \rho^2}} \exp\left(\frac{2 \rho \Phi^{-1}(u) \Phi^{-1}(v) - \rho^2 (\Phi^{-1}(u)^2 + \Phi^{-1}(v)^2)}{2 (1 - \rho^2)}\right). \end{aligned} \quad (7)$$

The logit transformation is often used in binary logistic regression to relate the probability of “success” (coded as 1, failure as 0) of the binary response variable with the linear predictor model that theoretically can take values over the whole real line. In diagnostic test accuracy studies, the “unobserved” sensitivities and specificities can range from 0 to 1 whereas their logits $= \log \frac{\pi_{ij}}{1 - \pi_{ij}}$ can take any real value allowing to use the normal distribution as follows

$$\text{logit}(\pi_{ij}) \sim N(\mu_j, \sigma_j) \Leftrightarrow \text{logit}(\pi_{ij}) = \mu_j + \varepsilon_{ij}, \quad (8)$$

where, μ_j is a vector of the mean sensitivity and specificity for a study with zero random effects, and ε_i is a vector of random effects associated with study i . Now u is the normal distribution function of $\text{logit}(\pi_{i1})$ with parameters μ_1 and σ_1 , v is the normal distribution function of $\text{logit}(\pi_{i2})$ with parameters μ_2 and σ_2 , Φ_2 is the distribution function of a bivariate standard normal distribution with correlation parameter $\rho \in (-1, 1)$ and Φ^{-1} is the quantile of the standard normal distribution. In terms of ρ , Kendall’s tau is expressed as $(\frac{2}{\pi})\arcsin(\rho)$.

With simple algebra the copula density in Equation 7 with normal marginal distributions simplifies to

$$c(u, v, \rho) = \frac{1}{\sqrt{1 - \rho^2}} \exp\left(\frac{1}{2(1 - \rho^2)} \left(\frac{2\rho(x - \mu_1)(y - \mu_2)}{\sigma_1\sigma_2} - \rho^2 \left(\frac{(x - \mu_1)^2}{\sigma_1^2} + \frac{(y - \mu_2)^2}{\sigma_2^2}\right)\right)\right). \quad (9)$$

The product of the copula density in Equation 9, the normal marginal of $\text{logit}(\pi_{i1})$ and $\text{logit}(\pi_{i2})$ in Equation 8 form a bivariate normal distribution which characterize the model by Reitsma *et al.* (2005), Arends *et al.* (2008), Chu and Cole (2006), and Riley *et al.* (2007a), the so-called bivariate random-effects meta-analysis (BRMA) model, recommended as the appropriate method for meta-analysis of diagnostic accuracy studies. Study level covariate information explaining heterogeneity is introduced through the parameters of the marginal and the copula as follows

$$\boldsymbol{\mu}_j = \mathbf{X}_j \mathbf{B}_j^\top. \quad (10)$$

\mathbf{X}_j is a $n \times p$ matrix containing the covariate values for the mean sensitivity ($j = 1$) and specificity ($j = 2$). For simplicity, assume that $\mathbf{X}_1 = \mathbf{X}_2 = \mathbf{X}$. \mathbf{B}_j^\top is a $p \times 1$ vector of regression parameters, and p is the number of parameters. By inverting the logit functions in Equation 8, we obtain

$$\pi_{ij} = \text{logit}^{-1}(\mu_j + \varepsilon_{ij}). \quad (11)$$

Therefore, the meta-analytic sensitivity and specificity obtained by averaging over the random study effect, is given by, for $j = 1, 2$

$$\mathbb{E}(\pi_j) = \mathbb{E}(\text{logit}^{-1}(\mu_j + \varepsilon_{ij})) = \int_{-\infty}^{\infty} \text{logit}^{-1}(\mu_j + \varepsilon_{ij}) f(\varepsilon_{ij}, \sigma_j) d\varepsilon_{ij}, \quad (12)$$

assuming that $\sigma_1^2 > 0$ and $\sigma_2^2 > 0$. The integration in Equation 12 has no analytical solution and therefore needs to be numerically approximated. Using simulations methods in the Bayesian framework, the meta-analytic estimates can be easily computed as well as the standard deviations and the credible intervals for $\mathbb{E}(\pi_j)$ with minimum effort.

In the frequentist framework, it is more convenient however to use numerical averaging by sampling a large number M of random-effects $\hat{\varepsilon}_{ij}$ from the fitted distribution and to estimate the meta-analytic sensitivity and specificity by (Molenberghs and Verbeke 2005), for $j = 1, 2$

$$\hat{\mathbb{E}}(\pi_j) = \frac{1}{M} \sum_{i=1}^M \text{logit}^{-1}(\hat{\mu}_j + \hat{\varepsilon}_{ij}). \quad (13)$$

However, inference is not straightforward in the frequentist framework since the standard errors are not available. When $\varepsilon_{ij} = 0$, then

$$\mathbb{E}(\pi_j | \varepsilon_{ij} = 0) = \text{logit}^{-1}(\mu_j). \quad (14)$$

Inference for $\mathbb{E}(\pi_j | \varepsilon_{ij} = 0)$, as expressed in Equation 14, can be done in both Bayesian and frequentist framework. The equation represents the mean sensitivity and specificity for a "central" study with $\varepsilon_{ij} = 0$. Researchers often seem to confuse $\mathbb{E}(\pi_j | \varepsilon_{ij} = 0)$ with $\mathbb{E}(\pi_j)$ but due to the non-linear logit transformations, they are clearly not the same parameter.

With the identity link function, no transformation on study-specific sensitivity and specificity is performed. A natural choice for u and v would be beta distribution functions with parameters (α_1, β_1) and (α_2, β_2) respectively. Since $\pi_{ij} \sim \text{beta}(\alpha_j, \beta_j)$, the meta-analytic sensitivity and specificity are analytically solved as follows

$$\mathbf{E}(\pi_j) = \frac{\alpha_j}{\alpha_j + \beta_j}. \quad (15)$$

After reparameterizing the beta distributions using the mean ($\mu_j = \frac{\alpha_j}{\alpha_j + \beta_j}$) and certainty ($\psi_j = \alpha_j + \beta_j$) or dispersion ($\varphi_j = \frac{1}{1 + \alpha_j + \beta_j}$) parameters different link functions introduce covariate information to the mean, certainty/dispersion and association (ρ) parameters. A typical model parameterization is

$$\begin{aligned} \boldsymbol{\mu}_j &= \text{logit}^{-1}(\mathbf{X}\mathbf{B}_j^\top), \\ \boldsymbol{\psi}_j &= g(\mathbf{W}\mathbf{C}_j^\top), \\ \boldsymbol{\alpha}_j &= \boldsymbol{\mu}_j \circ \boldsymbol{\psi}_j, \\ \boldsymbol{\beta}_j &= (\mathbf{1} - \boldsymbol{\mu}_j) \circ \boldsymbol{\psi}_j, \\ \boldsymbol{\rho} &= \tanh(\mathbf{Z}\mathbf{D}_j^\top) = \frac{\exp(2 \times \mathbf{Z}\mathbf{D}_j^\top) - 1}{\exp(2 \times \mathbf{Z}\mathbf{D}_j^\top) + 1}. \end{aligned} \quad (16)$$

\mathbf{X} , \mathbf{W} and \mathbf{Z} are $n \times p$ matrices containing the covariate values for the mean, dispersion and correlation which we will assume contain similar information and denote all by \mathbf{X} for simplicity purpose, p is the number of parameters, \mathbf{B}_j^\top , \mathbf{C}_j^\top and \mathbf{D}_j^\top are $p \times 1$ vectors of regression parameters relating covariates to the mean, variance and correlation respectively. $g(\cdot)$ is the log link mapping $\mathbf{X}\mathbf{C}_j^\top$ to the positive real number line and \circ is the Hadamard product.

Frank copula

This flexible copula in the so-called family of Archimedean copulas was introduced by Frank (1979). The functional form of the copula and the density which is plugged in Equation 3 is given by:

$$\begin{aligned} C(F(\pi_{i1}), F(\pi_{i2}), \theta) &= -\frac{1}{\theta} \log \left[1 + \frac{(e^{-\theta F(\pi_{i1})} - 1)(e^{-\theta F(\pi_{i2})} - 1)}{e^{-\theta} - 1} \right], \\ c(F(\pi_{i1}), F(\pi_{i2}), \theta) &= \frac{\theta (1 - e^{-\theta}) e^{-\theta (F(\pi_{i1}) + F(\pi_{i2}))}}{[1 - e^{-\theta} - (1 - e^{-\theta F(\pi_{i1})}) (1 - e^{-\theta F(\pi_{i2})})]^2}. \end{aligned} \quad (17)$$

Since $\theta \in \mathbb{R}$, both positive and negative correlation can be modeled, making this one of the more comprehensive copulas. When θ is 0, sensitivity and specificity are independent. For $\theta > 0$, sensitivity and specificity exhibit positive quadrant dependence and negative quadrant dependence when $\theta < 0$. The Spearman correlation ρ_s and Kendall's tau τ_k can be expressed in terms of θ as

$$\begin{aligned} \rho_s &= 1 - 12 \frac{D_2(-\theta) - D_1(-\theta)}{\theta}, \\ \tau_k &= 1 + 4 \frac{D_1(\theta) - 1}{\theta}, \end{aligned} \quad (18)$$

where $D_j(\delta)$ is the Debye function defined as

$$D_j(\delta) = \frac{j}{\delta^j} \int_0^\delta \frac{t^j}{\exp(t) - 1} dt, \quad j = 1, 2. \quad (19)$$

Covariate information is introduced in a similar manner as Equation 16. The identity link is used for the association parameter θ .

