ISSN: 2641-6336
DOI: 10.33552/ABBA.2023.05.000612
Annals of
Biostatistics & Biometric Applications
Iris Publishers

Research Article

A Nonparametric Random-Effects Meta-Analysis Model for Diagnostic Accuracy Studies with Multiple Thresholds of Quantitative Biomarkers

Aeilko H Zwinderman*

Department of Epidemiology and Data Science, Amsterdam University Medical Center, Netherland

***Corresponding author:** Aeilko H. Zwinderman, Department of Epidemiology and Data Science, Amsterdam University Medical Centers, Netherland.

Received Date: March 14, 2023

Published Date: May 30, 2023

Abstract

We introduce a nonparametric random-effects model for the meta-analysis of a series of diagnostic accuracy studies of a quantitative biomarker that reported sensitivity-and specificity-values at multiple thresholds of the biomarker. The model is based on the observed numbers of cases and controls in between cutoff-values of the biomarker.

Observed numbers of cases and controls within studies were modeled using multinomial-normal or multinomial-dirichlet distributions. Parameters of our new model were Estimated using an MCMC-algorithm in a Bayesian framework. We provide code to run our method within the R statistical software. With our new model two example datasets were analyzed and compared to the method implemented in the R package diagmeta.

With two example datasets our approach gave comparable results as the method implemented in diagmeta. Results are illustrated using estimated density and cumulative distribution function of the biomarker and associated credibility intervals. Transformations thereof such as ROC-curve, area under the ROC curve, and the Youden-index curve are also estimated together with credibility intervals.

We developed a new model for meta-analysis of diagnostic studies evaluating multiple thresholds of a quantitative biomarker. The new model provided comparable results as an existing method but with less assumptions.

Keywords: Diagnostic test; Meta-analysis; Multiple thresholds

Abbreviations: AUC: Area Under the Curve; CI: Common random Intercept; CS: Common random Slope; DICS: Different random Intercepts and Common random Slope; DIDS: Different random Intercepts and Different random Slopes; ELF: enhanced liver fibrosis; FENO: fractional exhaled nitric oxide; HA: hyaluronic acid; JAGS: Just Another Gibbs Sampler; MCMC: Multi Chain Monte Carlo; PIIINP: Procollagen III N-terminal Propeptide; ROC: Receiver Operator Curve; TIMP: tissue inhibitor of metalloproteinase-1

Introduction

Steinhauser et al [1] developed a method for the meta-analysis of multiple diagnostic test accuracy studies of a quantitative biomarker that reported sensitivity- and specificity-values at multiple thresholds of the biomarker. Their method is implemented in a user-friendly and free [2] and is very quick, depending a little bit on choice of the specific model and size of the data. Steinhauser et al made specific choices for the analysis model which validity may be questioned for some biomarkers. Their first choice concerned the distributions of the biomarker, which was assumed to be either normal or logistic. Obviously, this assumption may not be valid in particular cases. Second choice concerned the application of standard linear mixed-effects regression models of transformed sensitivity- and specificity-values on the biomarker-thresholds. The assumption about the

linear relationship between transformed sensitivity and specificity and thresholds need not be correct. Lastly, the residuals of the models for transformed sensitivities/specificities associated with different thresholds in the same study (i.e., e_{si} and f_{si} in the Steinhauser paper) were assumed to be independent. This assumption is peculiar because different realisations of the sensitivities/specificities in the same study are almost certainly correlated (with increasing threshold, sensitivity will -necessarily- decrease and specificity will similarly increase). A final concern is that the method is based on the normal distribution approximation of the binomial distribution of the numbers of true positives and true negatives (with the usual 0.5-trick when sensitivity or specificity is zero or 1). Steinhauser et al did point to this concern but did not as yet implement the binomial distribution of sensitivity/specificity in their R-package.

To address these issues we extended the work of Steinhauser et al for data of the type that was considered by Steinhauser et al. We suggest to use a nonparametric method based on the multinomial distribution. In the next sections we elaborate the type of (meta-) data we aim to analyse, we specify our statistical model and estimation method. To illustrate we show results of two data sets; a new set concerning results on the enhanced liver fibrosis (ELF) biomarker that might be useful to diagnose advanced liver fibrosis and a set that was published by [4] on diagnostic accuracy of fractional exhaled nitric oxide for the diagnosis of asthma. We finish with a discussion section.

Materials and Methods

Type of data

We consider a series of N diagnostic studies of a quantitative biomarker Y. In study i there were included n_{i1} 'cases' and n_{i0} 'controls'. In the report of study i the sensitivities and specificities to distinguish cases and controls were reported for k_i different cutoff-values $(k_i \geq 1): \xi_{i1}, \xi_{i2}, ..., \xi_{ik_i}$. Hence, for cutoff-value ξ_{ij} the observed sensitivity- and specificity-values were se_{ij} and sp_{ij} (see Table 1 with example data). From these reported results we calculated in every study the observed numbers of cases and controls with biomarker-values in between cutoff-values: $y_{i1j} = n_{i1}\left(se_{i,j-1} - se_{ij}\right)$ is the number of cases with $\xi_{i,j-1} < Y \leq \xi_{ij}$ and $y_{i0j} = n_{i0}\left(sp_{i,j+1} - sp_{ij}\right)$ is the number of controls with $\xi_{i,j-1} < Y \leq \xi_{ij}$. Sensitivity se_{i0} and specificity sp_{i,k_i+1} were defined as zero and could therefore be considered to be associated with cutoff-values $\xi_{i0} = -\infty$ and $\xi_{i,k_i+1} = \infty$, or any biological or theoretical lower- and upperbound of Y.

