UNIVERSITÄT
BIELEFELD
Medical School OWL

MED
Bio²
Biostatistics &
Medical Biometry

Meta-Analysis of Diagnostic Test Accuracy Studies: Statistical Perspectives and Practical Implications

Annika Hoyer[1]
Joined work with: Oliver Kuss[2] and Ferdinand V. Stoye[1]

Symposium Recent Advances in Meta-Analysis

28.05.2026

[1] Biostatistics and Medical Biometry, Medical School OWL, Bielefeld University
[2] Institute for Biometrics and Epidemiology, German Diabetes Center, Düsseldorf

Overview

Motivation

Meta-analysis of diagnostic studies

Meta-analysis of ROC curves

Applications

Outlook and Implications

Motivation

Example Type 2 Diabetes

Example Type 2 Diabetes

- Prevalence of diagnosed type 2 diabetes in Germany: 8%

Example Type 2 Diabetes

- Prevalence of diagnosed type 2 diabetes in Germany: 8%

→ Increased mortality compared to people without diabetes, late complications

Example Type 2 Diabetes

- Prevalence of diagnosed type 2 diabetes in Germany: 8%

→ Increased mortality compared to people without diabetes, late complications

- Prevalence of undiagnosed type 2 diabetes in Germany

Example Type 2 Diabetes

- Prevalence of diagnosed type 2 diabetes in Germany: 8%

→ Increased mortality compared to people without diabetes, late complications

- Prevalence of undiagnosed type 2 diabetes in Germany

→ Increased mortality compared to people without diabetes, late complications

Example Type 2 Diabetes

- Prevalence of diagnosed type 2 diabetes in Germany: 8%

→ Increased mortality compared to people without diabetes, late complications

- Prevalence of undiagnosed type 2 diabetes in Germany

→ Increased mortality compared to people without diabetes, late complications

→ BUT: Due to missing diagnosis no chance for treatment

Diagnostic tests

Diagnostic tests

- Enable the earliest possible diagnosis (e.g. by screening) or

Diagnostic tests

- Enable the earliest possible diagnosis (e.g. by screening) or
- Confirmation of a suspected disease (confirmation tests)

Diagnostic tests

- Enable the earliest possible diagnosis (e.g. by screening) or
- Confirmation of a suspected disease (confirmation tests)

→ Provide the basis for efficient treatment

Diagnostic tests

Diagnostic tests

1. Determination of a suitable diagnostic test

Diagnostic tests

1. Determination of a suitable diagnostic test

2. Determination of optimal thresholds

Diagnostic test accuracy studies

	Reference standard	
Index Test	diseased	non-diseased
	D^+	D^-
positive T^+	TP	FP
negative T^-	FN	TN

Diagnostic test accuracy studies

Index Test	**Reference standard** diseased D^+	 non-diseased D^-
positive T^+	TP	FP
negative T^-	FN	TN

- Sensitivity: $Se = P(T^+ \mid D^+) = \frac{TP}{TP+FN}$

Diagnostic test accuracy studies

	Reference standard	
Index Test	diseased D^+	non-diseased D^-
positive T^+	TP	FP
negative T^-	FN	TN

- Sensitivity: $Se = P(T^+ \mid D^+) = \frac{TP}{TP+FN}$
- Specificity: $Sp = P(T^- \mid D^-) = \frac{TN}{FP+TN}$

Diagnostic test accuracy studies

HbA1c	**OGTT** Type 2 diabetes D^+	 No type 2 diabetes D^-
positive T^+	65	82
negative T^-	50	1602

Diagnostic test accuracy studies

	OGTT	
HbA1c	Type 2 diabetes D^+	No type 2 diabetes D^-
positive T^+	65	82
negative T^-	50	1602

- Sensitivity: $Se = \frac{65}{65+50} \approx 56\%$

Diagnostic test accuracy studies

	OGTT	
HbA1c	Type 2 diabetes D^+	No type 2 diabetes D^-
positive T^+	65	82
negative T^-	50	1602

- Sensitivity: $Se = \frac{65}{65+50} \approx 56\%$
- Specificity: $Sp = \frac{1602}{1602+82} \approx 95\%$

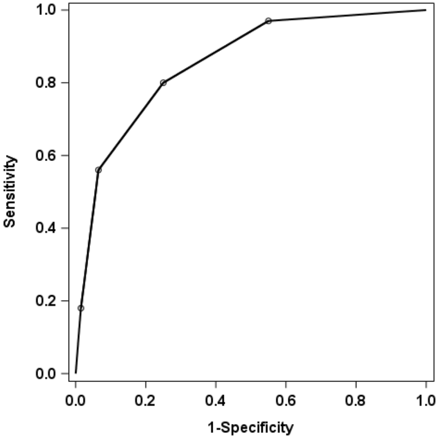
Common situation in clinical practice

Class	HbA1c (in %)	Type 2 diabetes	No type 2 diabetes
1	<5.5	150	900
2	5.5 to <6.0	850	600
3	6.0 to <6.5	1200	370
4	6.5 to <7.0	1900	100
5	≥ 7.0	900	30
	Sum	5000	2000

Common situation in clinical practice

Threshold	Sensitivity	Specificity
5.5	97%	45%
6.0	80%	75%
6.5	56%	94%
7.0	18%	99%

Receiver Operating Characteristic Curve



What is the best threshold?

Threshold	Sensitivity	Specificity	Youden-index
5.5	97%	45%	42
6.0	80%	75%	55
6.5	56%	94%	50
7.0	18%	99%	17

What is the best threshold?

Threshold	Sensitivity	Specificity	Youden-index
5.5	97%	45%	42
6.0	80%	75%	55
6.5	56%	94%	50
7.0	18%	99%	17

- Maximal Youden-index at 6.0

What is the best threshold?

Threshold	Sensitivity	Specificity	Youden-index
5.5	97%	45%	42
6.0	80%	75%	55
6.5	56%	94%	50
7.0	18%	99%	17

- Maximal Youden-index at 6.0
- Limitation: balancing sensitivity and specificity (due to equal weights)

What is the best threshold?