Farlie-Gumbel-Morgenstern copula (FGM)

This popular copula studied by Farlie (1960), Gumbel (1960) and Morgenstern (1956) is defined as

$$\begin{aligned} C(F(\pi_{i1}), F(\pi_{i2}), \theta) &= F(\pi_{i1}) F(\pi_{i2}) [1 + \theta (1 - F(\pi_{i1})) (1 - F(\pi_{i2}))], \\ c(F(\pi_{i1}), F(\pi_{i2}), \theta) &= [1 + \theta (2 F(\pi_{i1}) - 1) (2 F(\pi_{i2}) - 1)]. \end{aligned} \quad (20)$$

Because $\theta \in (-1, 1)$, the Spearman correlation and Kendall's tau are expressed in terms of θ as $\theta/3$ and $2\theta/9$ respectively, making this copula only appropriate for data with weak dependence since $|\rho_s| \leq 1/3$. In a similar manner as in Equation 16 the logit link, log/identity link and Fisher's z transformation can be used to introduce covariate information in modeling the mean, dispersion and association parameter.

Clayton copula

The Clayton copula function and density by Clayton (1978) is defined as

$$\begin{aligned} C(F(\pi_{i1}), F(\pi_{i2}), \theta) &= [F(\pi_{i1})^{-\theta} + F(\pi_{i2})^{-\theta} - 1]^{-\frac{1}{\theta}}, \\ c(F(\pi_{i1}), F(\pi_{i2}), \theta) &= (1 + \theta) F(\pi_{i1})^{-(1+\theta)} F(\pi_{i2})^{-(1+\theta)} \\ &\quad [F(\pi_{i1})^{-\theta} + F(\pi_{i2})^{-\theta} - 1]^{-\frac{(2\theta+1)}{\theta}}. \end{aligned} \quad (21)$$

Since $\theta \in (0, \infty)$, the Clayton copula typically models positive dependence; Kendall's tau equals $\theta/(\theta + 2)$. However, the copula function can be rotated by 90° or 270° to model negative dependence. The distribution and density functions following such rotations are given by

$$\begin{aligned} C_{90}(F(\pi_{i1}), F(\pi_{i2}), \theta) &= F(\pi_{i2}) - C(1 - F(\pi_{i1}), F(\pi_{i2}), \theta), \\ c_{90}(F(\pi_{i1}), F(\pi_{i2}), \theta) &= (1 + \theta) (1 - F(\pi_{i1}))^{-(1+\theta)} F(\pi_{i2})^{-(1+\theta)} \\ &\quad [(1 - F(\pi_{i1}))^{-\theta} + F(\pi_{i2})^{-\theta} - 1]^{-\frac{(2\theta+1)}{\theta}}, \end{aligned} \quad (22)$$

and

$$\begin{aligned} C_{270}(F(\pi_{i1}), F(\pi_{i2}), \theta) &= F(\pi_{i1}) - C(F(\pi_{i1}), 1 - F(\pi_{i2}), \theta), \\ c_{270}(F(\pi_{i1}), F(\pi_{i2}), \theta) &= (1 + \theta) F(\pi_{i1})^{-(1+\theta)} (1 - F(\pi_{i2}))^{-(1+\theta)} \\ &\quad [F(\pi_{i1})^{-\theta} + (1 - F(\pi_{i2}))^{-\theta} - 1]^{-\frac{(2\theta+1)}{\theta}}. \end{aligned} \quad (23)$$

The logit, log/identity and log/identity links can be used to introduce covariate information in modeling the mean (μ_j), certainty (ψ_j)/dispersion (φ_j) and association (θ) parameters respectively in the same way as in Equation 16.

Of course other copula functions that allow for negative association can be chosen. It is also an option to use known bivariate beta distributions. However, it is not always straightforward and analytically attractive to derive the corresponding copula function for all bivariate distributions. The use of existing bivariate beta distributions in meta-analysis of diagnostic accuracy studies has been limited because these densities model positive association (e.g., Libby and Novick 1982, Olkin and Liu 2003), or both positive and negative association but over a restricted range (e.g., Sarmanov 1966).

3. Software development and model diagnostics

3.1. The CopulaDTA package

The **CopulaDTA** package is an R (R Core Team 2017) package for modeling diagnostic test accuracy data using copula based bivariate beta-binomial distribution and providing estimates for the marginal mean sensitivity and specificity. It is an extension of **rstan** (Stan Development Team 2017), the R interface to Stan (Carpenter *et al.* 2017) for diagnostic test accuracy data. Stan is a probabilistic programming language which has implemented Hamilton Monte Carlo (HMC) and uses No-U-Turn sampler (NUTS) (Hoffman and Gelman 2014). The package facilitates easy application of complex models and their visualization within the Bayesian framework.

JAGS (Just Another Gibbs Sampler; Plummer 2003) is an alternative extensible general purpose sampling engine to Stan. Extending JAGS requires knowledge of C++ to assemble a dynamic link library (DLL) module or a shared library. From experience, configuring and building the module is a daunting and tedious task especially in the Windows operation system. The above short-comings coupled with the fact that Stan uses a more efficient and robust sampler than Gibbs sampling or Metropolis-Hasting, and its effectiveness in convergence from diffuse random initial values made us prefer the Stan MCMC sampling engine (Carpenter *et al.* 2017).

The **CopulaDTA** package is available via the Comprehensive R Archive Network (CRAN) at <http://CRAN.R-project.org/package=CopulaDTA>. With a working internet connection, the **CopulaDTA** package is installed and loaded in R with the following commands

```
R> install.packages("CopulaDTA", dependencies = TRUE)
R> library("CopulaDTA")
```

The **CopulaDTA** package provide functions to fit bivariate beta-binomial distributions constructed as a product of two beta marginal distributions and copula densities discussed in Section 2. The package also provides forest plots for a model with categorical covariates or with intercept only. Given the chosen copula function, a beta-binomial distribution is assembled up by the `cdtamodel` function which returns a `'cdtamodel'` object. The main function `fit` takes the `'cdtamodel'` object and fits the model to the given dataset and returns a `'cdtafit'` object for which `print`, `summary` and `plot` methods are provided.

3.2. Model diagnostics

To assess convergence, mixing and stationarity of the chains, it is necessary to check the

potential scale reduction factor \hat{R} , the effective sample size (ESS), the MCMC error and trace plots of the parameters. When all the chains reach the target posterior distribution, the estimated posterior variance is expected to be close to the within chain variance such that the ratio of the two, \hat{R} is close to 1 indicating that the chains are stable, properly mixed and likely to have reached the target distribution. A large \hat{R} indicates poor mixing and that more iterations are needed. The effective sample size indicates how much information one actually has about a certain parameter. When the samples are autocorrelated, less information from the posterior distribution of our parameters is expected than would be if the samples were independent. ESS close to the total post-warm-up iterations is an indication of less autocorrelation and good mixing of the chains. Simulations with higher ESS have lower standard errors and more stable estimates. Since the posterior distribution is simulated there is a chance that the approximation is off by some amount; the Monte Carlo (MCMC) error. MCMC error close to 0 indicates that one is likely to have reached the target distribution.

3.3. Model comparison and selection

Watanabe-Akaike information criterion (WAIC; [Watanabe 2010](#)), a recent model comparison tool to measure the predictive accuracy of the fitted models in the Bayesian framework, will be used to compare the models. WAIC can be viewed as an improvement of the Deviance information criterion (DIC) which, though popular, is known to have some problems ([Plummer 2008](#)). WAIC is a fully Bayesian tool, closely approximates Bayesian cross-validation, is invariant to reparameterization and can be used for simple as well as hierarchical and mixture models.

4. Datasets

4.1. Telomerase data

[Glas et al. \(2003\)](#) systematically reviewed the sensitivity and specificity of cytology and other markers including telomerase for primary diagnosis of bladder cancer. They fitted a bivariate normal distribution to the logit transformed sensitivity and specificity values across the studies allowing for heterogeneity between the studies. From the included 10 studies, they reported that telomerase had a sensitivity and specificity of 0.75 [0.71, 0.79] and 0.86 [0.71, 0.94] respectively. They concluded that telomerase was not sensitive enough to be recommended for daily use. This dataset is available within the package and the following commands

```
R> data("telomerase", package = "CopulaDTA")
R> telomerase
```

load the data into the R environment and generate the following output

	ID	TP	TN	FN	FP
1	1	25	25	8	1
2	2	17	11	4	3
3	3	88	31	16	16
4	4	16	80	10	3

```

5  5 40 137 17  1
6  6 38  24  9  6
7  7 23  12 19  0
8  8 27  18  6  2
9  9 14  29  3  3
10 10 37   7  7 22

```

ID is the study identifier, DIS is the number of diseased, TP is the number of true positives, NonDis is the number of healthy and TN is the number of true negatives.