So, for study i we translated published results into a vector of observed numbers of cases in the k_i+1 categories of Y in study i, denoted as $y_{i1} = \left(y_{i10}, ..., y_{i1k_i}\right)$, and a vector of observed numbers of controls in the same $k_i + 1$ categories of Y, denoted as $y_{i0} = \left(y_{i00}, ..., y_{i0k_i}\right)$. It is important to stress that in general the number of cutoff-values, k_i, (likely) varies between studies and also that the specific cutoff-values themselves, $\xi_{i1}, ..., \xi_{i,k_i}$, vary between studies. It is also important to stress that the biomarker Y is represented in this way as a categorical variable through the cutoff-values, and that the k_i+1 categories can be considered as ordered.

Statistical models

We considered the distributions of the vectors y_{i1} and y_{i0} to be multinomials given the numbers of cases and controls:

$$y_{i1} \sim multinomial\left(\tilde{\pi}_{i1}, n_{i1}\right)$$

$$y_{i0} \sim multinomial\left(\tilde{\pi}_{io}, n_{i0}\right) \quad (1)$$

where $\tilde{\pi}_{i1} = \left(\tilde{\pi}_{i10}, \tilde{\pi}_{i11}, ..., \tilde{\pi}_{i1k_i}\right)$ and

$\tilde{\pi}_{i0} = \left(\tilde{\pi}_{i00}, \tilde{\pi}_{i01}, ..., \tilde{\pi}_{i0k_i}\right)$ are vectors of length k_i+1 of probabilities: $\tilde{\pi}_{i1j}$ and $\tilde{\pi}_{i0j}$ are the probabilities that a random case or control in study i had biomarker value $\xi_{i,j-1} < Y \leq \xi_{i,j}$. Notice that in this model the covariance between two realisations y_{i1j} and y_{iil} in the group of cases in study i equals $-\pi_{i1j}\pi_{i1l}/n_{i1}$ (and similarly in the group of controls), and therefore the asymptotic covariance between for instance $se_{i1} = \tilde{\pi}_{i1}$ and $se_{i2} = \tilde{\pi}_{i1} + \tilde{\pi}_{i2}$ is equal to $se_{i1}(1 - se_{i2})/n_{i1}$, which is not necessarily zero (as is assumed in the diagmeta-package by Steinhauser et al).

Summed over all *N* studies we have m unique cutoff-values, say $\theta_1, \theta_2, ..., \theta_m$ and we next defined the probabilities that a random case or a random control in study i had biomarker value in between θ_{l-1} and θ_l as $\tilde{\pi}_{i1l}$ and $\tilde{\pi}_{i0l}$, respectively. Depending on the study-specific cutoff-values $\xi_{i1}, ..., \xi_{ik_i}$, we could write every probability π_{i1j} and every probability π_{i0j} as a sum of one or more probabilities π_{i0l} and π_{i0l}. Suppose for instance that there were $m = 10$ unique cutoff-values over all N studies, and that in study i there were two cutoff-values, ξ_{i1} and ξ_{i2}, and $\theta_4 = \xi_{i1}$ and $\theta_6 = \xi_{i2}$. Then π_{i11} is the probability that a random case in this study had $Y \leq \xi_{i1}$ or, equivalently, $Y \leq \theta_4$, thus $\tilde{\pi}_{i11} = \sum_{l=0}^{4}\tilde{\pi}_{i1l}$ and similarly, $\tilde{\pi}_{i12} = \sum_{l=5}^{6}\tilde{\pi}_{i1l}$ and $\tilde{\pi}_{i13} = \sum_{l=7}^{10}\tilde{\pi}_{i1l}$. More generally, we can write $\tilde{\pi}_{i1j} = \sum_{l=0}^{m}\pi_{i1l}z_{ijl}$ and $\pi_{i0j} = \sum_{l=0}^{m}\pi_{i0l}z_{ijl}$, where z_{ijl} equals $\xi_{i,j-1} \leq \theta_{l-1} < \theta_l \leq \xi_{i,j}$ and zero otherwise.

There are several options to further model $\pi_{i10}, \pi_{i11}, ..., \pi_{i1m}$ and $\pi_{i00}, \pi_{i00}, ..., \pi_{i0m}$ and we explored two ways, namely by mixing the multinomial distributions with a multivariate normal or with the Dirichlet distribution. For the multinomial-normal mixture we first transformed $\pi_{i10}, \pi_{i11}, ..., \pi_{i1m}$ and $\pi_{i00}, \pi_{i00}, ..., \pi_{i0m}$ using the soft-max transformation as:

$$\pi_{i1j} = \frac{\exp\left(a_{i1j}\right)}{1 + \sum_{l=1}^{m}\exp\left(a_{i1j}\right)}$$

$$\pi_{i0j} = \frac{\exp\left(a_{i0j}\right)}{1 + \sum_{l=1}^{m}\exp\left(a_{i0j}\right)} \quad (2)$$

(and $a_{i10} = a_{i00} = 0$), and then we assumed that the vector of parameters of study i, i.e. $\omega_i = (a_{i11},..., a_{i1m}, a_{i01},..., a_{i0m})$, was a random draw from the multivariate normal distribution with mean μ (of length 2m) and covariance matrix $\sum$ (of dimensions 2m). With this assumption our model is a random-effects meta-analysis model. A fixed-effects model is obtained by assuming $\sum = 0$, or actually by constraining $a_{i1j} = a_{1j}$ and $a_{i0j} = a_{0j}$ for all i.