Threshold	Sensitivity	Specificity	Youden-index
5.5	97%	45%	42
6.0	80%	75%	55
6.5	56%	94%	50
7.0	18%	99%	17

- Maximal Youden-index at 6.0
- Limitation: balancing sensitivity and specificity (due to equal weights)
- Screening: higher sensitivity

What is the best threshold?

Threshold	Sensitivity	Specificity	Youden-index
5.5	97%	45%	42
6.0	80%	75%	55
6.5	56%	94%	50
7.0	18%	99%	17

- Maximal Youden-index at 6.0
- Limitation: balancing sensitivity and specificity (due to equal weights)
- Screening: higher sensitivity
- Confirmation: higher specificity

What is the best threshold?

Threshold	Sensitivity	Specificity	Youden-index
5.5	97%	45%	42
6.0	80%	75%	55
6.5	56%	94%	50
7.0	18%	99%	17

- Maximal Youden-index at 6.0
- Limitation: balancing sensitivity and specificity (due to equal weights)
- Screening: higher sensitivity
- Confirmation: higher specificity

→ weighted Youden-index

Meta-analysis of diagnostic studies

Meta-analysis of diagnostic studies

Meta-analysis of diagnostic studies

Study	Sensitivity	Specificity
1	65/115	1602/1684
2	181/252	3877/5865
...	...	...

Meta-analysis of diagnostic studies

Study	Sensitivity	Specificity
1	65/115	1602/1684
2	181/252	3877/5865
...	...	...

Se_{MA}=... Sp_{MA}=...

negatively correlated

Statistical models

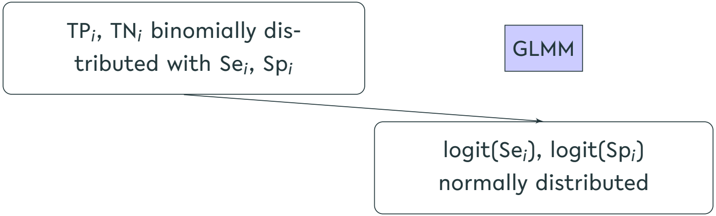
TP_i, TN_i binomially distributed with Se_i, Sp_i

Statistical models

TP_i, TN_i binomially distributed with Se_i, Sp_i

GLMM

Statistical models

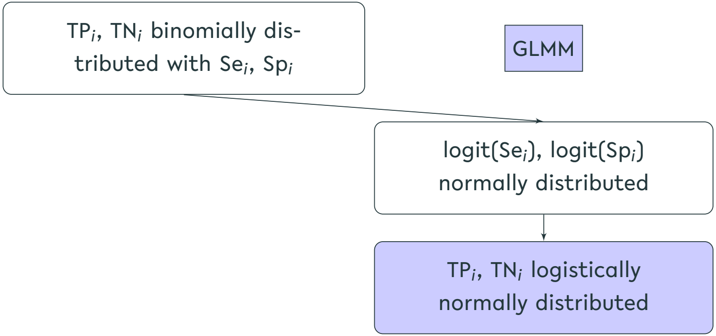


TP_i, TN_i binomially distributed with Se_i, Sp_i

GLMM

logit(Se_i), logit(Sp_i) normally distributed

Statistical models



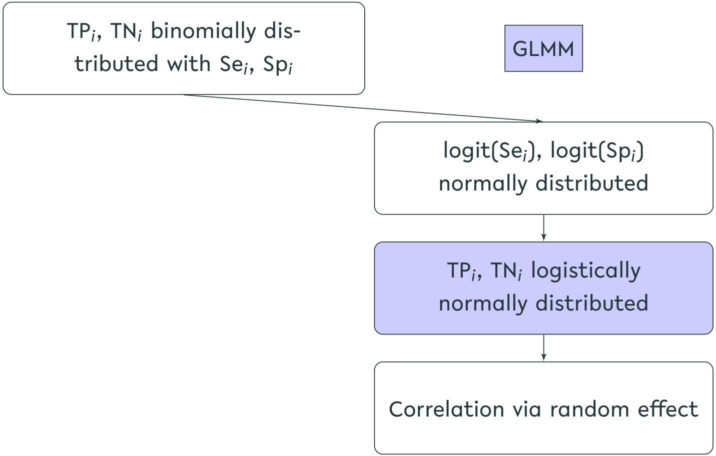
TP_i, TN_i binomially distributed with Se_i, Sp_i