4.2. ASCUS triage data

Arbyn *et al.* (2012) and Arbyn *et al.* (2013) performed a meta-analysis and Cochrane review on the accuracy of human papillomavirus testing and repeat cytology to triage of women with an equivocal Pap smear to diagnose cervical precancer. They fitted the BRMA model in SAS (SAS Institute Inc. 2013) using MetaDAS (Takwoingi and Deeks 2010) on 10 studies where both tests were used. They reported absolute sensitivity of 0.91 [0.86, 0.94] and 0.72 [0.63, 0.79] for HC2 and repeat cytology respectively. The specificity was 0.61 [0.54, 0.68] and 0.68 [0.60, 0.76] for HC2 and repeat cytology respectively. This data is used to demonstrate how the intercept-only model is extended in a meta-regression setting. This dataset is also available within the package and the following commands

```

R> data("ascus", package = "CopulaDTA")
R> ascus

```

load the data into the R environment and generate the following output

	Test	StudyID	TP	FP	TN	FN
1	RepC	Andersson 2005	6	14	28	4
2	RepC	Bergeron 2000	8	28	71	4
3	RepC	Del Mistro 2010	20	191	483	7
4	RepC	Kulasingham 2002	20	74	170	6
5	RepC	Lytwyn 2000	4	20	26	2
6	RepC	Manos 1999	48	324	570	15
7	RepC	Monsonogo 2008	10	18	168	15
8	RepC	Morin 2001	14	126	214	5
9	RepC	Silverloo 2009	24	43	105	10
10	RepC	Solomon 2001	227	1132	914	40
11	HC2	Andersson 2005	6	17	25	4
12	HC2	Bergeron 2000	10	38	61	2
13	HC2	Del Mistro 2010	27	154	566	2
14	HC2	Kulasingham 2002	23	115	129	3
15	HC2	Lytwyn 2000	4	19	33	1
16	HC2	Manos 1999	58	326	582	7
17	HC2	Monsonogo 2008	22	110	72	2
18	HC2	Morin 2001	17	88	253	2
19	HC2	Silverloo 2009	34	65	81	2
20	HC2	Solomon 2001	256	1050	984	11

Test is an explanatory variable showing the type of triage test, **StudyID** is the study identifier, **TP** is the number of true positives, **FP** is the number of false positives, **TN** is the number of true negatives, **FN** is the number of false negatives.

5. The intercept-only model

The **CopulaDTA** package has five different correlation structures that result in five different bivariate beta-binomial distributions to fit to the data. The correlation structure is specified by indicating `copula = "gauss"` or `"fgm"` or `"c90"` or `"c270"` or `"frank"` in the `fitcopula` function. The Gaussian copula bivariate beta-binomial distribution is fitted to the `telomerase` data with the following code

```
R> gauss.1 <- cdtamodel("gauss")
R> fitgauss.1 <- fit(gauss.1, data = telomerase, SID = "ID", iter = 28000,
+   warmup = 1000, thin = 30, seed = 3)
```

By default, `chains = 3` and `cores = 3` and need not be specified unless other values are desired. From the code above, 28000 samples are drawn from each of the 3 chains, the first 1000 samples are discarded and thereafter every 30th draw kept such that each chain has 900 post-warm-up draws making a total of 2700 post-warm-up draws. The seed value, `seed = 3`, specifies a random number generator to allow reproducibility of the results and `cores = 3` allows for parallel-processing of the chains by using 3 cores, one core for each chain. There were no initial values specified and in that case, the program randomly generates random values satisfying the parameter constraints. The trace plots in the top-left panel of Figure 1 produced with the code below show satisfactory mixing of the chains and convergence.

```
R> traceplot(fitgauss.1)
```

Next, obtain the model summary estimates as follows

```
R> print(fitgauss.1, digits = 2)
```

Posterior marginal mean and median sensitivity and specificity
with 95% credible intervals

	Parameter	Mean	Lower	Median	Upper	n_eff	Rhat
MUse[1]	Sensitivity	0.7538	0.69094	0.7485	0.814	237	1.0
MUsp[1]	Specificity	0.7965	0.65343	0.7982	0.911	67	1.0
ktau[1]	Correlation	-0.8502	-0.98314	-0.9115	-0.407	25	1.1
Varsen[1]	Var(Sens)	0.0065	0.00072	0.0062	0.019	1350	1.0
Varsp[1]	Var(Spec)	0.0504	0.01346	0.0470	0.110	52	1.0

Model characteristics

Copula function: gauss, sampling algorithm: NUTS(diag_e)

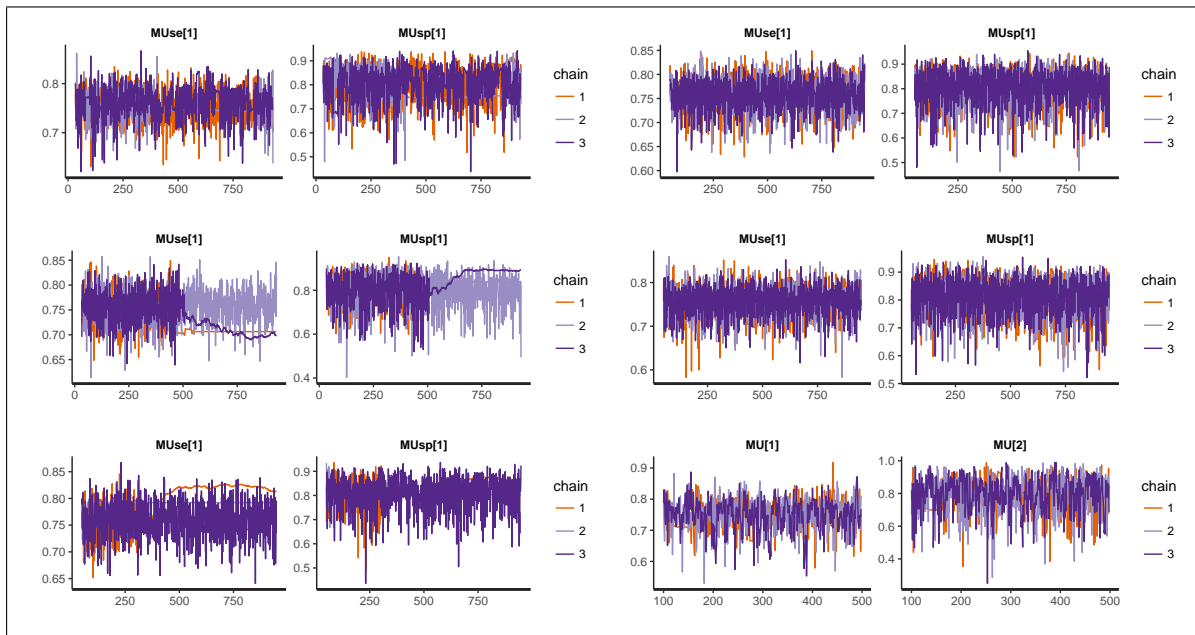


Figure 1: Trace plots of the posterior mean sensitivity and specificity for the `telomerase` data as estimated by the Gaussian, Clayton 90° (C90) and 270° (C270), Farlie-Gumbel-Morgenstern (FGM) and Frank copula based bivariate beta and bivariate normal (BRMA) distributions.

```
Formula(1): MUSe ~ 1
Formula(2): MUsp ~ 1
Formula(3): Omega ~ 1
3 chain(s)each with iter=28000; warm-up=1000; thin=30.
post-warmup draws per chain=900;total post-warmup draws=2700.
```

Predictive accuracy of the model

```
Log point-wise predictive density (LPPD): -37.93
Effective number of parameters: 5.95
Watanabe-Akaike information Criterion (WAIC): 87.75
```

From the output above, `n_eff` and `Rhat` both confirm proper mixing of the chains with little autocorrelation. The meta-analytic sensitivity `MUSe[1]` and specificity `MUsp[1]` is 0.75 [0.69, 0.81] and 0.80 [0.65, 0.91] respectively. The Kendall's tau correlation between sensitivity and specificity is estimated to be -0.85 $[-0.98, -0.41]$. The between-study variability in sensitivity and specificity is 0.01 [0.00, 0.02] and 0.05 [0.01, 0.11] respectively.