This model is an extension of the bivariate meta-analysis model. If there is only one cutoff-value (θ), then $\pi_{i10}, \pi_{i00}, \pi_{i11}, \pi_{i01}$ are the probabilities of false negative, true negative, true positive and false positive outcomes in study i, respectively, and a_{i11} and a_{i01} are the logit-transformed sensitivity and 1-specificity of study i. In such case the parameter μ represents the average logit-transformed sensitivity and 1-specificity in the population of diagnostic accuracy studies of this biomarker [3].

Like the bivariate model, our extended model can be generalized with (study-specific) covariates, and given the ordinal character of the biomarker categories it seems sensible to use a ordinal logistic model, such as for instance a proportional odds model:

$$\pi_{i1j} = \frac{\exp(a_{i1j} + b_{i1}x_i)}{1 + \sum_{l=1}^{m} \exp(a_{i1j} + b_{i1}x_i)}$$

$$\pi_{i0j} = \frac{\exp(a_{i0j} + b_{i0}x_i)}{1 + \sum_{l=1}^{m} \exp(a_{i0j} + b_{i0}x_i)} \quad (3)$$

where x_i is the observed value of a vector of covariates of study i. The vector of covariates may be different for cases and controls and the values of the covariates may even differ between biomarker-categories. The regression coefficients β_{i1} and β_{i0} may also vary over categories of the biomarker, but then the ordinal character of the biomarker-categories is not used in the statistical modeling. It may be useful too to assume that these regression coefficients are fixed parameters and not vary over studies (i.e., $\beta_{i1} = \beta_1$ and $\beta_{i0} = \beta_0$ for all studies). For now we will consider the situation without covariates.

As an alternative approach we considered the multinomial-Dirichlet mixture model. Here the study-specific probabilities $\pi_{i10},..., \pi_{i1m}, \pi_{i00},..., \pi_{i0m}$ were considered to be sampled from the Dirichlet distributions $f(\pi_{i10},..., \pi_{i1m}; \alpha_{10},..., \alpha_{1m})$ and $f(\pi_{i00},..., \pi_{i0m}; \alpha_{00},..., \alpha_{0m})$:

$$f(\pi_{i10},..., \pi_{i1m}; \alpha_{10},..., \alpha_{1m}) = \frac{1}{B(\alpha_{10},..., \alpha_{1m})} \prod_{j=0}^{m} \pi_{i1j}^{\alpha_{1j-1}}$$

$$f(\pi_{i00},..., \pi_{i0m}; \alpha_{00},..., \alpha_{0m}) = \frac{1}{B(\alpha_{00},..., \alpha_{0m})} \prod_{j=0}^{m} \pi_{i0j}^{\alpha_{0j-1}},$$

where B(...) are beta-functions. This mixture model is also a random-effects meta-analysis model. Notice that the study-specific probabilities of cases (π_{i1j}) and of controls (π_{i0j}) were supposed to be drawn from different Dirichlet distributions. This was chosen because of ease of computations, but is not absolutely necessary. It is different from the multinomial-normal mixture that is specified above, where a_{i0j} and a_{i0j} were sampled from a single multivariate normal distribution, but that choice can be amended too by drawing a_{i1j} and a_{i0j} from different normal distributions. The multinomial-Dirichlet mixture model can also be generalized to include covariates (by modeling α_{1j} and α_{0j} as functions of covariates x), but here we only considered the model without covariates.

Estimation

We first considered the multinomial-normal mixture model. Given $\omega_i = (a_{i1}, a_{i0})$ the conditional log likelihood of (y_{i1}, y_{i0}) equals

$$\begin{aligned} \log L_{ci}((y_{i1}, y_{i0}) | (a_{i1}, a_{i0})) &= \sum_{j=0}^{k_i} y_{i1j} \log(\tilde{\pi}_{i1j}) + y_{i0j} \log(\tilde{\pi}_{i0j}) \\ &= \sum_{j=0}^{k_i} y_{i1j} \log\left(\sum_{l=0}^{m} \pi_{i1l} z_{ijl}\right) + y_{i0j} \log\left(\sum_{l=0}^{m} \pi_{i0l} z_{ijl}\right) \end{aligned} \quad (5)$$

and the marginal likelihood of (y_{i1}, y_{i0}) equals

$$L_{mi}((y_{i1}, y_{i0})) = \int L_{ci}((y_{i1}, y_{i0}) | (a_{i1}, a_{i0})) \, g((a_{i1}, a_{i0}) | \mu, \Sigma) \, \partial(a_{i1}, a_{i0}) \quad (6)$$

where $g((a_{i1}, a_{i0}) | \mu, \Sigma)$ is the multivariate normal distribution function with mean and covariance matrix $\sum$. Summing contributions $\log L_{mi}((y_{i1}, y_{i0}))$ over all N studies gives the total marginal log-likelihood.

The integral in the marginal likelihood is multidimensional of dimension 2m and has no analytical solution. Optimizing the total log marginal likelihood is difficult if m is large, therefore based solely on convenience-arguments we decided to use MCMC-methods in a Bayesian framework to estimate the parameters of interest, i.e., $\mu = (\pi_0, \pi_1)$ and $\sum$ (where π_0 and π_1 are the vectors of category-probabilities averaged over studies for controls and cases, respectively). In this framework we used as hyperprior distributions

$$\mu = (\mu_1,..., \mu_{2m}) \sim multivariate\ normal(0, S)$$

$$\sum \sim Wishart(R, df = 2m) \quad (7)$$

where S was a diagonal symmetric matrix with entries "0.001", and R was a diagonal symmetric matrix with entries "1". These hyperprior distributions were considered to be (almost) uninformative.