GLMM

logit(Se_i), logit(Sp_i) normally distributed

TP_i, TN_i logistically normally distributed

Statistical models



TP_i, TN_i binomially distributed with Se_i, Sp_i

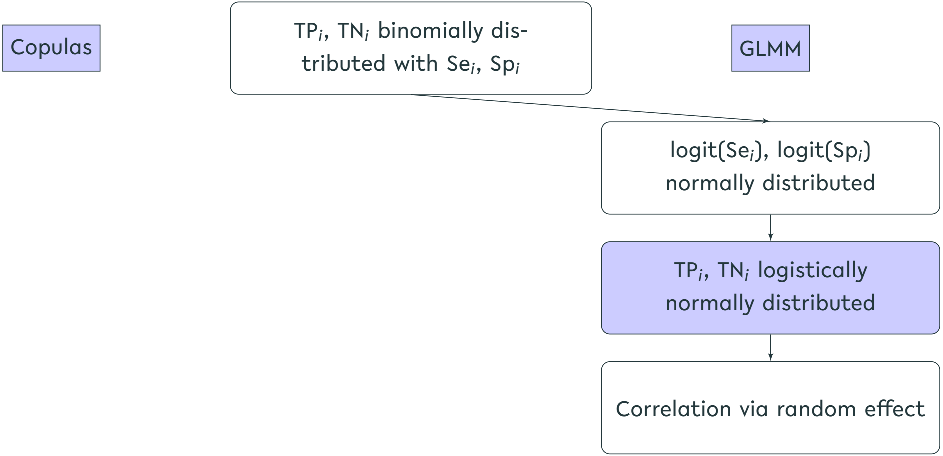
GLMM

logit(Se_i), logit(Sp_i) normally distributed

TP_i, TN_i logistically normally distributed

Correlation via random effect

Statistical models



Copulas

TP_i, TN_i binomially distributed with Se_i, Sp_i

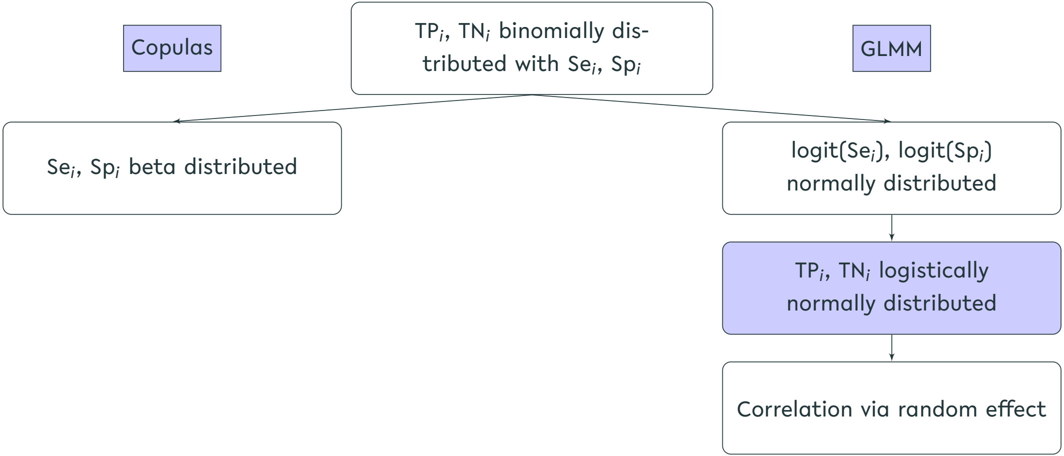
GLMM

logit(Se_i), logit(Sp_i) normally distributed

TP_i, TN_i logistically normally distributed

Correlation via random effect

Statistical models



Copulas

TP$_i$, TN$_i$ binomially distributed with Se$_i$, Sp$_i$

GLMM

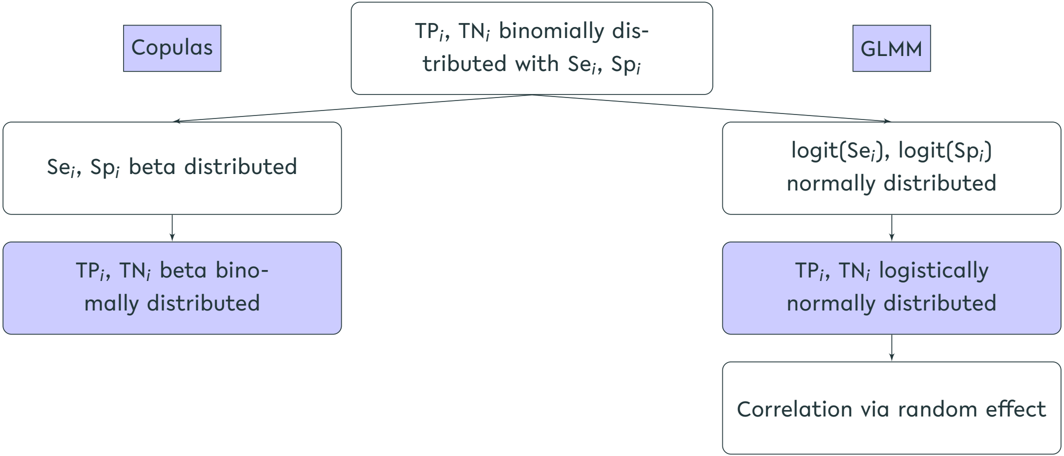
Se$_i$, Sp$_i$ beta distributed

logit(Se$_i$), logit(Sp$_i$) normally distributed

TP$_i$, TN$_i$ logistically normally distributed

Correlation via random effect

Statistical models



Copulas

TP_i, TN_i binomially distributed with Se_i, Sp_i

GLMM

Se_i, Sp_i beta distributed

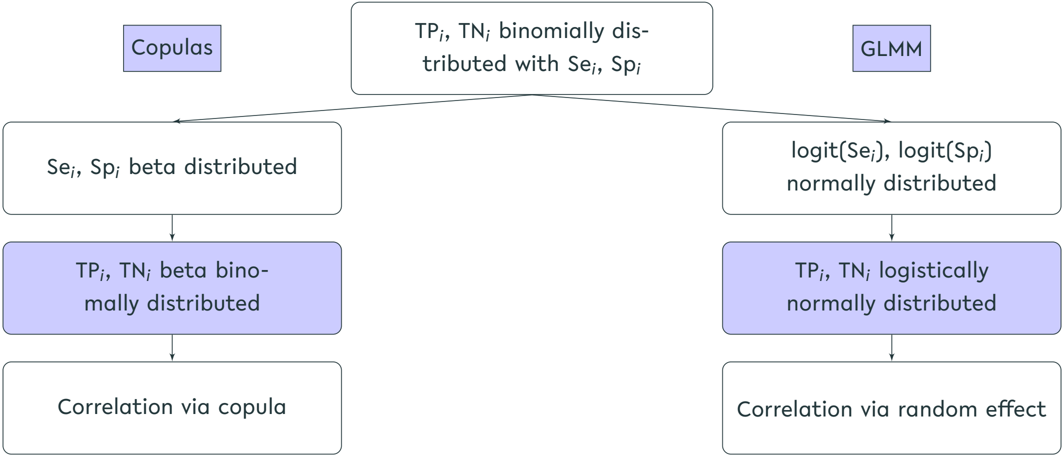
TP_i, TN_i beta binomally distributed

logit(Se_i), logit(Sp_i) normally distributed

TP_i, TN_i logistically normally distributed

Correlation via random effect

Statistical models



TP_i, TN_i binomially distributed with Se_i, Sp_i

Copulas

Se_i, Sp_i beta distributed

TP_i, TN_i beta binomally distributed

Correlation via copula

GLMM

logit(Se_i), logit(Sp_i) normally distributed

TP_i, TN_i logistically normally distributed

Correlation via random effect

Copula model: An overview

Copula model: An overview

- Distribution assumption:

$$TP_i \mid Se_i \sim Binomial(TP_i + FN_i, Se_i)$$

$$TN_i \mid Sp_i \sim Binomial(TN_i + FP_i, Sp_i)$$

Copula model: An overview

- Distribution assumption:

$$TP_i \mid Se_i \sim Binomial(TP_i + FN_i, Se_i)$$

$$TN_i \mid Sp_i \sim Binomial(TN_i + FP_i, Sp_i)$$

with

$$Se_i \sim Beta(\alpha_{Se}, \beta_{Se})$$

$$Sp_i \sim Beta(\alpha_{Sp}, \beta_{Sp})$$

Copula model: An overview

- Distribution assumption:

$$TP_i \mid Se_i \sim Binomial(TP_i + FN_i, Se_i)$$

$$TN_i \mid Sp_i \sim Binomial(TN_i + FP_i, Sp_i)$$

with

$$Se_i \sim Beta(\alpha_{Se}, \beta_{Se})$$

$$Sp_i \sim Beta(\alpha_{Sp}, \beta_{Sp})$$

→ TP_i and TN_i are beta binomially distributed

Copula model: An overview

- Distribution assumption:

$$TP_i \mid Se_i \sim Binomial(TP_i + FN_i, Se_i)$$

$$TN_i \mid Sp_i \sim Binomial(TN_i + FP_i, Sp_i)$$

with

$$Se_i \sim Beta(\alpha_{Se}, \beta_{Se})$$

$$Sp_i \sim Beta(\alpha_{Sp}, \beta_{Sp})$$

→ TP_i and TN_i are beta binomially distributed

- Modelling of dependencies using copulas

Copula model: An overview

Copula model: An overview

- Copula: Multivariate cumulative distribution function, constructed depending on univariate marginal distributions

Copula model: An overview

- Copula: Multivariate cumulative distribution function, constructed depending on univariate marginal distributions

→ Use beta binomial distributions of TP_i and TN_i as marginal distributions and construct bivariate joint distribution via copula

Copula model: An overview

- Copula: Multivariate cumulative distribution function, constructed depending on univariate marginal distributions

→ Use beta binomial distributions of TP_i and TN_i as marginal distributions and construct bivariate joint distribution via copula

- Sklar's Theorem: $F(x_1, x_2) = C(F_{Se}(x_1), F_{Sp}(x_2))$

Copula model: An overview

- Copula: Multivariate cumulative distribution function, constructed depending on univariate marginal distributions

→ Use beta binomial distributions of TP_i and TN_i as marginal distributions and construct bivariate joint distribution via copula

- Sklar's Theorem: $F(x_1, x_2) = C(F_{Se}(x_1), F_{Sp}(x_2))$
- Likelihood: Product of likelihoods of the copula and the marginal distributions

Example

Example

- Telomerase as tumor marker for bladder cancer (Glas et al. 2003)

Example

- Telomerase as tumor marker for bladder cancer (Glas et al. 2003)
- 10 studies included

Example

- Telomerase as tumor marker for bladder cancer (Glas et al. 2003)
- 10 studies included

Method	Sensitivity	Specificty
GLMM	76,7% [68,8%; 83,1%]	90,4% [71,5%; 97,2%]

Example

- Telomerase as tumor marker for bladder cancer (Glas et al. 2003)
- 10 studies included

Method	Sensitivity	Specificty
GLMM	76,7% [68,8%; 83,1%]	90,4% [71,5%; 97,2%]
Gaussian-Copula	79,4% [74,8%; 83,4%]	81,9% [65,7%; 91,4%]

Example

- Telomerase as tumor marker for bladder cancer (Glas et al. 2003)
- 10 studies included

Method	Sensitivity	Specificty
GLMM	76,7% [68,8%; 83,1%]	90,4% [71,5%; 97,2%]
Gaussian-Copula	79,4% [74,8%; 83,4%]	81,9% [65,7%; 91,4%]
Plackett-Copula	77,6% [72,2%; 82,2%]	86,5% [70,0%; 94,6%]

Extension options

Extension options

- Accounting for disease prevalence as third random variable

Extension options

- Accounting for disease prevalence as third random variable

→ trivariate model (doi: 10.1002/sim.6463)

Extension options

- Accounting for disease prevalence as third random variable

→ trivariate model (doi: 10.1002/sim.6463)

- Comparison of two diagnostic tests

Extension options

- Accounting for disease prevalence as third random variable

→ trivariate model (doi: 10.1002/sim.6463)

- Comparison of two diagnostic tests

→ quadrivariate Model (doi: 10.1002/sim.7556, doi: 10.1177/0962280216661587)

Meta-analysis of ROC curves

Common situation

Study	Threshold	Sensitivity	Specificity
1	5.3	89/115	1382/1684
1	5.5	72/115	1558/1684
1	5.6	65/115	1602/1684
2	5.8	207/252	2704/5865
2	5.9	196/252	3308/5865
...	...	...	...

Meta-analysis of ROC curves

Meta-analysis of ROC curves

- Challenge: ROC curves from single studies with different pairs of sensitivities and specificities at multiple thresholds

Meta-analysis of ROC curves

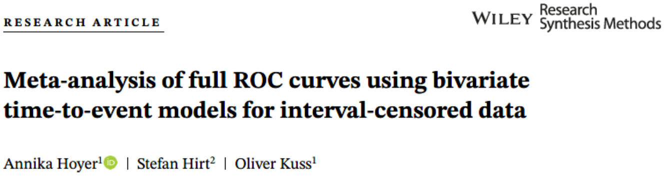
- Challenge: ROC curves from single studies with different pairs of sensitivities and specificities at multiple thresholds
- Additional challenge: Values and number of thresholds may vary between studies

Meta-analysis of ROC curves

- New approach motivated from survival analysis:

Meta-analysis of ROC curves

- New approach motivated from survival analysis:



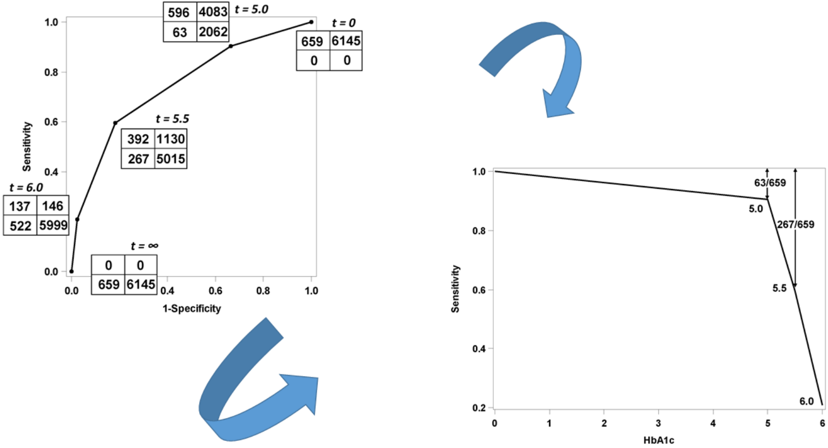
RESEARCH ARTICLE

WILEY Research Synthesis Methods

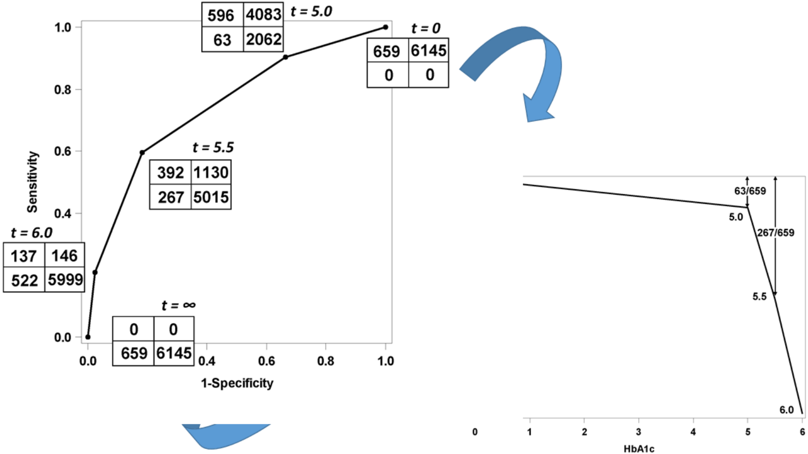
Meta-analysis of full ROC curves using bivariate time-to-event models for interval-censored data

Annika Hoyer[1] | Stefan Hirt[2] | Oliver Kuss[1]

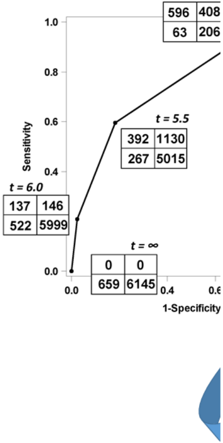
Bivariate time-to-event model in a nutshell

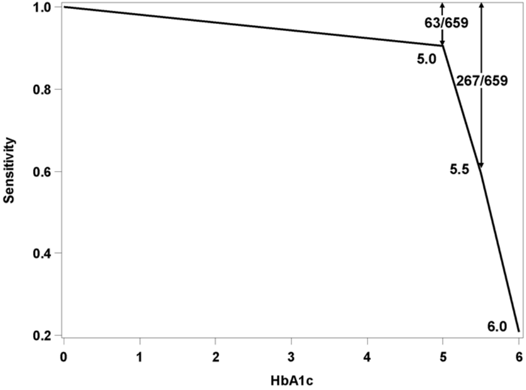


Bivariate time-to-event model in a nutshell



Bivariate time-to-event model in a nutshell





Bivariate time-to-event model in a nutshell

Bivariate time-to-event model in a nutshell

- Assumption of three different distributions for the diagnostic test values Y: Weibull, log-normal, log-logistic

Bivariate time-to-event model in a nutshell

- Assumption of three different distributions for the diagnostic test values Y: Weibull, log-normal, log-logistic
- Modeling of the log-transformed outcomes $\log(y_0)$ and $\log(y_1)$ (0 denotes non-diseased and 1 diseased study participants) simultaneously and connect both through a bivariate random effect:

Bivariate time-to-event model in a nutshell

- Assumption of three different distributions for the diagnostic test values Y: Weibull, log-normal, log-logistic
- Modeling of the log-transformed outcomes $\log(y_0)$ and $\log(y_1)$ (0 denotes non-diseased and 1 diseased study participants) simultaneously and connect both through a bivariate random effect:

$$\log(y_{0i}) = b_0 + \epsilon_0 + u_{0i},\ \log(y_{1i}) = b_1 + \epsilon_1 + u_{1i}, \tag{1}$$

Bivariate time-to-event model in a nutshell

- Assumption of three different distributions for the diagnostic test values Y: Weibull, log-normal, log-logistic
- Modeling of the log-transformed outcomes $\log(y_0)$ and $\log(y_1)$ (0 denotes non-diseased and 1 diseased study participants) simultaneously and connect both through a bivariate random effect:

$$\log(y_{0i}) = b_0 + \epsilon_0 + u_{0i},\ \log(y_{1i}) = b_1 + \epsilon_1 + u_{1i}, \tag{1}$$

$$\begin{pmatrix} u_{0i} \\ u_{1i} \end{pmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_0^2 & \rho\sigma_0\sigma_1 \\ \rho\sigma_0\sigma_1 & \sigma_1^2 \end{pmatrix}\right] \tag{2}$$

Parameter estimation

Parameter estimation

- Not directly interested in model parameters, but in S(ummary) ROC curve

Parameter estimation

- Not directly interested in model parameters, but in S(ummary) ROC curve
- Predict sensitivity and specificity at various thresholds based on BLUP-principle (best linear unbiased prediction)

Example

Example

- Population-based screening for type 2 diabetes

Example

- Population-based screening for type 2 diabetes
- 38 studies that evaluate HbA1c as diagnostic marker

Example

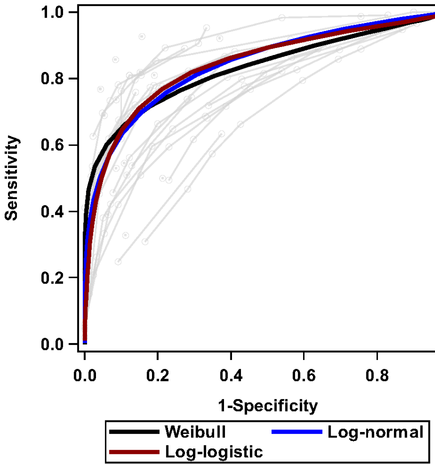
- Population-based screening for type 2 diabetes
- 38 studies that evaluate HbA1c as diagnostic marker
- In total: 124 pairs of sensitivity/specificity at 26 different thresholds