The command below produces the forest plot in Figure 2.

```
R> plot(fitgauss.1, graph = 3, title.3 = "" )
```

As observed in Figure 2, the posterior study-specific sensitivity and specificity are less extreme and variable than the “observed” study-specific sensitivity and specificity. In other words,

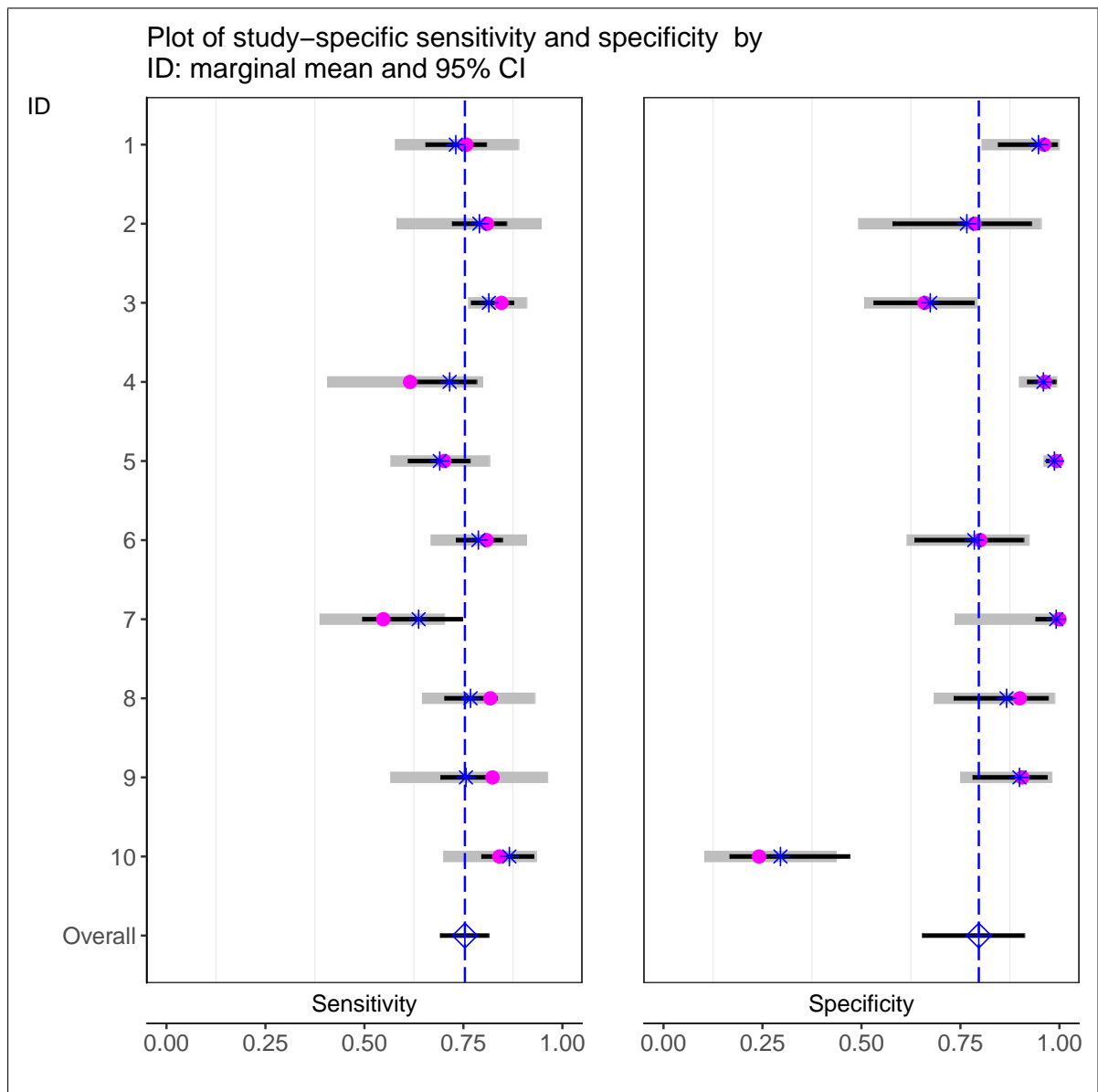


Figure 2: Plot of the study-specific sensitivity and specificity (magenta points) and their corresponding 95% exact confidence intervals (thick gray lines), superimposed with the posterior estimates (blue stars) and their corresponding 95% credible intervals (thin black lines). Posterior estimates from the Gaussian copula based bivariate beta distribution for the `telomerase` data.

there is “shrinkage” towards the overall mean sensitivity and specificity as studies borrow strength from each other in the following manner: the posterior study-specific estimates depend on the global estimate and thus also on all other studies.

The mean sensitivity and specificity as estimated by the other four copula based bivariate beta distributions are in Table 1 and graphically shown in Figure 3. Though not presented here, the full code of the other four fitted copula based bivariate beta distributions is in

the replication code. Figure 1 shows satisfactory chain mixing with little autocorrelation apart from the “Clayton270” model. The Clayton copula is known to be unstable when the correlation parameter is close to the boundaries (-1 or 0) and this could be the reason why sampling from the posterior distribution was difficult.

For comparison purpose, the current recommended model; the BRMA, which uses normal marginals is also fitted to the data though it is not part of the **CopulaDTA** package. The model is first expressed in Stan modeling language in the code below and is stored within the R environment as character string named BRMA1.

```
R> BRMA1 <- "
+ data {
+   int<lower = 0> Ns;
+   int<lower = 0> tp[Ns];
+   int<lower = 0> dis[Ns];
+   int<lower = 0> tn[Ns];
+   int<lower = 0> nondis[Ns];
+ }
+ parameters {
+   real etarho;
+   vector[2] mul;
+   vector<lower = 0>[2] sigma;
+   vector[2] logitp[Ns];
+   vector[2] logitphat[Ns];
+ }
+ transformed parameters {
+   vector[Ns] p[2];
+   vector[Ns] phat[2];
+   vector[2] MU;
+   vector[2] mu;
+   real rho;
+   real ktau;
+   matrix[2, 2] Sigma;
+   rho = tanh(etarho);
+   ktau = (2 / pi()) * asin(rho);
+   for (a in 1:2) {
+     for (b in 1:Ns) {
+       p[a][b] = inv_logit(logitp[b][a]);
+       phat[a][b] = inv_logit(logitphat[b][a]);
+     }
+     mu[a] = inv_logit(mul[a]);
+   }
+   MU[1] = mean(phat[1]);
+   MU[2] = mean(phat[2]);
+   Sigma[1, 1] = sigma[1]^2;
+   Sigma[1, 2] = sigma[1] * sigma[2] * rho;
+   Sigma[2, 1] = sigma[1] * sigma[2] * rho;
+   Sigma[2, 2] = sigma[2]^2;
```



```

+ }
+ model {
+   etarho ~ normal(0, 10);
+   mul ~ normal(0, 10);
+   sigma ~ cauchy(0, 2.5);
+   for (i in 1:Ns) {
+     logitp[i] ~ multi_normal(mul, Sigma);
+     logitphat[i] ~ multi_normal(mul, Sigma);
+   }
+   tp ~ binomial(dis, p[1]);
+   tn ~ binomial(nondis, p[2]);
+ }
+ generated quantities {
+   vector[2*Ns] loglik;
+   for (i in 1:Ns) {
+     loglik[i] = binomial_lpmf(tp[i] | dis[i], p[1][i]);
+   }
+   for (i in (Ns+1):(2*Ns)) {
+     loglik[i] = binomial_lpmf(tn[i-Ns] | nondis[i-Ns], p[2][i-Ns]);
+   }
+ }"

```

Next, prepare the data by creating as list as follows

```

R> datalist <- list(tp = telomerase$TP, dis = telomerase$TP + telomerase$FN,
+   tn = telomerase$TN, nondis = telomerase$TN + telomerase$FP, Ns = 10)

```

In the `data` block the dimensions and names of variables in the dataset are specified, here `Ns` indicates the number of studies in the dataset. The `parameters` block introduces the unknown parameters to be estimated. These are `etarho`, a scalar representing the Fisher's transformed form of the association parameter ρ ; `mul`, a 2×1 vector representing the mean of sensitivity and specificity on the logit scale for a central study where the random-effect is zero; `sigma`, a 2×1 vector representing the between-study standard deviation of sensitivity and specificity on the logit scale; `logitp`, a $Ns \times 2$ array of study-specific sensitivity in the first column and specificity in the second column on logit scale, and `logitphat`, a $Ns \times 2$ array of predicted sensitivity in the first column and predicted specificity in the second column on logit scale.

The parameters are further transformed in the `transformed parameters` block. Here, `p` is a $2 \times Ns$ array of sensitivity in the first column and specificity in the second column after inverse logit transformation of `logitp`, and `phat` is a $2 \times Ns$ array of predicted sensitivity in the first column and predicted specificity in the second column after inverse logit transformation of `logitphat` to be used in computing the meta-analytic sensitivity and specificity. `mu` is a 2×1 vector representing the mean of sensitivity and specificity for a certain study with a random effect equal to 0, `MU` is a 2×1 vector containing the meta-analytic sensitivity and specificity; `Sigma`, a 2×2 matrix representing the variance-covariance matrix of sensitivity and specificity on the logit scale, `rho` and `ktau` are scalars representing the Pearson's and Kendall's tau correlation respectively. The prior distributions for all parameters and the

data likelihood are defined in the `model` block. Finally, in the `generated quantities` block, `loglik` is a $(2Ns) \times 1$ vector of the log likelihood needed to compute the WAIC.