The model was implemented in JAGS [7] and was run through the R-interface of JAGS. Below is the JAGS-syntax in the Additional Material-section. We chose to apply 50.000 burn-in iterations and then checked convergence. When converged we drew 50.000 val-

ues from the posterior distributions with a thinning parameter of 10 to reduce autocorrelation between sampled values. Afterwards the posterior distributions were graphically illustrated and summarised with the mean, median and 2.5th and 97.5th percentiles. We also derived posterior distributions of transformations of (π_0, π_1), such as sensitivities, specificities and Youden index values.

Estimation of the multinomial-Dirichlet mixture was slightly easier because integrals involved in the likelihood can be solved analytically. Given $(\tilde{\pi}_{i10},..., \tilde{\pi}_{i1k_i})$ the conditional likelihood of $(y_{i10},..., y_{i1k_i})$ equals

$$L_{c1i}\left(y_{i10},..., y_{i1k_i} \mid \tilde{\pi}_{i10},..., \tilde{\pi}_{i1k_i}\right) = \prod_{j=0}^{k_i} \tilde{\pi}_{i1j}^{y_{i1j}} \quad (8)$$

and the marginal likelihood of the observations of study i can be written as:

$$L_{m1i}\left(y_{i10},..., y_{i1k_i}\right) = \int \left(\prod_{j=0}^{k_i} \tilde{\pi}_{i1j}^{y_{i1j}}\right) \frac{1}{B\left(\tilde{\alpha}_{i10},..., \tilde{\alpha}_{i1k_i}\right)} \prod_{j=0}^{k_i} \tilde{\pi}_{i1j}^{\tilde{\alpha}_{i1j}-1} \partial\pi_{i1} = \frac{B\left(\tilde{\alpha}_{i10}+y_{i10},..., \tilde{\alpha}_{i1k_i}+y_{i1k_i}\right)}{B\left(\tilde{\alpha}_{i10},..., \tilde{\alpha}_{i1k_i}\right)}, \quad (9)$$

where $\tilde{\alpha}_{i1j} = \sum_{l=0}^{m} \alpha_{1l} z_{ijl}$. The total marginal log-likelihood, L_{m1}, is the sum of $\log L_{m1i}(...)$ over pall studies which can be written as

$$\log L_{m1} = \sum_{i=1}^{N} \sum_{j=0}^{k_i} \sum_{r=1}^{y_{i1j}} \left(\sum_{l=0}^{m} \alpha_{1l} z_{ijl} + r - 1\right) - \sum_{r=1}^{\sum_{j=0}^{k_i} y_{i1j}} \left(\sum_{j=0}^{k_i} \sum_{l=0}^{m} \alpha_{1l} z_{ijl} + r - 1\right) \quad (10)$$

A similar total marginal log-likelihood was obtained for the data of the control-subjects in the *N* studies. These two likelihoods were independent, hence the decision to use two Dirichlet distributions led to separate analyses for the case- and control-data.

Like with the multinomial-normal mixture we used MC-MC-methods in a Bayesian frame-work to estimate the parameters of interest in this approach too, i.e., $\alpha_{00},..., \alpha_{0m}$ and $\alpha_{10},..., \alpha_{1m}$. Convenient is the possibility to obtain posterior distributions of any function of $\alpha_{10},..., \alpha_{1m}$ and $\alpha_{00},..., \alpha_{0m}$, such as averaged category-probabilities for cases and controls, sensitivities, specificities and Youden-index values. In this frame-work we used as hyperprior distributions independent gamma distributions: $\alpha_{1j} \sim gamma(0.01, 0.01)$, and $\alpha_{0j} \sim gamma(0.01, 0.01)$ for all $j = 0,..., m)$.

Results

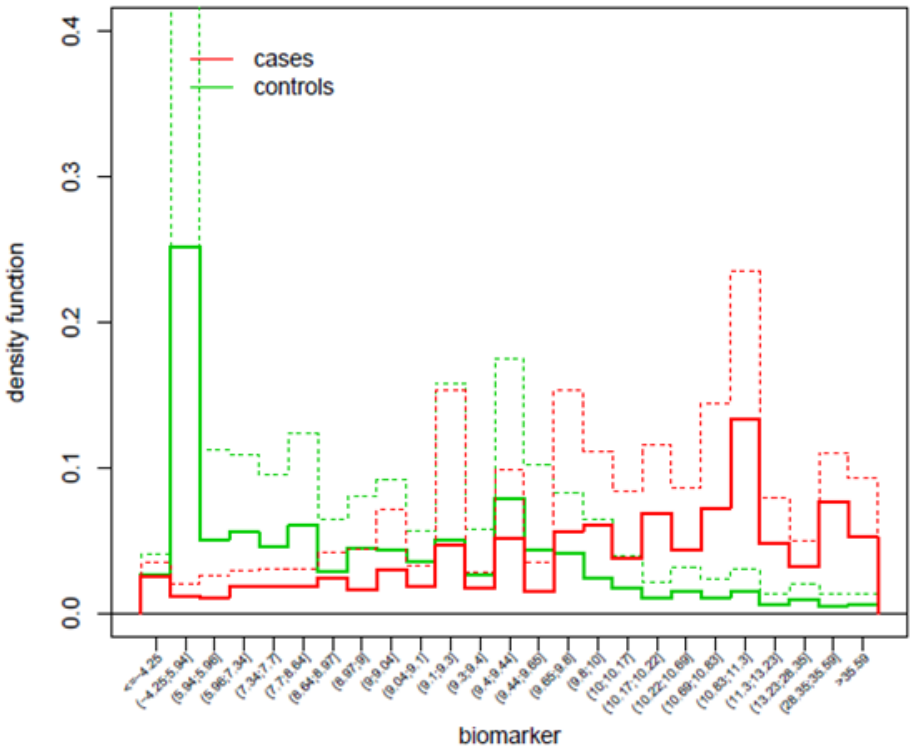
We analyzed data coming from *N* = 10 studies on the diagnostic value of the enhanced liver fibrosis (ELF) biomarker for diagnosis of advanced liver fibrosis. ELF is a weighted combination of type III procollagen peptide (PIIINP), hyaluronic acid (HA), and tissue inhibitor of metalloproteinase-1 (TIMP1) measured in blood. The data retrieved from the 10 published papers is summarised in Table 1. Results are available of 28 study-specific cutoffs and $m = 24$ unique thresholds of the ELF biomarker.