Example

- Population-based screening for type 2 diabetes
- 38 studies that evaluate HbA1c as diagnostic marker
- In total: 124 pairs of sensitivity/specificity at 26 different thresholds

→ Analysis with only 38 pairs would discard about 70% of the available observations

Example



1.0
0.8
0.6
0.4
0.2
0.0
Sensitivity
0.0
0.2
0.4
0.6
0.8
1.0
1-Specificity
Weibull
Log-normal
Log-logistic

Extension options

Extension options

- Usage of more flexible distributions

Extension options

- Usage of more flexible distributions

→ Generalized *F*-distribution doi: 10.1002/jrsm.1395

Extension options

- Usage of more flexible distributions

→ Generalized *F*-distribution doi: 10.1002/jrsm.1395

- Semi-parametric modelling

Extension options

- Usage of more flexible distributions

→ Generalized *F*-distribution doi: 10.1002/jrsm.1395

- Semi-parametric modelling

→ Piecewise-constant model doi: 10.1002/jrsm.1364

Extension options

- Usage of more flexible distributions

→ Generalized *F*-distribution doi: 10.1002/jrsm.1395

- Semi-parametric modelling

→ Piecewise-constant model doi: 10.1002/jrsm.1364

- Usage of bivariate copulas

Extension options

- Usage of more flexible distributions

→ Generalized *F*-distribution doi: 10.1002/jrsm.1395

- Semi-parametric modelling

→ Piecewise-constant model doi: 10.1002/jrsm.1364

- Usage of bivariate copulas
- Application of methods for discrete time-to-event data (GLMM with cloglog-link) doi: 10.1002/jrsm.1753

Extension options

- Usage of more flexible distributions

→ Generalized *F*-distribution doi: 10.1002/jrsm.1395

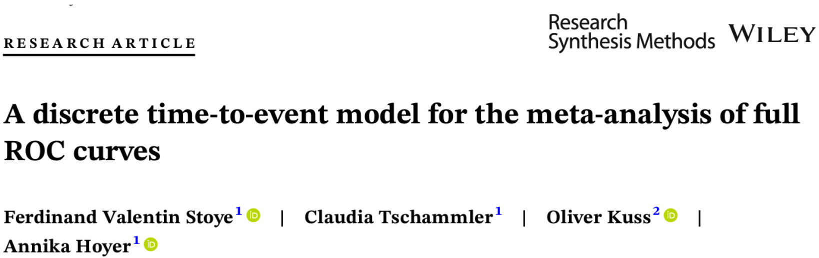
- Semi-parametric modelling

→ Piecewise-constant model doi: 10.1002/jrsm.1364

- Usage of bivariate copulas
- Application of methods for discrete time-to-event data (GLMM with cloglog-link) doi: 10.1002/jrsm.1753

→ funded by the German Research Foundation

Discrete time-to-event model



RESEARCH ARTICLE

Research Synthesis Methods WILEY

A discrete time-to-event model for the meta-analysis of full ROC curves

Ferdinand Valentin Stoye[1] | **Claudia Tschammler**[1] | **Oliver Kuss**[2] | **Annika Hoyer**[1]

Discrete time-to-event model

1. Transfer data from diagnostic studies into the desired format

Discrete time-to-event model

1. Transfer data from diagnostic studies into the desired format
2. Select distribution assumption and link function

Discrete time-to-event model

1. Transfer data from diagnostic studies into the desired format
2. Select distribution assumption and link function
3. Model estimation

$$y_k \sim Bern(\lambda_{c_k}), \quad \textit{(Bernoulli distribution)}$$

$$\lambda_{c_k} = h\left(\alpha_{0k} + \boldsymbol{x}_1^\top \cdot \boldsymbol{\beta}_1 + \boldsymbol{x}_1^\top \cdot \boldsymbol{\beta}_2 \cdot x_2 + \gamma_1 \cdot x_2 + \gamma_2 \cdot (1 - x_2)\right),$$

$$h(x) = 1 - \exp(-\exp(x)) \quad \textit{(inverse cloglog-link)}$$

$$\begin{pmatrix} \gamma_1 \\ \gamma_2 \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_h^2 & \rho\sigma_h\sigma_d \\ \rho\sigma_h\sigma_d & \sigma_d^2 \end{pmatrix}\right) \quad \textit{(Bivariate Normal distribution)}$$

Copula model

1. Transfer data from diagnostic studies into the desired format

Copula model

1. Transfer data from diagnostic studies into the desired format
2. Select copula

Copula model

1. Transfer data from diagnostic studies into the desired format
2. Select copula
3. Select marginal distributions
   - Mixture distributions

   → Distribution of test values, e.g. Weibull: $F_W(y, b, \lambda) = 1 - \exp\left(-(y \exp(-b))^{\frac{1}{\lambda}}\right)$

   → Distribution of events, e.g. Binomial
4. Formulate log-likelihood and estimate model parameters

Both models – Parameter → Sensitivity & Specificity → SROC

Final step: Calculate sensitivity and specificity from the estimated parameters

discrete GLMM:

$$\widehat{\text{sens}}(c_k) = \prod_{t=1}^{k} \exp(-\exp(\widehat{\beta_{1t}}))$$

$$\widehat{\text{spec}}(c_k) = 1 - \prod_{t=1}^{k} \exp(-\exp(\widehat{\beta_{1t}} + \widehat{\beta_{2t}}))$$