Next, call the function `stan` from the `rstan` package to translate the code into C++, compile the code and draw samples from the posterior distribution as follows

```
R> brma.1 <- stan(model_code = BRMA1, data = datalist, chains = 3,
+   iter = 5000, warmup = 1000, thin = 10, seed = 3, cores = 3)
```

The parameter estimates are extracted and the chain convergence and autocorrelation examined further with the following code

```
R> print(brma.1, pars = c("MU", "mu", "rho"), digits = 2,
+   prob = c(0.025, 0.975))
```

The above lines of code print the following output

```
Inference for Stan model: 6ecef28f7aded3dde67c4048dc0a6b2a.
3 chains, each with iter=5000; warmup=1000; thin=10;
post-warmup draws per chain=400, total post-warmup draws=1200.
```

	mean	se_mean	sd	2.5%	50%	97.5%	n_eff	Rhat
MU[1]	0.75	0.00	0.05	0.63	0.75	0.83	610	1.00
MU[2]	0.79	0.00	0.11	0.52	0.80	0.96	779	1.00
mu[1]	0.76	0.00	0.04	0.69	0.77	0.84	584	1.01
mu[2]	0.90	0.00	0.08	0.69	0.92	0.98	808	1.00
rho	-0.96	0.01	0.12	-1.00	-1.00	-0.64	393	1.00

Samples were drawn using NUTS(diag_e) at Mon Oct 16 02:02:18 2017. For each parameter, `n_eff` is a crude measure of effective sample size, and `Rhat` is the potential scale reduction factor on split chains (at convergence, `Rhat=1`).

The meta-analytic sensitivity (`MU[1]`) and specificity (`MU[2]`) and 95% credible intervals are 0.75 [0.63, 0.83] and 0.79 [0.52, 0.96] respectively. This differs from what the authors published (0.75 [0.71, 0.79] and 0.86 [0.71, 0.94]) in two ways. The authors fitted the standard bivariate normal distribution to the logit transformed sensitivity and specificity values across the studies allowing for heterogeneity between the studies as expressed in Equation 6 and disregarded the higher level of the hierarchical model expressed in Equation 5. Because of this the authors had to use a continuity correction of 0.5 since the seventh study had “observed” specificity equal to 1, a problem not encountered in the hierarchical model. Secondly the authors do not report the meta-analytic values but rather report the mean sensitivity (`mu[1]`) and specificity (`mu[2]`) for a particular, hypothetical study with random-effect equal to zero, which in our case is 0.76 [0.69, 0.84] and 0.90 [0.69, 0.98] respectively and is comparable to what the authors reported. This discrepancy between `MU` and `mu` will indeed increase with increase in the between-study variability. Here, `MU[1]` and `mu[1]` are similar because the between-study variability in sensitivity in the logit scale (`Sigma[1, 1]`) is small (0.36 [0.06, 1.29]). `MU[2]` is substantially smaller than `mu[2]` as a result of the substantial between-study heterogeneity in specificity in the logit scale `Sigma[2, 2] = 5.93 [1.59, 17.04]`.

Model	Parameter	Mean	Lower	Median	Upper	n_{eff}	\hat{R}	WAIC
Gaussian	Sensitivity	0.75	0.69	0.75	0.81	237.43	1.01	87.75
	Specificity	0.80	0.65	0.80	0.91	67.03	1.04	
	Correlation	-0.85	-0.98	-0.91	-0.41	25.22	1.06	
	Var(Sens)	0.01	0.00	0.01	0.02	1350.39	1.00	
	Var(Spec)	0.05	0.01	0.05	0.11	52.01	1.04	
C90	Sensitivity	0.74	0.69	0.74	0.81	11.74	1.33	114.92
	Specificity	0.81	0.64	0.84	0.90	28.57	1.11	
	Correlation	-0.47	-0.98	-0.62	0.00	7.56	1.71	
	Var(Sens)	0.01	0.00	0.01	0.02	11.35	1.18	
	Var(Spec)	0.04	0.02	0.04	0.11	30.96	1.09	
C270	Sensitivity	0.78	0.71	0.79	0.83	6.25	1.42	101.06
	Specificity	0.81	0.68	0.80	0.89	22.13	1.10	
	Correlation	-0.33	-0.98	0.00	0.00	2.73	1.76	
	Var(Sens)	0.00	0.00	0.00	0.02	3.62	1.32	
	Var(Spec)	0.03	0.02	0.02	0.09	4.71	1.27	
FGM	Sensitivity	0.76	0.68	0.76	0.82	2559.33	1.00	95.91
	Specificity	0.80	0.64	0.81	0.91	2700.00	1.00	
	Correlation	-0.19	-0.22	-0.22	0.22	2586.52	1.00	
	Var(Sens)	0.01	0.00	0.00	0.02	2612.21	1.00	
	Var(Spec)	0.04	0.01	0.04	0.11	2700.00	1.00	
Frank	Sensitivity	0.76	0.68	0.76	0.81	2700.00	1.00	90.47
	Specificity	0.81	0.67	0.82	0.91	2696.39	1.00	
	Correlation	-0.71	-0.85	-0.69	1.00	2700.00	NaN	
	Var(Sens)	0.01	0.00	0.01	0.02	2537.78	1.00	
	Var(Spec)	0.04	0.01	0.04	0.10	2700.00	1.00	
BRMA	Sensitivity	0.75	0.63	0.75	0.83	610.49	1.00	87.56
	Specificity	0.79	0.52	0.80	0.96	779.13	1.00	
	Correlation	-0.88	-1.00	-0.94	-0.44	188.21	1.02	
	Var(Sens)	0.36	0.06	0.26	1.29	75.47	1.03	
	Var(Spec)	5.93	1.59	4.77	17.04	1047.82	1.00	

Table 1: The posterior mean, median, 95% credible interval, effective sample size (n_{eff}) and potential scale reduction factor \hat{R} for the marginal means, between-study variability and Kendall's tau parameters as estimated by the Gaussian, Clayton 90° (C90) and 270° (C270), Farlie-Gumbel-Morgenstern (FGM) and Frank copula based bivariate beta and bivariate normal (BRMA) distributions for the `telomerase` data.

5.1. Model comparison

Table 1 shows that the correlation as estimated by the BRMA model and the Gaussian copula bivariate beta are closer to -1 and equivalent compared to the estimates from the Frank, 90°- and 270°-Clayton copula. The correlation estimate is lowest for the FGM copula bivariate beta model. This is due to the constraints on the association parameter in the FGM copula such that $|\tau_k| \leq 2/9$.

In Figure 3, the marginal mean sensitivity and specificity from the five bivariate beta distribu-