Table 1: Study cutoff values, and observed sensitivities and specificities in the various studies on the diagnostic value of the enhanced live fibrosis (ELF) biomarker.

Study nr	Author	Cutoff value	Number of controls	Specificity	Number of cases	Sensitivity
1	Dvorak 2014	5.96	39	0.974	17	0.882
1		5.94		0.923		0.941
2	Eddowes 2018	9.8	25	0.92	25	0.4
2		7.7		0.24		0.96
4	Lykiardopoulos 2016	7.7	120	0.683	38	0.737
4		9.8		1		0.184
8	Miele 2017	9.8	67	0.925	15	0.867
9	Guha 2008	10.17	148	0.899	44	0.795
9		-4.25		0.122		1
9		8.97		0.419		0.977
9		9.3		0.568		0.955
9		9.65		0.75		0.909
9		10.69		0.953		0.614
9		28.35		0.993		0.295
9		35.59		1		0.159
11	Anstee 2018	9.8	923	0.73	2260	0.74
11		11.3		0.98		0.2
12	Boursier 2018	8.64	250	0.424	167	0.898

12		10		0.896		0.473
13	Polyzos 2019	9	24	0.833	7	0.857
15	Stauber 2018	9.1	88	0.75	34	0.941
16	Welsh 2018	7.34	17	0	9	1
16		9.04		0.647		1
16		9.4		0.765		0.889
16		9.44		0.765		0.778
16		10.22		0.941		0.667
16		10.83		1		0.556
16		13.23		1		0

Using the multinomial-normal mixture approach, estimated averaged category-probabilities (i.e., 1 for cases and 0 for controls) were calculated and those are reported in the top panel of Figure 1: on visual inspection the biomarker did not seem to be normally distributed, especially not in the controls. The empirical cumulative distribution functions are illustrated in the lower panel of Figure 1, separately for the cases and the controls. In these figures every study is represented by colored thin lines, and the averaged cumulative distribution functions are represented by thick black lines.



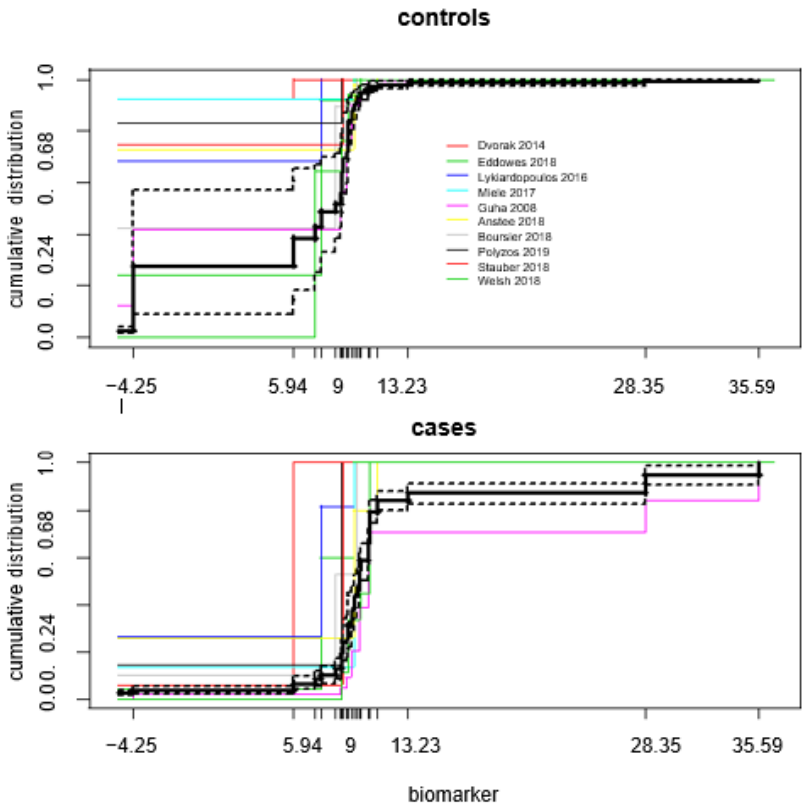


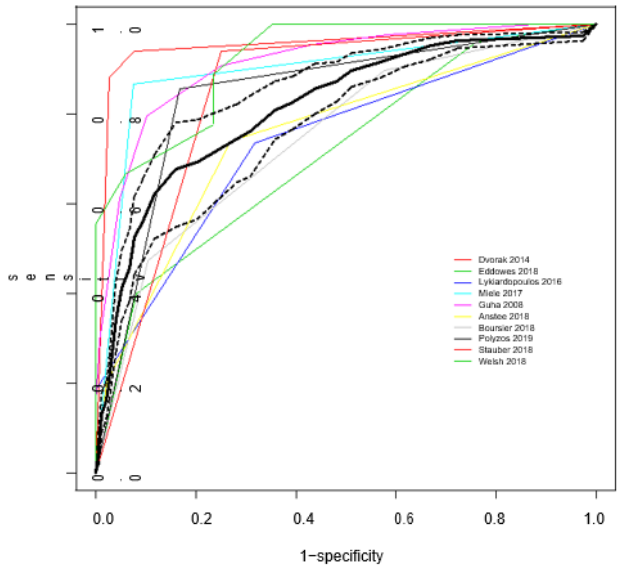
Figure 1: Density and cumulative functions.