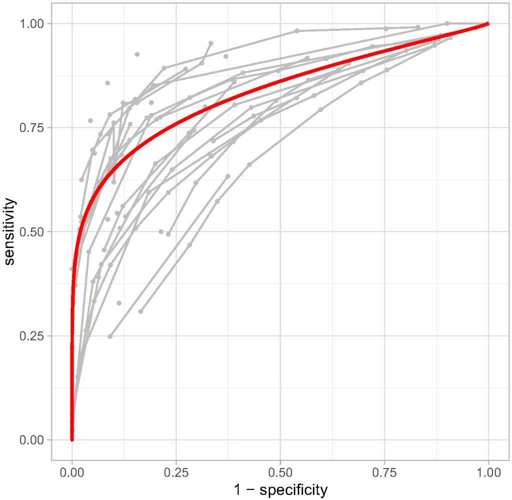
Copula model:

$$\widehat{\text{sens}}(y) = 1 - F_d(y \mid \hat{b}_d, \hat{\lambda}_d)$$

$$\widehat{\text{spec}}(y) = F_h(y \mid \hat{b}_h, \hat{\lambda}_h)$$

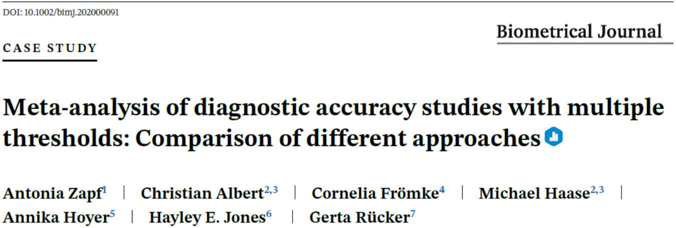
Code: `https://gitlab.ub.uni-bielefeld.de/stoyef/metaROC`

Both models – Parameter → Sensitivity & Specificity → SROC



Applications

Fair comparison of different approaches - Benchmark study



DOI: 10.1002/bimj.202000091

Biometrical Journal

CASE STUDY

Meta-analysis of diagnostic accuracy studies with multiple thresholds: Comparison of different approaches

Antonia Zapf[1] | Christian Albert[2,3] | Cornelia Frömke[4] | Michael Haase[2,3] |
Annika Hoyer[5] | Hayley E. Jones[6] | Gerta Rücker[7]

Fair comparison of different approaches

- Data set from nephrology

Fair comparison of different approaches

- Data set from nephrology
- Aim: Evaluate the diagnostic accuracy of neutrophil gelatinase lipocalin (NGAL) as a test for acute kidney injury (AKI)

Fair comparison of different approaches

- Data set from nephrology
- Aim: Evaluate the diagnostic accuracy of neutrophil gelatinase lipocalin (NGAL) as a test for acute kidney injury (AKI)
- Systematic review with 58 studies

Fair comparison of different approaches

- Data set from nephrology
- Aim: Evaluate the diagnostic accuracy of neutrophil gelatinase lipocalin (NGAL) as a test for acute kidney injury (AKI)
- Systematic review with 58 studies
- Sensitivity and specificity provided at three different thresholds such that:
    1. the estimated sensitivity is at least 95%

Fair comparison of different approaches

- Data set from nephrology
- Aim: Evaluate the diagnostic accuracy of neutrophil gelatinase lipocalin (NGAL) as a test for acute kidney injury (AKI)
- Systematic review with 58 studies
- Sensitivity and specificity provided at three different thresholds such that:
  1. the estimated sensitivity is at least 95%
  2. the estimated specificity is at least 95%

Fair comparison of different approaches

- Data set from nephrology
- Aim: Evaluate the diagnostic accuracy of neutrophil gelatinase lipocalin (NGAL) as a test for acute kidney injury (AKI)
- Systematic review with 58 studies
- Sensitivity and specificity provided at three different thresholds such that:
    1. the estimated sensitivity is at least 95%
    2. the estimated specificity is at least 95%
    3. the estimated sum of sensitivity and specificity is maximized (optimal Youden index)

Fair comparison of different approaches

Four methods are compared:

1. Random effects model by Steinhauser et al. (2016, frequentist approach)

Fair comparison of different approaches

Four methods are compared:

1. Random effects model by Steinhauser et al. (2016, frequentist approach)
2. Bayesian model by Jones et al. (2019)

Fair comparison of different approaches

Four methods are compared:

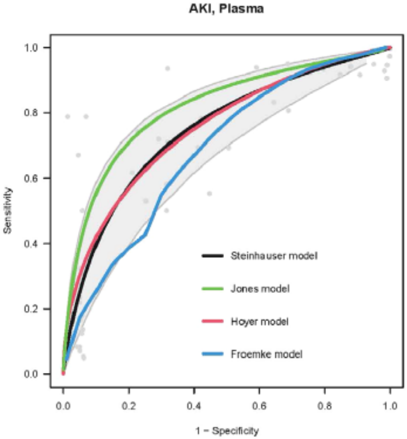
1. Random effects model by Steinhauser et al. (2016, frequentist approach)
2. Bayesian model by Jones et al. (2019)
3. Time-to-event model by Hoyer et al. (2018, frequentist approach)

Fair comparison of different approaches

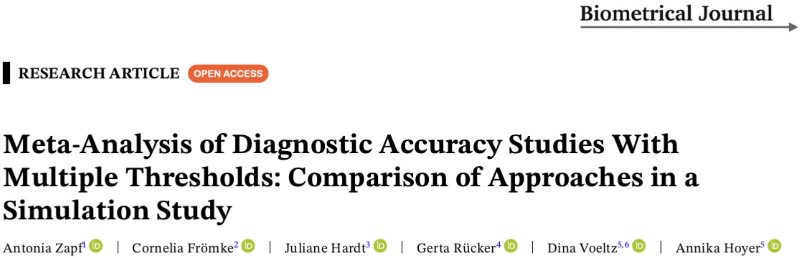
Four methods are compared:

1. Random effects model by Steinhauser et al. (2016, frequentist approach)
2. Bayesian model by Jones et al. (2019)
3. Time-to-event model by Hoyer et al. (2018, frequentist approach)
4. Nonparametric model by Frömke et al. (2020)

Fair comparison of different approaches



Fair comparison of different approaches - Simulation study



Biometrical Journal

RESEARCH ARTICLE OPEN ACCESS

Meta-Analysis of Diagnostic Accuracy Studies With Multiple Thresholds: Comparison of Approaches in a Simulation Study

Antonia Zapf[1] | Cornelia Frömke[2] | Juliane Hardt[3] | Gerta Rücker[4] | Dina Voeltz[5,6] | Annika Hoyer[5]

Fair comparison of different approaches

Three methods are compared:

1. Random effects model by Steinhauser et al. (2016, frequentist approach)

Fair comparison of different approaches

Three methods are compared:

1. Random effects model by Steinhauser et al. (2016, frequentist approach)
2. Time-to-event model by Hoyer et al. (2018, frequentist approach)

Fair comparison of different approaches

Three methods are compared:

1. Random effects model by Steinhauser et al. (2016, frequentist approach)
2. Time-to-event model by Hoyer et al. (2018, frequentist approach)
3. Nonparametric model by Frömke et al. (2020)

Fair comparison of different approaches

- Systematic simulation study, motivated by an example from medical research

Fair comparison of different approaches

- Systematic simulation study, motivated by an example from medical research

→ All three approaches worked partially well

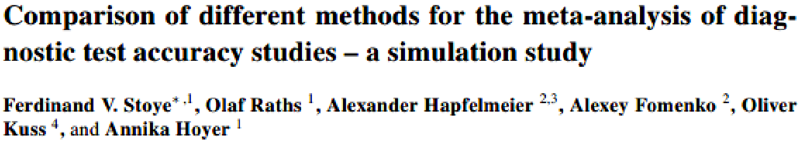
Fair comparison of different approaches

- Systematic simulation study, motivated by an example from medical research

→ All three approaches worked partially well

- The approach by Hoyer and colleagues was slightly superior in most scenarios and is recommended in practice

Fair comparison of different approaches - Work in progress



Comparison of different methods for the meta-analysis of diagnostic test accuracy studies – a simulation study

Ferdinand V. Stoye[*,1], **Olaf Raths**[1], **Alexander Hapfelmeier**[2,3], **Alexey Fomenko**[2], **Oliver Kuss**[4], and **Annika Hoyer**[1]

Fair comparison of different approaches

- Comparison of 5 single threshold methods and 6 multiple threshold methods (MTM)

Fair comparison of different approaches

- Comparison of 5 single threshold methods and 6 multiple threshold methods (MTM)
- Variation of 8 parameter dimensions, covering a broad range of real-life settings

Fair comparison of different approaches

- Comparison of 5 single threshold methods and 6 multiple threshold methods (MTM)
- Variation of 8 parameter dimensions, covering a broad range of real-life settings
- Model performances are comparable regarding bias, empirical coverage and convergence

Fair comparison of different approaches

- Comparison of 5 single threshold methods and 6 multiple threshold methods (MTM)
- Variation of 8 parameter dimensions, covering a broad range of real-life settings
- Model performances are comparable regarding bias, empirical coverage and convergence

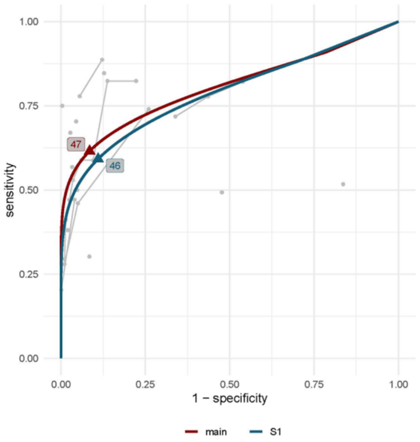
→ Main advantage of MTM: Threshold-dependent estimates of sensitivity and specificity

Application: Reassessment of HbA1c

Reassessment of the diagnostic accuracy of HbA_{1c} and glucose for type 2 diabetes: a systematic review and meta-analysis of observational studies

Christina Baechle[1,2] Ferdinand V. Stoye[3], Nafiseh Shokri-Mashhadi[1,2], Anna-Therese Klötzsch[1,2], Brenda Bongaerts[4], Annika Hoyer[3], Oliver Kuss[1,2], Rüdiger Landgraf[5], Wolfgang Rathmann[1], Sabrina Schlesinger[1,2]

Application: Reassessment of HbA1c



Outlook and Implications

Networks of diagnostic tests

- Extension of network meta-analysis of clinical trials

Networks of diagnostic tests

- Extension of network meta-analysis of clinical trials
- Comparison of at least two diagnostic tests with a common gold standard

Networks of diagnostic tests

- Extension of network meta-analysis of clinical trials
- Comparison of at least two diagnostic tests with a common gold standard
- Example type 2 diabetes: HbA1c in comparison to fasting plasma glucose

Precision diagnostics

- Fundamental requirement: specific individual must show different response to different actions

Precision diagnostics

- Fundamental requirement: specific individual must show different response to different actions
- In therapeutic research: “heterogeneity of treatment effect” (HTE)

Precision diagnostics

- Fundamental requirement: specific individual must show different response to different actions
- In therapeutic research: “heterogeneity of treatment effect” (HTE)

→ Missing corresponding concept of “heterogeneity of diagnostic accuracy” (HDA)

Precision diagnostics

- Fundamental requirement: specific individual must show different response to different actions
- In therapeutic research: “heterogeneity of treatment effect” (HTE)

→ Missing corresponding concept of “heterogeneity of diagnostic accuracy” (HDA)

- One could ask:
    1. whether making the diagnostic test is beneficial at all (“Test-HDA“)

Precision diagnostics

- Fundamental requirement: specific individual must show different response to different actions
- In therapeutic research: “heterogeneity of treatment effect” (HTE)

→ Missing corresponding concept of “heterogeneity of diagnostic accuracy” (HDA)

- One could ask:
  1. whether making the diagnostic test is beneficial at all (“Test-HDA“)
  2. whether personalised thresholds in testing should be applied (“Threshold-HDA”)

Precision diagnostics

- Fundamental requirement: specific individual must show different response to different actions
- In therapeutic research: “heterogeneity of treatment effect” (HTE)

→ Missing corresponding concept of “heterogeneity of diagnostic accuracy” (HDA)

- One could ask:
  1. whether making the diagnostic test is beneficial at all (“Test-HDA“)
  2. whether personalised thresholds in testing should be applied (“Threshold-HDA”)
  3. whether there are different diagnostic procedures from which a tested person would benefit differently (“Device-HDA”)

Summary

Summary

- Meta-analysis of ROC curves: statistical challenge

Summary

- Meta-analysis of ROC curves: statistical challenge

→ **BUT:** Standard approaches neglect the majority of observations

Summary

- Meta-analysis of ROC curves: statistical challenge

→ **BUT:** Standard approaches neglect the majority of observations

- New models motivated from survival analysis show plausible and reliable results in practical applications