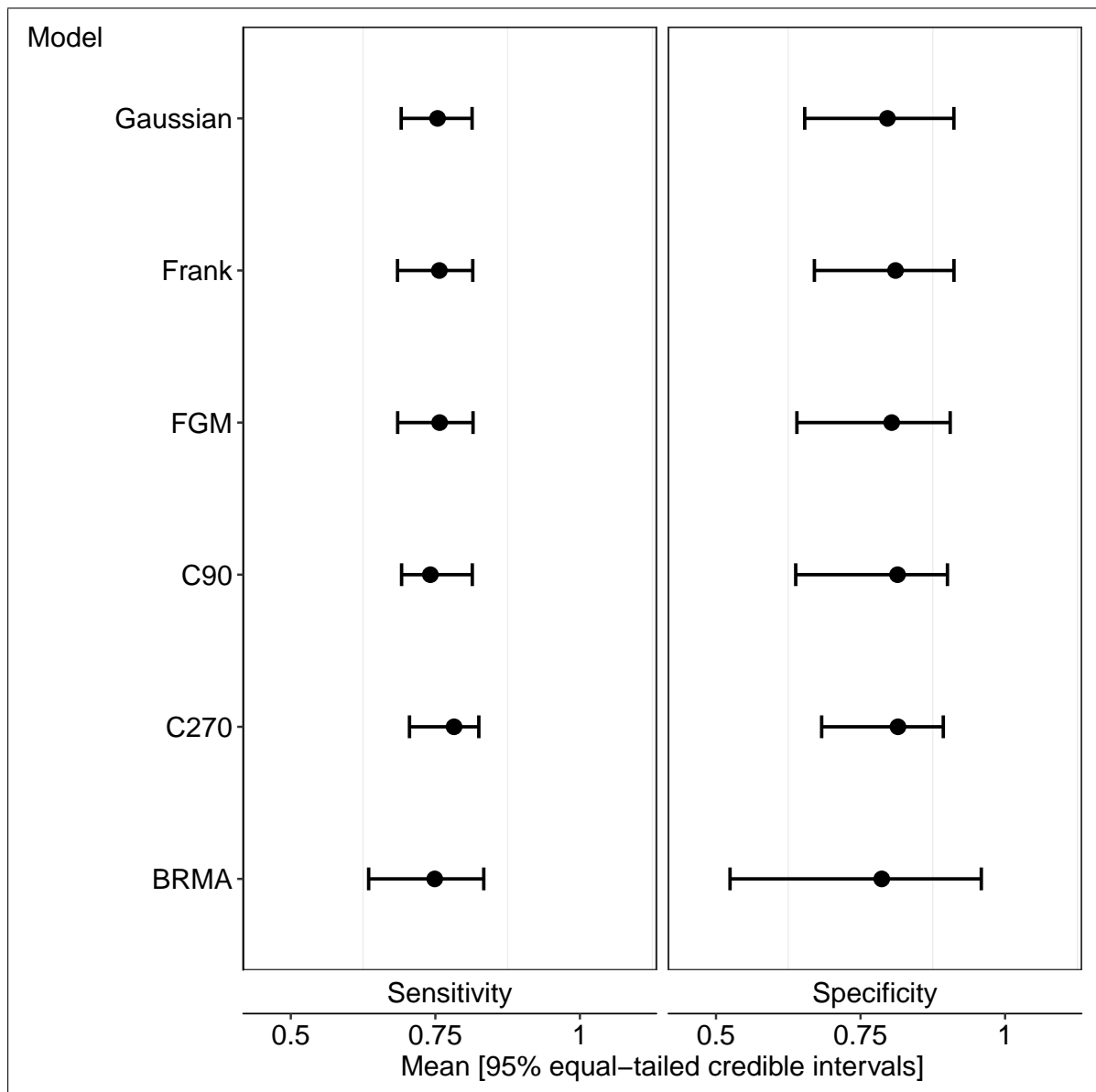


Figure 3: Plot of the posterior meta-analytic sensitivity (upper) and specificity (lower) and the corresponding 95% credible intervals) as estimated by the Gaussian, Clayton 90° (C90) and 270° (C270), Farlie-Gumbel-Morgenstern (FGM) and Frank copula based bivariate beta and bivariate normal (BRMA) distributions for the `telomerase` data.

tions are comparable with subtle differences in the 95% credible intervals despite differences in the correlation structure.

Glas *et al.* (2003) and Riley *et al.* (2007a) estimated the Pearson's correlation parameter in the BRMA model ρ as -1 within the frequentist framework. When there are only a few number of studies in the meta-analysis or when there is sparse data, the maximum likelihood estimation of the BRMA model often fails to converge or often estimates the Pearson's correlation as ± 1 (Riley *et al.* 2007b). Without convergence difficulties, the estimated Pearson's correlation

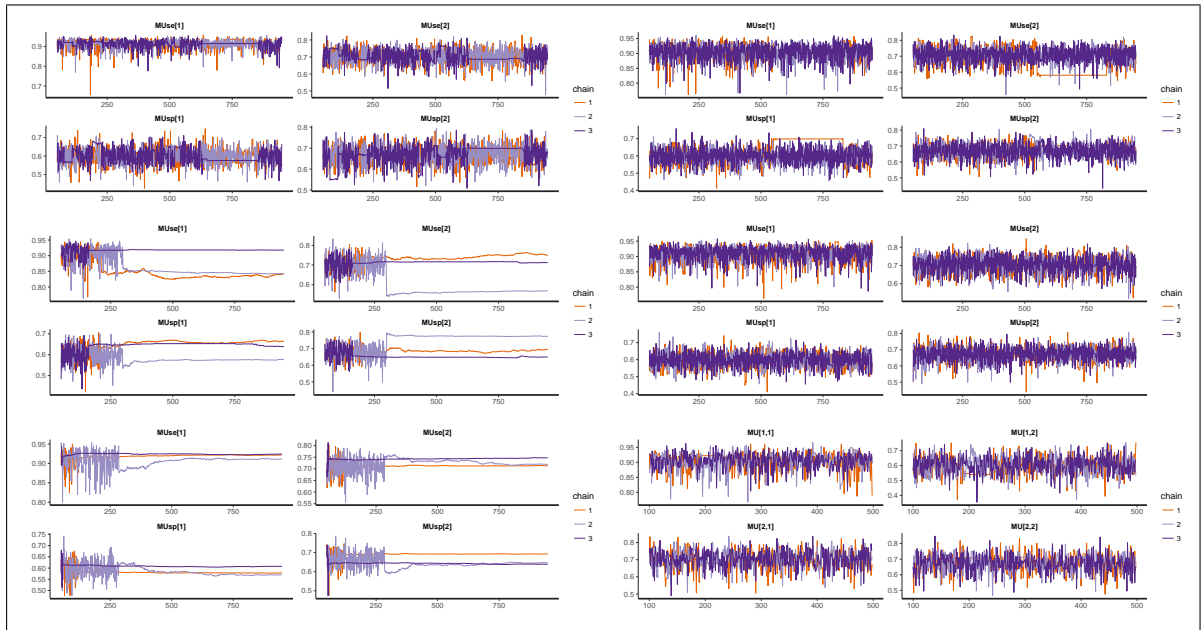


Figure 4: Trace plots of the posterior mean sensitivities and specificities for the `ascus` data as estimated by the Gaussian, Clayton 90° (C90) and 270° (C270), Farlie-Gumbel-Morgenstern (FGM) and Frank copula based bivariate beta and bivariate normal (BRMA) distributions.

from the BRMA model fitted within the Bayesian framework was $-0.96 [-1.00, -0.64]$. In general, Bayesian methods handle cases of small sample sizes better since the posterior is better behaved than the likelihood when suitable priors are defined.

Essentially, all six models are equivalent in the first level of hierarchy and differ in specifying the prior distributions for the “study-specific” sensitivity and specificity. As thus, the models should have the same number of parameters in which case it makes sense then to compare the log predictive densities. Upon inspection, the log predictive densities from the five copula-based models are practically equivalent (min = -38.83 , max = -37.82) but the effective number of parameters differed a bit (max = 5.95 , max = 18.63). The BRMA had 6 effective parameters but a lower log predictive density of -43.8 . Despite the lower predictive density, the last column of Table 1 indicates that the BRMA fitted the data best based on the WAIC.

6. Meta-regression

The `ascus` dataset has `Test` as a covariate. The covariate is used as it is of interest to study its effect on the joint distribution of sensitivity and specificity (including the correlation). The following code fits the FGM copula based bivariate beta-binomial distribution to the data.

```
R> fgm.2 <- cdtamodel(copula = "fgm",
+   modelargs = list(formula.se = StudyID ~ Test + 0))
R> fitfgm.2 <- fit(fgm.2, data = ascus, SID = "StudyID", iter = 19000,
+   warmup = 1000, thin = 20, seed = 3)
```

Figure 4 shows the trace plots for all six models fitted to the `ascus` data where all parameters,

Model	Test	Parameter	Mean	Lower	Median	Upper	n_{eff}	\hat{R}	WAIC
Gaussian	HC2	Correlation	-0.55	-0.97	-0.73	0.70	123.17	1.02	235.43
	Repc	Correlation	-0.87	-0.99	-0.92	-0.53	59.05	1.04	
C90	HC2	Correlation	-0.04	-0.58	0.00	0.00	56.03	1.04	230.86
	Repc	Correlation	-0.83	-0.98	-0.95	-0.51	2.20	1.87	
C270	HC2	Correlation	-0.02	-0.39	0.00	0.00	40.29	1.06	245.58
	Repc	Correlation	-0.79	-0.98	-0.88	-0.45	1.83	2.32	
FGM	HC2	Correlation	-0.06	-0.22	-0.22	0.22	37.80	1.08	246.63
	Repc	Correlation	-0.20	-0.22	-0.22	0.12	2671.83	1.00	
Frank	HC2	Correlation	-0.49	-0.83	-0.45	1.00	2700.00	NaN	238.56
	Repc	Correlation	-0.74	-0.86	-0.73	1.00	2700.00	NaN	
BRMA	Both	Correlation	-0.84	-0.99	-0.89	-0.42	45.67	1.06	233.15

Table 2: The posterior mean, median, 95% credible intervals, effective sample size, potential scale reduction factor of the correlation parameter(s) as estimated by the Gaussian, Clayton 90° (C90) and 270° (C270), Farlie-Gumbel-Morgenstern (FGM) and Frank copula based bivariate beta and bivariate normal (BRMA) distributions for the `ascus` dataset.

including the correlation parameter (except for BRMA) are modeled as a function of the covariate. There is proper chain mixing and convergence except for the case of the Clayton copula based bivariate beta. From the posterior relative sensitivity and specificity plotted in Figure 5, all the models that converged generally agree that repeat cytology was less sensitive than HC2 without significant loss in specificity.

The n_{eff} values in Table 2 indicate substantial autocorrelation in sampling the correlation parameters except for the Gaussian, FGM and Frank models. From the copula based bivariate beta distributions, it is apparent that the correlation between sensitivity and specificity in HC2 and repeat cytology is different.