From the estimated average distribution functions we calculated the average ROC curve and the average Youden index-values for all thresholds. Both are given in Figure 2. Youden index was highest for the cutoff-value Y=9.8 (Youden index=0.49, 95 % CI: 0.437-0.541), but a similar value was found for Y=9.0 (Youden index=0.488, 95% CI: 0.383-0.606). Estimated averaged sensitivities and specificities for all unique thresh-olds are reported in Table 2.

Table 2: Cutoff values and estimated averaged sensitivities and specificities.

Cutoff	Mean sensitivity	SD Sensitivity	Mean specificity	SD Specificity
-4.25	0.9693207	0.004777098	0.02789814	0.003948594
5.94	0.95045566	0.004134327	0.31382702	0.086766706
5.96	0.9360575	0.005980867	0.37505465	0.083353262
7.34	0.9069236	0.023658799	0.44013223	0.067984032
7.7	0.89515431	0.018423903	0.48990319	0.063187571
8.64	0.87031168	0.018667507	0.5368792	0.065490571
8.97	0.85492016	0.019085337	0.6155565	0.066996419
9	0.84142754	0.019298844	0.66581335	0.060351582
9.04	0.81165735	0.021849068	0.69732891	0.057131028
9.1	0.78030771	0.02309641	0.72151872	0.056034124
9.3	0.72299728	0.036643855	0.75566811	0.046710561
9.4	0.69375986	0.041391479	0.7949153	0.037427016
9.44	0.66072856	0.044795262	0.80896748	0.034383374
9.65	0.63548991	0.043959107	0.84238979	0.031644146
9.8	0.61669467	0.038855762	0.87308937	0.026219973
10	0.53590861	0.03514464	0.89893094	0.018627928
10.17	0.49255842	0.033277674	0.91116568	0.016986032

10.22	0.38469774	0.03696599	0.92685013	0.013769665
10.69	0.3154259	0.032938925	0.94335158	0.009243735
10.83	0.26738084	0.027964296	0.95765962	0.006180633
11.3	0.16151088	0.016187847	0.97100239	0.003835921
13.23	0.13173427	0.019854194	0.97830708	0.003519755
28.35	0.08007334	0.009776116	0.98546978	0.002306074
35.59	0.03750751	0.008358932	0.99144331	0.001860916



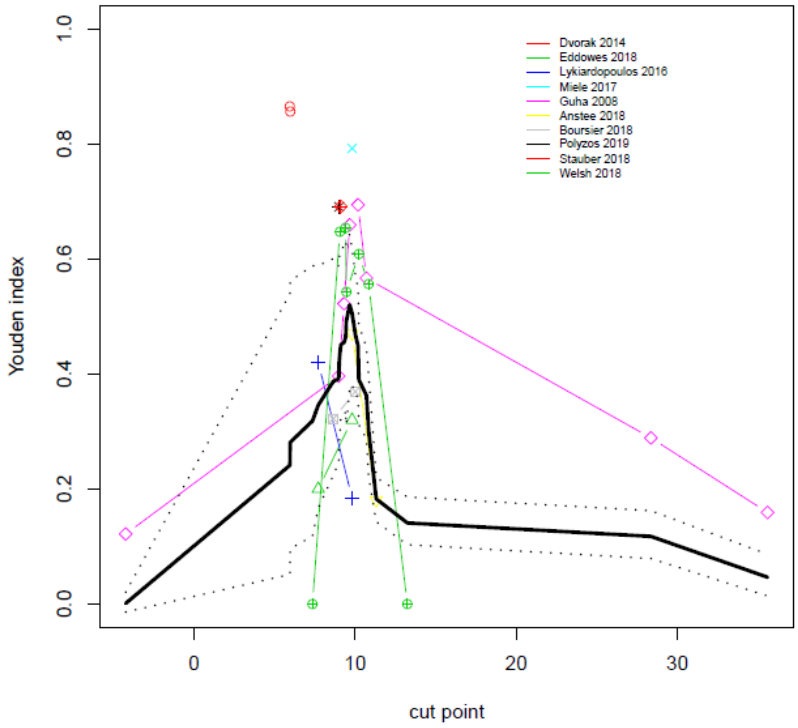


Figure 2: ROC and Youden-index curves.

The area under the ROC (AUC) was estimated as 0:816 with 95% credibility inter-val 0:788 0:851. Convergence of the model seemed to be sufficient; as an exam-ple the trace- and density-plots of the AUC-statistic are given in Figure 3. Using the multinomi-al-Dirichlet mixture we found similar results, although in general with wider credibility intervals; estimated area under the ROC was 0:792 with 95% credibility in-terval 0:700 0:866. Estimated area under the ROC was 0.796 (95% ci: 0.68-0.88) according to the least constrained DIDS model of Steinhauser et al, but the DIDS al-gorithm did not converge. Only with the simpler CS- and CI-mod-els did their algorithm converge but then the AUC estimates were 0.747 and 0.696, respectively.