The Clayton 90° model has the lowest WAIC even though sampling from the posterior distribution was difficult as seen in the trace plots in Figure 4 and the n_{eff} and \hat{R} values in Table 2. The difficulty in sampling from the posterior could be signaling over-parameterization of the correlation structure. It would thus be interesting to re-fit the models using only one correlation parameter and compare the models. WAIC is known to fail in certain settings and this examples shows that it is crucial to check the adequacy of the fit and plausibility of the model and not blindly rely on an information criterion to select the best fit to the data.

7. Discussion

Copula-based models offer great flexibility and ease but their use is not without caution. While the copulas used in this paper are attractive as they are mathematically tractable, Mikosch (2006) and Genest and Rémillard (2006) noted that it might be difficult to estimate copulas from data. Furthermore, the concepts behind copula models is slightly more complex and therefore require statistical expertise to understand and program them as they are not yet available as standard procedures/programs in statistical software.

In this paper, several advanced statistical models for meta-analysis of diagnostic accuracy studies were briefly discussed. The use of the R package **CopulaDTA** based on the flexible

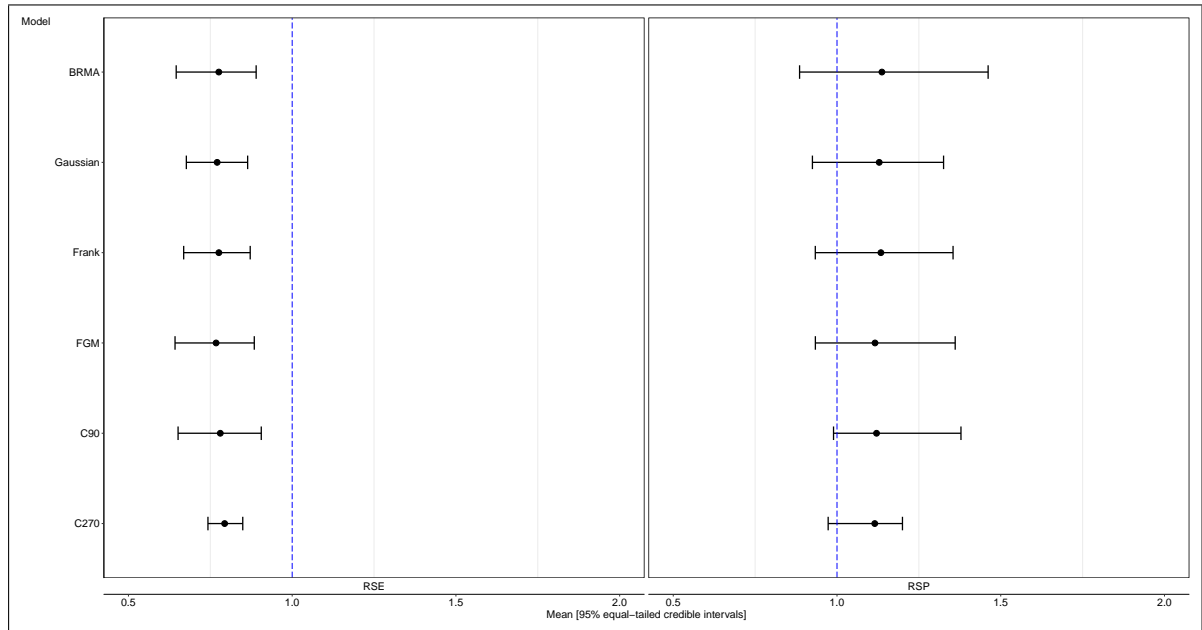


Figure 5: Pooled relative sensitivity (on top) and relative specificity (bottom) of repeat cytology (posterior mean and 95% credible intervals) compared to HPV testing with HC2 to detect cervical precancer in women with an atypical Pap smear estimated.

Stan interface was demonstrated and shows how complex models can be implemented in a convenient way.

In most practical situations, the marginal mean structure is of primary interest and the correlation structure is treated as nuisance making the choice of copula less critical. Nonetheless, an appropriate correlation structure is critical in the interpretation of the random variation in the data as well as obtaining valid model-based inference for the mean structure.

When the model for the mean is correct but the true distribution is misspecified, the estimates of the model parameters will be consistent but the standard errors will be incorrect (Agresti 2003). Nonetheless, the bivariate beta distribution has the advantage to allow direct joint modeling of sensitivity and specificity, without the need of any transformation, and consequently providing estimates with the appropriate meta-analytic interpretation but with the disadvantage of being more computationally intensive for some of the copula functions.

Leeflang *et al.* (2013) showed that the sensitivity and specificity often vary with disease prevalence. The models presented above can easily be extended and implemented to jointly model prevalence, sensitivity and specificity using tri-variate copulas.

There were some differences between the models in estimating the meta-analytic sensitivity and specificity and the correlation. Therefore, further research is necessary to investigate the effect of certain parameters, such as the number of studies, sample sizes and misspecification of the joint distribution on the meta-analytic estimates.

8. Conclusion

The proposed Bayesian joint model using copulas to construct bivariate beta distributions,

provides estimates with both the appropriate marginal as well as conditional interpretation, as opposed to the typical BRMA model which estimates sensitivity and specificity for specific studies with a particular value for the random-effects. Furthermore, the models do not have estimation difficulties with small sample sizes or large between-study variance because: (i) the between-study variances are not constant but depend on the underlying means and (ii) Bayesian methods are less influenced by small samples sizes.

The fitted models generally agree that the mean specificity was slightly lower than what [Glas et al. \(2003\)](#) reported and based on this we conclude that telomerase was not sensitive and specific enough to be recommended for daily use.

In the ASCUS triage data, conclusion based on the fitted models is in line with what the authors conclude: that HC2 was considerably more sensitive but slightly and non-significantly less specific than repeat cytology to triage women with an equivocal Pap smear to diagnose cervical precancer.

While the BRMA model had the lowest WAIC for both datasets, we still recommend modeling of sensitivity and specificity using bivariate beta distributions as they easily and directly provide meta-analytic estimates.

Acknowledgments

The authors declare that they have no competing interests.

M. Arbyn designed the OPSADAC project (Optimization of statistical procedures to assess the diagnostic accuracy of cervical cancer screening tests) of which this study is a part of. Victoria and M. Aerts conceptualized and initiated the study. Victoria wrote the code, analyzed the data and drafted the manuscript. M. Arbyn and M. Aerts edited the manuscript. All authors reviewed and approved the final manuscript.

V. Nyaga received financial support from the Scientific Institute of Public Health (Brussels) through the OPSADAC project. M. Arbyn was supported by the COHEAHR project funded by the 7th Framework Programme of the European Commission (grant No 603019) and the Belgian cancer center. M. Aerts was supported by the IAP research network nr P7/06 of the Belgian Government (Belgian Science Policy).

We are grateful to the editor and the anonymous reviewer for the valuable comments that helped to considerably improve the paper.

References

- Agresti A (2003). *Introduction to Generalized Linear Models*, chapter 4, pp. 115–164. 2nd edition. John Wiley & Sons, Hoboken. doi:10.1002/0471249688.ch4.
- Arbyn M, Roelens J, Simoens C, Buntinx F, Paraskevaidis E, Martin-Hirsch PPL, Prendiville WJ (2013). “Human Papillomavirus Testing Versus Repeat Cytology for Triage of Minor Cytological Cervical Lesions.” *The Cochrane Database of Systematic Reviews*, **3**(CD008054). doi:10.1002/14651858.cd008054.pub2.
- Arbyn M, Ronco G, Anttila A, Meijer CC, Poljak M, Ogilvie G, Koliopoulos G, Naucler P, Sankaranarayanan R, Peto J (2012). “Evidence Regarding Human Papillomavirus Testing