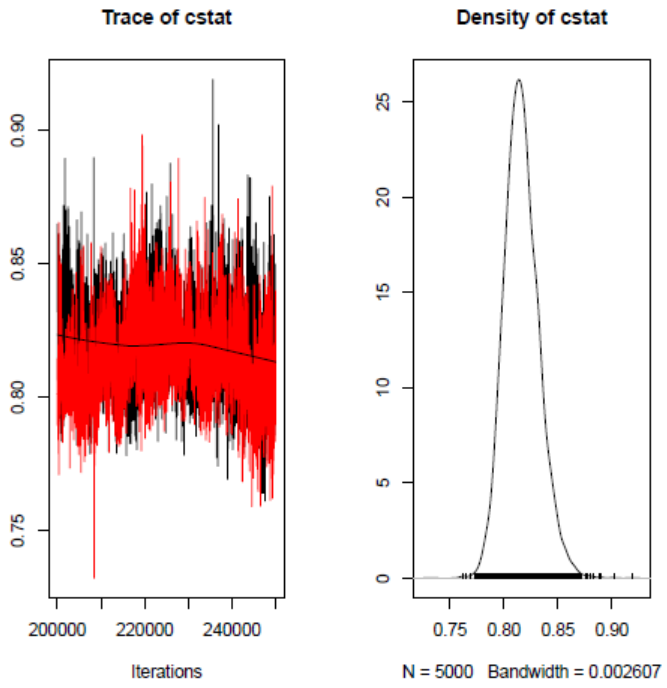
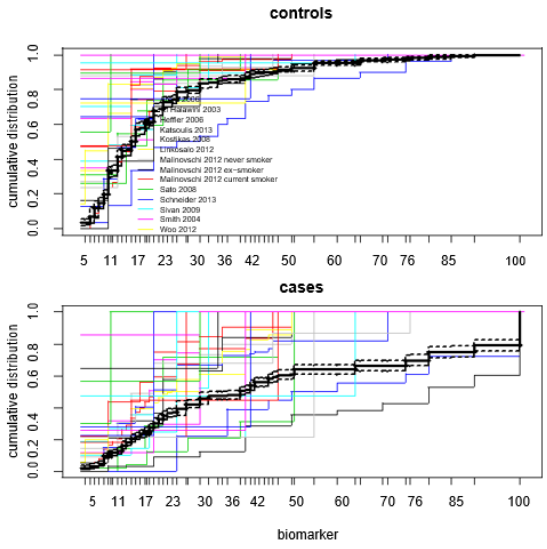


Figure 3: Trace- and density-plots of the c-statistic.

As a second example we re-analyzed the data of [4] on the di-agnostic accuracy of fractional exhaled nitric oxide (FENO) for the diagnosis of asthma. The data were from 29 studies reporting sen-sitivity and specificity results of 150 cut-offs, of which 53 thresh-olds were unique. The data is available on: https://data.mendeley.com/datasets/fndpn5bnps/1.



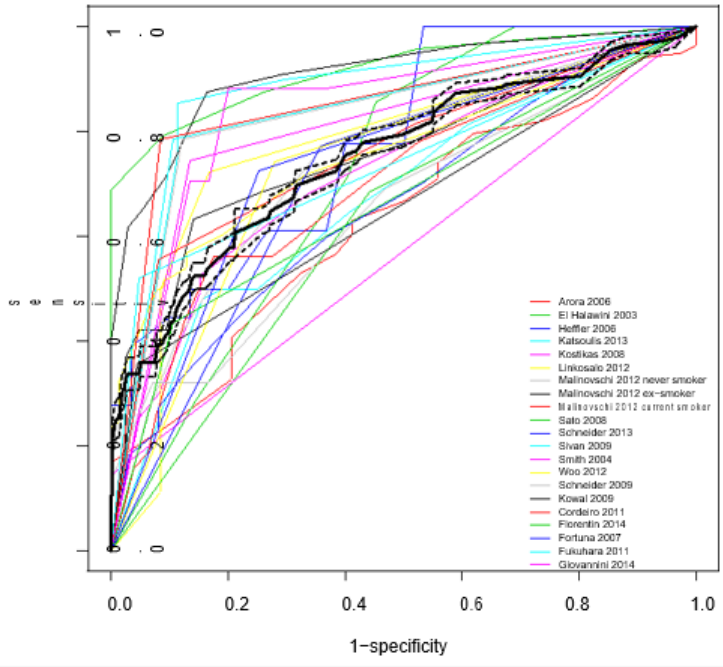


Figure 4: Density and cumulative functions of the various studies of the fractional ex-haled nitric oxide biomarker.

The random-effects model did not converge within a reasonable time-frame, and therefore we performed a fixed-effects analysis. The results of the multinomial-normal and of the multinomial-Dirichlet mixture models were again very comparable; we report therefore only the results of the multinomial-normal mixture model. Observed and estimated average cumulative distributions of FENO in controls and cases per study are reported in Figure 4 (top panel) and the associated ROC curves are reported in the lower panel. The thin coloured lines are associated with the individual studies, the thick black lines represent averages (together with 95% credibility intervals). Area-under-the-averaged-ROC-curve (AUC) was estimated as 0.750 (95% CI: 0.732-0.767). AUC was estimated as 0.778 (95% ci: 0.710-0.834) according to the DIDS-model of Steinhauser et al [5].

Discussion

We extended the work of Steinhauser et al for the meta-analysis of the results of a series of diagnostic studies of a quantitative biomarker Y where results were reported with sensitivity and specificity values at multiple and varying cutoff-values across the studies. Our model is based on the observed numbers of cases and controls in the various categories of the biomarker Y defined by the cutoff-values used by the different studies and we specified a random-effects, meta-analytic, structure. Our approach gave similar results as Steinhauser et al's method for the two example datasets that we considered, but that may not be the case in general. Our model requires less assumptions and therefore is likely more robust.

Both our and Steinhauser et al's method ignore the reasons why the specific cutoff-values were chosen in the studies, so, basically, the cutoff-values were considered to be 'fixed'. In practice cutoff-values are probably highly coincidental and the sampling-variation underlying these choices is ignored, which likely leads to too small credibility or confidence intervals.

That our approach does not require any distributional assumption for the biomarker is a nice feature, but it has as a downside that statistics (like Youden, sensitivity, specificity) can only be calculated for cutoff-values that have been used in at least one study. This is unlike Steinhauser et al's method where the underlying parametric model also allows estimation of performance statistics at thresholds that not have been used by any of the included studies. Our model can perhaps be smoothed in order to facilitate interpolation, for instance by penalizing squared differences between adjacent categories (i.e., $\left(\pi_{i1j} - \pi_{i1,j-1}\right)^2$), but that is outside the scope of the present paper.

Our approach is implemented in the standalone computerprogram JAGS and runs through the rjags-library in R. It is much less fast than the approach of Steinhauser et al, needing hours of computation time (versus seconds by Steinhauser et al's method). This is consequence of both the nonparametric character of our model and our decision to use a Bayesian estimation algorithm. Perhaps calculations can be sped up by using Hamiltonian MCMC (as implemented in the Stan program), but the large(r) number of parameters in our model that needs to be estimated is the main issue. For the liv-

er-biomarker example there were 24 different thresholds, meaning that there were 25 categories in the biomarker-distributions in cases and controls. Therefore, there were $2*(25-1)=48$ mean-parameters to be estimated (i.e., μ) and $48*(48-1)/2=1128$ variances and covariances. In Steinhauser et al's approach there were 4 mean and $4*(4-1)/2=6$ (co-)variance parameters in their largest DIDS-model which is best comparable to ours. Moreover, the number of parameters remains the same in their model whereas the number of parameters in our model is a direct function of the num-ber of different cutoff-values reported in the set of studies. It took a few seconds to analyse Schneider et all's FENO data, whereas it took several hours with our approach.

Computation time of our approach for the liver-biomarker data was about 30 minutes compared to a few seconds for Steinhauser et al's approach. However, Steinhauser et al's software reported convergence-problems for most of the complex models. It appeared that correlations between some of the random effects were close to unity. Models can be simplified in Steinhauser et al's approach but when doing that we found that results were quite varying. We found that AUC's varied between 0.696/0.747 according to Steinhauser et al's CI/CS models that converged without warnings to 0.796/0.809 according to the DIDS/DICS-models that failed to converge with singular fits. Because of the Bayesian algorithm that we used, such convergence problems did not arise with our approach. Convergence was fine as judged by trace-plots [6,7].

As described earlier, fixed-effects model-variants are easily formulated by dropping the study-subscript i from the transformed category-parameters a_{i0j} and a_{i1j} and actually specifying $a_{i1j}=a_{1j}$ and $a_{i0j}=a_{0j}$ for all $i=1,...,N$ and $j=1,...,m.$ This leads to a minor adaptation of the JAGS-syntax. The results of the fixed-effects model for the two examples were comparable, but the averaged density- and cumulative distribution functions were less smooth (and therefore the averaged ROC-curves too).

Conclusion

We developed a new model for meta-analysis of diagnostic studies evaluating multiple thresholds of a quantitative biomarker. The new model provided comparable results as an existing method but with less assumptions.

Acknowledgement

None.

Conflict of Interest

The author declares that he has no competing interests.

References

1. Steinhauser S, Schumacher M, Rucker G (2016) Modelling multiple thresholds in meta-analysis of diagnostic test accuracy studies. BMC Medical Research Methodology 16(1): 97.
2. Rucker G, Steinhauser S, Kolampally S, Schwarzer G. R-package Diagmeta: Meta-Analysis of Diagnostic Accuracy Studies with Several Cutpoints. Version 0.3-0. https://github.com/guido-s/diagmeta.
3. Patel PJ, Connoley D, Rhodes F, Srivastava A, Rosenberg W, et al. (2020) A review of the clinical utility of the Enhanced Liver Fibrosis test in multiple aetiologies of chronic liver disease. Ann Clin Biochem 57(1): 36-43.
4. Schneider A, Linde K, Reitsma JB, Steinhauser S, Rucker G, et al. (2017) A novel statistical model for analyzing data of a systematic review generates optimal cutoff values for fractional exhaled nitric oxide for asthma diagnosis. Journal of Clinical Epidemi-ology 92: 69-78.
5. Reitsma JB, Glas AS, Rutjes AW, Scholten RJ, Bossuyt PM, et al. (2005) Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. J Clin Epidemiol 58(10): 982-90.
6. Harbord RM, Deeks JJ, Egger M, Whiting P, Sterne JA, et al. (2007) A unification of models for meta-analysis of diagnostic accuracy studies. Biostatistics 8(2): 239-51.
7. Plummer M. JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling, 2003. Proceedings of the 3rd International Workshop on Distributed Statistical Computing (DSC 2003), March 2022, Vienna, Austria. http://mcmc-jags.sourceforge.net.