- in Secondary Prevention of Cervical Cancer.” *Vaccine*, **30**(Suppl 5), F88–F99. doi:10.1016/j.vaccine.2012.06.095.
- Arends LR, Hamza TH, van Houwelingen JC, Heijnenbrok-Kal MH, Hunink MGM, Stijnen T (2008). “Bivariate Random Effects Meta-Analysis of ROC Curves.” *Medical Decision Making*, **28**(5), 621–638. doi:10.1177/0272989x08319957.
- Carpenter B, Gelman A, Hoffman MD, Lee D, Goodrich B, Betancourt M, Brubaker M, Guo J, Li P, Riddell A (2017). “Stan: A Probabilistic Programming Language.” *Journal of Statistical Software*, **76**(1), 1–32. doi:10.18637/jss.v076.i01.
- Chu H, Cole SR (2006). “Bivariate Meta-Analysis of Sensitivity and Specificity with Sparse Data: A Generalized Linear Mixed Model Approach.” *Journal of Clinical Epidemiology*, **59**(12), 1331–1332. doi:10.1016/j.jclinepi.2006.06.011.
- Clayton DG (1978). “A Model for Association in Bivariate Life Tables and Its Application in Epidemiological Studies of Familial Tendency in Chronic Disease Incidence.” *Biometrika*, **65**(1), 141–151.
- Cong X, Cox DD, Cantor SB (2007). “Bayesian Meta-Analysis of Papanicolaou Smear Accuracy.” *Gynecologic Oncology*, **107**(1), S133–S137. doi:10.1016/j.ygyno.2007.08.080.
- Farlie DJG (1960). “The Performance of Some Correlation Coefficients for a General Bivariate Distribution.” *Biometrika*, **47**(3/4), 307–323. doi:10.2307/2333302.
- Frank MJ (1979). “On the Simultaneous Associativity of $F(X, y)$ and $x + y - F(X, y)$.” *Aequationes Mathematicae*, **19**(1), 194–226. doi:10.1007/bf02189866.
- Genest C, Rémillard B (2006). “Discussion of “Copulas: Tales and Facts”, by Thomas Mikosch.” *Extremes*, **9**(1), 27–36. doi:10.1007/s10687-006-0018-7.
- Glas AS, Roos D, Deutekom M, Zwinderman AH, Bossuyt PMM, Kurth KH (2003). “Tumor Markers in the Diagnosis of Primary Bladder Cancer. A Systematic Review.” *The Journal of Urology*, **169**(6), 1975–1982. doi:10.1097/01.ju.0000067461.30468.6d.
- Gumbel EJ (1960). “Bivariate Exponential Distributions.” *Journal of the American Statistical Association*, **55**(292), 698–707. doi:10.1080/01621459.1960.10483368.
- Hoffman MD, Gelman A (2014). “The No-U-Turn Sampler: Adaptively Setting Path Lengths in Hamiltonian Monte Carlo.” *Journal of Machine Learning Research*, **15**(4), 1593–1623.
- Leeftang MMG, Rutjes AWS, Reitsma JB, Hooft L, Bossuyt PMM (2013). “Variation of a Test’s Sensitivity and Specificity with Disease Prevalence.” *Canadian Medical Association Journal*, **185**(11), E537–E544. doi:10.1503/cmaj.121286.
- Libby DL, Novick MR (1982). “Multivariate Generalized Beta Distributions with Applications to Utility Assessment.” *Journal of Educational and Behavioral Statistics*, **7**(4), 271–294. doi:10.3102/10769986007004271.
- Meyer C (2013). “The Bivariate Normal Copula.” *Communications in Statistics – Theory and Methods*, **42**(13), 2402–2422. doi:10.1080/03610926.2011.611316.

- Mikosch T (2006). “Copulas: Tales and Facts – Rejoinder.” *Extremes*, **9**(1), 55–62. doi:[10.1007/s10687-006-0024-9](https://doi.org/10.1007/s10687-006-0024-9).
- Molenberghs G, Verbeke G (2005). “From Subject-Specific to Random-Effects Models.” In P Bickel, P Diggle, S Fienberg, U Gather, I Olkin, S Zeger (eds.), *Models for Discrete Longitudinal Data*, 1st edition, chapter 13, pp. 259–267. Springer-Verlag, NY, USA.
- Morgenstern D (1956). “Einfache Beispiele Zweidimensionaler Verteilungen.” *Mitteilungsblatt für Mathematische Statistik*, **8**(1), 234–235. doi:[10.1007/978-3-642-56018-7_4](https://doi.org/10.1007/978-3-642-56018-7_4).
- Moses LE, Shapiro D, Littenberg B (1993). “Combining Independent Studies of a Diagnostic Test into a Summary ROC Curve: Data-Analytic Approaches and Some Additional Considerations.” *Statistics in Medicine*, **12**(14), 1293–1316. doi:[10.1002/sim.4780121403](https://doi.org/10.1002/sim.4780121403).
- Nelsen RB (2006). *An Introduction to Copulas*. 2nd edition. Springer-Verlag. doi:[10.1007/0-387-28678-0](https://doi.org/10.1007/0-387-28678-0).
- Nikoloulopoulos AK (2015). “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**(29), 3842–3865. doi:[10.1002/sim.6595](https://doi.org/10.1002/sim.6595).
- Nyaga VN (2017). *CopulaDTA: Copula Based Bivariate Beta-Binomial Model for Diagnostic Test Accuracy Studies*. R package version 1.0.0, URL <https://CRAN.R-project.org/package=CopulaDTA>.
- Olkin I, Liu R (2003). “A Bivariate Beta Distribution.” *Statistics & Probability Letters*, **62**(4), 407–412. doi:[10.1016/s0167-7152\(03\)00048-8](https://doi.org/10.1016/s0167-7152(03)00048-8).
- Patton AJ (2006). “Modelling Asymmetric Exchange Rate Dependence.” *International Economic Review*, **47**(2), 527–556. doi:[10.1111/j.1468-2354.2006.00387.x](https://doi.org/10.1111/j.1468-2354.2006.00387.x).
- Plummer M (2003). “JAGS: A Program for Analysis of Bayesian Graphical Models Using Gibbs Sampling.” In K Hornik, F Leisch, A Zeileis (eds.), *Proceedings of the 3rd International Workshop on Distributed Statistical Computing*.
- Plummer M (2008). “Penalized Loss Functions for Bayesian Model Comparison.” *Biostatistics*, **9**(3), 523–539. doi:[10.1093/biostatistics/kxm049](https://doi.org/10.1093/biostatistics/kxm049).
- R Core Team (2017). *R : A Language and Environment for Statistical Computing*. Vienna, Austria. URL <https://www.R-project.org/>.
- Reitsma JB, Glas AS, Rutjes AW, Scholten RJ, Bossuyt PM, Zwinderman AH (2005). “Bivariate Analysis of Sensitivity and Specificity Produces Informative Summary Measures in Diagnostic Reviews.” *Journal of Clinical Epidemiology*, **58**(10), 982–990. doi:[10.1016/j.jclinepi.2005.02.022](https://doi.org/10.1016/j.jclinepi.2005.02.022).
- Riley RD (2009). “Multivariate Meta-Analysis: The Effect of Ignoring Within-Study Correlation.” *Journal of the Royal Statistical Society A*, **172**(4), 789–811. doi:[10.1111/j.1467-985x.2008.00593.x](https://doi.org/10.1111/j.1467-985x.2008.00593.x).
- Riley RD, Abrams KR, Lambert PC, Sutton AJ, Thompson JR (2007a). “An Evaluation of Bivariate Random-Effects Meta-Analysis for the Joint Synthesis of Two Correlated Outcomes.” *Statistics in Medicine*, **26**(1), 78–97. doi:[10.1002/sim.2524](https://doi.org/10.1002/sim.2524).

- Riley RD, Abrams KR, Sutton AJ, Lambert PC, Thompson JR (2007b). “Bivariate Random-Effects Meta-Analysis and the Estimation of Between-Study Correlation.” *BMC Medical Research Methodology*, **7**, 3. doi:10.1186/1471-2288-7-3.
- Rutter CM, Gatsonis CA (2001). “A Hierarchical Regression Approach to Meta-Analysis of Diagnostic Test Accuracy Evaluations.” *Statistics in Medicine*, **20**(19), 2865–2884. doi:10.1002/sim.942.
- Sarmanov OV (1966). “Generalized Normal Correlation and Two-Dimensional Fréchet Classes.” *Doklady Akademii Nauk SSSR*, **168**(1), 596–599.
- SAS Institute Inc (2013). *The SAS System, Version 9.4*. SAS Institute Inc., Cary, NC. URL <http://www.sas.com/>.
- Sklar M (1959). “Fonctions De Répartition à n Dimensions et Leurs Marges.” *Publications de l’Institut de Statistique de l’Université de Paris*, **8**, 229–231.
- Stan Development Team (2017). “**rstan**: R Interface to Stan.” R package version 2.16.2, URL <http://mc-stan.org/>.
- Takwoingi Y, Deeks J (2010). *MetaDAS: A SAS Macro for Meta-Analysis of Diagnostic Accuracy Studies*. SAS macro version 1.3, URL http://methods.cochrane.org/sites/methods.cochrane.org.sdt/files/public/uploads/METADAS_v1.3_txt.txt.
- Takwoingi Y, Guo B, Riley RD, Deeks JJ (2017). “Performance of Methods for Meta-Analysis of Diagnostic Test Accuracy with Few Studies or Sparse Data.” *Statistical Methods in Medical Research*, **26**(4), 1896–1911. doi:10.1177/0962280215592269.
- Watanabe S (2010). “Asymptotic Equivalence of Bayes Cross Validation and Widely Applicable Information Criterion in Singular Learning Theory.” *Journal of Machine Learning Research*, **11**(12), 3571–3594.

Affiliation:

Victoria N. Nyaga, Marc Arbyn
Unit of Cancer Epidemiology
Scientific Institute of Public Health
Juliette Wytsmanstraat 14
1050 Brussels, Belgium
E-mail: victoria.nyawiranyaga@wiv-isp.be, marc.arbyn@wiv-isp.be

Marc Aerts
Center for Statistics
Hasselt University
Agoralaan building D
3590 Diepenbeek, Belgium
E-mail: marc.aerts@uhasselt.be